

**REAL ACADEMIA DE CIENCIAS
EXACTAS, FÍSICAS Y NATURALES DE ESPAÑA**

**EL CÁNCER: UN PROCESO
DE EVOLUCIÓN DARWINIANA**

DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICA DE NÚMERO POR LA
EXCMA. SRA. D^a. NÚRIA LÓPEZ-BIGAS

Y CONTESTACIÓN DE LA
EXCMA. SRA. D^a. MARÍA ÁNGELA NIETO TOLEDANO

EL DÍA 24 DE ABRIL DE 2024



MADRID
Domicilio de la Academia: Valverde, 22
www.rac.es

**REAL ACADEMIA DE CIENCIAS
EXACTAS, FÍSICAS Y NATURALES DE ESPAÑA**

**EL CÁNCER:
UN PROCESO DE
EVOLUCIÓN DARWINIANA**

DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICA DE NÚMERO POR LA

EXCMA. SRA. D^a NÚRIA LÓPEZ-BIGAS

Y CONTESTACIÓN DE LA

EXCMA. SRA. D^a MARÍA ÁNGELA NIETO TOLEDANO

EL DÍA 24 DE ABRIL DE 2024



MADRID
Domicilio de la Academia
Valverde, 22

ISSN: 0214-9540
ISBN: 978-84-87125-84-3
Depósito legal: M-9463-2024

ÍNDICE

Agradecimientos	5
Introducción	7
1. Evolución darwiniana	7
2. El cáncer y su incidencia	9
3. Evolución dentro de nuestro cuerpo	10
4. Barreras contra el cáncer	13
a. Barreras celulares	14
b. Arquitectura de nuestros tejidos como barrera contra el cáncer	14
c. El sistema inmunitario como barrera contra el cáncer	16
5. La paradoja de Peto	17
6. El cáncer está causado por alteraciones en el genoma	18
7. Secuenciación de genomas de tumores	20
8. El genoma del tumor contiene trazas de su historia	21
Variación: ¿Cómo ocurren las mutaciones?	27
9. Daño, errores y reparación del ADN	27
10. Firmas mutacionales	28
11. Variación de la tasa de mutaciones en el genoma	30
a. De cómo la replicación afecta a la tasa de mutaciones	31
b. Empaquetando 2 metros de ADN en un espacio de 6 micrómetros	31
c. Transcribiendo y reparando	32
12. Variación local en la tasa de mutación	33
a. Factores de transcripción	34
b. Nucleosomas	36
c. De exones e intrones	41
13. Mutaciones causadas por la quimioterapia	43
14. Estimación de la tasa de mutagénesis neutral	47

Selección: ¿Qué mutaciones causan cáncer?	49
15. Selección positiva para identificar genes de cáncer	49
16. Hacia un Compendio de Genes de Cáncer	51
17. Otras alteraciones genómicas en tumores	55
18. ¿Drivers en el genoma no codificante?	56
19. ¿Cómo identificar las mutaciones específicas que causan cáncer?	59
20. ¿Cuántas mutaciones driver se requieren para causar un cáncer?	63
21. ¿Cómo ayuda este conocimiento a la mejora de los tratamientos contra el cáncer?	64
Nuevas fronteras: variación y selección en tejidos sanos	67
22. Variación y selección en tejidos sanos	67
23. Hematopoyesis clonal	68
a. Descubriendo los genes de hematopoyesis clonal	69
b. Identificando las mutaciones causantes de hematopoyesis clonal	71
c. ¿Cómo afecta la quimioterapia a la hematopoyesis clonal? ...	71
24. Expansiones clonales en tejidos sólidos	73
25. Las mutaciones drivers son necesarias pero no suficientes	75
26. ¿Por qué la probabilidad de que tengamos cáncer aumenta con la edad?	76
27. Carcinógenos y factores de riesgo	78
28. Midiendo la estructura clonal para estudiar el riesgo de cáncer	81
29. Prevención del cáncer	84
Conclusiones y mirada al futuro	87
Bibliografía	91
Grupo	99
Contestación de la Excma. Sra. D ^a María Ángela Nieto Toledano ...	101

Excelentísimo Señor Presidente,
Excelentísimas Señoras y Señores Académicos,
Queridos colegas, familiares y amigos, Señoras y Señores

AGRADECIMIENTOS

Empezaré mi discurso mostrando mi absoluta sorpresa y agradecimiento a los académicos por proponer mi nombre y apoyar mi nominación para formar parte de esta academia. Agradecimiento por valorar tan positivamente mis méritos y aportaciones como para considerarme con el honor de formar parte de esta institución histórica. Sorpresa porque jamás en mi vida se me había ni pasado por la cabeza que algún día formaría parte de esta institución.

Mi mayor agradecimiento es para mis padres, quienes nos animaron, a mí y a mis hermanos, con sus palabras y su ejemplo a dedicarnos a lo que nos gustara e intentar hacerlo lo mejor posible. Uno de sus principales objetivos fue darnos la oportunidad de una educación universitaria que ellos no tuvieron. Gracias a su esfuerzo y a la educación pública, que debemos preservar a toda costa, yo pude estudiar la carrera de biología.

No fue hasta casi al final de la carrera que empecé a entender lo que significaba ser científica y a imaginarme que yo quizás podría dedicarme a ello. Fue gracias a conocer a Anna Bigas, prima de mi madre. Anna, que en ese momento estaba haciendo un postdoctorado en Seattle, fue mi principal referencia y apoyo para esos primeros pasos en el mundo de la investigación. ¡Gracias!

Muchas otras personas me han acompañado, formado y guiado en este camino, en particular mis directores de tesis: Xavier Estivill y Mariona Arbonés, y Christos Ouzounis y Roderic Guigó, en cuyos laboratorios aprendí bioinformática durante mi época postdoctoral.

Debo ser muy honesta y decir que todo, absolutamente todo, lo que los académicos han valorado para mi nominación y lo que se explica en este discurso, no ha sido un logro individual; ha sido un esfuerzo colectivo. El mérito es compartido con un grupo fantástico de personas con quien he

tenido la suerte de trabajar en el laboratorio que dirijo, tanto en nuestro inicio en la Universidad Pompeu Fabra (2006 - 2016), como más recientemente en el IRB Barcelona (des del 2016). Con el propósito de subrayar la naturaleza colaborativa de este trabajo, menciono a todas estas personas excepcionales al término de mi discurso escrito.

En particular quiero hablar de Abel González-Pérez, con quien comparto conversaciones científicas, ideas, dirección de proyectos y de Tesis doctorales desde hace más de 14 años. Trabajar con Abel es un lujo absoluto y ciertamente a mi me hace mejor científica. Otras personas en nuestro laboratorio son y han sido claves en varios momentos, y para no alargarme no mencionaré nombres concretos aquí, aunque ellas saben quienes son.

También quiero agradecer a los pacientes y familiares que han compartido generosamente muestras y datos de su tumor u otros tejidos para la investigación. Sin su generosidad desinteresada nuestra investigación no sería posible.

Finalmente agradecer a mi familia, mi pareja Alba, y mis hijas Clara y Jana. Alba me acompaña en mi vida desde antes de acabar la Tesis doctoral, y ha sido un pilar fundamental durante toda mi trayectoria científica. Sin su constante apoyo no estaría hoy aquí.

INTRODUCCIÓN

Este discurso se centra en el estudio del cáncer como un proceso de evolución darwiniana. Variación y selección están en la base de este fenómeno. En nuestros órganos y tejidos, existe variación entre las células que los componen debido a múltiples procesos que causan cambios en el ADN (mutaciones) en cada una de las células a lo largo de la vida. De esta forma, cada célula es ligeramente distinta a sus células vecinas. La mayoría de estas mutaciones no tienen un efecto importante en el funcionamiento celular, pero ocasionalmente una de estas mutaciones en un contexto concreto, confiere una ventaja selectiva a una célula determinada para que se divida más eficientemente que las células adyacentes. Como resultado se incrementa el número de células descendientes de la célula con la mutación, lo que llamamos una expansión clonal. La acumulación de varias mutaciones que pueden dar ventaja selectiva a células de este clon puede llevar a un crecimiento descontrolado causando un cáncer. Esta evolución darwiniana persiste durante procesos de invasión tumoral y metástasis, e incluso durante la terapia, causando en demasiadas ocasiones la resistencia al tratamiento y la reaparición del tumor.

El estudio del cáncer bajo el prisma de la evolución darwiniana nos permite encontrar respuestas a los interrogantes fundamentales acerca de esta enfermedad. En este discurso examinaré en detalle los dos aspectos claves de la evolución darwiniana en nuestros tejidos. **Variación:** ¿cómo ocurren las mutaciones en nuestras células? Y **Selección:** ¿Qué mutaciones dan una ventaja selectiva hacia el desarrollo de un cáncer?

1. Evolución darwiniana

En 1832 un joven naturalista inglés de 22 años emprendió un viaje trascendental a bordo del HMS Beagle. Charles Darwin viajó alrededor del mundo durante casi cinco años con la misión de recabar información sobre la geología, la fauna y la flora de las zonas exploradas. Sus meticulosas observaciones sobre la diversidad de la vida en distintas regiones del mundo, así como sus intercambios de ideas con varios naturalistas de la época, incluida su correspondencia con Alfred Russel Wallace, le permitieron desarrollar una teoría que cambiaría el curso de la historia de la ciencia y del pensamiento (Darwin 1859)¹.

La teoría es sencilla. Todas las especies tienen un origen común, y han evolucionado por variación y selección natural. Cualquier población de individuos de una especie es diversa, es decir hay diferencias entre unos y otros. Las características hereditarias que resultan en menos probabilidades de supervivencia en el entorno o peores oportunidades reproductivas a sus portadores acaban desapareciendo de la población (selección negativa). Las características hereditarias que hacen que un individuo sobreviva y se reproduzca más exitosamente se expanden en la población, pudiendo llegar a fijarse como una característica de la especie (selección positiva). Con el tiempo, a través de la variación y selección natural, las especies van cambiando y adaptándose a su entorno.

Faltaban casi 100 años para conocer los detalles de cómo se transmite esta variación de una generación a la siguiente. Sin embargo, los principios fundamentales de la teoría ya eran acertados.

Ahora sabemos que esta variación se genera por alteraciones en el ADN (ácido desoxirribonucleico) que se transmiten a la siguiente generación. Estas alteraciones, llamadas mutaciones, hacen que los organismos sean únicos y diversos, proporcionando la variedad necesaria para que la selección natural actúe. Todos los organismos de la Tierra (con la única excepción de los virus de ARN, ácido ribonucleico) compartimos el ADN como base de nuestro genoma. En el caso de los organismos unicelulares y algunos pluricelulares, la transferencia del ADN a la siguiente generación ocurre por una simple división celular o gemación, donde cada descendiente hereda una copia relativamente fiel del ADN de la célula progenitora, con ciertas variaciones. La mayoría de plantas y animales, en cambio, se propagan por reproducción sexual, donde el ADN de dos individuos de la especie se combina para generar un nuevo organismo descendiente. En este caso unas pocas células especializadas, los gametos femeninos y masculinos, son las responsables de transmitir la información hereditaria a la progenie. El resto de células del organismo, encargadas de formar todos los tejidos del organismo, son las llamadas células somáticas y no contribuyen con material genético a la siguiente generación.

Estudiar ecología, fisiología, embriología y medicina bajo el prisma de la evolución proporciona una base desde la cual entender procesos aparentemente muy complejos y dispares. Ya lo decía el genetista Theodosius Dobzhansky en 1973, “nada en biología tiene sentido excepto a la luz de la evolución”². El cáncer no es una excepción.

2. El cáncer y su incidencia

Cáncer es el nombre que le damos a un conjunto de enfermedades que tienen en común la proliferación descontrolada de células de nuestro cuerpo. Es una enfermedad muy frecuente, en España aproximadamente un 33% de hombres y un 23% de mujeres la sufren a lo largo de su vida³. Sin embargo, es una enfermedad mayoritariamente asociada a la edad (Figura 1A).

Es esclarecedor observar esta gráfica y apreciar que la incidencia de cáncer empieza su ascenso justo después de la edad reproductiva. Desde el prisma de la evolución darwiniana es lógico pensar que la selección natural tiene menos efecto en edad posreproductiva. Nuestra especie, como todas las demás, ha evolucionado para maximizar la probabilidad de dejar descendencia y perpetuar la especie. Por esta razón la presión selectiva hasta la edad reproductiva es mayor. También es importante mencionar que durante la mayor parte de nuestra historia evolutiva como especie *Homo sapiens*, la probabilidad de supervivencia más allá de los 30 años era baja en cualquier caso. En la antigüedad la mayoría de humanos morían por infecciones, depredadores u otras causas antes de llegar a esta edad. En el siglo XVIII la esperanza de vida en Francia era de menos de 30 años⁵ (Figura 1B).

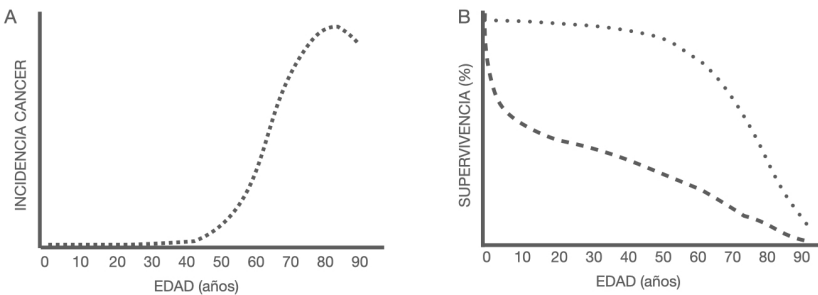


Figura 1. A. Gráfica esquemática del incremento de incidencia acumulada del cáncer con la edad. Basada en datos reales obtenidos del Instituto Nacional del Cáncer de Estados Unidos⁴. **B.** Gráfica esquemática de supervivencia acumulada con la edad en el siglo XVIII (línea discontinua) y la época actual (línea de puntos). Adaptada de⁵

La acción de la selección natural, a lo largo de millones de años, ha determinado la aparición de adaptaciones en el funcionamiento de nuestras células, tejidos y órganos que previenen el desarrollo del cáncer a edades tempranas y en especial antes de la edad reproductiva, como veremos en la sección 4. Sin embargo, la selección natural ha sido mayoritariamente ciega a adaptaciones en edades posteriores, lo que explica que la incidencia de cáncer aumenta con la edad.

Según un simple modelo matemático formulado en 1954 por los epidemiólogos británicos, Peter Armitage y Richard Doll⁶, el incremento exponencial de la incidencia de muchos tipos de cáncer se podría explicar considerando el cáncer como el resultado de un proceso con múltiples etapas, entre 5 y 7 según el tipo de cáncer: cada una de estas etapas serían cambios celulares discretos, no reversibles y aleatorios, que se deben dar en la misma célula para que ésta devenga cancerosa. Podríamos interpretar cada etapa como un salto a los distintos obstáculos o barreras impuestas por la evolución de nuestra especie para evitar que una célula de nuestro cuerpo se des controle hasta el punto de causar un cáncer. Los mismos autores apuntaron a que estos cambios podrían ser mutaciones somáticas, es decir mutaciones acumuladas en células somáticas a lo largo de la vida del individuo.

El cáncer está directamente relacionado con la multicelularidad. La mayoría de los animales puede sufrir cáncer, incluso hay evidencias fósiles de sarcomas en dinosaurios. Desde el momento en que se inició la multicelularidad, se hizo necesario un control exquisito de la proliferación celular. Las células somáticas (es decir todas las células de nuestro cuerpo exceptuando las células germinales: óvulos y espermatozoides) están al servicio del individuo, y no de sí mismas. La pérdida de control de la proliferación celular y la competencia en lugar de la cooperación entre células es lo que lleva al cáncer. Por eso cuando nos preguntan si el cáncer va a desaparecer, la respuesta es no. Podemos entender mejor cómo se origina, diagnosticarlo antes, desarrollar mejores tratamientos, incluso disminuir su incidencia con medidas preventivas, pero no será posible erradicar esta enfermedad ya que está estrechamente ligada a nuestra naturaleza de organismos pluricelulares.

3. Evolución dentro de nuestro cuerpo

Nuestros tejidos están sujetos de forma continuada a procesos de evolución darwiniana. Todas las células de nuestro cuerpo (unos 30 billones) derivan de una célula ancestral, el cigoto. Desde la primera división celular se acumulan mutaciones en el ADN de nuestras células, generando variación hereditaria entre ellas. Hereditaria en el sentido que se transmiten a todas las células somáticas descendientes, que no debemos confundir con las mutaciones que se transmiten a la siguiente generación, que solo son aquellas que ocurren en las células de la línea germinal (óvulos y espermatozoides). Aquellas mutaciones que ocurren en las primeras divisiones celulares de la formación de un nuevo organismo pueden llegar a estar presentes en un porcentaje alto de las células de un individuo, es lo que llamamos mosaicismosomático. Las mutaciones que ocurren más tarde en el desarrollo o

en el adulto, quedan relegadas a un porcentaje pequeño de las células de un tejido, y a un linaje concreto. En definitiva, nuestro cuerpo es un mosaico de clones somáticos algunos iniciados más tempranamente en el desarrollo y otros que continuamente se van acumulando durante nuestra vida.

De hecho, una gran parte de nuestras células, en particular aquellas que forman parte de tejidos de división relativamente rápida, como los epiteliales (las capas de células que recubren superficies), se sustituyen por otras cada pocos días o pocos meses. En el transcurso de los años, la mayor parte de las células que conformaron el organismo han sido reemplazadas por nuevas células. Lo que nos lleva a recordar la paradoja del barco de Teseo. Después de haber sido reemplazadas todas las piezas que conforman el barco, ¿puede seguir considerándose el mismo barco?, ¿o deberíamos considerarlo un barco diferente al inicial? Algunas células de nuestro cuerpo son, en cambio, muy longevas, y nos acompañan durante toda nuestra vida, como es el caso, por ejemplo, de las neuronas y las células musculares cardíacas.

En este contexto de multicelularidad y mosaicismo somático la evolución darwiniana opera de manera análoga a como lo hace en poblaciones de bacterias y organismos unicelulares. Si alguna de las mutaciones confiere una ventaja selectiva a la célula que la adquiere, esta puede sobrevivir y dividirse más que otras células a su alrededor. Como resultado se puede desencadenar una expansión clonal, que resultará en un número mayor de células descendientes de esta célula con la mutación ventajosa. Llamamos a estas mutaciones que confieren una ventaja selectiva “mutaciones *drivers*”, tomando prestado el término del inglés, y así me referiré a ellas a lo largo de este texto.

La ventaja selectiva que pueden proporcionar las mutaciones *drivers* depende del medio o del contexto donde se encuentra la célula que las contiene. Por ejemplo, la misma mutación puede dar una ventaja selectiva a una célula en el esófago, y no proporcionar ninguna ventaja a una célula de la piel. Igualmente esta ventaja selectiva dependerá del momento y del entorno específico dentro del tejido en que se encuentre esta célula. La evolución de las especies descrita por Darwin opera exactamente de la misma forma, es totalmente dependiente del ambiente. Por ejemplo, una jirafa con el cuello más largo puede acceder a las hojas más altas de los árboles para alimentarse. Pero esto solo es una ventaja selectiva si las hojas más cercanas se agotan. En este contexto, una variación que proporciona un cuello más largo a una jirafa le da una ventaja selectiva en comparación al resto de jirafas. En los tejidos de nuestro cuerpo ocurre lo mismo: una mutación puede ser o no ventajosa para una célula dependiendo de múltiples factores del ambiente.

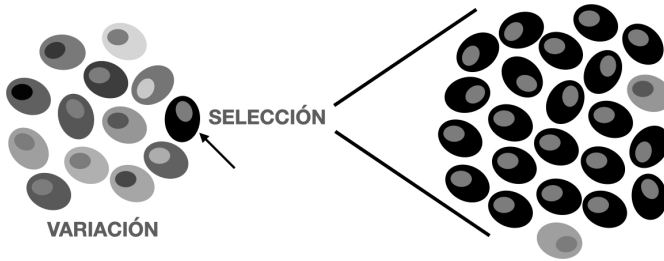


Figura 2. Dibujo esquemático que representa la variación entre las células de nuestro organismo. Cada óvalo representa una célula que por el tono de gris vemos que es ligeramente distinta a otras. Cuando una célula adquiere una mutación driver que proporciona una ventaja selectiva a esta célula (señalada con la flecha), ésta puede dar lugar a una expansión clonal.

Las células con mutaciones ventajosas pueden liderar una expansión clonal, que al acumular nuevas alteraciones genómicas y más olas de expansión clonal podría llegar a un descontrol celular total, y convertirse en células cancerosas, que tal como nos cuenta Carlos López-Otín, son células egoístas, inmortales y viajeras⁷. Estas células tumorales se expanden descontroladamente, invadiendo tejidos, viajando y colonizando otros órganos (lo que llamamos metástasis), e incluso en ocasiones llevando a la muerte del individuo en el que se han originado. Y aquí una paradoja: la ventaja selectiva a nivel de células individuales conduce a la desaparición de las mismas con la muerte del individuo. Afortunadamente, salvo casos muy particulares^{8,9}, estas células tumorales no pueden colonizar tejidos de otra persona. Las células cancerosas, tanto de cánceres humanos como de la mayoría de cánceres de otras especies, son incapaces de transmitirse de un individuo a otro. Entre otras razones, debido a que el sistema inmunitario las detectaría como ajenas y las eliminaría (ver sección 4c).

Pero hay algunas excepciones, son los que llamamos cánceres transmisibles. Algunos de los ejemplos más llamativos son los tumores transmisibles del diablo de Tasmania¹⁰. Son tumores faciales y orales malignos que se transmiten por mordeduras que se infligen entre los animales durante el apareamiento y alimentación. La primera identificación de este tumor fue en 1996, aunque estudios genómicos y evolutivos de los tumores han determinado que apareció unos 10 años antes en un diablo “fundador”, y a partir de este se ha transmitido por todo el noreste de Tasmania. La población de diablos en esta región ha disminuido drásticamente y la especie está en peligro de extinción. Existe un linaje tumoral adicional, que se calcula que apareció en una hembra alrededor del 2011. El hecho de que se

hayan observado dos linajes independientes indica que la probabilidad de que esto ocurra en esta especie no es ínfima.

Otro ejemplo es un tumor venéreo transmisible de perros. En este caso, el tumor se desarrolla en el área genital y se transmite de un animal a otro durante el apareamiento. Se calcula que este cáncer se originó en un perro “fundador” hace varios miles de años, y actualmente afecta a perros de todo el mundo. El estudio de este tumor desde el punto de vista evolutivo es fascinante. La secuenciación de cientos de muestras de este tumor obtenidas de perros en diferentes partes del mundo permite reconstruir un árbol filogenético que explica la relación entre los tumores de distintos perros, y como consecuencia estudiar la variación y selección ocurrida en este linaje tumoral durante miles de años. Incluso permite explorar migraciones de perros con seres humanos alrededor del mundo en los últimos siglos¹¹.

También se han descubierto cánceres transmisibles en varios moluscos bivalvos, como almejas, berberechos y mejillones. Recientemente un estudio de la Universidad de Santiago de Compostela, liderado por el biólogo José Tubio, ha estudiado más de 7000 muestras de un tipo de leucemia que afecta a los berberechos, pudiendo reconstruir árboles filogenéticos de este tumor y relacionarlos con su expansión geográfica¹².

Es importante remarcar que estos cánceres transmisibles son excepciones. La mayoría de los tumores de animales se inician en una célula normal en un tejido de un individuo y desencadenan una trepidante carrera evolutiva hacia la nada. Cuanto más exitosas sean las células de este clon tumoral en saltarse todas las barreras y seguir proliferando descontroladamente, más probabilidades tienen de afectar mortalmente a su huésped, y por tanto causar su propia desaparición.

4. Barreras contra el cáncer

Como comentaba en la sección 2, el cáncer es una enfermedad asociada a la edad. Millones de años de evolución han resultado en un organismo robusto para evitar el cáncer a edades tempranas. La evolución en cambio ha sido mayoritariamente ciega en etapas posteriores, más allá de los 40 o 50 años. Esta robustez se debe a un conjunto de características de nuestras células, nuestros órganos y tejidos y nuestro sistema inmunitario. Sin pretender ser exhaustiva mencionaré algunas de estas barreras en las siguientes subsecciones.

a. Barreras celulares

Cada célula de nuestro cuerpo desempeña una función específica y sólo debe dividirse cuando es necesario. Si una célula está dañada o ya no cumple su función, debe someterse a un proceso de eliminación para el bien del organismo. La apoptosis, una forma de suicidio o muerte celular programada, se activa en estos casos. Las señales que desencadenan la apoptosis pueden provenir tanto del entorno celular como del propio interior de la célula.

Además, la mayoría de nuestras células tienen un límite en la cantidad de divisiones que pueden experimentar; sólo unas pocas, las llamadas células madre, de las que hablaremos más en la siguiente sección, conservan la capacidad de dividirse de manera ilimitada¹³. En cada división, los extremos de los cromosomas, conocidos como telómeros, se acortan, y este acortamiento sirve como un mecanismo para contabilizar las divisiones y alertar cuando se alcanza el límite¹⁴. En este momento las células entran en un estado que llamamos de “senescencia” y no se dividen más.

Por otro lado, el contacto entre células juega un importante papel en la regulación de su comportamiento. En particular, las células epiteliales debido a interacciones entre ellas tienen limitada su capacidad de moverse y migrar, formando así un tejido compacto y bien ordenado. Se ha demostrado que las células epiteliales en su transformación hacia células tumorales modifican estas características adquiriendo propiedades migratorias, que les permite invadir otros tejidos colindantes y eventualmente formar una metástasis. Ángela Nieto, junto con otros prominentes investigadores españoles (Amparo Cano, Eduard Batlle y Antonio Garcia de Herreros) identificaron hace más de 20 años los mecanismos moleculares de esta transformación^{15,16}.

Cualquier célula tumoral ha superado, de alguna manera, estas y otras barreras que normalmente evitan la proliferación descontrolada, una característica fundamental del cáncer¹⁷⁻¹⁹. Por ejemplo, muchas células cancerosas pierden el control del mecanismo de apoptosis, de manera que aunque reciban señales de que les toca morir se resisten a ello. Además, la mayoría de células tumorales activan la reconstrucción de los telómeros para posibilitar su división ilimitada.

b. Arquitectura de nuestros tejidos como barrera contra el cáncer

Ya en 1975 John Cairns propuso que la arquitectura de nuestros tejidos ha evolucionado para constituir barreras que eviten el desarrollo tumoral²⁰. En

nuestros tejidos reinan exquisitas jerarquías de diferenciación, donde la mayor parte de células que hacen las funciones específicas del tejido son células diferenciadas incapaces de dividirse. Solo unas pocas células mantienen la capacidad de proliferar indefinidamente, y generar descendencia. A estas las llamamos células madre.

Desde el punto de vista evolutivo esta jerarquía tiene sentido, ya que se reduce la población efectiva de células que pueden propagar mutaciones a sus descendientes. La mayor parte de nuestras células desaparecen y las mutaciones que estas reciben son por tanto irrelevantes para el riesgo de cáncer. Las células madre residen en zonas específicas del tejido o nichos específicos que les proporcionan un grado de protección contra factores ambientales adversos. Por ejemplo, las células madre de la sangre se encuentran en la médula ósea, en el interior de los huesos, protegidas del efecto mutagénico de la luz ultravioleta²¹. En la piel, el colon y el esófago, las células madre están en la base del epitelio, alejadas de la parte exterior o del lumen. Además algunos tejidos de multiplicación rápida están organizados de manera que las células madre están sujetas a una fuerte restricción espacial y su progenie no puede desplazarse a otras partes del tejido. Estas estructuras minimizan la posibilidad de competición entre células, y por tanto limitan la expansión de una población que pudiera adquirir una ventaja selectiva. Este es el caso por ejemplo del epitelio del colon, que forma unas vellosidades y criptas para aumentar la superficie de epitelio de absorción. Estas criptas son estructuras clonales con una pequeña población de células madre en la base de la cripta, morfológicamente muy compartimentadas, lo cual limita la posibilidad de competición entre células madre de diferentes criptas.

La tasa de evolución de cualquier población está determinada por: a) la velocidad a la que se generan nuevas mutaciones, b) la probabilidad de que nuevos mutantes se extiendan por la población y c) el tiempo que les lleva hacerlo. La velocidad con la que se generan nuevas mutaciones en una población celular es un simple producto entre dos factores: i) la tasa de mutaciones por célula y ii) el número de células susceptibles de adquirirlas. Reducir el número de células que pueden transmitir nuevas mutaciones a la descendencia (sólo las células madre), y mantenerlas en nichos protegidos, como hemos visto, es una estrategia efectiva para disminuir la tasa de mutación poblacional. El hecho que estas células madre además residan en territorios delimitados, reduciendo la competición entre ellas, claramente reduce la posibilidad de propagación de linajes mutantes ventajosos.

c. El sistema inmunitario como barrera contra el cáncer

Nuestro sistema inmunitario es fascinante. Está formado por una compleja red de células, tejidos, moléculas y órganos bien orquestada con la función de detectar y eliminar invasores en nuestro cuerpo. Las células del sistema inmunitario, en particular los linfocitos, están entrenados para diferenciar lo propio de lo ajeno. Por ejemplo, pueden detectar la presencia de un virus en nuestro cuerpo ya que identifican proteínas del virus como ajenas. Pero las funciones del sistema inmunitario no se limitan a protegernos de peligros externos.

Ya a finales del siglo XIX se sugirió que el sistema inmunológico además de combatir organismos invasores tiene un rol en la detección y eliminación de células tumorales. Actualmente esta función está totalmente demostrada e incluso es la base de una nueva familia de tratamientos muy prometedores que utilizan el sistema inmunitario para atacar el cáncer, la inmunoterapia.

El rol del sistema inmunitario en la detección y eliminación de posibles células tumorales es probablemente una de las barreras más importantes para evitar el cáncer en nuestros tejidos durante buena parte de nuestra vida. Tanto la inmunidad innata, no específica, como la inmunidad adquirida o específica participan en este proceso. Los linfocitos T citotóxicos por ejemplo, son capaces de detectar fragmentos de proteínas anómalas en la superficie de las células tumorales, e inducir la muerte de estas células por apoptosis.

El desarrollo de un cáncer implica que el tumor, de una forma u otra, ha conseguido saltarse esta barrera y esconderse de la vigilancia del sistema inmunitario, o incluso inactivarlo, para evitar ser reconocido como anómalo y atacado. Hay varias estrategias que utilizan los tumores para conseguirlo. Un ejemplo son las proteínas CTLA-4, y PD-1, proteínas expresadas en la membrana de los linfocitos T que sirven como freno de su actividad inmunitaria. En su evolución, algunos tumores desarrollan adaptaciones para activar este freno. Las inmunoterapias más exitosas consisten en anticuerpos que bloquean estos frenos (anti-CTLA-4 y anti-PD-1) para así reactivar la función del sistema inmunitario contra el cáncer. En 2018 James P. Allison y Tasuku Honjo recibieron el premio Nobel de fisiología o medicina por el descubrimiento de las funciones de estas proteínas que han llevado a la invención de la inmunoterapia.

En resumen, la evolución de nuestra especie nos ha dotado de múltiples barreras, tanto a nivel celular, de organización de tejidos, como del sistema

inmunitario, que son efectivas para evitar que células mutadas proliferen descontroladamente durante buena parte de nuestra vida. Pero en algunos de nosotros, algunas células lograrán superar todas las barreras y darán origen a un cáncer. Un constante juego de variación y selección se libra continuamente en nuestros tejidos durante toda nuestra vida.

5. La paradoja de Peto

Como he mencionado previamente, todos los organismos pluricelulares son susceptibles a desarrollar cáncer. Dado el constante proceso de evolución darwiniana en nuestros tejidos, y considerando que cada división celular implica un riesgo de adquirir nuevas mutaciones, podríamos anticipar que los animales de mayor tamaño, y por ende con un mayor número de células, tendrían una probabilidad más elevada de que alguna de estas células inicie un proceso tumoral. Por otro lado, por el mismo razonamiento, especies con una mayor esperanza de vida deberían tener una mayor incidencia de cáncer.

Ciertamente, la incidencia de cáncer, como ya hemos visto, sí que se incrementa con la edad en humanos, y de una forma similar ocurre en otras especies de mamíferos. También hay una relación entre tamaño corporal e incidencia de cáncer entre individuos de una misma especie, que está bien documentada en humanos y en perros. Los perros de mayor tamaño tienen un riesgo mayor de sufrir cáncer.

Pero curiosamente esta relación no se mantiene entre distintas especies de mamíferos. Hay grandes diferencias en la esperanza de vida entre distintas especies (por ejemplo un ratón vive menos de 2 años y un humano unos 70-80 años) y también hay diferencias enormes en tamaño corporal (un ratón pesa aproximadamente unos 20 gramos y un humano aproximadamente 70 kilogramos). Sin embargo, no hay una relación entre estos dos parámetros, tamaño corporal y longevidad, y la incidencia de cáncer. Richard Peto planteó esta paradoja hace aproximadamente 50 años y, como sucede con todo en biología, solo podemos empezar a comprender su significado al contemplarlo a la luz de la evolución²².

La paradoja de Peto se explica si entendemos que a lo largo de la evolución, la selección natural ha favorecido la aparición de más barreras contra el cáncer y/o barreras más efectivas en especies con mayor tamaño corporal y mayor esperanza de vida, para evitar la alta incidencia de cáncer. Falta mucho para conocer los detalles de cómo cada especie logra

mantener una incidencia de cáncer que permita su supervivencia, pero hay algunas pistas interesantes. Los elefantes, por ejemplo, tienen hasta 20 copias del gen TP53 en su genoma²³. Un estudio reciente ha demostrado que la tasa de mutaciones somáticas varía entre especies de forma inversamente proporcional con su longevidad, es decir las especies que más viven consiguen mantener una tasa de mutación somática más baja²⁴, aunque esta relación inversa no se mantiene con el tamaño corporal.

La paradoja de Peto es un enigma evolutivo interesante que nos puede ayudar a comprender cómo la evolución de las especies ha favorecido la aparición de múltiples barreras contra el cáncer en las diferentes especies, adaptando la incidencia del cáncer al tamaño corporal y longevidad de cada especie.

6. El cáncer está causado por alteraciones en el genoma

En papiros egipcios del año 1600 a.C. ya se hace referencia al cáncer. Los escritos de Hipócrates del siglo IV a.c. describen unas lesiones ulcerosas a las que se refiere como “karkinos”, por su forma semejando las patas de un cangrejo, de ahí el origen etimológico de la palabra “cáncer”. Pero sus causas se mantuvieron misteriosas hasta mucho más tarde. Algunas de las primeras evidencias científicas de las causas del cáncer derivan de observaciones de la asociación de ciertas profesiones con ciertos tipos de cáncer específicos. En 1775, el cirujano británico Percival Pott se percató de la alta frecuencia de cáncer de escroto entre los deshollinadores, asociando la neoplasia al hollín y al alquitrán que se incrustaba en la ropa y en el escroto. A finales del siglo XIX, en el pueblo minero de Scheneeborgen, al este de Alemania, el médico de la mina junto con el médico del distrito se percataron de que los mineros de las minas de uranio de Scheneeborgen acostumbaban a morir jóvenes y frecuentemente de cáncer de pulmón. Más tarde se descubrió que el gas radón, un gas radioactivo que se produce naturalmente como producto de la desintegración del uranio, es el causante directo de estos cánceres de pulmón. Estas y otras asociaciones permitieron empezar a entender el efecto de factores ambientales en el desarrollo del cáncer.

Aunque, como en los ejemplos anteriores, la mayoría de casos de cáncer no tienen un componente hereditario, aproximadamente el 5-10% si lo tienen. El estudio de cánceres hereditarios permitió intuir las causas genéticas del cáncer y más tarde identificar algunos de los primeros genes causantes del cáncer. En 1866, el cirujano francés Paul Broca publicó por pri-

mera vez un caso de cáncer de mama hereditario. Describió el árbol genealógico de la familia de su mujer, con cuatro generaciones de mujeres afectadas por cáncer de mama. A estas observaciones se sumaron las de Theodor Boveri, quien a finales del siglo XIX, propuso que un cáncer se origina de una única célula, cuyos cromosomas están alterados, lo que provoca la división incontrolada de la célula. No fue hasta mucho tiempo después que sus hipótesis se vieron confirmadas por evidencias genéticas y celulares.

Sin pretender hacer una revisión completa de las primeras evidencias genéticas del cáncer, mencionaré sólo algunos momentos claves que contribuyeron de forma definitiva a demostrar las bases genéticas de esta enfermedad.

En 1911, Peyton Rous, patólogo americano, consiguió transferir un tumor maligno (concretamente un sarcoma) que crecía en una gallina doméstica a otra ave simplemente al exponer al ave sana a un filtrado que no contenía células, pero que aún así conservaba elementos capaces de transmitir la capacidad de formar tumores²⁵. Esto evidenció que unidades más pequeñas que las células eran responsables de la tumorigénesis. Este misterioso filtrado contenía lo que ahora conocemos como el Virus del Sarcoma de Rous. En 1966, ya con 87 años de edad, Rous recibió el premio Nobel de fisiología o medicina por este hallazgo. Durante un tiempo se creyó que los tumores eran causados mayoritariamente por virus; más tarde se ha visto que estos casos son en realidad una minoría.

En 1976 Harold Varmus y Michael Bishop, trabajando con el mismo virus del Sarcoma de Rous, descubrieron que el oncogén presente en el ADN del virus no es un gen de origen viral, sino un gen del genoma humano, el oncogén SRC, que el virus había adquirido durante la replicación en la célula huésped, y que se transmite a través del virus. Este descubrimiento les valió a Varmus y Bishop el premio Nobel de fisiología o medicina en 1989.

La identificación de la primera mutación somática en un gen de cáncer humano la debemos a Mariano Barbacid y otros investigadores en el año 1982²⁶. En concreto identificaron la mutación G12V (cambio de un aminoácido por otro en la posición 12 de proteína) en el gen *HRAS* en una línea celular de carcinoma de vejiga humana. Otros genes de cáncer detectados tempranamente se identificaron primero por sus mutaciones en la línea germinal en familias con cáncer familiar, como *RBI*, o los genes familiares asociados al cáncer de mama *BRCA1* y *BRCA2*. El gen *BRCA1* lo identifi-

có la genetista americana Mary-Claire King y su equipo en 1990²⁷. El gen *BRCA2* fue descubierto por el grupo de Mike Stratton y Richard Wooster en Londres en 1994²⁸.

Estos primeros estudios establecieron una evidencia clara de que mutaciones somáticas o de la línea germinal en algunos genes son las causantes del cáncer, y supusieron el inicio de la carrera para identificar otros genes relacionados con el cáncer.

7. Secuenciación de genomas de tumores

En el año 2002 se completó el primer borrador de la secuencia del genoma humano^{29,30}. A partir de aquí y gracias a mejoras en las tecnologías de secuenciación del ADN se empezaron a plantear proyectos destinados a secuenciar genomas de tumores para identificar las mutaciones causantes del cáncer.

Un primer estudio pionero liderado por Mike Stratton y Richard Wooster en 2005 secuenció 518 genes, escogidos meticulosamente por el hecho de codificar por proteínas con actividad de señalización celular (quinasas), en 25 tumores de mama³¹. Al año siguiente, el grupo de Bert Vogelstein secuenció 13 023 genes en 11 tumores de mama y 11 tumores colorrectales³². En 2007 un nuevo estudio del grupo de Mike Stratton amplió la secuenciación de genes que codifican por proteínas con actividad quinasa a 210 tumores humanos de varios tipos³³. En 2008 se logró la primera secuenciación del genoma completo de una leucemia³⁴ y posteriormente de tumores de otros tejidos³⁵⁻³⁷. Una de las principales sorpresas de estos primeros estudios fue la identificación de un gran número de mutaciones en cada genoma tumoral (miles en los casos que se secuenció el genoma completo). Quizás se esperaba la identificación de unos pocos eventos mutacionales en genes causantes del cáncer. Pero pronto se vio que la gran mayoría de las mutaciones y de los genes mutados no tienen ninguna implicación en el desarrollo tumoral. Así, el siguiente reto que se hizo evidente fue el de identificar aquellas mutaciones específicas causantes del cáncer entre las miles de alteraciones que contiene un tumor.

Otra observación importante de estos estudios pioneros fue que no todos los tipos de mutaciones se detectan con las mismas frecuencias. De las cuatro bases nitrogenadas que componen el ADN, adenina (A), citosina (C), guanina (G) y timina (T), se observó una clara preferencia por muta-

ciones de Cs que cambian a Ts en contextos de CGs (C seguida de una G). Además estas preferencias mutacionales varían según el tejido e incluso según el tipo de tumor. Más adelante veremos que esta observación abrió una puerta a la comprensión de los distintos procesos mutacionales que operan en nuestros tejidos.

Estos primeros estudios allanaron el camino para el lanzamiento de grandes iniciativas de secuenciación de tumores en varios países, como *The Cancer Genome Atlas* (TCGA), una iniciativa financiada por Estados Unidos cuyo objetivo fue secuenciar los exomas (es decir el conjunto de regiones del genoma que codifican proteínas) de cientos de tumores de 33 tipos de cáncer frecuentes³⁸. A medida que las tecnologías de secuenciación fueron avanzando, proyectos más ambiciosos, muchos de ellos agrupados bajo el paraguas del Consorcio Internacional del Genoma del Cáncer (ICGC)³⁹, se fijaron como objetivo secuenciar el genoma completo de miles de muestras de tumores. En España, como parte del consorcio internacional ICGC, se inició un proyecto liderado por los científicos Elías Campo y Carlos López-Otín, para secuenciar más de 500 genomas completos de leucemia linfática crónica⁴⁰⁻⁴².

Han pasado más de 15 años desde el inicio de la secuenciación de genomas de tumores, y hasta la fecha la comunidad investigadora, a través de diferentes iniciativas de secuenciación y análisis genómico de tumores, ha cosechado datos de decenas de miles de genomas tumorales de pacientes con cáncer de todo el mundo. Estos genomas han aportado datos únicos, ricos y valiosos para abordar el estudio del cáncer desde una perspectiva evolutiva.

8. El genoma del tumor contiene trazas de su historia

El genoma de un tumor y sus miles de mutaciones contienen trazas de toda su historia. A veces en mis ponencias científicas hago el símil con la ciudad de Roma. Cuando visitamos Roma, nos encontramos con una ciudad compleja, vibrante y moderna, donde viven casi 3 millones de personas. Como turistas probablemente recorreremos el Coliseo, las ruinas del Foro Romano, el Castillo de Sant'Angelo y algunas iglesias. Muchísimas otras ruinas, templos, iglesias y monumentos son testimonio de la larga y rica historia de la ciudad. Como visitantes nos puede parecer un conjunto caótico de ruinas, pero los expertos, estudiando meticulosamente cada uno de estos restos con sofisticadas técnicas de arqueología, pueden reconstruir en detalle la historia de la ciudad a través de los siglos.

Tenemos una sensación similar cuando observamos el genoma de un tumor y nos encontramos con miles de mutaciones y otras alteraciones genómicas complejas. Inicialmente puede parecer completamente caótico, pero equipados con sofisticadas técnicas de “arqueología molecular” podemos llegar a reconstruir partes importantes de la historia evolutiva del tumor.

Esto es una gran diferencia en comparación con el estudio de patrones de expresión de genes o proteínas, así como con los estudios de la estructura de la cromatina. Estos nos proporcionan información sobre el momento puntual en que se obtuvo la muestra, en lugar de revelar su historia evolutiva. Por ejemplo, al analizar una muestra tumoral y estudiar el nivel de expresión de los genes, ya sea en el conjunto del tumor o a nivel de célula única, estamos examinando el estado de esas células en el instante específico en el que se obtuvo la muestra.

Para revelar la historia evolutiva del cáncer a través de sus mutaciones, el primer paso es obtener una muestra del tumor – normalmente se obtiene de una biopsia o durante la cirugía para extirpar el tumor – extraer su ADN y secuenciarlo para, de esta manera, identificar las mutaciones del tumor. Pero aquí nos enfrentamos con un problema: el genoma de cada uno de nosotros es ligeramente distinto y de hecho cada persona tiene variantes heredadas que difieren del genoma humano de referencia. Aunque menos del 0.1% del genoma varía entre una persona y otra, si hacemos unos números rápidos (el genoma tiene 3200 millones de pares de nucleótidos) nos damos cuenta de que esto significa que hay millones de nucleótidos (o bases) que difieren entre dos personas. Para poder identificar las mutaciones únicas del tumor y diferenciarlas de las variantes heredadas por la persona, es necesario secuenciar el genoma de una segunda muestra de tejido sano de la misma persona. Todas aquellas variantes heredadas están en todas las células de la persona, y por tanto estarán compartidas entre el genoma de las dos muestras. Las mutaciones ocurridas en el linaje tumoral, en cambio, las observaremos sólo en el genoma del tumor. En el caso de tumores sólidos lo más común es utilizar la sangre como muestra de tejido sano, ya que es una muestra relativamente fácil de obtener. En el caso de leucemias u otras neoplasias hematológicas se obtiene una muestra de otro tejido, como la piel, o una muestra del paciente en el momento de máxima respuesta al tratamiento cuando el porcentaje de células tumorales en sangre es mínimo.

Desde la primera división del cigoto y en todas las divisiones sucesivas, ocurren mutaciones que se transmiten a la progenie, es decir a todas las células descendientes de la célula que recibe la mutación (Figura 3). Los

tumores se han expandido a partir de una sola célula de un tejido en su día era una célula normal. Esta célula contiene mutaciones que le dieron una ventaja selectiva (mutaciones *drivers*) para dividirse más de lo que debería produciendo un gran número de células descendientes, que pueden formar un tumor. Todas las células del tumor comparten todas las mutaciones que esta célula originaria ha adquirido durante toda su historia, ya que todas ellas se han transmitido a su progenie.

Durante la progresión tumoral se continúan acumulando nuevas mutaciones en cada célula del tumor. Si una de estas nuevas mutaciones proporciona una mayor ventaja selectiva, la progenie de la célula con esta nueva mutación puede acabar dominando el tumor. Sucesivas expansiones clonales pueden ocurrir durante la evolución de un tumor. Cuando secuenciamos el genoma de un tumor vemos las mutaciones de la célula ancestral causante de la expansión clonal completa más reciente (en inglés *MRCA: Most Recent Common Ancestor*) como mutaciones clonales, es decir, compartidas por todas las células del tumor. En cambio aquellas mutaciones que han ocurrido en células del tumor después de la última expansión clonal completa no están en todas las células tumorales, sólo en un subconjunto. Aquellas que forman parte de un clon suficientemente grande se pueden observar como mutaciones subclonales.

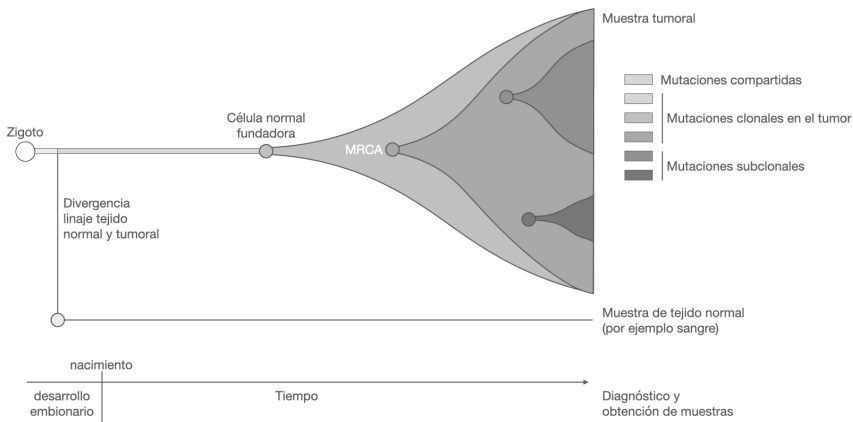


Figura 3. Esquema de la evolución de un tumor. Las mutaciones heredadas de los progenitores y aquellas que ocurren durante las primeras divisiones antes de la divergencia entre los linajes de los tejidos de origen del tumor y de la muestra normal están compartidas entre las dos muestras, y por tanto no se identifican como mutaciones específicas del tumor. El resto de mutaciones que han ocurrido hasta el ancestro común más reciente (MRCA según siglas en inglés) son mutaciones clonales en el tumor. Mutaciones que han ocurrido en células del tumor después de la última expansión clonal completa no están en todas las células tumorales, sólo en un subconjunto. Aquellas que forman parte de un clon suficientemente grande se pueden observar como mutaciones subclonales.

Cuando secuenciamos el ADN de un tumor leemos cada posición del genoma múltiples veces, por ejemplo 100 veces (100X). Las mutaciones que vemos en estas 100 lecturas y el número de lecturas en que las vemos nos da información muy valiosa sobre la historia del linaje celular que ha dado lugar al tumor. De alguna manera nos proporciona una ventana al pasado. Por un lado, vemos las mutaciones que tenía la célula ancestral común más reciente, es decir vemos las mutaciones de una célula única que vivió en el pasado, se expandió y dio origen a todas las células actuales del tumor. Como cada célula tiene dos copias del genoma y la mayoría de mutaciones se encuentran sólo en una de las dos copias, las mutaciones de esta célula ancestral las observamos aproximadamente en la mitad de las lecturas (frecuencia alélica alrededor de 0.5). Por otro lado, también podemos ver mutaciones que ocurrieron en una célula descendiente de la célula ancestral común, y que propició una expansión clonal lo bastante grande. Por ejemplo, si una célula dio origen al 30% de las células actuales del tumor, veremos sus mutaciones en aproximadamente el 15% de las lecturas (frecuencia alélica alrededor de 0.15). Mutaciones más recientes, que son exclusivas de una célula o compartidas por unas pocas células del tumor, pueden pasar inadvertidas y, en todo caso, son difíciles de distinguir de los errores de secuenciación. Al umbral de frecuencia alélica que permite diferenciar con alta fiabilidad las mutaciones verdaderas de los errores de secuenciación lo denominamos “límite de detección” (Figura 4). Se estima que la tasa de error de una secuenciación estándar está alrededor de 0.1% (un error cada 1000 bases leídas). Si secuenciamos 3200 millones de bases, encontraríamos unos 3.2 millones de errores, un número mucho más alto que el número real de mutaciones en un genoma. Si secuenciamos todo el genoma a una cobertura de 100X encontraríamos 320 millones de errores.

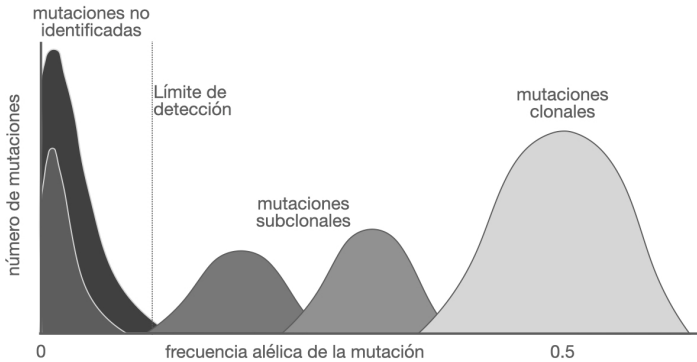


Figura 4. Ejemplo de distribución de mutaciones a distinta frecuencia alélica en un tumor. La frecuencia alélica de la mutación es la proporción de lecturas en la que vemos la mutación. En cada tumor el número de mutaciones y la distribución de frecuencias alélicas es distinta

Para evitar estos errores, tomamos como válidas sólo mutaciones que hemos visto en múltiples lecturas. Por ejemplo, en una secuenciación a 100X, las mutaciones clonales tendrán el soporte de aproximadamente 50 lecturas, así que difícilmente las confundiremos con errores. Las mutaciones en un subclón que ocupe el 30% de la población tumoral tendrán soporte de unas 15 lecturas. Pero aquellas variantes para las que tengamos soporte de menos de 5 o 3 lecturas podrían ser errores, y por eso normalmente las ignoramos y decimos que están por debajo del límite de detección (Figura 4).

Es importante destacar, que la secuenciación de un tumor nos da una ventana al pasado, pero en cambio no nos permite ver mutaciones ocurridas recientemente.

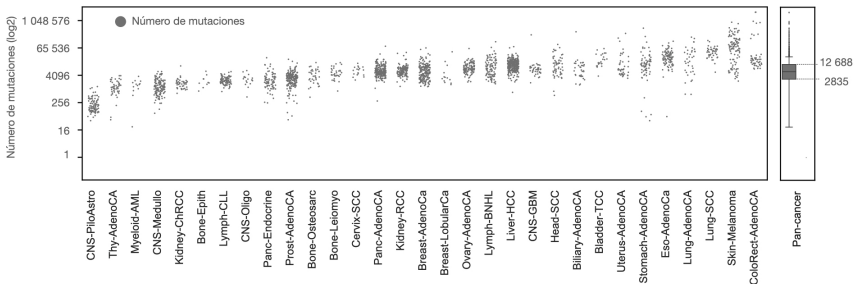


Figura 5. Número de mutaciones identificadas en el genoma completo en más de 2500 tumores de distintos tipos de cáncer. Figura adaptada de44. Las abreviaturas de los tipos de cáncer son las siguientes: Biliary-AdenoCA, adenocarcinoma biliar; Bladder-TCC, carcinoma de células transicionales de vejiga; Bone-Epith, neoplasia ósea, epitelioide; Bone-Leiomyo, Leiomioma óseo; Bone-Osteosarc, sarcoma óseo; Breast-AdenoCA, adenocarcinoma de mama; Cervix-SCC, carcinoma de células escamosas de cérvix; CNS-GBM, glioblastoma del sistema nervioso central; CNS-Oligo, oligodendroglioma del sistema nervioso central; CNS-Medullo, meduloblastoma del sistema nervioso central; CNS-PiloAstro, astrocitoma pilocítico del sistema nervioso central; ColoRect-AdenoCA, adenocarcinoma colorrectal; Eso-AdenoCA, adenocarcinoma de esófago; Head-SCC, carcinoma de células escamosas de cabeza y cuello; Kidney-ChRCC, carcinoma renal de células cromóforas; Kidney-RCC, carcinoma renal de células renales; Liver-HCC, carcinoma hepatocelular de hígado; Lung-AdenoCA, adenocarcinoma de pulmón; Lung-SCC, carcinoma de células escamosas de pulmón; Lymph-CLL, leucemia linfocítica crónica; Lymph-BNHL, linfoma de células B maduras; Lymph-NOS, linfoma no especificado de células linfoides; Myeloid-AML, leucemia mieloide aguda; Ovary-AdenoCA, adenocarcinoma de ovario; Panc-AdenoCA, adenocarcinoma pancreático; Panc-Endocrine, tumor neuroendocrino pancreático; Prost-AdenoCA, adenocarcinoma de próstata; Skin-Melanoma, melanoma cutáneo; Stomach-AdenoCA, adenocarcinoma gástrico; Thy-AdenoCA, adenocarcinoma tiroideo de bajo grado; y Uterus-AdenoCA, adenocarcinoma de útero.

De las miles de mutaciones que contiene el genoma de un tumor (Figura 5), sólo unas pocas (menos de una decena), son las causantes del cáncer. Como he comentado, a estas las llamamos mutaciones *driver* del cáncer. Al

resto de mutaciones, la gran mayoría, que no están implicadas en el proceso tumoral las llamamos mutaciones pasajeras, porque acompañan a las mutaciones causantes del cáncer en la célula que inicia la expansión clonal, pero no tienen ningún rol en ésta⁴³.

El patrón de mutaciones del tumor contiene trazas de toda la historia de su linaje celular. Esto incluye tanto trazas de los procesos mutacionales a los que ha estado sujeto, como de las distintas olas de expansión clonal causadas por mutaciones *drivers*. Los datos de miles de genomas acumulados durante los últimos 15 años proporcionan una oportunidad única para entender cómo ocurren las mutaciones en cada uno de nuestros tejidos, e identificar los genes y las mutaciones causantes del cáncer.

VARIACIÓN: ¿CÓMO OCURREN LAS MUTACIONES?

9. Daño, errores y reparación del ADN

Las mutaciones se producen diariamente en las células de nuestros tejidos debido a diversos procesos mutacionales. Estos incluyen lesiones en el ADN causadas por factores externos, como la radiación ultravioleta del sol o los carcinógenos presentes en el tabaco. También se generan por factores internos, como reacciones químicas que ocurren en nuestras células, por ejemplo la hidrólisis o la oxidación. Además, las mutaciones pueden originarse por errores durante la replicación del ADN.

Nuestras células están armadas con un conjunto de proteínas encargadas de la reparación del ADN. Estas supervisan continuamente el ADN para identificar lesiones y errores de la replicación, para repararlos antes de que se fijen como mutaciones somáticas. Por ejemplo la vía de reparación de bases incorrectamente apareadas (*Mismatch Repair* en inglés) se encarga mayoritariamente de reparar errores durante la replicación; en concreto identifica pares de bases que no siguen la complementariedad de Watson y Crick, es decir A-T y C-G, a los que llamamos *mismatches* en inglés. Por otro lado, la vía de reparación por escisión de nucleótidos puede reconocer y reparar entre otras las lesiones causadas por la luz ultravioleta en el ADN. La reparación por escisión de bases puede reparar, entre otras, las bases oxidadas.

La mayoría de estas vías de reparación no son en absoluto exclusivas de los humanos, sino que son compartidas entre la mayoría de organismo vivos, incluso bacterias. Desde el punto de vista evolutivo, cierta tasa de mutación puede ser tolerada, incluso favorecida, por los mecanismos de selección darwiniana, dado que es totalmente necesaria para generar variación. Pero una tasa de mutación demasiado alta es perjudicial. Como resultado de la evolución durante millones de años de los organismos de la Tierra, ha aparecido este arsenal de vías de reparación del ADN para mantener la tasa de mutaciones adecuada compatible con la vida y a la vez permitir cierta variación entre células e individuos, como motor de la evolución.

Estas vías de reparación funcionan también en las células de nuestra línea germinal, evitando que se transmitan demasiadas mutaciones a la siguiente generación. Se calcula que en cada generación, en condiciones normales, se acumulan unas 50 mutaciones nuevas en el genoma. En las células somáticas de nuestro organismo la continua reparación del ADN

evita que se acumulen demasiadas mutaciones en nuestros tejidos a medida que nos vamos haciendo mayores. Por ejemplo, en las células de la sangre se calcula que se acumulan unas 14 y 21 mutaciones en el genoma de cada célula cada año^{45,94}, de manera que a lo largo de nuestra vida se acumulan cientos de mutaciones en cada célula de la sangre. Las células del epitelio del colon son de las que tienen una tasa de división mayor y también una tasa de mutación mayor, acumulando unas 50 mutaciones por año⁴⁶. En un adulto, por tanto, fácilmente se habrán acumulado aproximadamente unas 2000 mutaciones por célula en el colon.

10. Firmas mutacionales

Cada uno de estos procesos mutacionales puede crear un tipo de lesiones muy concretas. Por ejemplo la luz ultravioleta causa lesiones preferentemente en dímeros de pirimidinas, en concreto CC y CT. Cuando estas lesiones no son reparadas a tiempo, en el siguiente ciclo de replicación pueden generar una mutación de una C a una T. De manera que las mutaciones causadas por la luz ultravioleta tienen un patrón muy concreto, son Cs que mutan a Ts en el contexto de CC o CT.

La desaminación espontánea de las citosinas metiladas es un potente proceso mutacional, que cambia Cs a Ts cuando la C está metilada. Como la metilación ocurre preferencialmente en CGs, este proceso mutacional también tiene un patrón muy concreto, son cambios de C a T mayoritariamente en CGs.

Estos son algunos de los ejemplos más conocidos y mejor estudiados, pero de aquí se puede extrapolar la idea de que cada proceso mutacional produce cambios con una preferencia por ciertos tipos de mutaciones en determinados di o trinucleótidos. Este fue el razonamiento que hicieron Mike Stratton, Ludmil Alexandrov, Serena Nick-Zainal y otros científicos para generar un catálogo de firmas mutacionales primero en tumores de cáncer de mama⁴⁷, que seguidamente extendieron a tumores originados en múltiples tejidos^{48,49}.

Para generar este catálogo se requiere una colección cuanto más amplia mejor de mutaciones somáticas observadas en tumores. Las sustituciones se pueden agrupar en 6 clases, C>A, C>G, C>T, T>A, T>C, T>G. Como una vez una mutación está fijada no podemos saber si ocurrió una sustitución de una C>T o su base complementaria, una G>A, por convención las anotamos siempre a partir de la pirimidina (Cs y Ts). Si incorporamos tam-

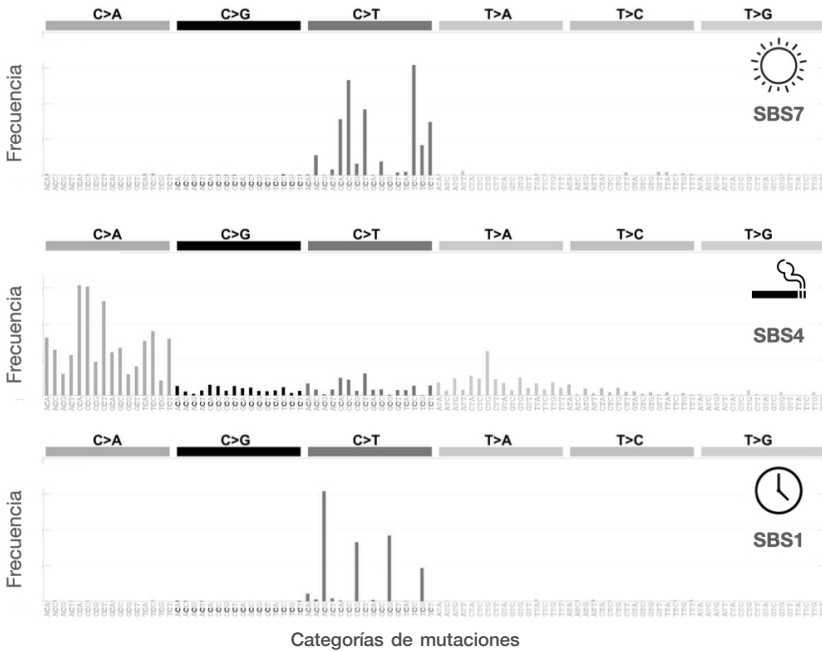


Figura 6. Ejemplos de firmas mutacionales, mostrando la frecuencia de cada una de las 96 categorías de mutaciones. Las categorías de mutaciones consideran el tipo de sustitución y el tipo de nucleótido presente antes y después de la sustitución.

bién el contexto inmediato, es decir la bases adyacentes al nucleótido mutado, hay 96 categorías distintas de mutaciones (Figura 6). El razonamiento que aplican Mike Stratton y su equipo es el siguiente: asumiendo que el recuento de mutaciones de cada tipo en cada muestra se pudiera explicar como la superposición de un número limitado de perfiles mutacionales, mediante algoritmos de factorización positiva de matrices se podrían inferir los patrones o firmas mutacionales que mejor explicarían la variabilidad del conjunto de mutaciones en los tumores estudiados. El primer repertorio publicado en 2013 contenía 22 firmas mutacionales. Algunas de ellas recapitulan el conocimiento previo sobre procesos mutacionales y patrones de mutaciones (Figura 6). Por ejemplo la firma 7 corresponde a la mutagénesis causada por la luz del sol. La firma 4 corresponde a la mutagénesis por tabaco, que ya se conocía previamente por generar mayoritariamente cambios de C a A. La firma 1 corresponde al patrón esperado por la desaminación espontánea de las citosinas. Esta firma es especialmente interesante ya que es una firma ubicua, es decir, se encuentra en todos los tumores y tejidos de todas las personas, animales y otros organismos. El número de mutaciones de este tipo, junto con el de otra firma para la que no se conoce su

etiología (firma 5), se correlaciona fuertemente con la edad, es decir, va incrementando de manera lineal con la edad de la persona, así que la podemos utilizar como un reloj que nos marca el tiempo desde la primera división del cigoto (*clock signature* o firma reloj). Este proceso mutacional también está activo en la línea germinal, con lo cual mutaciones con este patrón se transmiten a las siguientes generaciones.

En el repertorio actual hay más de 100 firmas mutacionales catalogadas. Para muchas de ellas no se conoce la etiología concreta. En ocasiones sabemos que ocurre mayoritariamente en un tipo de tejido o en otro, o en tumores con una determinada alteración genética o en personas con una cierta exposición a algún compuesto. Pero aún faltan muchos estudios y experimentos para caracterizar cada uno de estos procesos mutacionales en detalle. Lo que es cierto es que esta manera de estudiar los procesos mutacionales a partir de datos de miles de tumores nos proporciona un tipo de información de incalculable valor para entender la mutagénesis en nuestros tejidos desde la primera división del cigoto.

Aparte de obtener firmas mutacionales de sustituciones de bases, se han aplicado aproximaciones matemáticas similares para la obtención de firmas mutacionales de otros tipos de alteraciones en el genoma, como inserciones y deleciones de pocas bases, variaciones estructurales, y cambios en el número de copia de fragmentos genómicos⁴⁹⁻⁵¹.

11. Variación de la tasa de mutaciones en el genoma

Además de evidenciar que los distintos procesos mutacionales tienen preferencia para generar ciertos tipos de mutaciones en ciertos contextos, el estudio de mutaciones también ha desvelado que la tasa de mutaciones es muy variable a lo largo del genoma. Es decir hay regiones del genoma que acumulan gran número de mutaciones, mientras en otras la tasa de mutaciones es mucho menor (Figura 7). Estas diferencias van ligadas a cómo el genoma se replica, se empaqueta, se transcribe, y se repara.

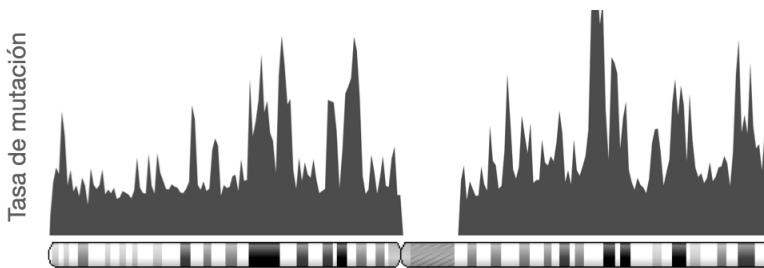


Figura 7. Tasa de mutación somática en cáncer de colon por megabase en el cromosoma 1.

a. De cómo la replicación afecta a la tasa de mutaciones

Antes de dividirse, cada célula debe replicar todo su genoma, es decir, los 3200 millones de pares de bases, para garantizar que las células hijas contengan toda la información en su núcleo. Este proceso se llama replicación del ADN. El inicio de la replicación del ADN ocurre simultáneamente en varios puntos, más o menos fijos, a lo largo de la doble cadena de ADN, los llamamos orígenes de replicación. Las regiones cercanas a estos orígenes de replicación se sintetizan antes, y acumulan menor número de mutaciones, mientras que las que están más alejadas se replican más tarde, y acumulan un mayor número de mutaciones^{52,53}. Podemos especular que esto puede estar relacionado con mayor tiempo para la reparación de errores cometidos durante la replicación en aquellos fragmentos que se replican tempranamente. Aunque otros factores, como la disponibilidad de nucleótidos libres para la síntesis, pueden también estar implicados en esta correlación.

En cada uno de estos puntos de inicio de la replicación, la doble helice del ADN se abre como una cremallera y las dos cadenas se replican utilizando mecanismos ligeramente distintos. Esta asimetría se ve reflejada en una tasa de mutaciones diferencial entre las dos cadenas en algunos procesos mutacionales⁴⁹.

b. Empaquetando 2 metros de ADN en un espacio de 6 micrómetros

Una hélice de ADN tiene 3200 millones de pares de bases agrupados en 23 cromosomas. Cada nucleótido en la doble cadena ocupa aproximadamente 3,4 Å (es decir 3,4E-10 metros)⁵⁴. Una simple multiplicación entre estos dos números nos revela que un genoma humano entero, si lo estiráramos, ocuparía aproximadamente un metro. En cada célula de nuestro cuerpo guardamos 2 genomas, uno que heredamos de la madre y otro del padre, por lo tanto, tenemos unos 2 metros de unas finísimas moléculas de ADN. Estos están empaquetados en el núcleo de la célula, un espacio minúsculo de unos 6 micrómetros de diámetro. Un reto importante de la biología es entender cómo se logra empaquetar 2 metros de ADN en este espacio tan minúsculo, y aun así mantener su funcionalidad y accesibilidad a las distintas regiones del genoma para hacer todas sus actividades: replicar, transcribir y reparar.

La unidad básica de este empaquetado son los nucleosomas. Unos 146 nucleótidos de la cadena de ADN se envuelven alrededor de unas proteínas llamadas histonas, dando más de una vuelta y media al núcleo de histonas.

A continuación, hay una parte de la cadena de ADN no envuelta alrededor de histonas llamada región *linker* que a su vez es seguida por otro nucleosoma.

Los nucleosomas se empaquetan entre ellos, uno encima del otro, para formar una fibra altamente condensada llamada cromatina. Distintas regiones del genoma están más o menos empaquetadas, en parte dependiendo de en qué medida la célula requiera acceder a la información de la región. El grado de compactación del genoma tiene una relación directa con la tasa de mutaciones. En general, aquellas regiones más compactadas acumulan más mutaciones que las regiones más accesibles. Esto ocurre tanto para las mutaciones germinales como para las mutaciones somáticas. Varios estudios han asociado la tasa de mutaciones con determinadas modificaciones de histonas que regulan la accesibilidad al ADN^{52,55,56}. Para poder reparar lesiones que ocurren en el ADN, es necesario que las proteínas encargadas de esta reparación puedan acceder a la zona dañada. Aquellas zonas con alta compactación son menos accesibles a los mecanismos de reparación lo que hace que finalmente acumulen más mutaciones. Aparte de la accesibilidad, también hay un rol más activo de ciertas modificaciones de histonas para reclutar proteínas encargadas de la reparación del ADN, como veremos por ejemplo en la sección 12c. Esto también contribuye a la correlación entre las características de la cromatina y la tasa de mutaciones.

En resumen, la accesibilidad a las diferentes regiones del genoma debido a los distintos grados de compactación de la cromatina, junto con roles activos de las modificaciones de histonas, explican parte de la variabilidad en la tasa de mutaciones a lo largo del genoma.

c. Transcribiendo y reparando

La transcripción es una de las acciones más importantes que ocurren en el ADN. Cada vez que se requieren proteínas concretas para el funcionamiento de una célula, se debe acceder a la información del gen en el ADN que codifica esta proteína y transcribir la información del ADN al ARN mensajero para que esta información viaje hasta los ribosomas, unas “fábricas” de proteínas que sintetizan nuevas copias de la proteína en cuestión.

Hay procesos de reparación del ADN ligados a la transcripción. La transcripción de un gen crea la posibilidad de que su secuencia sea esca-

nada por el propio complejo proteico que la transcribe para identificar posibles lesiones en el ADN. Si hay alguna lesión, se detiene la transcripción y se reclutan proteínas encargadas de repararla. Este proceso se llama reparación por escisión de nucleótidos acoplada a la transcripción, y es capaz de reparar lesiones como los dímeros de pirimidinas modificados por la luz ultravioleta del sol. Visto a través de la lente evolutiva esto tiene sentido: millones de años de evolución han resultado en mecanismos para proteger a las regiones que contienen la información más valiosa, los genes, de la acumulación excesiva de mutaciones. Aquellos genes que se expresan más se reparan más eficientemente^{53,57}. Así, se observa una clara relación inversa entre la tasa de mutaciones y el nivel de expresión de los genes.

En el momento de la transcripción, la doble cadena del ADN se abre como una cremallera, de manera similar a como lo hace para la replicación. Una de las dos hebras de la doble cadena de ADN es la que contiene la información para codificar el gen, y sirve de molde para copiar la información en el ARN. Es en esta cadena en la que se detectan y reparan de manera más efectiva las lesiones en los nucleótidos. Como resultado, se acumula un menor número de mutaciones en la cadena molde transcrita que en la no transcrita. Esto se conoce como sesgo de hebra transcripcional (*transcriptional strand bias*) en el patrón de mutaciones, y es una clara indicación de reparación asociada a la transcripción en ese gen.

12. Variación local en la tasa de mutación

En la sección anterior hemos visto que la tasa de mutación varía en distintas regiones del genoma, y cómo estas variaciones van ligadas a la forma en la que el genoma se replica, se empaqueta, se transcribe, y se repara. Gracias a los datos de miles de tumores de pacientes con cáncer compartidos por la comunidad científica, es posible dilucidar esta tasa de mutación, así como sus determinantes, en regiones relativamente grandes del genoma. Por ejemplo, partiendo el genoma en fragmentos relativamente grandes (por ejemplo, una megabase, es decir un millón de nucleótidos), podemos medir la tasa de mutaciones en cada partición (Figura 8) y estudiar la correlación entre la tasa de mutaciones y las estructuras o funciones diferenciales del genoma en dichos fragmentos.

Es lógico pensar que puede haber variabilidad en la tasa de mutaciones a escalas más pequeñas, lo que llamamos variación local. Estas son más difíciles de estudiar dada la relativamente baja densidad de mutaciones en

relación al total de tumores de pacientes secuenciados hasta el momento. En una región concreta del genoma de unas 100 o 200 bases habremos observado ninguna, una, o muy pocas mutaciones, con lo que no se puede obtener una tasa fiable por región. Una alternativa sencilla es agrupar varias regiones con características funcionales similares para estimar la tasa de mutación en ese tipo de regiones, por ejemplo, exones o regiones de unión a factores de transcripción. En mi grupo durante los últimos 7-8 años hemos abordado el estudio de esta variación local en la tasa de mutaciones⁵⁸. Hemos descubierto cómo ciertas estructuras locales del genoma interfieren con la reparación del ADN, modificando la tasa de mutación en regiones muy concretas.

a. Factores de transcripción

En 2015 iniciamos un proyecto en el laboratorio para identificar mutaciones driver en regiones no codificantes (como explicaré en la sección 18). Empezamos estudiando las mutaciones en melanomas, un tipo de cáncer de piel. Al analizar las mutaciones en las zonas reguladoras cerca del inicio de la transcripción (llamadas promotores) nos encontramos con un resultado inesperado. Este resultado nos hizo sospechar de un incremento en la tasa de mutación en regiones reguladoras de la expresión génica que no se había observado hasta el momento. Como estas regiones se distinguen porque a ellas se unen proteínas reguladoras de la expresión génica, como factores de transcripción, formulamos como hipótesis que la unión de dichas proteínas al ADN afectaba de alguna manera a la mutagénesis en estas regiones. Para estudiarlo en detalle recopilamos regiones del genoma con evidencias de contener sitios de unión a factores de transcripción. Los sitios de unión de estas proteínas en el ADN reciben el nombre en inglés de *Transcription Factor Binding sites* (TFBS). Alineamos los TFBS del genoma de las células en que se origina el melanoma, melanocitos, según su posición central, y contamos las mutaciones observadas en cada posición de la alineación.

Observamos, tal como habíamos hipotetizado, un incremento de más de 5 veces en la tasa de mutaciones en las regiones del TFBS comparado con las regiones colindantes⁵⁹. Para ver si este incremento se debía a diferencias en la composición de la secuencia en TFBS, calculamos la tasa de mutaciones esperable teniendo en cuenta la composición de la secuencia y los procesos mutacionales específicos de las muestras de melanoma, confirmando que el incremento no se debía a la secuencia de los TFBS (Figura 8a).

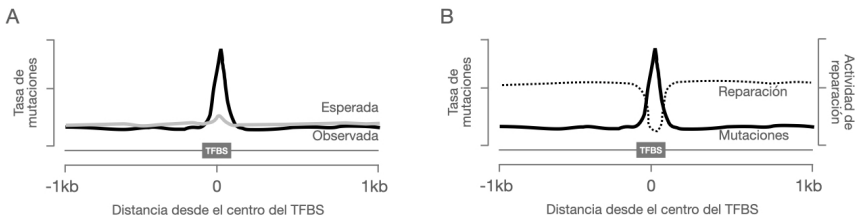


Figura 8. A. Dibujo esquemático que muestra el incremento en la tasa de mutaciones en sitios de unión a factores de transcripción (TFBS). La línea negra indica la tasa de mutaciones observada. La línea gris indica la tasa de mutaciones esperadas según la composición de la secuencia. **B.** Dibujo esquemático mostrando la disminución de la actividad de reparación en TFBS.

Este incremento de mutaciones en TFBS podría ser debido tanto a un incremento de las lesiones en el ADN como a una disminución en la reparación de estas lesiones en estas regiones. La mayoría de las mutaciones en melanomas están causadas por la luz ultravioleta, por la exposición de la piel al sol. Estas lesiones son reparadas por la vía de reparación de escisión de nucleótidos, y de hecho justo cuando nosotros observamos este incremento de mutaciones se publicó un mapa de la actividad de reparación por escisión de nucleótidos a lo largo del genoma⁶⁰ del laboratorio de Aziz Sancar. Utilizando los datos de este mapa y centrándolos sobre las secuencias con TFBS observamos una clara disminución de la actividad de reparación en los sitios de unión a factores de transcripción (Figura 8b). Ese mismo año el profesor Aziz Sancar recibió el premio Nobel de química por su contribución a elucidar los mecanismos de reparación de las lesiones en el ADN causadas por la luz ultravioleta.

Aparte de la observación del incremento de la tasa de mutación en TFBS en promotores, también observamos un gran incremento en otros sitios de unión de proteínas al ADN, como los sitios de unión de la proteína CTCF. CTCF es una proteína muy importante y altamente conservada que regula la estructura tridimensional de la cromatina. Puede unir dos cadenas de ADN formando grandes lazos. En estas regiones vimos que la tasa de mutación está altamente incrementada y la actividad de reparación disminuida. Estudios posteriores, incluyendo estudios de nuestro propio grupo, han confirmado estos resultados y han evidenciado un panorama aún más complejo, en que la unión de algunas proteínas al ADN, no sólo interfieren con la reparación, sino que también afecta a la probabilidad de crear lesiones por la luz ultravioleta en los dímeros de pirimidina⁶¹⁻⁶³.

También observamos que la tasa de mutaciones mostraba una periodicidad de aproximadamente 170 pares de bases. Nos dimos cuenta de que los máximos de esta función periódica coincidían con las posiciones en las que el ADN se envuelve alrededor de las histonas. Es decir, las regiones del ADN cubiertas por nucleosomas reciben más mutaciones que las regiones entre nucleosomas (regiones *linker*). En la siguiente sección entro en más detalle en esta cuestión.

En resumen, en este trabajo demostramos que las proteínas unidas al ADN interfieren con la maquinaria de reparación por escisión de nucleótidos, causando un incremento local en la tasa de mutaciones en estas regiones⁵⁹.

b. Nucleosomas

Como he explicado anteriormente, los nucleosomas son la unidad básica del empaquetamiento de la cromatina. Unos 146 nucleótidos de la cadena de ADN se enrollan alrededor de las histonas, dando 1.65 vueltas alrededor de ellas. Entre dos nucleosomas contiguos, hay una región libre de histonas, llamada *linker*, seguida de otro nucleosoma. La mayor parte del genoma humano (un 75-90%) está cubierto por nucleosomas.

La mayoría de las moléculas de ADN tienen una tendencia natural a plegarse alrededor del nucleosoma. En general los nucleosomas en el genoma humano no están siempre posicionados en las mismas regiones, sino que su posicionamiento es dinámico. Pero sí hay ciertas regiones del genoma donde los nucleosomas se posicionan de manera fija y conservada (*well-positioned nucleosomes*). Se ha observado una clara tendencia a que el ADN plegado alrededor de las histonas en estas posiciones conservadas tenga unas características determinadas en su secuencia. En particular, las zonas en que el surco menor del ADN se encuentra dirigido hacia el nucleosoma tienden a ser ricas en As y Ts, mientras que las regiones en que el surco menor apunta hacia fuera del nucleosoma tienden a ser ricas en Gs y Cs, generando así una periodicidad de As y Ts cada 10.3 pares de bases (una vuelta de la hélice del ADN). A esta periodicidad de secuencia se le ha llamado “periodicidad WW”. Poco después de la secuenciación del genoma de levadura ya se observó esta periodicidad y se asoció al ADN enrollado alrededor de las histonas⁶⁴. Durante mi estancia postdoctoral en el Instituto Europeo de Bioinformática compartía grupo y conversaciones con Benjamin Audit, un físico francés que utilizaba técnicas de reconocimiento de patrones en los primeros genomas

secuenciados de animales y plantas. Él y otros investigadores confirmaron que esta periodicidad WW se observa en el genoma de múltiples especies eucariotas y que va ligada a la presencia de nucleosomas bien posicionados⁶⁵. La interpretación que se ha dado a esta observación es que el surco menor del ADN en secuencias ricas en A-T es más susceptible a la torsión que el correspondiente a secuencias ricas en G-C, y esta periodicidad habría pues aparecido por presión selectiva al favorecer el posicionamiento de nucleosomas claves en el genoma. Nuestros hallazgos sobre la efectividad en la reparación del ADN en distintas regiones de la molécula que rodea el nucleosoma, como veremos a continuación, nos permitieron proponer una explicación muy distinta al origen de esta periodicidad WW.

En el estudio de las mutaciones de melanomas ya habíamos observado una periodicidad en la tasa de mutaciones entre cromosomas y *linkers*. Quisimos indagar más en el efecto de los nucleosomas en la tasa de mutaciones también en otros tejidos. Confirmamos que este incremento de mutaciones en nucleosomas no sólo ocurre en mutaciones causadas por la luz ultravioleta, sino que también se observa en tejidos sometidos a otros procesos mutacionales (Figura 9)⁶⁶.

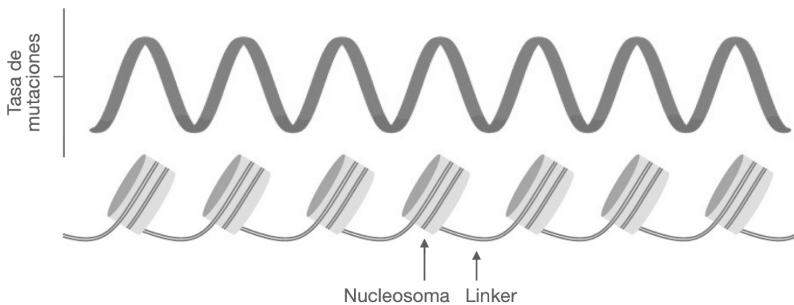


Figura 9. Dibujo esquemático mostrando la periodicidad en la tasa de mutación siguiendo la presencia de nucleosomas y linkers.

Observando detalladamente las gráficas de la tasa de mutación en nucleosomas nos percatamos de un curioso patrón en la región del ADN que envuelve cada nucleosoma. Nos dimos cuenta de que hay otra periodicidad, aunque con una amplitud menor, en el patrón de mutaciones, en este caso de aproximadamente 10 pares de bases. Inmediatamente me hizo pensar en la periodicidad WW en la secuencia de eucariotas que observó mi compañero Benjamin y nos pusimos a estudiarlo en detalle (Figura 10).

Demostramos que la manera como el ADN interacciona con las histonas tiene un efecto importante en cómo este recibe lesiones y en cómo se repara, generando un patrón de tasa de mutaciones periódico, con un periodo promedio de 10.3 pares de bases⁶⁶. Además, encontramos que la polaridad de este periodo depende del proceso mutacional. Por ejemplo, la periodicidad de las mutaciones causadas por la luz ultravioleta tiene su máximo en las regiones en las que el surco menor se orienta hacia afuera del nucleosoma, y su mínimo en las regiones en que se orienta hacia las histonas. Pudimos demostrar que esto se debe a un incremento en la tasa de lesiones causadas por luz ultravioleta en el surco menor cuando el mismo se orienta hacia afuera. Creemos que la diferente conformación de los grupos químicos que sobresalen de cada dipirimidina cuando el surco menor se orienta hacia afuera o hacia dentro (debido a la diferente torsión que conllevan ambas conformaciones) es la causa de la diferencia observada en la tasa de generación de lesiones. Específicamente se sabe que el ángulo que forman estos grupos químicos en estas dos conformaciones es distinto: esto afecta directamente a su tendencia a formar una lesión por luz ultravioleta.

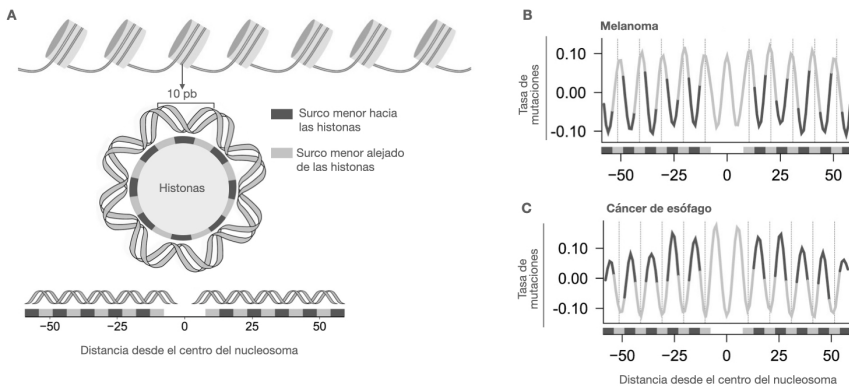


Figura 10. **A.** Dibujo esquemático del ADN que envuelve las histonas en un nucleosoma. El surco menor del ADN está encarado hacia las histonas, o se aleja de las histonas, cada 10 pares de bases (pb) aproximadamente, una vuelta de la estructura del ADN. **B.** Tasa de mutaciones observadas en melanoma en el ADN que envuelve las histonas, centrado en el centro del nucleosoma (diada). Observamos un mayor número de mutaciones cuando el surco menor está orientado hacia afuera de las histonas. **C.** En cambio en el esófago observamos una señal periódica con el mismo periodo (10.3pb) pero completamente invertida, es decir, con una tasa de mutaciones más alta cuando el surco menor está orientado hacia las histonas.

En otros tumores, como los de esófago, y en otros procesos mutacionales observamos una señal periódica con el mismo periodo (10.3pb) pero completamente invertida, es decir con el máximo en las regiones en que el

surco menor está orientado hacia las histonas. A continuación, analizamos mapas de actividad de la reparación de algunos tipos de daño en el ADN, como los que he mencionado en la sección anterior, y demostramos que esta tasa diferencial de mutaciones es debida a una reducción en la accesibilidad de las proteínas de reparación a la región dañada cuando el surco se orienta hacia las histonas. Esto provoca que se reduzca la eficiencia de la reparación del daño en estas regiones. Por lo tanto se acumula un mayor número de mutaciones en estas regiones.

Esta periodicidad la observamos inicialmente en mutaciones somáticas en tumores. Pero razonamos que como los mismos procesos de reparación actúan también en células de la línea germinal, podría del mismo modo observarse en las variantes transmitidas a la siguiente generación. Para esto estudiamos la distribución de las tasas de variantes germinales comunes entre poblaciones humanas y de una planta gramínea (*Arabidopsis thaliana*) en el ADN que cubre los nucleosomas. Estas variantes se distribuyen siguiendo la misma periodicidad de 10.3pb, con los máximos en las regiones menos accesibles a la reparación, es decir, cuando el surco menor se enfrenta a las histonas. La misma periodicidad se observa en las variantes fijadas en la evolución a más largo plazo, en la divergencia de la secuencia del genoma entre tríos de especies, por ejemplo, entre los humanos y otros dos primates, o entre tres especies evolutivamente cercanas de plantas gramíneas.

Estos resultados conducen a una conclusión ineludible: si el posicionamiento de los nucleosomas en el ADN provoca que la variación heredable se distribuya de una manera no uniforme, que depende de la orientación del surco menor del ADN con respecto a las histonas, es concebible que los millones de años de evolución transcurridos desde la aparición de los nucleosomas (posteriormente a la separación del tronco común de las arqueobacterias y los eucariotas de las bacterias) hayan dejado huellas en la composición del genoma de los eucariotas y que puedan ser detectadas en la actualidad (Figura 11).

La mayoría de las mutaciones que se transmiten de una generación a otra en la mayoría de las especies son C/G>T/A, debido a que los procesos mutacionales que operan en la línea germinal de los mamíferos, pero también en levaduras y otras especies, están dominados por la firma 1 causada por la desaminación espontánea de la citosina y la 5-metilcitosina. Si la mayoría de mutaciones tienden a modificar C/Gs a T/As, el genoma con el tiempo iría incrementando su composición de A/Ts en detrimento de la

proporción de C/Gs. Esto es así, la mayoría de los genomas eucariotas son ricos en A/Ts. A la vez, los resultados obtenidos hasta este momento nos condujeron a plantear que aquellas regiones del ADN que se reparan con menos eficiencia y por tanto reciben más mutaciones, con el tiempo tendrán una mayor tendencia a enriquecerse, de manera relativa, en A/Ts. Esto es lo que habría ocurrido en las regiones con nucleosomas bien posicionados. La mayor tasa de mutaciones cuando el surco menor se enfrenta a las histonas, sostenida durante millones de años, ha dejado una marca en el genoma en que cada 10.3pb hay una proporción mayor de A/Ts. Esta huella es la periodicidad WW.

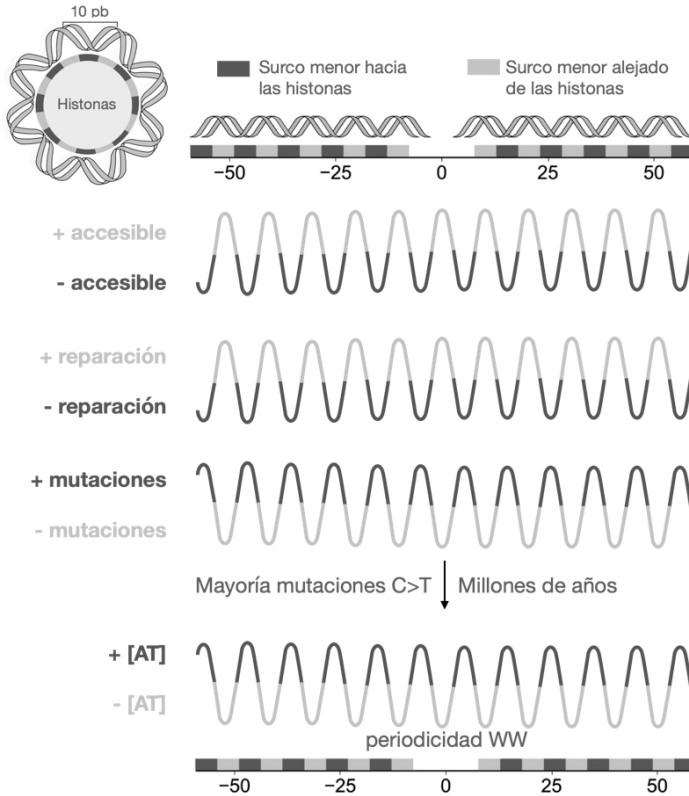


Figura 11. Creación y mantenimiento de la periodicidad WW de 10 pb en eucariotas. Las regiones de ADN en que el surco menor está orientado hacia las histonas tienen menor accesibilidad que cuando el surco menor está orientado hacia afuera. Esta diferencia en accesibilidad produce una tasa diferencial de reparación del ADN, que a la vez produce una tasa diferencial de mutaciones cada 10 pb. Las regiones menos accesibles, se reparan menos y por tanto reciben más mutaciones. Como las mutaciones que se transmiten de una generación a otra son mayoritariamente C>T, aquellas regiones con más mutaciones tenderán a tener más Ts y As que las regiones con menor tasa de mutación, generando así la periodicidad de 10 pb en el contenido de ATs (periodicidad WW).

Aunque esta periodicidad WW en los genomas eucariotas se ha relacionado con la curvatura favorable del ADN alrededor de los nucleosomas, nuestros resultados apuntan a que la generación y el mantenimiento de la periodicidad WW pueden explicarse por esta periodicidad en la eficiencia de reparación del ADN que causa una tasa diferencial de mutaciones (mayoritariamente Cs a Ts) cada 10.3pb.

A mi me fascina que observaciones hechas inicialmente en mutaciones somáticas en un grupo de tumores humanos, nos haya llevado a arrojar luz sobre un fenómeno de la composición del genoma que está presente en todos los organismos eucariotas. Pensar en los efectos de estas diferencias en la reparación del ADN a escala evolutiva nos ayuda a entender la evolución de los genomas de eucariotas.

c. De exones e intrones

La mayoría de los genes en el genoma humano no están codificados de manera continua en la secuencia de nucleótidos, sino que tienen espacios de regiones no codificantes (intrones) en medio de regiones codificantes (exones). Cuando la secuencia de un gen se transcribe a ARN, se transcriben también los intrones, pero a continuación ocurre un proceso de corte y empalme (*splicing*) para eliminar los intrones y así generar una secuencia continua de ARN que contiene toda la secuencia codificante de proteína sin interrupciones. Los nucleosomas en exones e intrones están marcados de manera distinta. En particular, los nucleosomas en los exones tienen niveles más altos de trimetilación en la lisina 36 de la histona 3 (H3K36me3). Esta marca está directamente involucrada en la reparación de *mismatches* del ADN (MMR). En concreto la proteína MutS α , que es la encargada de la identificación de *mismatches* durante la replicación del ADN, tiene un dominio, llamado PWWP, que se une directamente a la región de la histona con esta modificación⁶⁷.

Razonamos que la diferencia en esta marca de histonas podría conllevar que los exones fueran reparados más eficientemente y por tanto contuvieran menos mutaciones que lo esperado en comparación con los intrones. Para comprobar esta hipótesis nos centramos en tumores de colon que tienen una mutación en una de las proteínas encargadas de la síntesis de las cadenas de ADN en la replicación del genoma, la polimerasa epsilon (Pol ϵ). Debido a esta mutación, la polimerasa pierde la capacidad de corrección de errores y por tanto la síntesis de las cadenas del ADN deja un número de sustituciones (*mismatches*) mucho mayor que en condiciones normales. Estos tumores son

los que llamamos supermutantes, ya que adquieren hasta 50 veces más mutaciones que los tumores de colon sin esta mutación en Pol ϵ . Las mutaciones observadas en estos tumores tienen una firma mutacional muy concreta (firma 10) dominada por TCTs que cambian a TATs.

En estos tumores, las proteínas encargadas de la reparación de *mismatches* se ven desbordadas por el gran número de los mismos. Aunque probablemente reparan una gran parte de errores, los que dejan sin reparar se pueden fijar como mutaciones. Decidimos estudiar el patrón de mutaciones comparando exones y sus intrones adyacentes en tumores de colon con mutaciones en la polimerasa. Observamos una clara reducción en el número de mutaciones en los exones comparado con sus intrones adyacentes. Además, como los exones tienen una composición de nucleótidos distinta (relativamente más ricos en Cs y Gs), el número de mutaciones que esperamos en los exones es en realidad mayor que en los intrones (Figura 12). Comprobamos también que esta disminución en el número de mutaciones ocurre tanto en mutaciones sinónimas (es decir que no cambian la secuencia de aminoácidos) como en mutaciones no sinónimas. Esto indica que se trata de un problema de reparación y no de selección negativa o menor tolerancia de mutaciones en exones. Esta disminución de mutaciones en exones es general y se observa también en otros tipos de cáncer. En resumen, observamos una clara reducción en la tasa de mutación en los exones, que la asociamos a una mayor actividad de reparación de mismatches⁶⁸.

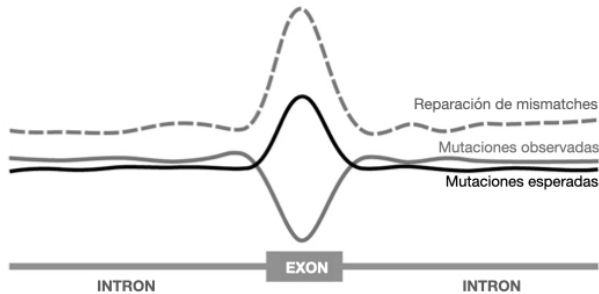


Figura 12. Dibujo esquemático mostrando la tasa de mutaciones observadas en exones e intrones adyacentes en tumores colorrectales con mutaciones en Pol ϵ . La tasa de mutación esperada se calcula teniendo en cuenta la composición de la secuencia. La línea discontinua muestra la inferencia del incremento de la actividad de reparación de mismatches.

Para comprobar que realmente se debe a la diferencia de reparación de *mismatches* entre intrones y exones, tuvimos la oportunidad de estudiar las mutaciones en tumores de niños con el síndrome de cáncer por deficiencia

bi-alélica en la reparación de *mismatches* en el ADN (bMMRD, en sus siglas en inglés). Estos niños nacen con mutaciones en los genes implicados en la reparación de *mismatches*, y no tienen la capacidad de repararlos en ninguna de las células de sus tejidos. Como consecuencia, desarrollan distintos tipos de cáncer a edades tempranas, y desgraciadamente casi nunca llegan a la edad adulta. Frecuentemente los tumores que se desarrollan en estos niños adquieren, con el tiempo, mutaciones en el gen de Pol ϵ , de manera que confluyen el gran número de errores causados por la polimerasa con la incapacidad de estas células para repararlos. Estos tumores acumulan más de 1000 veces más mutaciones que otros tumores, en algunos casos hasta 1 millón de mutaciones en un tumor. Los llamamos hipermutantes. Razonamos que si la diferencia en la tasa de mutaciones entre exones e intrones se debía exclusivamente a la reparación diferencial de *mismatches*, en los tumores con deficiencia de reparación del ADN no debería haber diferencia entre exones e intrones. Los datos de secuenciación del genoma de los tumores de estos niños nos permitieron confirmar que en ausencia de esta vía de reparación no hay ninguna diferencia entre la tasa de mutación observada y la esperada en exones e intrones, por tanto confirmamos que el factor causante es la actividad diferencial de la reparación de *mismatches*.

Es interesante plantear que la evolución ha dado con una solución para proteger mejor las regiones exónicas de los genes, que son las que en promedio contienen más bases susceptibles de alterar indirectamente la conformación final de la proteína a partir de una mutación.

13. Mutaciones causadas por la quimioterapia

La quimioterapia es el tratamiento contra el cáncer más utilizado y salva millones de vidas cada año. Aunque se están desarrollando cada vez más tratamientos de precisión para atacar específicamente las células del tumor, y la inmunoterapia está mostrando resultados muy prometedores en ciertos tipos de tumores, la quimioterapia y la cirugía siguen siendo la primera línea de tratamiento para muchos tipos de cáncer.

En general las quimioterapias interfieren de una manera u otra con los procesos de división celular. Como las células cancerosas se dividen mucho y muy rápido, estos tratamientos son eficaces contra el cáncer. Pero como son tratamientos sistémicos, es decir que llegan a todas las partes del cuerpo, pueden también afectar a otras células sanas de nuestro cuerpo que requieren dividirse. Este impacto sobre el tejido sano es lo que causa los conocidos efectos secundarios durante el tratamiento con quimioterapia.

Aparte de los efectos secundarios que se producen durante el tratamiento, hay evidencias epidemiológicas de efectos de la quimioterapia a largo plazo, incluso años después del tratamiento. Por ejemplo, las personas que han recibido quimioterapias tienen un mayor riesgo de padecer tumores secundarios (como leucemias), problemas cardiovasculares y, en general, envejecimiento prematuro. Esto es particularmente relevante para los niños tratados con quimioterapia, ya que una vez han sobrevivido al cáncer infantil gracias al tratamiento, tienen aún toda la vida por delante. Las causas moleculares de los efectos secundarios a largo plazo de la quimioterapia no se conocen en la mayoría de los casos. Tan sólo se tienen evidencias epidemiológicas, es decir, estudios estadísticos sobre la población, que demuestran que los supervivientes de cáncer tienen incidencias más altas de algunas condiciones clínicas como las mencionadas.

Algunas quimioterapias son agentes que dañan el ADN (por ejemplo el cisplatino y la temozolomida) o interfieren con su replicación (como la capecitabina). Las células que sobreviven a la quimioterapia podrían haber adquirido mutaciones como resultado de su acción sobre el ADN. Como la quimioterapia es un tratamiento sistémico, estas mutaciones podrían acumularse no sólo en el tumor, sino en los tejidos sanos del paciente. Y quizás podrían tener relación con algunos de los efectos secundarios a largo plazo de estos tratamientos. En 2017 iniciamos una línea de investigación en mi laboratorio con el objetivo de entender cómo las quimioterapias afectan a los tejidos sanos del paciente.

Nuestro razonamiento fue que, de la misma manera que otros procesos mutacionales dejan patrones específicos en el ADN (firmas mutacionales), las quimioterapias que causan mutaciones dejarían una huella distintiva en el patrón de mutaciones en los pacientes tratados. La primera oportunidad de abordar esta pregunta la tuvimos gracias a los datos que nos cedió la fundación Hartwig (*Hartwig Medical Foundation*), una entidad centrada en la secuenciación de tumores en Holanda. La mayoría de los tumores secuenciados son de pacientes con cáncer metastásico, muchos de los cuales han recibido tratamiento contra el cáncer antes de la obtención de la muestra tumoral. Además de los datos genómicos (genoma completo del tumor y de una muestra de sangre), esta cohorte está bien anotada clínicamente, incluyendo detalles de qué tratamiento recibieron los pacientes y cuánto tiempo transcurrió entre el tratamiento y la aparición de la metástasis.

Razonamos que si los tratamientos dejaban una huella mutacional específica en los genomas de las células con mutaciones, las muestras de metás-

tasis obtenidas después de tratamiento podrían ser una buena oportunidad para detectarlas. Como las mutaciones ocurren de manera particular en cada célula, para poderlas detectar es necesario que haya una expansión clonal de algunas de estas células de manera que las mutaciones del tratamiento queden por encima del límite de detección (ver Figura 4). La continua evolución del tumor, con posibles expansiones clonales, durante el tratamiento y la metástasis podría proporcionar esta ventana de oportunidad.

Analizamos el genoma de más de 3500 tumores metastásicos y extrajimos firmas mutacionales (usando un método análogo al descrito en la sección 9). Intuitivamente, esperamos que aquellas firmas mutacionales asociadas al tratamiento estén presentes en los tumores de pacientes tratados y ausentes en pacientes no tratados. Aunque en la práctica el análisis es más complicado de lo que parece debido a la concurrencia de varios factores de confusión, como por ejemplo el hecho que los pacientes frecuentemente no reciben una sola quimioterapia si no una combinación de varias. Desarrollamos un método estadístico capaz de identificar aquellas firmas mutacionales asociadas a tratamientos específicos que permite eliminar el efecto de dichos factores⁶⁹. En concreto identificamos firmas mutacionales distintivas del cisplatino, oxaliplatino, carboplatino, capecitabina (5 fluorouracilo), y temozolomida. Observamos que las mutaciones causadas por las quimioterapia son mayoritariamente subclonales y que estos tratamientos producen del orden de entre cientos y miles de mutaciones en las células tumorales⁶⁹.

Con este estudio pudimos demostrar que las células tumorales adquieren mutaciones causadas por la quimioterapia, pero ¿acumulan también mutaciones el resto de células del paciente, en particular las células de los tejidos sanos? Contestar esta pregunta es más difícil por dos razones: i) no es fácil disponer de muestras de tejidos sanos y ii) normalmente en los tejidos sanos no ocurren expansiones clonales como la de los tumores que nos permitan identificar las mutaciones con las tecnologías de secuenciación que he descrito antes.

Razonamos que los casos de neoplasias secundarias al tratamiento nos podrían ayudar a dar respuesta a esta pregunta. Colaboramos con clínicos del departamento de hematología del Hospital de Sant Pau en Barcelona para obtener ADN de leucemias secundarias al tratamiento, y además recopilamos otros genomas de leucemias secundarias publicados anteriormente (en todos los casos, leucemias mieloides agudas). En estas personas las células de la sangre son sanas en el momento del tratamiento del tumor primario (sólido en todos los casos incluidos en el estudio) con quimio-

rapia, pero meses o años más tarde los pacientes desarrollan una leucemia. La leucemia es, como todos los procesos tumorales, una expansión clonal a partir de una célula sana. Si esta célula inicial recibió mutaciones del tratamiento y se expandió a continuación en mayor o menor medida, deberíamos poder detectar la huella mutacional del tratamiento en el genoma de la leucemia por medio de métodos convencionales de secuenciación. En los 8 casos de leucemias secundarias de pacientes previamente tratados con cisplatino detectamos números similares de mutaciones con el patrón mutacional del cisplatino (entre cientos y miles) a los que observamos en las metástasis. Sin embargo, en las leucemias secundarias las mutaciones del tratamiento eran clonales, lo que indica que se desarrollaron a partir de una expansión clonal completa de una célula única después del tratamiento (ver Figura 3)⁷⁰. En otras palabras, la célula ancestral común más reciente (MRCA) de estas leucemias es una célula madre hematopoyética que existía durante el tiempo de la exposición a la quimioterapia, y que inició la expansión durante o después del tratamiento.

¿Y el resto de los tejidos sanos adquieren también mutaciones? Para poder responder a esta pregunta, a través de una colaboración con el Hospital pediátrico de Sant Joan de Déu de Barcelona, pudimos recientemente acceder al ADN de múltiples tejidos de una niña que murió de cáncer y cuyos padres generosamente autorizaron la autopsia para apoyar la investigación. Esta niña había recibido cisplatino y, en tejidos obtenidos a partir de seis órganos diferentes detectamos el patrón mutacional del tratamiento⁷¹. Para poder detectar las mutaciones en este caso recurrimos a una tecnología de secuenciación distinta, ya que en estos tejidos no hubo una expansión clonal, y las mutaciones del cisplatino quedan por debajo del límite de detección de una secuenciación más convencional (ver Figura 4). Aplicamos un método de secuenciación de alta precisión que consiste en secuenciar las dos cadenas de una molécula de ADN (*duplex sequencing*⁷²), de manera que podemos distinguir mutaciones verdaderas (aquellas que vemos en las dos cadenas) de los errores del proceso de secuenciación que se encuentran solo en una de las dos cadenas.

En resumen, con este conjunto de estudios identificamos firmas mutacionales para algunas de las quimioterapias más utilizadas en la clínica, y demostramos que algunos de estos tratamientos dejan entre cientos y unos pocos miles de mutaciones en células tumorales y células sanas en los distintos tejidos de las personas tratadas. Esto es sólo el primer paso, aún queda un largo camino para entender si hay alguna relación entre estas mutaciones y los efectos secundarios de la quimioterapia a largo plazo.

14. Estimación de la tasa de mutagénesis neutral

Hemos visto cómo varían las tasas de mutaciones a gran escala (a nivel de megabase) y a escala más local en regiones concretas como sitios de unión a factores de transcripción, nucleosomas y exones. También hemos visto cómo distintos contextos de secuencia (por ejemplo trinucleótidos), tienen probabilidades distintas de mutar según el proceso mutacional (firmas mutacionales). Pero aún nos queda un largo camino para poder estimar con precisión la probabilidad relativa de que de cada una de los 3200 millones de bases del genoma mute.

Como veremos en la siguiente parte del discurso, estimar correctamente la tasa de mutaciones en distintas regiones del genoma y en distintos tejidos y células, es clave para poder estimar la selección positiva e identificar genes y mutaciones causantes del cáncer. Todos los conocimientos generados en los últimos años sobre procesos mutacionales y tasas de mutaciones en distintas partes del genoma son importantes para poder estimar la mutagénesis de manera suficientemente precisa para detectar selección positiva correctamente.

Soñamos con poseer, en un futuro, un modelo matemático completo capaz de estimar con suficiente precisión la probabilidad de cualquier posible mutación en el genoma, dadas una serie de características y condicionantes. Tenemos que tener en cuenta que esta probabilidad depende de la persona (en particular de su exposición a diferentes procesos mutacionales a lo largo de la vida, y la capacidad de sus células de reparar el ADN) y del tejido y tipo celular (no es lo mismo una hepatocito del hígado que una célula epitelial del pulmón). Así que en realidad este modelo matemático estimaría una probabilidad de mutación distinta para cada base en el genoma según el tejido y tipo celular y según la exposición a diferentes mutágenos.

El tejido y tipo celular afecta a esta probabilidad de mutación de dos maneras distintas. Por un lado, distintos tejidos están expuestos de manera diferente a procesos mutacionales. Por ejemplo, la luz ultravioleta sólo afecta a tejidos expuestos a la luz del sol, como la piel, pero no a órganos internos. Los carcinógenos de los cigarrillos de tabaco afectan particularmente a las células del pulmón. Por otro lado, cada tejido y tipo celular tiene una tasa de división, una conformación de la cromatina determinada, unos factores de transcripción específicos, y un conjunto de genes transcritos distintos. Como todos estos elementos afectan a la tasa de reparación del ADN y en última instancia a la tasa de mutación, incluso para procesos mutacionales que ocurren en todos los tejidos, como la desaminación es-

pontánea de la citosina, la estimación de la tasa de mutación por base será necesariamente distinta en cada tipo celular de cada tejido.

Una gran dificultad para llegar a estimaciones precisas de la tasa de mutación a resolución de nucleótido individual es la relativa escasez de mutaciones observadas en los 3200 millones de pares de bases (pb) del genoma humano entre los datos disponibles. Actualmente hay datos de secuenciación de unos 7500 genomas completos (datos generados por múltiples proyectos internacionales y disponibles para investigación), y aunque cada genoma contiene miles de mutaciones, el número de mutaciones observadas hasta el momento es todavía muy escaso en comparación con el número total de bases del genoma. Esto nos impide medir directamente la variabilidad en la tasa de mutación a escalas menores de 10Kb. Si en un futuro dispusiéramos de millones de genomas de tumores este problema se vería resuelto en gran medida. Debemos asumir pues que nuestras estimaciones actuales de la tasa de mutación por base tienen un cierto nivel de incertidumbre. Es importante poder al menos estimar cuál es esta incertidumbre, es decir, estimar cual es el nivel de error que estamos cometiendo. En un estudio reciente en el laboratorio utilizamos mutaciones en tumores que no son drivers (pasajeras) recurrentes, es decir aquellas observadas en más de un paciente, para estimar el nivel de error que cometemos con las estimaciones de la tasa de mutación a nivel de base⁷³. Para algunos procesos mutacionales, como la desaminación espontánea de la citosina metilada (firma 1), las estimaciones teóricas basadas en el contextos de trinucleótido y el estado de metilación de las Cs en contexto CG en líneas celulares arroja resultados consistentes con la observación, mientras que para otros todavía hay un proporción significativa de la observación que los modelos teóricos no pueden explicar. Esto nos indica que todavía hay características del ADN o de la cromatina que afectan a la tasa de mutación local de estos procesos mutacionales que desconocemos.

En resumen, el estudio de mutaciones somáticas en miles de tumores ha llevado a la identificación de firmas mutacionales características de cada uno los principales procesos mutacionales que operan en nuestros tejidos, y a entender cómo los sistemas de reparación del ADN están afectados por la estructura de la cromatina, los nucleosomas, las proteínas unidas al ADN, la transcripción y la replicación. Todo este conocimiento es clave para poder estimar la tasa de mutaciones en distintas regiones del genoma y simular el proceso de mutagénesis neutral en nuestros tejidos. Medir la mutagénesis neutral es el primer paso para detectar señales de selección positiva, que nos ayuda a identificar genes y mutaciones causantes del cáncer, como veremos en la siguiente parte del discurso.

SELECCIÓN: ¿QUÉ MUTACIONES CAUSAN CÁNCER?

15. Selección positiva para identificar genes de cáncer

Desde la secuenciación de los primeros genomas de tumores, se evidenció la necesidad de desarrollar métodos estadísticos rigurosos que permitieran identificar qué genes contienen mutaciones cuasantes o *drivers* del cáncer entre los cientos o miles de genes mutados en tumores. En este punto, comenzaba a ser evidente que observaríamos mutaciones en todos los genes del genoma si secuenciáramos muestras de tumores de suficientes pacientes. ¿Cómo identificar, pues, aquellos cuyas mutaciones son claves para el desarrollo del tumor?

La perspectiva evolutiva nos puede ayudar en este problema. Primero, es importante comprender que hay varias trayectorias evolutivas para desarrollar un tipo de cáncer. Es decir, el juego de variación y selección entre las células de un tejido humano tiene múltiples soluciones que pueden llegar a un mismo resultado, un tumor. Esto explica, por ejemplo, que tumores de colon de distintas personas tengan una combinación distinta de mutaciones *drivers*. Pero podemos asumir que en un tejido concreto hay un número finito de trayectorias, y que algunas de estas son más frecuentes por conducir de manera más eficiente hacia el tumor. Así, si analizamos las mutaciones de un conjunto de tumores del mismo tipo de cáncer podemos esperar ver ciertos patrones recurrentes, por ejemplo múltiples tumores con mutaciones en el mismo gen.

Como hemos visto en las secciones anteriores, la variación entre las células de nuestros tejidos se debe a mutaciones que ocurren continuamente por múltiples procesos mutacionales. Estas mutaciones ocurren en una célula, al azar, sin una dirección determinada. Esto es lo que llamamos variación neutral, o mutagénesis neutral. La selección positiva ocurre gracias a mutaciones *driver* que confieren ventaja selectiva a la célula fundadora del tumor o célula de origen. Así en las células que conforman el tumor observaremos tanto las mutaciones *driver* de ese tumor, como el conjunto de mutaciones pasajeras que ocurrieron durante la historia del linaje de la célula de origen del tumor. Distinguir entre mutaciones *driver* y pasajeras es complicado si sólo contamos con un tumor. Pero cuando analizamos mutaciones en un conjunto lo suficientemente grande de tumores del mismo tipo, emerge un patrón distinto en los genes con mutaciones *drivers* que revela las huellas de la acción de la selección positiva, es decir, obser-

vamos un patrón distinto al que esperaríamos por un proceso de mutagénesis neutral. En cambio, los genes que sólo contienen mutaciones pasajeras mantienen el patrón compatible con la variación neutral.

Identificar señales de selección positiva – es decir, características cuantificables que permitan dilucidar desviaciones o anomalías con respecto a la variación neutral – en el patrón de mutaciones observadas en un conjunto de tumores es una buena estrategia para identificar genes causantes del cáncer. Una de estas señales es la frecuencia de mutaciones. En los genes de cáncer observamos una frecuencia de mutaciones más alta que la esperada según la mutagénesis neutral en ese tejido. También observamos que estos genes contienen mutaciones distribuidas en su secuencia de manera inesperada según la mutagénesis neutral, por ejemplo mutaciones agrupadas en ciertos dominios o regiones concretas del gen. Por último, en los genes de cáncer frecuentemente observamos un sesgo hacia mutaciones que alteran la función de la proteína (Figura 13).

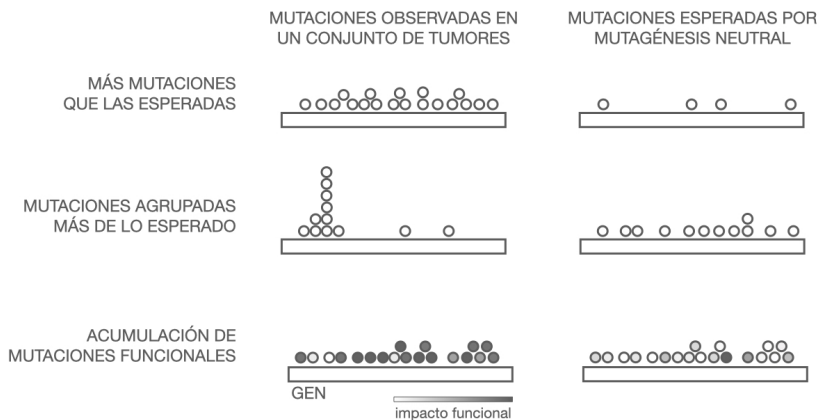


Figura 13. Identificación de selección positiva en el patrón de mutaciones en un conjunto de tumores. Tres patrones indicativos de selección positiva son el observar más mutaciones de lo esperado por la mutagénesis neutral, observar mutaciones agrupadas en regiones concretas del gen más de lo esperado, y observar una desviación hacia mutaciones que alteran la función de la proteína.

Una línea importante de investigación en mi laboratorio se ha centrado en el desarrollo y mejora de métodos capaces de identificar genes causantes del cáncer detectando estas señales de selección positiva en el patrón de mutaciones⁷⁴⁻⁷⁹. Ya en 2012 publicamos uno de los primeros métodos para identificar genes de cáncer a partir del patrón de mutaciones⁷⁹. Estas metodologías de detección de señales de selección positiva han resultado ser

muy efectivas. Partiendo simplemente de las mutaciones somáticas observadas en un conjunto de tumores, estos métodos son capaces de identificar aquellos genes causantes del cáncer con una alta precisión. En general todos los genes de cáncer conocidos tienen uno u otro, o en muchas ocasiones varios de estos patrones mutacionales. De esta forma, estas metodologías identifican los genes de cáncer previamente ya conocidos, pero además permiten la posibilidad de encontrar genes nuevos, previamente no descritos como causantes de tumores.

A lo largo de los años hemos visto que para obtener la lista más completa y rigurosa de genes de cáncer en un conjunto de tumores, la mejor estrategia es combinar varias señales de selección positiva.

16. Hacia un Compendio de Genes de Cáncer

Desde la identificación de la primera mutación causante del cáncer de vejiga en 1982, uno de los objetivos principales de la investigación del cáncer ha sido la identificación de todos los genes implicados en la tumorigénesis. Esta información es clave para entender los mecanismos de la transformación de células normales en células tumorales y para desarrollar nuevos tratamientos de precisión contra el cáncer. De hecho, algunas proteínas codificadas por genes de cáncer son posibles dianas terapéuticas.

En 2004 se recopiló el primer Censo de Genes de Cáncer⁸⁰ que en ese momento constaba de 291 genes. Actualmente este censo ya tiene más de 700 genes. Este censo está recopilado por un grupo de personas expertas que revisa la literatura científica de manera sistemática y, siguiendo una serie de criterios estrictos, deciden qué genes entran en la lista, y anotan el gen con el tipo de cáncer al que están asociados y otra información relevante. Este censo incluye tanto genes que están alterados por mutaciones somáticas puntuales, como con otro tipo de alteraciones estructurales en el genoma, como veremos en la siguiente sección. Ciertamente, tras más de 40 años de investigación del cáncer, la literatura científica es muy rica en información de genes implicados en tumorigénesis en distintos tejidos, pero hay que tener en cuenta que está altamente sesgada. Algunos genes han centrado mucho interés y podemos encontrar miles de estudios científicos que detallan su rol en varios tipos de cáncer, mientras que otros genes apenas han sido estudiados, y por tanto la literatura sobre ellos es muy pobre o nula.

Con la acumulación de datos de secuenciación de genomas de tumores, se abrió la oportunidad de identificar genes de cáncer directamente a partir

del análisis de mutaciones, de una manera sistemática y evitando sesgos. Es decir, usando una metodología que diera la misma oportunidad a todos los genes de entrar o no en la clasificación como genes de cáncer sin nuestras preconcepciones de cuál podría ser o no importante. Ya en 2013 iniciamos en nuestro laboratorio el proyecto IntOGen⁷⁴ para generar un Compendio de Genes de Cáncer lo más completo posible. Para esto desarrollamos un proceso computacional para analizar sistemáticamente todas las cohortes de tumores para las que disponemos de datos de exoma o genomas de tumores. El análisis incluye, en primer lugar, la identificación de señales de selección positiva en cada gen y cada cohorte para cada tipo de tumor de forma independiente. Seguidamente se combinan todas las señales de selección positiva detectadas por diferentes métodos computacionales complementarios⁷⁵ para obtener un consenso de genes driver en la cohorte. El resultado es una identificación robusta de genes con señales de selección positiva según el patrón de mutaciones que, por tanto, son clasificados como genes de cáncer. En la versión del compendio que publicamos en 2020 procesamos mutaciones en genomas de 28 076 tumores e identificamos 568 genes de cáncer en 66 tipos de tumores. En el compendio hay, por ejemplo, 99 genes involucrados en cáncer de mama, 42 genes en adenocarcinoma de pulmón y 20 genes en neuroblastoma. Es importante mencionar que identificar un gen cuyas mutaciones pueden causar cáncer, solo es el primer paso para entender cómo se forma un tumor. Falta un largo trecho para entender cuál es el rol de cada una de las proteínas codificadas por estos genes en el desarrollo del cáncer. Nosotros publicamos los resultados del Compendio en la web IntOGen.org para que otros puedan explorar y seguir la investigación para comprender cómo cada una de estas proteínas mutadas participa en el proceso tumoral.

En este artículo del 2020 hacemos una revisión histórica de la identificación de genes de cáncer, detallamos la construcción del compendio, reflexionamos sobre lo que hemos aprendido hasta el momento, y apuntamos a retos futuros⁷⁵. Detallaré aquí algunas reflexiones relevantes.

Este compendio, al estar generado de manera sistemática a partir de datos de mutaciones en miles de tumores, nos permite abordar ciertas preguntas evitando el sesgo de la literatura científica. Por un lado, es importante remarcar que el compendio identifica la gran mayoría de genes conocidos del cáncer, ya que todos ellos tienen señales de selección positiva en las mutaciones observadas en tumores: esto es un buen indicador de la corrección de los resultados arrojados por nuestra metodología. Por otro lado, esta aproximación sistemática tiene la oportunidad de identificar ge-

nes nuevos que no conocíamos como genes de cáncer. El compendio publicado en 2020 contenía 568 genes de los cuales 416 coincidían con genes en el censo de la literatura⁷⁵. De manera que 152 eran genes no anotados en este censo.

Además de la lista completa de genes *driver*, el compendio incluye la asociación entre cada uno de estos genes y cada uno de los 66 tipos de tumores para los que tenemos datos. En otras palabras, cada tipo de tumor tiene asociado un conjunto de genes con mutaciones *driver*. Esta metodología sistemática aporta muchas nuevas asociaciones que desconocíamos previamente. Es decir, aunque sepamos por la literatura que un gen causa un tipo de cáncer, esta aproximación logra completar la implicación de este gen en otros tipos de cáncer en el que aún no se ha estudiado. Para ser más específica, de las 1892 interacciones entre genes y tipos de cáncer en el compendio que publicamos en el 2020, solo 388 aparecen en el censo a partir de la literatura, es decir, 1504 son interacciones nuevas de genes implicados en nuevos tipos de tumores.

La aproximación sistemática nos permite abordar la pregunta de la especificidad de los genes de cáncer en distintos tipos tumorales. Observando los resultados del compendio nos damos cuenta de que la mayoría de los genes de cáncer son relativamente específicos de uno o pocos tejidos. Solo una decena de los genes son causantes de cáncer en más de 20 tumores distintos. Evidentemente entre estos están algunos de los genes de cáncer más conocidos como *TP53*, *KRAS*, *PIK3CA*, y *ARID1A*, siendo *TP53* el campeón absoluto, con señales de selección positiva en la mayoría de tipos de cáncer (52 de 66 en el compendio del 2020). El gen *TP53* codifica la proteína p53, que es considerada “el guardián del genoma” por su papel esencial como coordinadora de la respuesta al daño genómico y otros problemas celulares.

El compendio además de ser una lista de genes por cada tipo de cáncer, también contiene los patrones mutacionales de cada uno de los genes en cada tipo de tumor. Estos patrones apuntan a los mecanismos por los cuales las mutaciones alteran la proteína para proveer a la célula de la ventaja selectiva y participar en el desarrollo del cáncer. Estudiando las características de estos patrones de mutaciones podemos, por ejemplo, clasificar los genes en oncogenes (es decir, aquellos en los que las mutaciones activan la proteína) y genes supresores de tumores (aquellos en los que las mutaciones causantes del cáncer inactivan el gen). Estas características de los patrones mutacionales de cada gen de cáncer también son claves para poder

identificar qué mutaciones específicas en el gen tienen la capacidad de causar el cáncer, como veremos más adelante.

Es importante aclarar que esta metodología identifica genes que causan cáncer a través de mutaciones puntuales o pequeñas inserciones o deleciones. Estos son los tipos de mutaciones que analizamos en IntOGen y por los que podemos identificar muy eficientemente señales de selección positiva. Hay otros tipos de alteraciones genómicas, como translocaciones, aumento del número de copias del gen, pérdida de regiones cromosómicas, o mutaciones germinales, entre otras, que no están incluidas en el Compendio de IntOGen. Estas las abordaremos en la siguiente sección.

¿Cómo de completo es el compendio? ¿Nos faltan muchos genes de cáncer por identificar? ¿O ya hemos identificado la mayoría? Creemos que si nos centramos en aquellos tipos de cáncer más comunes y mejor estudiados (para los que existen más tumores con el genoma secuenciado), es probable que ya hayamos identificado todos o casi todos los genes que pueden causar el cáncer con mutaciones con una cierta frecuencia. Por ejemplo, en la versión actual de IntOGen (v2023) hemos analizado las mutaciones somáticas en todos los genes de 3201 tumores de cáncer de mama y hemos identificado 117 genes con señales de selección positiva. Algunos de estos están mutados con una alta frecuencia, como *TP53* y *PIK3CA* que se encuentran mutados en alrededor del 30% de tumores de mama. Mientras que en otros genes hemos observado mutaciones en tan solo unos pocos tumores, por ejemplo *MTOR* y *EGFR* que están mutados en menos del 0.1% en cáncer de mama. Nuestra capacidad para identificar señales de selección positiva depende del número de tumores para los que tenemos datos, y de la frecuencia mutacional del gen. Por esta razón creemos que los genes mutados con cierta frecuencia en aquellos tipos de tumores con más muestras secuenciadas ya están recogidos en el compendio. Es decir, lo que falta por identificar son genes de cáncer que afectan a un número reducido de pacientes, y genes causantes de tumores poco frecuentes, para los que hasta el momento se han secuenciado solo pocos casos. Para esto estamos colaborando con grupos que se centran en tumores poco frecuentes, como mesotelioma⁸¹, sarcomas y tumores neuroendocrinos pulmonares. Otro tema a considerar es que la mayoría de los estudios de secuenciación de tumores se han hecho en ciertas regiones del mundo (EUA, Europa, Japón, China). Los tumores de personas de otras etnicidades y de personas que han habitado otras regiones del planeta podrían presentar diferencias significativas en los genes implicados, pero sólo podremos contestar esta pregunta cuando dispongamos de suficientes secuencias de tumores de personas de estas otras partes del mundo.

En resumen, el avance desde la identificación de los primeros genes de cáncer hace unos 40 años es enorme. Tenemos un compendio de genes causantes del cáncer y sabemos qué tipos de tumores pueden causar cuando reciben mutaciones. Sin embargo, aún falta un largo camino para completar el compendio, sobre todo para identificar genes mutados poco frecuentemente, esclarecer cuales son los genes *driver* de tumores poco frecuentes y establecer las posibles diferencias en genes de cáncer entre distintas etnias y en diferentes partes del mundo.

17. Otras alteraciones genómicas en tumores

Hasta el momento cuando he hablado de mutaciones me he referido mayoritariamente a mutaciones puntuales, es decir cambios de un nucleótido a otro, y pequeñas inserciones y deleciones (de unos pocos nucleótidos). Son los tipos de mutaciones para los que podemos medir de manera rigurosa la selección positiva con métodos estadísticos como los que utilizamos en IntOGen, básicamente por el conocimiento y las posibilidades prácticas de modelar la mutagenesis neutral para esas mutaciones. En cambio, para otros tipos de alteraciones genómicas, actualmente no es posible medir correctamente la tasa de mutación neutral y por tanto medir selección positiva. Por esta razón es más difícil identificar qué alteraciones genómicas de estos otros tipos están implicadas en cáncer, a menos que la frecuencia en que vemos estos eventos en tumores sea muy alta. Pero el genoma puede alterarse de muchas maneras distintas, y la secuenciación de genomas de tumores nos ha permitido observar muchos eventos distintos de alteraciones genómicas e incluso empezar a entender sus mecanismos.

Uno de estos tipos de alteraciones, conocidas como translocaciones, ocurre cuando hay una rotura de la doble cadena de ADN y ésta se une de manera errónea a otra región del genoma. En algunos tumores encontramos un gran número de ellas. Algunas son claramente translocaciones causantes del cáncer, como el caso del cromosoma Filadelfia, que de hecho fue la primera alteración genética adquirida identificada como causante del cáncer, antes incluso que la mutación de *HRAS* en 1982. En concreto fue identificada como causante de la leucemia mieloide crónica en niños por la genetista estadounidense Janet Rowley en 1973⁸². Esta translocación une dos fragmentos de los cromosomas 9 y 22, creando un gen nuevo que fusiona parte de dos genes, *ABL* (del cromosoma 9) y *BCR* (del cromosoma 22). Este gen de fusión mantiene la actividad quinasa del gen *ABL*, que queda activada de manera constitutiva e independiente de la presencia de factores de activación, transmitiendo una señal a la célula para dividirse

continuamente y causando así la leucemia. El 90% de los enfermos con leucemia mieloide crónica tienen esta translocación. De hecho este tipo de cáncer es en el que se desarrolló la primera terapia de precisión, Imatinib, como veremos más adelante (sección 21), y que ha cambiado completamente el curso de esta enfermedad.

Aparte de roturas en la doble cadena de ADN, los genomas de tumores presentan frecuentemente una duplicación entera del genoma, así como pérdidas y ganancias de regiones concretas. Desde la pérdida de un brazo entero de un cromosoma (por ejemplo el brazo corto del cromosoma 3 se pierde en la mayoría de tumores de riñón), a regiones más pequeñas que se amplifican (es decir, se generan varias copias) o se pierden. En general, una característica muy común entre tumores es lo que llamamos inestabilidad genómica, caracterizada por múltiples alteraciones genómicas, que muchas veces va asociada a la pérdida de función de la proteína p53.

Incluso se han detectado fenómenos en que un solo evento catastrófico destroza una parte del genoma, que se recompone en múltiples fragmentos de uno o más cromosomas reordenados: es lo que llamamos cromotripsis⁸³. También se han observado eventos, a veces asociados a fenómenos de cromotripsis, en que un trozo del genoma se separa de los cromosomas y se circulariza, creado un “minicromosoma circular” (llamados “doble-minute” o ecDNA) que se replica de manera autónoma y se segrega de manera asimétrica, pudiéndose generar múltiples copias de esta región. Algunos de estos eventos contienen oncogenes, la amplificación de los cuales es ventajosa para la célula. Por ejemplo, en algunos meduloblastomas, un tipo de cáncer cerebral que afecta a niños, se identificaron del orden de cientos de copias del oncogén *MYCN*, y se ha comprobado que se encuentra en estas estructuras de ADN circular⁸⁴. En resumen, debido a la inestabilidad genómica frecuentemente asociada a la carcinogénesis, múltiples alteraciones más o menos complejas se detectan en genomas de tumores, algunas de las cuales pueden ser eventos directamente causantes del cáncer.

18. ¿Drivers en el genoma no codificante?

El Compendio de Genes de Cáncer de IntOGen se centra en mutaciones en genes que codifican para proteínas. Estos representan poco más del 1% del genoma. Una pregunta crítica es si hay mutaciones causantes del cáncer en regiones no codificantes, por ejemplo en regiones reguladoras de la expresión de los genes o en genes no codificantes.

El ejemplo más conocido y frecuente de mutaciones driver no codificantes son las mutaciones en el promotor de *TERT*^{85,86}. *TERT* es el gen que codifica para la telomerasa, la proteína que sintetiza los telómeros. Los telómeros son las estructuras de DNA que protegen los extremos de los cromosomas y, en ausencia de telomerasa, se acortan en cada división celular sirviendo así de controladores del número de divisiones celulares. La telomerasa está activada de una manera u otra en la mayoría de tumores, y así es como las células tumorales se saltan una de las barreras anti tumorigénicas, la limitación del número de divisiones (como hemos visto en la sección 4a). Con la capacidad de alargar los telómeros, las células tumorales pueden dividirse de manera ilimitada. En una proporción relativamente alta de tumores de diferentes tipos de cáncer se observan mutaciones en el promotor del gen *TERT*. Estas mutaciones provocan la activación de la expresión del gen de manera continua.

Otros ejemplos de mutaciones *drivers* no codificantes son las mutaciones en la región no codificante del gen *NOTCH1* en leucemia linfática crónica⁴². Estas fueron identificadas por primera vez en el proyecto Español del consorcio internacional del genoma del cáncer. Estas mutaciones provocan un *splicing* aberrante del gen generando una proteína NOTCH1 con mayor estabilidad.

¿Hay otras mutaciones causantes del cáncer en regiones no codificantes? Esta es una pregunta muy difícil de contestar, ya que el genoma es muy extenso y la función de la mayor parte de estas regiones es desconocida. En su momento pensamos que adaptar metodologías de detección de señales de selección positiva para identificar regiones no codificantes con selección positiva sería una buena estrategia para abordar esta pregunta. Para eso necesitábamos dos cosas, genomas completos de tumores, y métodos adaptados a identificar señales de selección positiva en regiones no codificantes. Recibimos financiación de una de las fuentes más competitivas y prestigiosas de Europa (*European Research Council* - ERC) para abordar esta pregunta.

Lo primero que hicimos fue adaptar nuestras metodologías de detección de señales de selección positiva en genes para funcionar en regiones reguladoras de la transcripción, como los promotores (regiones al inicio de transcripción de los genes) y los *enhancers* (regiones reguladoras lejos del inicio de la transcripción). Para eso, en lugar de calcular el impacto funcional de las mutaciones a partir de su consecuencia en el cambio de la secuencia de la proteína, nos centramos en calcular el efecto según la modi-

ficación de regiones de unión a factores de transcripción (*transcription factor binding sites*, TFBS). Los promotores y *enhancers* contienen regiones muy concretas, de unos 20 pares de bases, a las que se unen proteínas reguladoras de la expresión génica. Estas regiones tienen un motivo conocido en la secuencia, y podemos calcular el impacto funcional según si la mutación observada ocurre en un TFBS y modifica sustancialmente el motivo. Por ejemplo, la mutación puede modificar la unión del factor de transcripción al ADN y por tanto alterar la regulación del gen, como en el caso de las mutaciones en el promotor de *TERT*.

La primera sorpresa vino cuando aplicamos este nuevo método al estudio de mutaciones en promotores en un tipo de cáncer de piel, melanoma. Vimos que la mayoría de promotores tenían un sesgo hacia mutaciones con impacto funcional alto. Es decir, una gran parte de las mutaciones observadas en tumores afectan a regiones de unión de factores de transcripción. Inmediatamente nos dimos cuenta que era imposible que todos los promotores tuvieran mutaciones causantes del cáncer, y empezamos a plantearnos si nuestra asunción de cómo ocurren las mutaciones antes de la selección natural es distinta a como creíamos. Fue así como descubrimos que las mutaciones causadas por la luz ultravioleta ocurren más frecuentemente en regiones de unión a proteínas⁵⁹, tal como he explicado en la sección 11a, y como iniciamos una línea de investigación en el grupo destinada a mejorar nuestra comprensión de los procesos mutacionales. Este descubrimiento nos hizo darnos cuenta de que no podemos asumir que la alta frecuencia de mutaciones en ciertas regiones no codificantes sea una indicación de selección positiva, si no que es importante estudiar en primer lugar si pudiera ser causada por sesgos en la tasa de mutaciones de ciertos procesos mutacionales.

Dado que el estudio de procesos mutacionales es complejo, especialmente en zonas no codificantes del genoma, la identificación de mutaciones causantes del cáncer en estas regiones sigue siendo difícil. Gracias al análisis de mutaciones en miles de tumores, ahora sabemos que no hay ninguna otra mutación en la parte no codificante del genoma tan frecuente como las mutaciones en el promotor de *TERT*. Aún así, es posible que, a frecuencias menores, otras mutaciones en regiones reguladoras o en genes no codificantes contribuyan al desarrollo de tumores. Su identificación sigue siendo complicada por dos razones principales: i) la falta de conocimiento de las funciones de las regiones no codificantes del genoma y ii) la dificultad para determinar con precisión la tasa de mutación neutral. Aunque hemos hecho avances significativos en identificar características del

genoma que afectan a la tasa de mutación, el cálculo preciso de la frecuencia de mutaciones esperadas en cada nucleótido individual sigue siendo un reto no resuelto, que dificulta la identificación de mutaciones driver en regiones no codificantes del genoma.

19. ¿Cómo identificar las mutaciones específicas que causan cáncer?

Con el Compendio de Genes de Cáncer de IntOGen, como he mencionado antes, creemos que tenemos una lista más o menos completa de aquellos genes más frecuentemente mutados en los tipos de tumores más comunes. Pero otro reto esencial no solventado es identificar qué mutaciones específicas en estos genes son las que causan cáncer.

Y es que no todas las mutaciones en genes de cáncer son mutaciones *driver*. Incluso en un gen de cáncer tan frecuentemente mutado como TP53, podemos observar mutaciones pasajeras en tumores. Distinguir entre mutaciones *driver* y mutaciones pasajeras es un reto importante para la interpretación de datos genómicos de un tumor en la práctica clínica, como veremos más adelante.

Para interpretar la importancia de una mutación en un gen de cáncer que se haya detectado en un tumor, habitualmente buscamos información en la literatura o en las bases de datos sobre su posible efecto en la función del gen, y si se conoce si es o no oncogénica. Si no encontramos ninguna información sobre la mutación, ésta se clasifica como variante de significado desconocido o incierto (*Variant of Uncertain Significance VUS* en inglés). Si buscamos en estas bases de datos las mutaciones que encontramos en tumores en genes de cáncer nos damos cuenta que la mayoría de estas acababan clasificadas como mutaciones de significado incierto (VUS).

Nos propusimos contribuir al reto de distinguir mutaciones *driver* de pasajeras en genes de cáncer con una idea similar a cómo abordamos el desarrollo del Compendio de Genes de Cáncer. En lugar de basarnos en el conocimiento acumulado en la literatura sobre variantes de cáncer, que aunque muy valioso es incompleto y está altamente sesgado, queríamos abordar el problema aprendiendo directamente a partir de los datos de mutaciones observadas en miles de tumores de pacientes.

El reto consiste en desarrollar una metodología computacional que pueda clasificar cualquier mutación en un gen de cáncer como mutación *driver*

o pasajera. Decidimos utilizar técnicas de aprendizaje automático (*machine learning*) que puedan discernir la combinación de características mutacionales que distinguen un conjunto de mutaciones *driver* y un conjunto de mutaciones pasajeras. Con este modelo entrenado a partir de un conjunto de mutaciones suficientemente representativo podríamos abordar la clasificación de cualquier posible mutación en el gen. Llamamos a esta metodología computacional BoostDM⁸⁷.

Una consideración importante es que en lugar de desarrollar un clasificador general, decidimos que era necesario desarrollar un clasificador para cada gen. La función de cada proteína en la célula es distinta, y cada una se altera de formas diferentes en la tumorigénesis. Por tanto la combinación de características que puede definir qué mutaciones son *drivers* serán distintas de un gen a otro. Otra consideración es que siempre que fuera posible queríamos desarrollar predictores específicos del tipo de tumor. La razón es el hecho que hay ejemplos de genes cuyo patrón de mutaciones es muy distinto según el tipo tumoral. Este es el caso, por ejemplo, del gen *EGFR*, cuyas mutaciones pueden causar glioblastoma y cáncer de pulmón. En ambos tipos de tumores es un gen mutado en una frecuencia relativamente alta, pero el patrón de mutaciones que observamos en un tejido y otro es muy distinto, indicando que probablemente el mecanismo molecular en que la proteína debe alterarse para causar glioblastoma o cáncer de pulmón es distinta.

Para entrenar un clasificador mediante aprendizaje automático necesitamos un conjunto de casos positivos (mutaciones *drivers*) y un conjunto de casos negativos (mutaciones pasajeras). Pero en nuestro caso carecemos de ello, ya que es precisamente el problema que estamos intentando solucionar. Tampoco nos serviría utilizar un conjunto de mutaciones en el gen anotadas en la literatura o en bases de datos como *drivers* o pasajeras, ya que entonces estaríamos creando un clasificador que aprendería los sesgos de la literatura.

Aquí los conceptos de evolución darwiniana vienen al rescate una vez más. Desarrollamos un clasificador inspirado en biología evolutiva, en el que nos guiamos directamente por los datos, sin introducir ningún sesgo. En concreto, razonamos que con el cálculo de selección positiva podemos estimar indirectamente la proporción de mutaciones observadas que son *drivers* en cada gen. Si bien no podemos saber qué mutaciones concretas son *driver*, para aquellos genes en que esta proporción está por encima de un cierto umbral, por ejemplo el 85%, las mutaciones observadas serían un

conjunto de casos positivos suficientemente enriquecido en *drivers* para la tarea de aprendizaje automático.

Para generar el conjunto de casos negativos, o mutaciones “pasajeras”, lo que hacemos es simular computacionalmente el proceso de mutagénesis neutral. Generamos el mismo número de mutaciones a las observadas, siguiendo la probabilidad de cada mutación de ocurrir según la preferencia de trinucleótidos (como en las firmas mutacionales, ver sección 9) en el tejido en cuestión. Estas mutaciones serán mayoritariamente pasajeras.

Diseñamos una estrategia de aprendizaje que tiene en cuenta que el conjunto de casos positivos y negativos no es perfecto, y comprobamos su efectividad para aprender a identificar las mutaciones *driver* y pasajeras, con la gran ventaja de que no estamos introduciendo ningún sesgo.

Para identificar la combinación de características que definen las mutaciones *driver* tenemos que anotar todas las mutaciones con un conjunto de características que nos parezcan útiles para la clasificación. En la primera versión de BoostDM utilizamos 18 características⁸⁷. Estas incluyen, por ejemplo, la consecuencia de la mutación en la secuencia codificante de la proteína, su nivel de conservación, el dominio de la proteína donde se encuentra la mutación, y si la mutación se encuentra en una región con un enriquecimiento significativo de mutaciones (lo que conocemos como “clusters de mutaciones”), entre otras. Actualmente estamos trabajando en una nueva versión que incluye nuevas características, algunas derivadas de la estructura tridimensional de la proteína según las reconstrucciones conformacionales *in silico* de *AlphaFold* (un un modelo de aprendizaje automático que predice la estructura 3D de una proteína a partir de su secuencia de aminoácidos⁸⁸).

Para obtener modelos de BoostDM de buena calidad, es decir con buena capacidad predictiva, es necesario haber observado un número suficiente de mutaciones en el gen en cuestión, que sean representativas del conjunto de posibles mutaciones *driver* en el gen. Desarrollamos métricas para estimar esta representatividad y determinar el número mínimo de mutaciones necesarias para el aprendizaje. Este número es distinto en cada gen, pero para dar una idea, consideramos que no podemos crear modelos suficientemente fiables si no hemos observado al menos 30 tumores con mutaciones en el gen en cuestión. Para algunos genes este número ya es suficiente para crear un buen modelo, pero para otros requerimos más mutaciones observadas. Pudimos crear modelos para 84 genes de los más de

500 genes de cáncer en el Compendio de genes de cáncer. Para el resto no hemos visto suficientes mutaciones todavía. Esperamos que con el avance de la secuenciación de tumores podamos crear modelos para la mayoría de genes de cáncer en un futuro próximo. En cualquier caso estos modelos ya son capaces de clasificar un número mucho mayor de las mutaciones observadas en tumores, reduciendo el número de mutaciones clasificadas como VUS (significancia incierta).

Una vez tenemos un modelo de BoostDM para un gen concreto en un tipo de tumor determinado, por ejemplo el gen *EGFR* en adenocarcinoma de pulmón, podemos aplicar el modelo para clasificar todas las posibles mutaciones en el gen (*In silico saturation mutagenesis* en inglés) y así crear un mapa de todas las posibles mutaciones *drivers* en el gen, incluso aquellas que aún no hemos observado en tumores.

Consideramos este trabajo como una prueba de concepto de las posibilidades de este tipo de aprendizaje automático, y tenemos aspiraciones de mejorar estos modelos en un futuro. Por ejemplo, sabemos, como comentaba en la introducción, que una mutación no proporciona ventaja selectiva siempre, sino que depende del tejido y de su contexto. En la versión actual BoostDM ya tiene en cuenta el tejido, pero otros elementos del contexto son ignorados. Por ejemplo, aunque sabemos que el cáncer está causado por una combinación de eventos *driver*, BoostDM analiza cada mutación por separado. Versiones futuras de BoostDM podrían tener en cuenta la combinación de mutaciones *driver* en la célula para determinar su efecto.

Hay mutaciones que proporcionan una ventaja selectiva durante el tratamiento contra el cáncer, son las que llamamos mutaciones de resistencia al tratamiento. Estas se identifican como *drivers* por BoostDM si los datos de aprendizaje incluyen tumores de pacientes tratados, pero ahora mismo los modelos no tienen información de qué pacientes han recibido o no tratamiento, y por tanto no puede distinguir entre mutaciones implicadas en el desarrollo tumoral y mutaciones implicadas en la resistencia al tratamiento. Nos podemos imaginar un modelo de aprendizaje en el futuro que pueda asociar las mutaciones *driver* al contexto en el que son *drivers*.

En resumen, el aprendizaje automático a partir de mutaciones observadas en miles de tumores nos permite, siguiendo unos razonamientos de biología evolutiva, obtener modelos de calidad capaces de clasificar todas las posibles mutaciones en un gen de cáncer como mutaciones *driver* o pasajeras. A medida que tengamos acceso a más datos de genomas de tu-

mores podremos generar modelos para más genes, y avances futuros en esta tecnología podría permitir no solo identificar de manera binaria mutaciones *drivers* y pasajeras, sino también asociarlas al contexto en que proporcionan esta ventaja selectiva.

20. ¿Cuántas mutaciones *driver* se requieren para causar un cáncer?

Al principio del discurso veíamos como Armitage y Doll en 1954, con un simple modelo matemático modelando el incremento exponencial de la incidencia del cáncer con la edad, estimaron que se requieren entre 5 y 7 cambios celulares discretos en una célula para desarrollar un cáncer, que podrían ser mutaciones somáticas⁶. Ahora tenemos información del genoma completo de miles de tumores y podemos identificar mutaciones *drivers* presentes en cada uno de ellos. ¿Podemos aproximarnos a una respuesta a esta pregunta? ¿Cuántas mutaciones *driver* se requieren para causar un cáncer?

En 2017, como parte del estudio del consorcio internacional del genoma del cáncer (ICGC) centrado en genomas completos de tumores (*Pan Cancer Analysis of Whole Genomes - PCAWG*), desde mi laboratorio lideramos el análisis para obtener un panorama lo más completo posible de todas las alteraciones causantes del cáncer (*drivers*) en más de 2500 tumores de 37 tipos de cáncer^{44,89}. Este estudio consiste en identificar todos los eventos genómicos que podríamos catalogar como *drivers* en cada uno de los tumores. No es una tarea fácil, ya que no conocemos todos los genes que causan cáncer, ni todas las posibles variantes estructurales que dan una ventaja selectiva a las células. Algunos de los eventos que clasificamos como *driver* en un tumor podrían no serlo en ese contexto concreto. Este estudio no incluye eventos epigenéticos que también son heredables y podrían conferir una ventaja selectiva y ser eventos *driver*. Aún así, este panorama es el más completo hecho hasta el momento, e incluye mutaciones puntuales codificantes y no codificantes, alteraciones del número de copias y otros reordenamientos genómicos de origen somático, y variantes de la línea germinal potencialmente de predisposición al cáncer. En más del 90% de los tumores detectamos al menos un evento *driver*, y creemos que en el 10% que no los identificamos es probablemente debido a nuestra falta de conocimiento de todos los posibles eventos *driver*. Esto nos indica que los eventos genómicos están en la raíz de prácticamente todos los tumores. En cada tumor detectamos en promedio 4,6 eventos *drivers*. Hay diferencias entre unos tipos de tumores y otros, pero aun así los números no difieren mucho: van de una

media de aproximadamente 2 eventos *driver* en leucemia mieloide aguda a 7-8 en cáncer de vejiga, colon y pulmón⁴⁴.

Aun teniendo en cuenta que consideramos este panorama como una primera aproximación a catalogar todos los eventos *driver* por tumor, y que podrían haber ciertas variaciones, creemos que los números se mantendrán más o menos en este orden, entre 2 y 8 eventos por tumor. Es sorprendente cuánto se acercan los números a estimaciones completamente ortogonales hechas hace unos 70 años, casi treinta años antes de la descripción de la primera mutación *driver* identificada en cáncer de vejiga.

21. ¿Cómo ayuda este conocimiento a la mejora de los tratamientos contra el cáncer?

Uno de los objetivos de la investigación del cáncer es desarrollar nuevos tratamientos dirigidos específicamente contra las células tumorales. Una estrategia en esta línea es desarrollar fármacos que puedan bloquear aquellas proteínas alteradas específicamente en el tumor. Así pues, algunas de las proteínas codificadas por genes de cáncer se consideran dianas terapéuticas.

Un ejemplo de este concepto es el fármaco Imatinib (comercializado como Gleevec), que inhibe la actividad tirosina quinasa de la proteína de fusión resultado del cromosoma Filadelfia en leucemia mieloide crónica, que ya he mencionado en la sección 16. La aprobación de este fármaco en 2001, ha cambiado totalmente la esperanza de vida de pacientes con esta enfermedad⁹⁰. Este fármaco también inhibe la actividad quinasa de otras proteínas y actualmente se administra a pacientes con otros tipos de cáncer además de la leucemia mieloide crónica.

Otro ejemplo es un fármaco que se aprobó en 2011 para tratar un tipo de melanoma causado por una mutación en el gen que codifica para la proteína BRAF, en concreto la mutación en la posición 600 de la proteína (V600)⁹¹. Este cambio en el gen *BRAF* produce una proteína de señalización constitutivamente activada que hace que las células reciban la señal de dividirse continuamente. El fármaco Vemurafenib se desarrolló para inhibir específicamente la actividad de la proteína BRAF con la mutación V600. Este fármaco no sirve para todos los pacientes con melanoma, solo para aquellos que tienen la mutación BRAF V600. Así, antes de decidir administrar este fármaco a un paciente es necesario hacer una prueba diagnóstica para comprobar si el melanoma del paciente tiene o no esta mutación.

El cáncer de pulmón es uno de los tipos de cáncer en que frecuentemente se utiliza información genómica para tomar decisiones de tratamiento. Varias proteínas codificadas por genes con mutaciones *driver* de cáncer de pulmón son posibles dianas terapéuticas, y para algunas de estas existen inhibidores que bloquean su función. Por ejemplo, algunos tumores de pulmón tienen mutaciones activadoras en el gen que codifica para EGFR, el receptor del factor de crecimiento epidérmico. Estas mutaciones pueden llevar a la activación constante de la vía de señalización de EGFR, que es crucial para el crecimiento y la supervivencia celular, dando así una ventaja selectiva a estas células, y causando una expansión clonal hacia el desarrollo tumoral. Se han desarrollado inhibidores de EGFR como Erlotinib, Gefitinib, Afatinib y otros, que están especialmente indicados para aquellos tumores con mutaciones activadoras en el gen.

Otro ejemplo ilustrativo lo encontramos en los inhibidores de PARP, que son fármacos desarrollados para bloquear una de las vías de reparación del ADN. Este tratamiento es especialmente efectivo en pacientes con cáncer de ovario con mutaciones que causan deficiencia en algunos mecanismos de reparación de ADN paralelos a la vía de PARP, como por ejemplo mutaciones en los genes *BRCA1* o *BRCA2*.

Al conjunto de estos tratamientos los llamamos tratamientos dirigidos (*targeted therapies* en inglés). Y se han desarrollado con la esperanza de que, al dirigirse directamente a las células tumorales, sean más efectivos, menos tóxicos y tengan menos efectos secundarios. En los últimos 20-25 años ha habido un gran esfuerzo para el desarrollo de este tipo de tratamientos y ya hay una batería de más de 80 fármacos dirigidos aprobados. Cada uno de ellos está aprobado para indicaciones específicas muy concretas, frecuentemente asociadas a la presencia de alteraciones genómicas determinadas en el tumor.

Es importante reconocer que, a pesar de aplicar tratamientos con fármacos dirigidos a las causas moleculares del tumor, por desgracia los tumores frecuentemente desarrollan resistencia a los fármacos utilizados. Este proceso de resistencia al fármaco no es más que una continuación del proceso de evolución darwiniano dentro del tumor. Durante el tratamiento, si alguna célula tumoral tiene alguna alteración heredable que le confiere resistencia al fármaco tendrá una ventaja selectiva y podrá resistir al tratamiento, proliferando y provocando una recaída del tumor. Procesos de resistencia al tratamiento ocurren tanto para tratamientos dirigidos como para quimioterapia. Una manera de indagar los mecanismos de resistencia es

secuenciar el genoma del tumor antes y después del tratamiento y estudiar su evolución. Junto con los grupos de Ana Bigas y Josep Maria Ribera, recientemente estudiamos la evolución de leucemias linfoblásticas agudas de células T durante el tratamiento, confirmado que la recaída después del tratamiento se debe a alteraciones heredables que confieren resistencia⁹².

Cada vez más en la clínica se complementa la información del tumor con datos moleculares y datos de secuenciación del ADN tumoral para afinar la decisión de qué tratamiento es el más adecuado para un paciente. Uno de los retos principales para implementar este diagnóstico molecular es la complejidad de los datos genómicos y la dificultad en interpretar la información del genoma del tumor para tomar la mejor decisión para el paciente. Discriminar cuáles de todas las alteraciones genéticas presentes en un tumor son las relevantes y cuáles pueden ser dianas terapéuticas es un elemento fundamental para avanzar hacia estos nuevos abordajes terapéuticos en pacientes con cáncer.

En esta línea, hace ya más de siete años nos propusimos contribuir a mejorar este paso de interpretación de genomas del cáncer. Desarrollamos una herramienta llamada *Cancer Genome Interpreter* (CGI) que analiza todas las mutaciones en un tumor e identifica cuál de ellas es una mutación *driver* del tumor y cuáles son mutaciones pasajeras. También consulta bases de datos de biomarcadores de tratamiento y aporta la información relevante sobre si la información genómica del tumor proporciona indicación para algún fármaco dirigido contra ese tumor específico.

Es importante destacar que, para realizar la interpretación de mutaciones, CGI incluye las herramientas computacionales y de aprendizaje automático que hemos desarrollado en el grupo, como IntOGen y BoostDM. De esta forma lo que aprendemos con el análisis sistemático de miles de tumores ayuda a la interpretación de las mutaciones en un nuevo tumor. Tenemos la aspiración de seguir mejorando el aprendizaje automático a partir de tumores de miles de pacientes para mejorar la interpretación del genoma tumoral de un nuevo tumor, para así afinar el diagnóstico molecular y optimizar el tratamiento de futuros pacientes.

NUEVAS FRONTERAS: VARIACIÓN Y SELECCIÓN EN TEJIDOS SANOS

22. Variación y selección en tejidos sanos

Hasta ahora hemos hablado de mutaciones somáticas en tumores. Y es que en los últimos 15 años, el análisis de mutaciones en miles de tumores, como hemos visto, ha aportado grandes avances en nuestro conocimiento de cómo ocurren las mutaciones y cuáles de ellas están implicadas en el desarrollo tumoral. Pero tal como he comentado desde el principio, el proceso de variación y selección es continuo en nuestros tejidos, desde la primera división celular. Un tumor es el resultado final de un largo proceso de evolución darwiniana en el conjunto de células que conforman los tejidos de nuestro cuerpo.

Si pudiéramos estudiar el proceso de evolución darwiniana en tejidos no tumorales, es decir medir la variación y la selección en tejido sano, podríamos obtener pistas de cómo se inicia el cáncer, incluso antes de que haya ocurrido. Pero este estudio es complicado por varias razones. Por un lado, porque así como obtener una muestra del tejido tumoral es una práctica clínica habitual para el diagnóstico de la enfermedad, obtener una muestra de tejido sano no lo es, con algunas excepciones, como por ejemplo muestras de sangre para analíticas. Por otro lado, como en el tejido sano no se ha producido una expansión clonal completa, como en el caso de un tumor, las mutaciones están limitadas a una célula o son compartidas entre pocas células. Estas quedan por debajo del límite de detección con una secuenciación estándar: o bien no las vemos, o las vemos en una o pocas lecturas y no las podemos distinguir de los múltiples errores de secuenciación (ver sección 7). Pero, recientemente, se han producido avances notables con el desarrollo de nuevas aproximaciones que abren la posibilidad de identificar mutaciones en tejido sano de manera acurada, allanando el camino para nuevas investigaciones y descubrimientos que presenciaremos en los años venideros, tal como explicaré más adelante.

23. Hematopoyesis clonal

La sangre es la muestra de tejido sano que más comúnmente se obtiene por razones médicas, y que por tanto se ha estudiado más. También es

uno de los tejidos cuyo desarrollo se conoce mejor. Todas las células de la sangre derivan de las células madre hematopoyéticas que residen en la médula ósea, un tejido esponjoso que se encuentra en el interior de los huesos. Estas células madre hematopoyéticas se dividen, y producen células progenitoras que se diferencian en los distintos tipos de células de la sangre, incluyendo los eritrocitos (glóbulos rojos) que transportan oxígeno, y los linfocitos que participan en la defensa inmunológica, entre muchas otras.

Ya en la década de 1960 se encontraron las primeras evidencias de lo que conocemos como hematopoyesis clonal, un fenómeno estrechamente asociado con la edad. La hematopoyesis clonal se caracteriza por una proliferación y expansión de una población determinada de células sanguíneas derivadas de un ancestro celular común en la médula ósea. Este ancestro, una célula madre hematopoyética, por alguna razón, tiene una ventaja selectiva y prolifera más eficientemente que el resto, dando lugar a un mayor número de células de la sangre. La primera observación de este fenómeno se descubrió mediante estudios de inactivación no aleatoria del cromosoma X en mujeres. Las mujeres tenemos dos cromosomas X, mientras que los hombres solo uno. En todas las células femeninas durante las primeras etapas del desarrollo se inactiva uno de los dos aleatoriamente. Así en la sangre, como en el resto de tejidos, deberíamos encontrar aproximadamente la mitad de las células con uno de los dos cromosomas inactivados y la otra mitad con el otro. Pero en mujeres de edad avanzada se observaron poblaciones de células sanguíneas clonales con una inactivación preferencial de uno de los dos cromosomas X.

En 2014, llegó la evidencia de que la hematopoyesis clonal es un proceso normal de la edad en la población humana causada por mutaciones en el ADN de las células madre hematopoyéticas. Estudios del ADN sanguíneo de miles de personas que se habían colectado inicialmente con el objetivo de estudiar variantes hereditarias asociadas a esquizofrenia, enfermedad bipolar y diabetes, evidenciaron que entre el 10% y el 20% de personas de más de 80 años tienen hematopoyesis clonal, definida como más de un 4% de células de la sangre compartiendo una mutación concreta. En personas menores de 40 años en cambio, raramente se detectan estos clones. La presencia de hematopoyesis clonal va asociada a riesgos para la salud, como el desarrollo de neoplasias hematopoyéticas (por ejemplo leucemia) y una mayor incidencia de enfermedades cardiovasculares⁹³, entre otras patologías.

También en esta época se detectaron los primeros genes implicados en el proceso, como *DNMT3A* y *TET2*, entre otros. Mutaciones concretas en estos genes proporcionan una ventaja selectiva a las células hematopoyéticas que las tienen, que hace que se dividan más y produzcan un mayor número de células sanguíneas, en comparación con otras células madre hematopoyéticas. Es un típico proceso de evolución darwiniana igual que el que hemos explicado para la aparición del cáncer, pero en un tejido sano, sin llegar a desarrollar una patología. Las mutaciones detectadas en la sangre están enriquecidas en C>T y siguen un patrón al que llamamos firma mutacional hematopoyética, que se explica por la combinación de las firmas mutacionales 1 y 5, entre otras⁴⁵.

Teniendo en cuenta que cada año se acumulan unas 14 o 21 mutaciones nuevas en las células de la sangre, y que tenemos del orden de 200 000 células madre hematopoyéticas^{45,94}, podemos calcular fácilmente el nivel de variación que existe en la población de células madre de la sangre de una persona de 10, 40 o 80 años. Esta variación es el primer paso para la evolución darwiniana. Aquellas mutaciones que confieren una ventaja selectiva a estas células para generar más descendencia, con el tiempo pueden acabar representando un porcentaje relativamente elevado de la población de células de la sangre. Dado que las células sanguíneas circulan libremente entre la médula ósea y el torrente sanguíneo, no existen barreras físicas que limiten la competición entre clones, a diferencia de lo que sucede en los tejidos sólidos.

a. Descubriendo los genes de hematopoyesis clonal

Aunque en los últimos años se han identificado algunos de los genes implicados en hematopoyesis clonal, se observan muchos casos de expansión clonal en la sangre para los que no se detectan mutaciones en ningún gen conocido implicado en el proceso. Esto apunta a que probablemente hay otros genes, que cuando están mutados pueden liderar un proceso de expansión clonal en la sangre. Nos planteamos que podíamos contribuir a descubrir estos genes con una aproximación similar a la que habíamos desarrollado para la identificación de genes de cáncer, es decir, estimando la mutagénesis neutral y midiendo selección positiva en el patrón de mutaciones en la sangre⁹⁵.

Para eso lo primero que necesitábamos eran mutaciones somáticas en muestras de sangre de un número importante de individuos. Identificar mutaciones somáticas en la sangre es más complicado que en el tumor.

La sangre es un tejido sano y no hay una expansión clonal completa como en el tumor, por tanto las mutaciones estarán sólo en un subconjunto de las células. El hecho que haya hematopoyesis clonal es lo que nos permite identificar algunas de estas mutaciones, ya que en este caso las podemos ver en un número de lecturas suficiente para no confundirlas con errores de secuenciación. Aun así, tenemos que ir con cuidado de no confundirlas con variantes germinales, las que están en todas las células (frecuencia alélica alrededor de 0.5). Por tanto tenemos un margen bastante limitado para identificar las mutaciones somáticas de la sangre, no podemos detectar mutaciones con muy baja frecuencia alélica, por el riesgo de confundirlas con errores de secuenciación, pero si tienen una frecuencia alélica muy alta las podemos confundir con variantes de la línea germinal.

Razonamos que los datos de miles de pacientes con cáncer que se habían generado para los estudios de genómica del cáncer proporcionaban una oportunidad única para identificar mutaciones somáticas en la sangre. En todos los casos de tumores en tejido no hematopoyético, se había obtenido y secuenciado el ADN de una muestra de sangre del mismo paciente para utilizar como control de variantes germinales en la persona y poder identificar las mutaciones específicas del tumor. Podíamos pues hacer una identificación de mutaciones al revés, es decir utilizando el tumor como control de variantes en la línea germinal para identificar mutaciones somáticas en la sangre. Con una estrategia rigurosa de identificación de mutaciones y utilizando el ADN del tumor como control de variantes en la línea germinal detectamos miles de mutaciones en la sangre y comprobamos que seguían la firma de mutagénesis hematopoyética.

El siguiente paso para identificar genes *drivers* de hematopoyesis clonal era identificar patrones de selección positiva en las mutaciones somáticas en la sangre. Para eso utilizamos la metodología computacional de IntOGen, e identificamos 64 genes con estas señales, que catalogamos como genes *drivers* de hematopoyesis clonal⁹⁵. Entre ellos están todos los genes conocidos implicados en este proceso además de nuevos genes.

En resumen, igual que en el caso de los tumores, detectar señales de selección positiva en el patrón de mutaciones somáticas en el tejido sano, la sangre en este caso, es una manera eficiente de identificar genes cuyas mutaciones confieren ventaja selectiva a células en ese tejido para liderar una expansión clonal.

b. Identificando las mutaciones causantes de hematopoyesis clonal

Igual que en el estudio del cáncer, aunque tengamos un compendio de genes causantes de hematopoyesis clonal, un reto distinto es identificar las mutaciones específicas de estos genes que proporcionan la ventaja selectiva a las células para liderar el proceso de expansión clonal (mutaciones *drivers*). Solo un subconjunto de mutaciones en *DNMT3A* o *TET2* son *drivers* de hematopoyesis clonal; muchas otras no tienen esta capacidad. Hasta el momento, para distinguir unas de otras se utilizan reglas hechas *ad hoc* basadas en el conocimiento que tenemos de cómo funcionaban estas proteínas y las mutaciones que hemos observado, pero igual que para cáncer este conocimiento es sesgado e incompleto. Nos planteamos aprender directamente de los datos de mutaciones observadas en la sangre de miles de individuos con la misma aproximación computacional que diseñamos para identificar mutaciones *driver* de cáncer. Entrenamos modelos de BoostDM para identificar mutaciones *driver* de clonal hematopoyesis en 12 genes implicados en el proceso⁹⁶.

En resumen, al emplear las mismas metodologías de análisis de mutaciones somáticas en tumores para examinar mutaciones somáticas en tejido sano, estamos explorando la evolución darwiniana en estos tejidos. Podemos medir y simular la mutagénesis neutral, detectar señales de selección positiva, y hacer modelos de aprendizaje automático para identificar mutaciones *driver*. Este enfoque nos brinda una amplia oportunidad para realizar nuevos descubrimientos, especialmente ahora que disponemos de metodologías precisas para identificar mutaciones somáticas en tejido sano, como veremos más adelante.

c. ¿Cómo afecta la quimioterapia a la hematopoyesis clonal?

Antes hemos visto cómo algunas quimioterapias afectan al ADN y causan mutaciones tanto en células del tumor como en otros tejidos. En general, las quimioterapias, de una forma u otra, interfieren con la división celular, y en muchos casos causan la muerte de un gran número de células. Desde el punto de vista de evolución darwiniana el tratamiento con quimioterapia implica un cambio drástico en el entorno de los tejidos, que altera las reglas de la selección natural. En el tumor, logra matar muchas células, pero si alguna de ellas tiene alguna alteración heredable que le confiere resistencia al tratamiento, esta célula y su descendencia adquieren una ventaja selectiva en ese nuevo entorno, provocando una recaída.

La quimioterapia, al ser un tratamiento sistémico, también tiene efectos importantes en tejidos sanos. Provoca la muerte de un gran número de células en tejidos de alta proliferación, como el epitelio intestinal o la sangre. Siguiendo con un razonamiento de evolución darwiniana, el cambio en el ambiente en estos tejidos durante el tratamiento podría cambiar su evolución, favoreciendo aquellas células que tienen una ventaja selectiva ante el tratamiento de quimioterapia.

Nos planteamos estudiar el efecto de la quimioterapia en la evolución del tejido hematopoyético. Para ello analizamos las mutaciones identificadas en la sangre de pacientes que habían recibido quimioterapia y las comparamos con otras de pacientes de cáncer que no habían recibido dichos tratamientos. Confirmamos un aumento de la prevalencia de hematopoyesis clonal asociada al tratamiento. Es decir, las personas que han recibido quimioterapia tienen mayor probabilidad de experimentar una expansión clonal en la sangre que sea detectable por medio de la secuenciación. Además, las mutaciones causantes de estas expansiones clonales durante el tratamiento no son las mismas que acostumbramos a ver asociadas a la edad, como *DNMT3A* o *TET2*, sino que preferentemente se seleccionan mutaciones en los genes *PPM1D*, *CHEK2* y *TP53*^{95,97,98}. Estos tres genes codifican proteínas relacionadas con la respuesta al estrés celular y, en particular al daño en el ADN. Podemos especular que aquellas células con mutaciones en estos genes, adquieren una ventaja selectiva durante el tratamiento con fármacos citotóxicos y que causan daño en el ADN. Estas células mutadas pueden sobrevivir y proliferar mejor que otras células de la sangre durante el tratamiento, dando lugar a una expansión clonal detectable tiempo después del tratamiento.

Como la hematopoyesis clonal se encuentra asociada a ciertas condiciones clínicas ligadas al envejecimiento, algunas de las cuales también se observan en personas supervivientes del cáncer, es tentador especular que algunos de los efectos a largo plazo del tratamiento podrían tener relación con este incremento de la hematopoyesis clonal causada por el tratamiento. Esto es especialmente importante para los niños supervivientes del cáncer, ya que una vez superada la enfermedad tienen una larga vida por delante. Por ello estamos estudiando el efecto de la quimioterapia en tejido sano tanto en adultos como en niños. Aún queda un largo camino, pero se esperan nuevos avances para los próximos años no solo en la comprensión de cómo afecta el tratamiento contra el cáncer al ADN creando nuevas mutaciones (como hemos visto en la sección 12), sino también en cómo cambia el juego de la selección modificando el curso de la evolución normal de los tejidos de la persona que recibe el tratamiento.

24. Expansiones clonales en tejidos sólidos

Aunque la sangre es el tejido mejor estudiado, la evolución darwiniana ocurre en muchos otros tejidos de nuestro cuerpo. Una diferencia muy importante entre los tejidos sólidos y la sangre, es que así como las células del sistema hematopoyético pueden viajar y expandirse por todo el torrente sanguíneo, la expansión de clones en tejidos sólidos está altamente restringida por las características espaciales y anatómicas de cada órgano. Por esta razón los clones estarían confinados en una área localizada más o menos pequeña. Esta característica se puede utilizar como una ventaja para estudiar la estructura clonal de estos tejidos.

Una de las primeras evidencias de la existencia de clones somáticos en tejidos sólidos se obtuvo en 2015 de un estudio de mutaciones somáticas en la piel liderado por Iñigo Martincorena y Peter Campbell en Cambridge, Inglaterra⁹⁹. Para solventar el problema de disponibilidad de tejido utilizaron muestras de piel del párpado de cuatro personas que se habían sometido a una cirugía para reparar el problema de caída del párpado (blefaroplastia). Para poder identificar las mutaciones que ocurren solo en un subconjunto de células de la piel, realizaron pequeñas biopsias (de 0.8 a 4.7 milímetros cuadrados), extrajeron el ADN y secuenciaron 74 genes de cáncer con hasta 500 lecturas (500X). Tal como he explicado en la sección 7, aquellas variantes observadas con soporte de varias lecturas tienen mayores probabilidades de ser mutaciones verdaderas, mientras que las observadas en una o muy pocas lecturas podrían ser errores de secuenciación. Con una secuenciación de 500X pudieron identificar las mutaciones de clones de la piel que son suficientemente grandes en relación a la zona de tejido obtenida. Las mutaciones observadas eran mayoritariamente C>T, lo que coincide con lo que esperaríamos dado la exposición a la luz ultravioleta de esta parte de la piel (firma mutacional 7).

Este estudio confirmó que la piel es un mosaico de miles de clones que contienen mutaciones en genes implicados en cáncer. Usando el mismo tipo de aproximación que para el caso de tumores descrita en el apartado 15, se puede observar que el patrón de mutaciones en ciertos genes, como *TP53* y *NOTCH1*, es claramente indicativo de selección positiva.

En estudios de otros tejidos, como los epitelios del esófago^{100,101} y la vejiga¹⁰², se han obtenido resultados similares, confirmando que somos un mosaico de clones expandidos por mutaciones en ciertos genes que confieren selección positiva. Estos clones incrementan en número y en tamaño con la edad, llegando en algunos casos, como en el esófago, a colonizar la

mayor parte del epitelio. Muchos de los genes con selección positiva coinciden con genes de cáncer en tumores del mismo tejido. Pero en algunos casos la frecuencia cambia. Por ejemplo, en el esófago, se observa una frecuencia relativa mayor de clones con mutaciones en el gen *NOTCH1* en tejido sano que la frecuencia que se observa en tumores de esófago. Esto nos hace cuestionar si mutaciones en este gen son *drivers* del cáncer o bien si las mutaciones de este gen que vemos en tumores son un vestigio de su expansión en tejido sano. Se ha sugerido que en realidad las mutaciones del gen *NOTCH1* podrían ser protectoras contra el cáncer en este epitelio porque, al expandirse, estarían ganando la competencia a otros clones potencialmente malignos.

En otros tejidos, como el colon, se pueden secuenciar criptas individuales, confirmando que cada una de ellas es un clon que evoluciona de manera independiente. Pero la gran mayoría de estas no contienen mutaciones *driver*¹⁰³. La estructura del tejido, en que cada cripta se forma a partir de unas pocas células madre en su base, favorece la deriva genética por encima de la selección positiva. En personas con enfermedad inflamatoria intestinal crónica, como la enfermedad de Crohn y la colitis ulcerosa, el panorama cambia completamente¹⁰⁴⁻¹⁰⁶. En este caso se observan clones que se expanden en varias criptas, con mutaciones *driver* en algunos genes de cáncer de colon y otros genes que no causan cáncer. El continuo ambiente inflamatorio del intestino deriva en una evolución completamente distinta del tejido de colon en estos pacientes comparado con personas sin esta condición. En el hígado también se ha observado una estructura clonal distinta en pacientes con enfermedad hepática crónica, como la enfermedad del hígado graso, que en personas sin esta condición¹⁰⁷. La estructura clonal del epitelio de la vejiga también presenta diferencias importantes entre distintas personas, en unos individuos acaban seleccionados preferentemente clones con mutaciones en unos genes mientras que en otros se seleccionan mutaciones en otros genes¹⁰².

En resumen, la posibilidad de secuenciar el ADN de pequeñas estructuras del tejido sano ha abierto una ventana a observar la evolución darwiniana en nuestros tejidos, confirmado que somos un mosaico de clones. Estos clones a menudo contienen mutaciones *drivers* que les confieren ventaja selectiva, y aumentan en número y tamaño con la edad. La estructura clonal es diferente entre distintas personas, en algunos casos estas divergencias se asocian a condiciones de inflamación crónica, mientras que en otros casos aún desconocemos las razones de estas diferencias interindividuales.

25. Las mutaciones *drivers* son necesarias pero no suficientes

La evidencia de miles de clones con mutaciones *drivers* en nuestros tejidos sanos nos plantea una paradoja. Sabemos que las mutaciones *drivers* son necesarias para el desarrollo del tumor porque hemos comprobado que todos los tumores derivan de una célula única con varias mutaciones *driver* que le confieren una ventaja selectiva. Pero a la vez vemos que estas mutaciones no son suficientes para desarrollar un tumor. Miles de células en nuestros tejidos sanos contienen mutaciones *drivers*, que les han proporcionado la ventaja selectiva necesaria para expandirse en forma de clon, pero la gran mayoría de estos nunca desarrollarán un cáncer.

¿Qué diferencia pues al tejido normal con mutaciones *driver* de un tumor? ¿Qué hace que una célula que era capaz de mantener la homeostasis normal se descontrola e inicie una trepidante carrera hacia el desarrollo tumoral?

Para contestar estas preguntas es importante tener en cuenta que las mutaciones por sí solas no tienen ningún efecto. Una mutación puede hacer que un gen codifique para una proteína defectuosa. No obstante, el efecto que esta proteína defectuosa tiene en la célula y el tejido depende completamente del tipo celular donde ha aparecido la mutación, del estado de esta célula, y del contexto en el que se encuentra en el tejido.

Es importante recordar que cada tipo celular constituye un contexto determinado caracterizado por la forma, estructura y función de la célula. Además una célula se encuentra en un espacio determinado de un tejido, en tejidos sólidos comúnmente rodeada por la matriz extracelular, que proporciona un entorno físico y químico para el funcionamiento, comunicación y regulación de las células. Las fuerzas físicas ejercidas por la matriz extracelular y las células vecinas, así como la disponibilidad de espacio, juegan también un papel importante en este contexto celular para determinar si una célula con una cierta mutación *driver* experimentará o no una expansión clonal. La competición entre células también desempeña un papel significativo, ya que la ventaja selectiva de una célula no se define únicamente por sus propias características, sino que se establece en comparación con la capacidad adaptativa de las células vecinas. De la misma manera que la competición entre células juega un papel importante en el desarrollo y mantenimiento de nuestros tejidos, tal como nos cuenta Ginés Morata, también tiene un rol muy importante en la emergencia y desarrollo del cáncer¹⁰⁸.

Así, podríamos tener células en nuestros tejidos con mutaciones “*driver*”, que nunca llegarán a formar un tumor por no encontrarse en la célula y el contexto adecuado. Reproduciendo una metáfora utilizada por la investigadora Mina Bissell, “los genes son como las teclas de un piano: aunque son esenciales, es el contexto el que hace la música”^{109,110}.

Otro elemento importante es que los tumores, tal como hemos visto en la sección 17, tienen múltiples eventos *driver* (una media de 4,3), incluyendo en muchas ocasiones variantes estructurales. En cambio la mayoría de células con mutaciones en tejido sano tienen una o máximo dos mutaciones *driver*. Estas todavía mantienen la homeostasis y su comportamiento normal dentro del tejido. La acumulación de más mutaciones *driver* en la célula adecuada en el contexto adecuado es lo que podría llevar a un crecimiento descontrolado hacia un tumor. Este inicio de crecimiento descontrolado a la vez puede llevar a inestabilidad genómica causando más alteraciones en el genoma. Una manera de verlo es considerar que para que se desarrolle un tumor es probablemente necesario que una célula haya adquirido el conjunto de alteraciones suficientes para saltarse todas las barreras contra el cáncer.

En resumen, las mutaciones *driver*, aunque necesarias, no son suficientes para causar un cáncer. La acumulación de una combinación de alteraciones *driver* en la célula adecuada en el contexto adecuado es lo que puede proporcionar a una célula y sus descendientes las características necesarias para saltarse todas las barreras contra el cáncer.

26. ¿Por qué la probabilidad de que tengamos cáncer aumenta con la edad?

Como hemos visto en la sección 2, la probabilidad de sufrir un cáncer antes de los 40 años es muy pequeña, pero aumenta de manera exponencial a partir de esa edad. Considerando el cáncer como el resultado de un proceso de 5-7 etapas, hemos visto como un simple modelo matemático propuesto por Armitage y Doll podría explicar este incremento exponencial³². Probablemente la realidad es más compleja, y requiere que consideremos tanto la acumulación de mutaciones, como la expansión clonal causada por algunos de estos eventos.

Lo primero que debe ocurrir es que una futura célula tumoral adquiera la primera mutación *driver*. La probabilidad de que una célula adquiera una mutación determinada aumenta con la edad, y se puede ver incremen-

tada por la exposición a factores externos como la luz ultravioleta o el tabaco.

La probabilidad de adquirir una segunda mutación *driver* en una célula con una primera *driver* se ve incrementada en caso que la primera mutación proporcione una ventaja selectiva a esa célula en ese contexto. Si esto ocurre, la célula prolifera a una velocidad mayor que las células colindantes, liderando una expansión clonal, e incrementando por tanto el número de células con una primera mutación *driver* que pueden recibir una segunda mutación *driver*. Y así sucesivamente para adquirir una tercera y cuarta mutación *driver*.

La expansión clonal también requiere tiempo, más o menos tiempo dependiendo de la tasa de crecimiento del clon. Utilizando muestras seriadas de sangre se ha podido calcular la tasa de crecimiento de clones en la sangre con mutaciones en *TET2*, *DNMT3A* y en otros genes¹¹¹. Los clones con mutaciones en *DNMT3A* tienen una tasa de crecimiento de aproximadamente un 5% al año, y los clones con mutaciones en *TET2* del 10%. Aunque por el momento no tenemos datos sobre la tasa de crecimiento de clones con mutaciones *driver* en tejidos sólidos, en cualquier caso es lógico asumir que requiere un cierto tiempo.

Ya hemos visto que las expansiones clonales derivadas de la adquisición de una mutación *driver* son eventos frecuentes en tejidos como la piel, la sangre, el esófago, la vejiga y otros, y que las células con mutaciones *driver* pueden llegar a colonizar un porcentaje alto del tejido a edades avanzadas. Esto es pues un juego de probabilidades de variación y selección, en que las barreras contra el cáncer (ver sección 4) impiden o minimizan mucho la probabilidad de que tengamos cáncer antes de la edad reproductiva, pero a edades posteriores esta probabilidad va aumentando. Si logramos alargar la vida muchos años la probabilidad de desarrollar un tumor será en principio todavía mayor.

Además del tiempo que requiere la acumulación de mutaciones y el crecimiento de los clones que puede explicar en gran medida que el cáncer sea una enfermedad de la edad adulta, otros factores que cambian con la edad podrían también tener algún rol. Por ejemplo, la capacidad del sistema inmunitario de responder tanto a agentes externos como de detectar y eliminar células tumorales decrece con la edad. Hay teorías que apuntan a que el envejecimiento de los tejidos en sí facilita la ventaja selectiva de clones con mutaciones *drivers* a edades avanzadas¹¹².

27. Carcinógenos y factores de riesgo

La incidencia de cáncer varía de manera sustancial entre distintas poblaciones humanas. Por ejemplo, la incidencia de cáncer de esófago muestra diferencias de hasta 30 veces entre distintos países del mundo (Figura 14). También hay diferencias regionales en muchos otros tipos de cáncer, como cáncer de hígado, de pulmón, de mama, y de próstata, entre otros. Muchas de estas diferencias no son de origen genético sino ambiental, ya que se ha observado que las personas migrantes adquieren la tasa de incidencia de su nuevo hogar en 1 o 2 generaciones¹¹³.

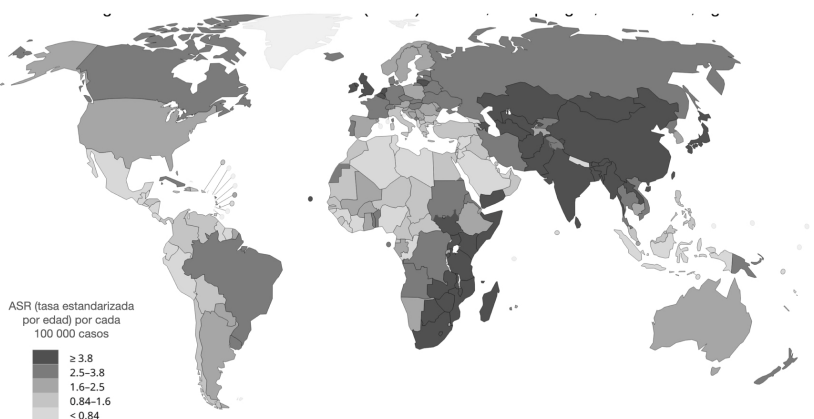


Figura 14. Tasa estandarizada por edad de la incidencia del cáncer de esófago en distintos países del mundo. Datos de GLOBOCAN 2020, incluyendo ambos sexos, y edades entre 0 y 74 años. Mapa obtenido de IARC, WHO (<http://gco.iarc.fr>).

Estudios epidemiológicos han identificado diversos factores que aumentan el riesgo de que una persona desarrolle cáncer. A nivel individual, estos factores no son predictivos, es decir, no pueden predecir con exactitud quién desarrollará cáncer y quién no. No obstante, a nivel poblacional, son altamente precisos en el cálculo de la incidencia de la enfermedad en la población en general. Por ejemplo, las personas fumadoras tienen un mayor riesgo de sufrir varios tipos de cáncer, especialmente cáncer de pulmón. Las personas con obesidad tienen un riesgo mayor de sufrir cáncer de mama, de colon o de endometrio, entre otros tumores. Asimismo, la exposición a una serie de sustancias y agentes externos puede incrementar la probabilidad de sufrir un cáncer. A estos los llamamos carcinógenos. Por ejemplo la luz ultravioleta, algunas sustancias de los cigarrillos de tabaco, el hollín de las chimeneas, y el gas radón, entre muchos otros.

¿Qué determina que unas poblaciones y grupos de personas tengan mayor riesgo de un tipo de cáncer? Siguiendo el razonamiento de evolución darwiniana podría deberse a un aumento de la variación o a un cambio de la selección. Por ejemplo, un factor de riesgo que aumente la tasa de mutaciones en un tejido, como es el caso del tabaco en el epitelio pulmonar, podría incrementar el riesgo de cáncer en ese tejido, cáncer de pulmón en el caso del tabaco. Otros factores en cambio podrían no incrementar la tasa de mutaciones pero aumentar la probabilidad de expansión clonal de células con mutaciones *driver*, elevando también así la probabilidad de desarrollar cáncer.

La creencia más extendida es que el efecto fundamental de los carcinógenos es su efecto mutagénico. Sin embargo, un estudio reciente se secuenció el ADN de tumores de ratones causados por la exposición a 20 carcinógenos distintos¹¹⁴. Se esperaba encontrar un incremento de mutaciones y una firma mutacional para estos carcinógenos. Pero para sorpresa de los investigadores no se observó un aumento en las mutaciones en los tumores causados por 17 de los 20 carcinógenos, en comparación con los tumores que se desarrollaron en el grupo de control de ratones no expuestos a carcinógenos, ni se detectó ninguna firma mutacional específica. Esto apunta a que estos carcinógenos no son mutágenos. Si no son mutágenos, ¿cómo incrementan el riesgo de cáncer?

Nuestra especulación en el artículo de opinión “*Are carcinogens direct mutagens?*”¹¹⁵ es que estos carcinógenos cambian el ambiente del tejido (el contexto), proporcionando una ventaja selectiva a células con mutaciones ya existentes. Así, sin causar nuevas mutaciones, estas sustancias pueden incrementar la probabilidad de que ocurra un cáncer, promoviendo una expansión clonal de células con mutaciones *drivers* (Figura 15).

Esta perspectiva sería equivalente al modelo propuesto por Berenblum y Shubik en 1947 en base a estudios realizados en ratones¹¹⁶. En este modelo propusieron que el desarrollo del cáncer tiene dos etapas, iniciación y promoción. La iniciación equivale a la adquisición de mutaciones en una célula, y la promoción involucra la estimulación del crecimiento de las células iniciadas. Desde el prisma de la evolución darwiniana, la iniciación equivaldría a la generación de la variación y la promoción a la selección positiva y expansión clonal de células mutantes (Figura 16). Un ejemplo de carcinógeno que actúa como promotor sería la polución ambiental, evidencias recientes indican que la polución causa cáncer de pulmón no porque aumente la tasa de mutaciones en las células de pulmón sino porque genera inflamación, la cual permite que los clones mutantes pre-existentes se expandan¹¹⁷.

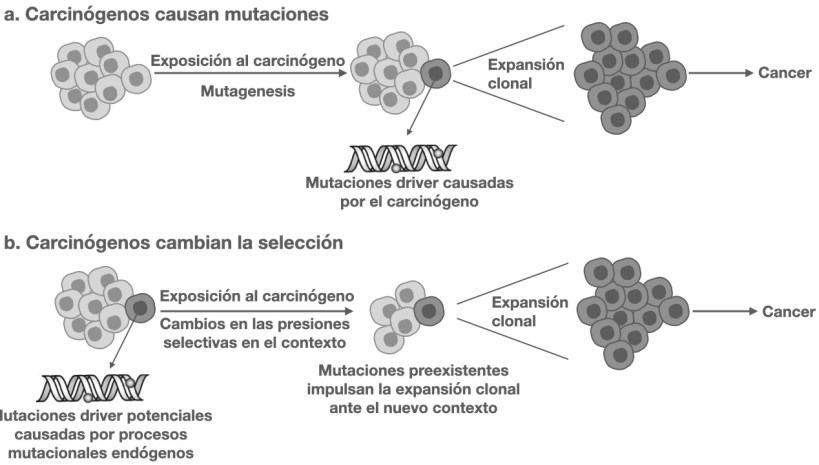


Figura 15. Modelos de cómo actúan los carcinógenos. Adaptado de Lopez-Bigas y Gonzalez-Perez 2020¹¹⁵.

Un carcinógeno, es decir un agente que incrementa el riesgo de cáncer, puede actuar de dos formas distintas, incrementando el número de mutaciones, en este caso sería un mutágeno, o modificando el ambiente en un tejido promoviendo la expansión de células con mutaciones preexistentes, en cuyo caso sería un promotor (Figura 16). Un carcinógeno puede actuar como mutágeno y promotor a la vez, este sería, por ejemplo el caso del tabaco.

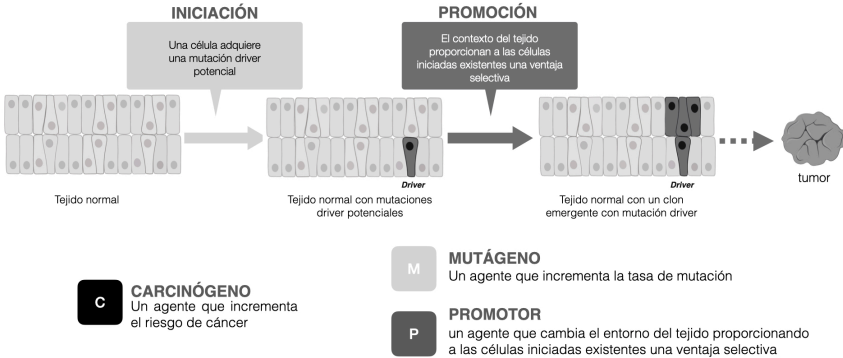


Figura 16. Modelo de iniciación y promoción tal como lo proponemos en el proyecto PROMINENT. Según este modelo un carcinógeno, puede actuar como mutágeno y/o como promotor.

Actualmente estamos trabajando en el proyecto PROMINENT, financiado dentro del marco de los *Cancer Grand Challenges*, para estudiar el efecto de los carcinógenos y factores de riesgo como promotores. Este proyecto está co-liderado por Allan Balmain (Universidad de California San Francisco), Paul Brenan (Centro Internacional de Investigaciones sobre el Cáncer), y yo misma.

El primer problema con el que nos encontramos es que así como hay múltiples maneras de medir el efecto de agentes mutágenos (como la prueba de Ames¹¹⁸, o el estudio de firmas mutacionales como hemos visto en la sección 10 y 12) no existen actualmente pruebas diseñadas para medir el efecto de promotores. Siguiendo la definición de promotor (Figura 18), estos actúan modificando el contexto o entorno en un tejido promoviendo la expansión de células con mutaciones *driver* preexistentes. Una forma adecuada de evaluar el impacto de los promotores sería medir la expansión clonal de mutaciones *drivers* en tejido sano de personas, animales o tejidos en cultivo expuestos a posibles promotores. Actualmente estamos desarrollando este tipo de pruebas como parte del proyecto PROMINENT.

28. Midiendo la estructura clonal para estudiar el riesgo de cáncer

La expansión de ciertos clones en tejido sano podría ser un indicador de riesgo de cáncer. Esto ocurre por ejemplo en la sangre, en que la detección de hematopoyesis clonal es un indicador de riesgo de leucemia. ¿Ocurre lo mismo en tejidos sólidos? ¿Ciertas expansiones clonales correlacionan con riesgo a sufrir un cáncer en ese tejido? Un objetivo importante del proyecto PROMINENT es determinar la relación entre estructura clonal del tejido sano con el riesgo de desarrollar un cáncer en ese tejido.

Los estudios que he descrito en la sección 23 han abierto el camino para medir la estructura clonal de un tejido sano, pero son muy laboriosos y costosos, y por tanto difíciles de escalar. Esto se debe a que se necesita primero obtener pequeñas biopsias de tejido, y realizar la secuenciación de cada una de ellas para detectar los clones y determinar su tamaño. Hasta la fecha, estos estudios se han llevado a cabo en un número limitado de individuos, y en cada caso, se han secuenciado relativamente pocas biopsias.

Sería ideal tener una tecnología que nos permitiera obtener una visión más completa y global de la estructura clonal de los tejidos sanos en cien-

tos o miles de individuos. Esto nos permitiría estudiar si esta estructura clonal se correlaciona con el riesgo de desarrollar distintos tipos de cáncer, e indagar cómo los factores de riesgo o la exposición a carcinógenos modifica la estructura clonal del tejido sano.

Con el objetivo de identificar mutaciones en tejido sano estamos utilizando secuenciación duplex, que ya he mencionado en la sección 12. Es un método de secuenciación de alta precisión que consiste en secuenciar las dos cadenas de una molécula de ADN, de manera que podemos distinguir mutaciones verdaderas (aquellas que vemos en las dos cadenas) de los errores del proceso de secuenciación que se encuentran solo en una de las dos cadenas (*duplex sequencing*)⁷². Así, podemos obtener una muestra extensa del tejido sin recurrir a pequeñas biopsias que podrían contener sólo uno o pocos clones. De este modo, podemos observar cientos de mutaciones *drivers* en múltiples clones. Por ejemplo, en colaboración con la investigadora Rosana Risques, de la Universidad de Washington en Seattle, estamos estudiando la estructura clonal del epitelio de la vejiga de decenas de personas. Las muestras han sido obtenidas con un raspado del epitelio de la vejiga de donantes durante la autopsia. Extraemos ADN del epitelio y hacemos una secuenciación duplex profunda a 5000X, centrándonos en 15 genes implicados en cáncer de vejiga o que se han observado anteriormente como *drivers* de expansión clonal en tejido sano de vejiga. Esto nos permite estudiar una región de la vejiga relativamente amplia, donde pueden haber cientos o miles de clones independientes.

El poder estadístico para detectar selección positiva en este tipo de estudios es muy grande. Es interesante notar que cuando analizamos tumores de vejiga, o de cualquier otro tejido, en cada muestra de un paciente, un tumor, vemos la culminación de una sola expansión clonal. Por tanto, para detectar miles de mutaciones *drivers* en un gen deberíamos estudiar miles de tumores. Pero una secuenciación de alta resolución en un tejido sano como vejiga nos permite identificar centenares o miles de expansiones clonales en una sola muestra.

Con este estudio confirmamos la presencia de expansiones clonales en el epitelio de vejiga, con fuertes señales de selección positiva en genes como *TP53*, *KDM6A*, *ARID1A*, *STAG2* y otros.

En un estudio preliminar de muestras de tan solo unas 25 personas, hemos identificado un orden de magnitud más de mutaciones *driver* en al-

gunos genes en epitelio de vejiga de los que hemos identificado con 15 años de secuenciación de tumores de vejiga. Por ejemplo, en la versión actual de IntOGen (v2023) tenemos datos de 905 genomas de vejiga, y hemos observado 251 mutaciones en el gen *ARID1A*, mientras que en el estudio de muestras de tejido sano de 25 personas ya hemos identificado más de 4000 mutaciones en este gen. Nos podemos imaginar pues que secuenciado muestras de más personas y/o a más profundidad podríamos llegar a observar todas las mutaciones *driver* de la proteína.

Con estos datos confirmamos diferencias muy importantes en la estructura clonal entre individuos. Por ejemplo, mientras en unos individuos se observa una clara preferencia por expansiones clonales con mutaciones en *KDM6A*, en otros las expansiones clonales preferencialmente tienen mutaciones en *STAG2*. ¿Qué es distinto entre estas personas para que la evolución del epitelio de su vejiga siga cursos tan dispares? ¿Se debe a la exposición de estas personas a distintos promotores o factores de riesgo? ¿Y, qué relación tienen estas diferencias en la estructura clonal con el riesgo de desarrollar un cáncer de vejiga?

En PROMINENT proponemos hacer este tipo de estudio para varios órganos con cientos de personas de varias regiones del mundo donde existe distinto riesgo de cáncer en el órgano en cuestión. Las muestras que estamos usando en este estudio forman parte de un biobanco colectado a lo largo de mucho tiempo por el Centro Internacional de Investigaciones sobre el Cáncer (IARC), que forma parte de la Organización Mundial de la Salud de las Naciones Unidas. Estas muestras están muy bien anotadas con información relevante sobre exposiciones y factores de riesgo de cada una de las personas participantes. Con este estudio, que acaba de empezar, esperamos abrir camino a otros que permitan estimar por qué unas personas expuestas a unos factores tienen más o menos riesgo de cáncer.

En el proyecto PROMINENT, en colaboración con Allan Balmain en la Universidad de California en San Francisco y Calvin Kuo, en la Universidad de Stanford, también estamos haciendo este estudio en ratones y organoides. En estos casos podemos controlar muy bien la exposición de los animales a carcinógenos y podemos observar directamente el efecto que estos tienen en generar mutaciones y en promover la expansión clonal de ciertas mutaciones.

29. Prevención del cáncer

En las últimas cuatro décadas, se ha realizado una considerable inversión y esfuerzo en mejorar los tratamientos contra el cáncer, logrando avances notables. Algunos tratamientos dirigidos, como los que he mencionado en la sección 20, y avances en las inmunoterapias han arrojado resultados prometedores. Sin embargo, es importante recordar que en muchos casos, estos tratamientos tan solo logran aumentar el tiempo de vida del paciente unos meses, a menudo con toxicidades elevadas que afectan la calidad de vida en este período. También es importante considerar que los costos asociados a estos tratamientos son muy elevados, planteando desafíos para su acceso equitativo en todo el mundo.

Seguramente estaríamos todos de acuerdo que la mejor estrategia sería evitar que el cáncer ocurra. Estudios epidemiológicos indican que el 40% de la incidencia de cáncer se explica por factores de riesgo conocidos. Los más importantes son el tabaco y la obesidad, seguidos por la exposición a la luz ultravioleta, ciertas exposiciones ocupacionales, algunas infecciones, el consumo de alcohol y la contaminación atmosférica, entre otros. Reduciendo la exposición de la población a estos factores de riesgo tiene el potencial de reducir el número de casos de cáncer de manera sustancial. Esto se ha comprobado en el caso del tabaco. Las medidas antitabaco en los últimos años han tenido un efecto importante en la reducción del riesgo de cáncer de pulmón. Otras medidas como la vacunación contra la hepatitis B o el papiloma están reduciendo la incidencia de cáncer de hígado y cáncer cervical, respectivamente.

¿Cuántos más casos de cáncer serían evitables si se redujera la exposición a factores de riesgo aún desconocidos? Estudios basados en la incidencia diferencial entre poblaciones de un determinado tipo de cáncer, como los propuestos por Richard Doll y Richard Peto en 1977 y 1981 pueden estimar el porcentaje de casos que serían en principio evitables^{119,120}. Esta aproximación se basa en la asunción que hay una tasa basal, no evitable, de un tipo de cáncer, que debería ser igual entre distintos grupos poblacionales. Las diferencias en incidencia, por tanto, indicarían factores que afectan diferencialmente a distintas poblaciones y, al menos una parte de estos factores podrían ser evitables, en principio. Estos estudios, tanto el de 1981 como otro más reciente del 2022¹²¹, han llegado a estimar que hasta el 80% de los casos de cáncer serían evitables en principio. Para una gran parte de éstos no se conocen las causas, y por tanto a nivel práctico, a día de hoy, no lo son. Sin embargo, esto indica que avanzar en comprender las razones de estas diferencias pobla-

les en la incidencia del cáncer podría tener efectos notables para la sociedad en el futuro.

Tenemos la aspiración de que el tipo de estudios propuestos en el proyecto PROMINENT permitan avanzar en nuestra comprensión de cómo los factores y compuestos a los que estamos expuestos a lo largo de nuestra vida modifican el curso de la evolución de nuestros tejidos y cómo esto se asocia a más o menos riesgo de cáncer e incluso de otras enfermedades.

Para esto, en lugar de esperar a que ocurra un cáncer, proponemos estudiar directamente la composición clonal del tejido sano como fenotipo intermedio, y explorar su utilidad como indicador del riesgo de cáncer. Por el momento, en el proyecto PROMINENT, estamos estudiando sangre, tejidos sólidos sanos obtenidos de biopsias, resecciones durante la cirugía para extirpar un cáncer, y durante las autopsias. Sin embargo, en el futuro, para llevar a cabo estudios a nivel poblacional, sería ideal disponer de formas menos invasivas y fácilmente escalables para obtener tejido sano.

Esta nueva era de estudios epidemiológicos en el tejido sano tienen el potencial de impulsar nuestra comprensión de cómo ocurre el cáncer en sus primeras etapas y de cómo distintos factores a los que estamos expuestos afectan este proceso. Esto podría llevar a estrategias más precisas de prevención del cáncer, mejoras en la detección precoz o la identificación de personas en riesgo alto, e incluso desarrollar fármacos que podrían revertir el efecto de promoción del cáncer en etapas tempranas, lo que denominamos promolíticos en el contexto del proyecto PROMINENT.

CONCLUSIONES Y MIRADA AL FUTURO

Durante el discurso he detallado como el estudio del cáncer desde la perspectiva de evolución darwiniana nos permite comprender este conjunto de enfermedades aparentemente tan dispares y complejas. He explicado cómo ha avanzado en los últimos años nuestro conocimiento del cáncer, desde la identificación de las primeras mutaciones relacionadas con el desarrollo tumoral, pasando por la secuenciación de los primeros genomas de tumores, hasta llegar a la compilación de los genes y mutaciones causantes de los principales tipos de cáncer. También he detallado los avances en la comprensión de cómo se producen las mutaciones en el genoma de nuestras células.

Todo este viaje lo hemos hecho a través del estudio de mutaciones somáticas en tumores y más recientemente en tejido sano. Las mutaciones somáticas nos permiten reconstruir la historia del tejido, tanto para identificar los procesos mutacionales que han operado en estas células a lo largo de toda su historia, como los eventos de selección que han ocurrido.

A lo largo del discurso hemos reflexionado sobre el cáncer desde distintas perspectivas evolutivas. Desde la perspectiva macroevolutiva, que opera a escalas de tiempo de millones de años, considerando el origen de la multicelularidad, la evolución del genoma de eucariotas, y la aparición de barreras contra el cáncer. Hasta la perspectiva de la evolución dentro de nuestros tejidos, que opera a escalas de meses y años. Es interesante observar cómo contrastando escalas evolutivas tan distintas nos ayuda a comprender fenómenos biológicos como la tasa de mutaciones en nuestras células, la incidencia del cáncer a distintas edades, e incluso la evolución del genoma.

Muchas preguntas quedan todavía abiertas en cada uno de los capítulos de esta historia y será importante abordarlas en el futuro. ¿Qué otros factores del genoma y la cromatina afectan a la acumulación de mutaciones en nuestros cromosomas? ¿Podemos llegar a una medición precisa de la mutagénesis neutral en cada base del genoma para cada uno de nosotros en cada uno de nuestros tejidos? ¿Seremos capaces no sólo de identificar qué genes y mutaciones son causantes del cáncer, sino de descubrir en qué contexto lo hacen y exactamente cómo lo hacen? ¿Podemos dilucidar la posible cooperación entre mutaciones driver en el proceso de tumorigénesis?

La evolución darwiniana es un proceso continuo que se inicia con la primera división celular de un organismo. Sin embargo, ha sido sólo muy recientemente que somos capaces de observar la estructura clonal y la evolución en tejidos sanos. Esto abre una ventana al futuro para avanzar en nuestra comprensión de las primeras etapas del desarrollo tumoral, y ofrece una perspectiva esperanzadora para mejorar la prevención del cáncer.

Dado el gran número de mutaciones que ocurren por azar en cada una de nuestras células, y el gran número de células en nuestros tejidos, es interesante considerar la gran variación genética presente dentro de cada uno de nosotros. Con un simple cálculo, considerando por ejemplo 1000 mutaciones por célula, dado que el genoma tiene 3200 millones de bases y que tenemos 7,2 billones (7.2×10^{12}) de células nucleadas¹²² nos damos cuenta de que podríamos tener más de 2 millones de células con cada una de todas las posibles mutaciones en el genoma. Es decir toda la variación genética potencial ya existe en nuestros tejidos. Si nos centramos, por ejemplo, solo en las células madre del intestino grueso (se calcula que hay unos 66 millones de células) o del intestino delgado (unos 433 millones) nos damos cuenta que cada posible mutación del genoma humano podría estar presente en unas 20 o 135 células en estos tejidos respectivamente. Este cálculo es una burda simplificación, en que no se tiene en cuenta ni la variación en la tasa de mutaciones por base, ni la jerarquía y relación entre células. Aun así nos ayuda a ser conscientes del nivel de variación genética dentro de nosotros.

En un futuro sería interesante poder calcular con precisión el número de células mutadas que esperamos en cada tejido de cada persona considerando tan solo la mutagénesis neutral. De esta forma podríamos calcular en una persona de cierta edad (ya que el número de mutaciones depende de la edad), con ciertas exposiciones a mutágenos (como tabaco o quimioterapia), cuántas células con ciertas mutaciones estarían presentes en sus tejidos. Con las nuevas tecnologías de secuenciación de alta precisión podemos observar directamente las mutaciones presentes en tejido sano de estas personas, y comparar con lo esperado, lo que nos permitirá reconstruir de manera precisa la historia evolutiva del tejido sano, teniendo en cuenta las mutaciones que ocurren por azar y aquellas que se seleccionan positivamente durante la evolución del tejido. Incluso, estos estudios también nos permitirán explorar el otro lado de la selección, hasta ahora mucho menos accesible: la selección negativa, es decir aquellas mutaciones que no son tolerables para una célula y por tanto las observamos menos de lo esperado por mutagénesis neutral.

Para aquellos tejidos con un nivel alto de selección positiva, como vejiga, sangre, piel y esófago, entre otros, podemos plantear la opción de secuenciar de manera suficientemente profunda como para identificar todas las mutaciones *driver* de expansiones clonales en un gen. Sería como una mutagénesis por saturación natural. Esto nos permitiría incluso calcular la ventaja (o desventaja) selectiva proporcionada por cada mutación en un tejido concreto.

Hasta hace muy poco solo teníamos la capacidad de identificar mutaciones en tumores, lo que nos ha permitido medir selección positiva e identificar genes y mutaciones que han sido seleccionados positivamente durante toda la historia del linaje celular que da lugar al cáncer. Con la identificación de mutaciones en tejido sano y la posibilidad de hacer medidas precisas de selección antes de la aparición del tumor, podemos empezar a tener una visión más precisa de qué genes y mutaciones están directamente implicados en las etapas iniciales o etapas finales del desarrollo de un tumor. Incluso identificar mutaciones que, aunque frecuentemente observadas en cáncer, en realidad podrían ser mutaciones protectoras contra el cáncer, ya que su frecuencia en tejido sano es aún mayor. La definición de qué genes y mutaciones son causantes del cáncer se verá refinada en los próximos años utilizando información de tejido sano.

Una cuestión relevante para abordar en un futuro cercano es determinar si existe correlación en la estructura clonal entre distintos tejidos. Por ejemplo, ¿hay alguna correlación entre la estructura clonal del epitelio bucal con la del esófago o estómago en la misma persona?, ¿las personas tratadas con quimioterapia que tienen mayor hematopoyesis clonal tienen también más expansiones clonales en otros tejidos no hematopoyéticos como la vejiga o el colon? En caso que hubiera una cierta correlación, podríamos medir el grado de expansión clonal en un tejido como indicador de la estructura clonal en otros.

Podemos imaginar un futuro en el que contemos con tecnologías eficientes y mínimamente invasivas implementadas en la práctica clínica que nos permitan medir parámetros relacionados con la evolución clonal de nuestros tejidos, asociándolos al riesgo de desarrollar ciertas enfermedades. En lo que respecta a la sangre esto ya está cerca de convertirse en una realidad. Existen departamentos especializados en la medición de la hematopoyesis clonal en ciertos hospitales de vanguardia¹²³, que aspiran a brindar recomendaciones clínicas para el cuidado de la salud del paciente según el resultado de la medición de la evolución clonal de la sangre.

Sin duda, se necesita realizar más investigación científica y hacer nuevos descubrimientos para avanzar en este camino. No obstante, los progresos de los últimos años generan optimismo respecto a los avances que podrían tener lugar en los próximos años en esta área. En algún momento habremos respondido, al menos en parte a estas preguntas ¿Cómo afecta nuestro estilo de vida y la exposición a lo largo de nuestra vida a ciertos agentes a la evolución somática de nuestros tejidos?, ¿Qué parámetros de evolución clonal se relacionan con el riesgo de desarrollar cáncer y otras enfermedades relacionadas con el envejecimiento?

El estudio del cáncer desde el prisma de la evolución darwiniana nos acompaña en el diseño de experimentos, análisis de mutaciones e interpretación de los resultados. Tal como mencionaba en la introducción “nada en biología tiene sentido excepto a la luz de la evolución”².

BIBLIOGRAFÍA

1. Darwin, C. & QC, D. W. *On the Origin of Species*. (Natural History Museum, London, 2019).
2. Dobzhansky, T. Nothing in Biology Makes Sense except in the Light of Evolution. *Am. Biol. Teach.* **35**, 125–129 (1973).
3. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
4. Surveillance, Epidemiology, and End Results Program. *SEER* <https://seer.cancer.gov/index.html>.
5. Meslé, F. & Vallin, J. Reconstitution de tables annuelles de mortalité pour la France au XIXe siècle. *Popul. Fr. Ed.* **44**, 1121–1158 (1989).
6. Armitage, P. & Doll, R. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
7. Otín, C. L. *Egoístas, inmortales y viajeras: Las claves del cáncer y de sus nuevos tratamientos: conocer para curar*. (Ediciones Paidós, 2021).
8. Greaves, M. & Hughes, W. Cancer cell transmission via the placenta. *Evol. Med. Public Health* **2018**, 106–115 (2018).
9. Greaves, M. *Cancer: The Evolutionary Legacy*. (Oxford University Press, Oxford, 2002).
10. Stammnitz, M. R. *et al.* The evolution of two transmissible cancers in Tasmanian devils. *Science* **380**, 283–293 (2023).
11. Baez-Ortega, A. *et al.* Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science* **365**, eaau9923 (2019).
12. Bruzos, A. L. *et al.* Somatic evolution of marine transmissible leukemias in the common cockle, *Cerastoderma edule*. *Nat. Cancer* **4**, 1575–1591 (2023).
13. Hayflick, L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* **37**, 614–636 (1965).
14. Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat. Rev. Genet.* **6**, 611–622 (2005).
15. Cano, A. *et al.* The transcription factor Snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.* **2**, 76–83 (2000).
16. Battle, E. *et al.* The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89 (2000).
17. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
18. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

19. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
20. Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255**, 197–200 (1975).
21. Kapp, F. G. *et al.* Protection from UV light is an evolutionarily conserved feature of the haematopoietic niche. *Nature* **558**, 445–448 (2018).
22. Peto, R., Roe, F. J., Lee, P. N., Levy, L. & Clack, J. Cancer and ageing in mice and men. *Br. J. Cancer* **32**, 411–426 (1975).
23. Abegglen, L. M. *et al.* Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA* **314**, 1850–1860 (2015).
24. Cagan, A. *et al.* Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).
25. Rous, P. A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS. *J. Exp. Med.* **13**, 397–411 (1911).
26. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300**, 149–152 (1982).
27. Hall, J. M. *et al.* Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. *Science* **250**, 1684–1689 (1990).
28. Wooster, R. *et al.* Localization of a Breast Cancer Susceptibility Gene, BRCA2, to Chromosome 13q12-13. *Science* **265**, 2088–2090 (1994).
29. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
30. The Sequence of the Human Genome | Science. <https://www.science.org/doi/10.1126/science.1058040>.
31. Stephens, P. *et al.* A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat. Genet.* **37**, 590–592 (2005).
32. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
33. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
34. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
35. Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).

36. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
37. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
38. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
39. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
40. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011).
41. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
42. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
43. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
44. Sabarinathan, R. *et al.* The whole-genome panorama of cancer drivers. *BioRxiv* 190330 (2017).
45. Osorio, F. G. *et al.* Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. *Cell Rep.* **25**, 2308–2316.e4 (2018).
46. Moore, L. *et al.* The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
47. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
48. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
49. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
50. Drews, R. M. *et al.* A pan-cancer compendium of chromosomal instability. *Nature* **606**, 976–983 (2022).
51. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).
52. Stamatoyannopoulos, J. A. *et al.* Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
53. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
54. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

55. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
56. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).
57. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
58. Gonzalez-Perez, A., Sabarinathan, R. & Lopez-Bigas, N. Local determinants of the mutational landscape of the human genome. *Cell* **177**, 101–114 (2019).
59. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
60. Hu, J., Adar, S., Selby, C. P., Lieb, J. D. & Sancar, A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev.* **29**, 948–960 (2015).
61. Frigola, J., Sabarinathan, R., Gonzalez-Perez, A. & Lopez-Bigas, N. Variable interplay of UV-induced DNA damage and repair at transcription factor binding sites. *Nucleic Acids Res.* **49**, 891–901 (2021).
62. Mao, P. *et al.* ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nat. Commun.* **9**, 2626 (2018).
63. Fredriksson, N. J. *et al.* Recurrent promoter mutations in melanoma are defined by an extended context-specific mutational signature. *PLOS Genet.* **13**, e1006773 (2017).
64. Herzel, H., Weiss, O. & Trifonov, E. N. 10-11 bp periodicities in complete genomes reflect protein structure and DNA folding. *Bioinforma. Oxf. Engl.* **15**, 187–193 (1999).
65. Audit, B., Vaillant, C., Arneodo, A., d’Aubenton-Carafa, Y. & Thermes, C. Long-range Correlations between DNA Bending Sites: Relation to the Structure and Dynamics of Nucleosomes. *J. Mol. Biol.* **316**, 903–918 (2002).
66. Pich, O. *et al.* Somatic and germline mutation periodicity follow the orientation of the DNA minor groove around nucleosomes. *Cell* **175**, 1074–1087 (2018).
67. Huang, Y., Gu, L. & Li, G.-M. H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J. Biol. Chem.* **293**, 7811–7823 (2018).
68. Frigola, J. *et al.* Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* **49**, 1684–1692 (2017).

69. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat. Genet.* **51**, 1732–1740 (2019).
70. Pich, O. *et al.* The evolution of hematopoietic cells under cancer therapy. *Nat. Commun.* **12**, 4803 (2021).
71. Sánchez-Guixé, M. *et al.* Origins of second malignancies in children and mutational footprint of chemotherapy in normal tissues. *Cancer Discovery.* **14**, 1-12 (2024).
72. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci.* **109**, 14508–14513 (2012).
73. Arnedo-Pac, C., Muiños, F., Gonzalez-Perez, A. & Lopez-Bigas, N. Hotspot propensity across mutational processes. *Mol. Syst. Biol.* **20**, 6–27 (2024).
74. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
75. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
76. Arnedo-Pac, C., Mularoni, L., Muiños, F., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. *Bioinforma. Oxf. Engl.* **35**, 4788–4790 (2019).
77. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **17**, 128 (2016).
78. Tamborero, D., Gonzalez-Perez, A. & Lopez-Bigas, N. Oncodrive-CLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* **29**, 2238–2244 (2013).
79. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169–e169 (2012).
80. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
81. Mangiante, L. *et al.* Multiomic analysis of malignant pleural mesothelioma identifies molecular axes and specialized tumor profiles driving intertumor heterogeneity. *Nat. Genet.* **55**, 607–618 (2023).
82. Rowley, J. D. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature* **243**, 290–293 (1973).
83. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144**, 27–40 (2011).

84. Koche, R. P. *et al.* Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).
85. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
86. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
87. Muiños, F., Martínez-Jiménez, F., Pich, O., Gonzalez-Perez, A. & Lopez-Bigas, N. In silico saturation mutagenesis of cancer genes. *Nature* **596**, 428–432 (2021).
88. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
89. Aaltonen, L. A. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
90. Druker, B. J. *et al.* Activity of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in the Blast Crisis of Chronic Myeloid Leukemia and Acute Lymphoblastic Leukemia with the Philadelphia Chromosome. *N. Engl. J. Med.* **344**, 1038–1042 (2001).
91. Bollag, G. *et al.* Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **11**, 873–886 (2012).
92. Sentís, I. *et al.* The evolution of relapse of adult T cell acute lymphoblastic leukemia. *Genome Biol.* **21**, 284 (2020).
93. Ahmad, H., Jahn, N. & Jaiswal, S. Clonal Hematopoiesis and Its Impact on Human Health. *Annu. Rev. Med.* **74**, 249–260 (2023).
94. Abascal, F. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).
95. Pich, O., Reyes-Salazar, I., Gonzalez-Perez, A. & Lopez-Bigas, N. Discovering the drivers of clonal hematopoiesis. *Nat. Commun.* **13**, 4267 (2022).
96. Demajo, S. *et al.* Identification of Clonal Hematopoiesis Driver Mutations through In Silico Saturation Mutagenesis. 2023.12.13.23299893 Preprint at <https://doi.org/10.1101/2023.12.13.23299893> (2023).
97. Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).
98. Hagiwara, K. *et al.* Dynamics of Age- versus Therapy-Related Clonal Hematopoiesis in Long-term Survivors of Pediatric Cancer. *Cancer Discov.* **13**, 844–857 (2023).
99. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
100. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).

101. Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
102. Lawson, A. R. J. *et al.* Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* **370**, 75–82 (2020).
103. Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).
104. Olafsson, S. *et al.* Somatic Evolution in Non-neoplastic IBD-Affected Colon. *Cell* **182**, 672–684.e11 (2020).
105. Kakiuchi, N. *et al.* Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **577**, 260–265 (2020).
106. Nanki, K. *et al.* Somatic inflammatory gene mutations in human ulcerative colitis epithelium. *Nature* **577**, 254–259 (2020).
107. Brunner, S. F. *et al.* Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
108. Morata, G. & Calleja, M. Cell competition and tumorigenesis in the imaginal discs of *Drosophila*. *Semin. Cancer Biol.* **63**, 19–26 (2020).
109. Bissell, M. J. & Hines, W. C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* **17**, 320–329 (2011).
110. Nelson, C. M. & Bissell, M. J. Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu. Rev. Cell Dev. Biol.* **22**, 287–309 (2006).
111. Fabre, M. A. *et al.* The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature* **606**, 335–342 (2022).
112. DeGregori, J. & Weinberg, R. A. *Adaptive Oncogenesis: A New Understanding of How Cancer Evolves inside Us.* (Harvard University Press, Cambridge, Massachusetts London, England, 2018).
113. Parkin, D. M. & Khlat, M. Studies of cancer in migrants: rationale and methodology. *Eur. J. Cancer Oxf. Engl.* **1990** **32A**, 761–771 (1996).
114. Riva, L. *et al.* The mutational signature profile of known and suspected human carcinogens in mice. *Nat. Genet.* **52**, 1189–1197 (2020).
115. Lopez-Bigas, N. & Gonzalez-Perez, A. Are carcinogens direct mutagens? *Nat. Genet.* **52**, 1137–1138 (2020).
116. Berenblum, I. & Shubik, P. A New, Quantitative, Approach to the Study of the Stages of Chemical Carcinogenesis in the Mouse's Skin. *Br. J. Cancer* **1**, 383–391 (1947).
117. Hill, W. *et al.* Lung adenocarcinoma promotion by air pollutants. *Nature* **616**, 159–167 (2023).
118. Ames, B. N., Durston, W. E., Yamasaki, E. & Lee, F. D. Carcinogens are mutagens: a simple test system combining liver homogenates for

- activation and bacteria for detection. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2281–2285 (1973).
119. Doll, R. Strategy for detection of cancer hazards to man. *Nature* **265**, 589–596 (1977).
120. Doll, R. & Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl. Cancer Inst.* **66**, 1191–1308 (1981).
121. Brennan, P. & Davey-Smith, G. Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed. *JNCI J. Natl. Cancer Inst.* **114**, 353–360 (2022).
122. Hatton, I. A. *et al.* The human cell count and size distribution. *Proc. Natl. Acad. Sci.* **120**, e2303077120 (2023).
123. Sella, T. *et al.* Patient perspectives on testing for clonal hematopoiesis of indeterminate potential. *Blood Adv.* **6**, 6151–6161 (2022).

GRUPO

Inicié mi grupo de investigación en abril del 2006, cuando empezó mi contrato de investigadora Ramón y Cajal en la Universidad Pompeu Fabra. En noviembre del 2016 nos mudamos al Instituto de Investigación Biomédica de Barcelona (IRBB). Durante estos 18 años he tenido la suerte de trabajar con un grupo excepcional de personas que me han inspirado y con quienes hemos hecho conjuntamente los descubrimientos explicados en este discurso.

Con el propósito de subrayar la naturaleza colaborativa de este trabajo, a continuación listo las personas que han formado parte del grupo de investigación en algún momento de su historia.

Abel González-Pérez	Iker Reyes Salazar
Albert Cortés	Inés Sentís
Albert Mascarell	Izar Villasante
Alberto Santos	Joan Enric Ramis
Andrea Grilli	Joan Frigola
Armand Gutierrez	Joe Usset
Axel Rosendahl	Jordi Deu Pons
Carla Castigniani	José Bonet Giné
Carlos Lopez	Josep Miquel Porcar
Carlota Rubio Perez	Judith Bello
Christian Perez Llamas	Khademul Islam
Claudia Arnedo	Katyayani Anshu
Claudia Serrano	Konstantinos Alexiou
David Martinez Millán	Laia Abelló
David Tamborero	Lenka Segura
Davide Polizzi	Loris Mularoni
Electra Tapanari	Magda Guardiola
Elisabet Figuerola	Mari Carmen Lara
Enrique Cerdeira	Maria Andrianova
Erika López Arribillaga	Maria Subirana
Federica Brando	Martina Gasull
Fernando Benito	Medya Shikhagaie
Ferran Muiños	Meritxell Novillo
Ferriol Calvet	Meritxell Rovira Clusellas
Francisco Martínez-Jiménez	Michael P. Schroeder
Gunes Gundem	Miguel Grau
Giuseppe Marceddu	Mònica Sanchez-Guixé
Hanna Kranas	Morena Pinheiro

Núria Samper
Olivia Dove
Olivia Tort
Oriol Pich
Paula Gomis
Ramon Amat
Raquel Blanco
Ricard Lambea
Rocio Chamorro
Ruslana Shadrina
Sabarinathan Radhakrishnan

Santiago Demajo
Santiago Gonzalez
Simon Furney
Sarah Aitken
Sophia Derdak
Stefano Pellegrini
Victor Gonzalez Huici
Tomasz Kisiel
Winona Oliveros
Xavier Rafael-Palou

Agradecimientos extras:

Gracias a Abel Gonzalez-Perez, Anna Bigas, Ferran Muiños, Iñigo Martincorena, Martina Gasull, Francisco Martinez-Jiménez, Rosana Risques, Santiago Demajo y Claudia Arnedo por leer versiones preliminares del discurso, y por sus valiosas opiniones y contribuciones durante su elaboración.

**CONTESTACIÓN
DE LA
EXCMA. SRA. D^a MARÍA ÁNGELA NIETO TOLEDANO**

Excmos Srs. Presidente, miembros de la mesa presidencial y académicos, y Excma. Sra. Núria López Bigas y familia, colegas, amigos, señoras y señores.

Es un honor para mí dar la bienvenida a la Real Academia de Ciencias en nombre de sus miembros a Núria López-Bigas. Candidatura inicialmente presentada por los académicos Carlos López-Otín, David Ríos y yo misma, y recibida con entusiasmo por el Pleno, me corresponde a mí el honor hoy de resumir sus principales logros y cualidades excepcionales. Recibe hoy la medalla número 59, de nueva creación, siendo la undécima mujer que toma posesión como académica de número desde la fundación de esta Real Academia en 1847. Tanto el aumento en el número de académicos como la creciente incorporación de científicas obedece al ánimo decidido de esta institución de avanzar en ambos aspectos, según consta en los Estatutos aprobados el 15 de Diciembre de 2020, de los que todos nos sentimos muy orgullosos. Antes de caer en la tentación de hacer reflexiones al respecto, déjenme decirles que, como mujer, como científica y académica, tengo el orgullo de dar la bienvenida hoy en nombre de la Real Academia de Ciencias a una de las mentes más claras y privilegiadas de la ciencia de nuestro país. Si tuviéramos que definir la personalidad científica de Núria López-Bigas en tres palabras éstas serían visión estratégica, clarividencia y originalidad. Visión estratégica entendida como la habilidad para anticipar y comprender cambios; clarividencia, en su primera acepción del diccionario de la lengua española, la facultad para comprender y discernir claramente las cosas; y originalidad, definida como la aproximación única a la resolución de problemas. Estas cualidades se aprecian de inmediato al seguir y revisar las contribuciones científicas que han situado a la Dra. López-Bigas a la vanguardia de la Biología Computacional dentro y fuera de nuestras fronteras. Mi admiración por esas cualidades únicas ha venido acompañada de la más reciente, la que corresponde al conocimiento de la vertiente más humana de nuestra nueva Académica. Con una personalidad discreta, desprende la tranquilidad y la seguridad de un entorno amable, no solo profesionalmente sino por encima de todo, en su vida familiar.

Núria López-Bigas nació en Monistrol de Montserrat, una villa del siglo XIV con antecedentes históricos desde el siglo IX, en las faldas de la impresionante montaña de Montserrat, en el interior, a menos de una hora de Barcelona. Sus padres, Dolors y Pere, creyeron firmemente en la importancia de la educación de sus hijos, Núria y sus tres hermanos, trabajaron duro para darles las mejores oportunidades de acuerdo con sus inquietudes, y confiaron en el sistema público. Su padre ya no puede acompañarla hoy, pero seguro que le siente presente porque ha sido un pilar en su vida, como lo ha sido y lo es Dolors, su madre. Me dice mi querida amiga Anna Bigas, prima de Dolors y actualmente directora de CIBERONC (Instituto de investigación del cáncer en red del Instituto de Salud Carlos III), que la familia es una piña, piña compacta de la que parte fundamental son Alba, que camina la vida con Núria desde hace más de 20 años, y sus hijas Clara y Jana. Todas ellas están aquí hoy junto con el resto de la familia y del grupo de investigación, que en el caso de los científicos es también casi familia. Enhorabuena a todos. Esta piña es el motivo por el que Núria vive a menos de 20 minutos de Monistrol y a menos de una hora de donde trabaja, a pesar de las numerosas ofertas que ha recibido para trasladarse a alguno de los centros de investigación más prestigiosos del mundo. Es también de justicia añadir que donde trabaja, el Instituto de Investigación Biomédica (Institut de Recerca Biomèdica, IRB) de Barcelona, es un centro de Excelencia, como muestra la acreditación de Centro Severo Ochoa recién obtenida por cuarta vez consecutiva. Enhorabuena una vez más a su director, Francesc Posas, y a Núria y demás miembros del IRB.

Como ya se desprende de lo mencionado, Núria siempre ha tenido una visión clara de hacia dónde quería ir, y para ello estudió Biología en la Universidad de Barcelona. Antes de acabar la carrera visitó el Fred Hutchinson Cancer Research Centre en Seattle, donde Anna Bigas realizaba su estancia postdoctoral. En estas estancias cortas como estudiante de verano Núria adquirió su pasión y determinación por la ciencia. Se doctoró también por la Universidad de Barcelona tras su trabajo en bases genéticas de la sordera hereditaria trabajando, como hemos oído, con los Drs. Xavier Estivill y Mariona Arbonés. Cuando acababa el doctorado tomó una decisión que determinaría el curso de su carrera científica. Tras un periodo de recabar información y de reflexión profunda, Núria le dice a Anna Bigas que ha decidido dejar el laboratorio experimental de poyata y que va a dirigir sus pasos a la biología computacional. Anna recibió esta noticia con sorpresa, pero en la confianza de que, viniendo de Núria, era una decisión meditada y seguramente acertada. Y así fue. Se traslada al *European Bioinformatics Institute* (EBI), parte del laboratorio Europeo de Biología Mole-

cular (EMBL), Institución en la que tengo en honor de ser Vicepresidenta de su Consejo y representar a España. La misión del EBI es “*ayudar a los científicos a aprovechar el potencial de los macrodatos en biología, explotando información compleja para hacer descubrimientos que beneficien a la humanidad*”. Nada describe mejor el trabajo de Núria López-Bigas, quien tras su paso por el EBI trabajando con uno de los pioneros en genómica funcional, Christos Ouzounis, regresa a Barcelona para trabajar con Roderic Guigó, en el Centro de Regulación Genómica (CRG). Su estancia en el laboratorio del Dr. Guigó, le permitió a Núria entrar aún más en contacto estrecho con grandes consorcios internacionales de biología computacional del genoma humano como la *Encyclopedia of DNA Elements* (ENCODE) y posteriormente, ya como investigadora independiente desde 2006 en la Universidad Pompeu Fabra y desde 2016 en el IRB, en consorcios de cáncer como el *Pan-Cancer Analysis of Whole Genomes* (PCAWG), el *International Cancer Genome Consortium* (ICGC) o el *The Cancer Genome Atlas* (TCGA).

Sus trabajos tienen gran reconocimiento internacional como muestran las más de 37.000 citas que figuran en *Google Scholar*, con más de 4.000 citas al año en los últimos años (ver https://scholar.google.com/citations?view_op=list_works&hl=es&hl=es&user=IQcK7oAAAAJ&pagesize=80), sus numerosas invitaciones a prestigiosos congresos internacionales, y sus múltiples premios, incluida la obtención de los proyectos internacionales más prestigiosos y competitivos como los del European Research Council (ERC), y más recientemente, el Cancer Grand Challenge Award (<https://cancergrandchallenges.org/teams/prominent>); la obtención de la membresía de la Organización Europea de Biología Molecular (EMBO), el XI premio del Banco de Sabadell, el premio de la Ciudad de Barcelona o el *Innovator Award de la Sociedad Internacional de Biología Computacional* (ISCB).

Me gustaría ahora detenerme en algunos de sus **hallazgos**, en particular en aquellos **que trascienden** desde mi punto de vista lo esperado de la Biología computacional. Desde los inicios del análisis de los genomas de eucariotas se habían observado patrones intrigantes en la composición de la secuencia genómica, incluyendo una curiosa periodicidad en la secuencia del ADN que envuelve los nucleosomas. Pues bien, como hemos apreciado en su discurso, gracias al trabajo de Núria y su equipo ahora sabemos por qué, y que quizás de forma inesperada, la razón es puramente física, espacial, por una propiedad intrínseca del empaquetamiento del ADN en la célula, y por la disposición más accesible a enzimas de reparación en unas

regiones respecto a otras. Estas conclusiones, obtenidas tras el análisis de mutaciones somáticas presentes en melanomas, cáncer de colon y otros tumores, han permitido describir marcadores para análisis mutacionales evolutivos y descifrar la razón de las tasas mutacionales diferenciales, pero no sólo para el estudio del cáncer, sino para el entendimiento de **mecanismos universales** del empaquetamiento y regulación de la homeostasis del material genético en eucariotas. Estos descubrimientos emocionan, y son el resultado de indagar y persistir en el análisis hasta los mecanismos más íntimos que explican el funcionamiento del proceso en estudio.

Es bien conocido que los tratamientos contra el cáncer como la quimio o la radioterapia no están exentos de efectos secundarios, frecuentemente limitantes para los pacientes mientras se producen. Núria y su equipo fueron más allá y se preguntaron si la quimioterapia dejaba alteraciones en el ADN de las células tumorales, como parecían sugerir evidencias circunstanciales de la aparición más frecuente de tumores secundarios, y encontraron en las células tumorales mutaciones específicas asociadas a tratamientos diferentes. Pero una vez más, estos resultados llevaron a formular una pregunta aún más relevante, ya que los tratamientos quimioterapéuticos son sistémicos. ¿Cuál es el impacto de estos tratamientos en las células no tumorales? La respuesta es que **los tejidos sanos también acumulan las mismas mutaciones** que las células cancerosas. De aquí por tanto inmediatamente surge la necesidad de hacer un análisis mutacional en tejidos sanos ya independiente de la quimioterapia, sumándose a estudios pioneros del análisis de la aparición y acumulación de mutaciones somáticas en distintos tejidos a lo largo de la vida.

Utilizando una aproximación similar a la que habían utilizado para la identificación de genes del cáncer, mutagénesis neutral y selección positiva de mutaciones, realizan un análisis de los genes en un tejido en principio sano pero que muestra tendencias de selección clonal, o expansión de poblaciones específicas procedentes de un ancestro común, fenómeno observado en la sangre y asociado a la edad. Encuentran lo que ahora se considera un fenómeno general, la aparición de clones conteniendo mutaciones específicas, que se expanden por presentar alguna ventaja selectiva, haciendo que con la edad cada tejido esté constituido por un mosaico de clones. El hecho de que algunos de los genes mutados en estos clones de tejido sano sean genes de cáncer nos trae una nueva sorpresa, la conclusión de que los **genes inductores de tumores (*drivers*) son necesarios pero no suficientes para el desarrollo tumoral**. Entonces, ¿qué más hace falta?

En medio de la pandemia, a finales de 2020, aparece un trabajo de los grupos de Allan Balmain (Universidad de California en San Francisco, UCSF) y David Adams (Wellcome Sanger Institute en el Reino Unido) mostrando que los distintos agentes carcinogénicos no parecían inducir mutaciones, en contra de lo que se había pensado durante décadas. Núria escribe un comentario en la misma revista, *Nature Genetics*, apuntando que los **carcinógenos**, en lugar de actuar como mutágenos actúan como **promotores del cáncer en células que ya presentan mutaciones** previas, las conocidas asociadas a tumores. Esto constituye un **cambio de paradigma**, e incita a un estudio exhaustivo de la aparición de mutaciones en tejido sano. Este es el empeño de un gran proyecto en colaboración internacional liderado por Alain Balmain y Núria López-Bigas llamado PROMINENT. Con él visualizamos una **nueva forma de enfrentarse a la lucha contra el cáncer**.

Núria y su grupo también han desarrollado metodologías muy relevantes para el estudio de la genómica del cáncer. Como muy bien apunta nuestro compañero Carlos López-Otín, *“la aplicación de estos métodos al análisis detallado del genoma de más de 28.000 tumores malignos humanos ha facilitado la creación de un censo de genes de cáncer en diferentes tumores (<http://www.intogen.org>). Más recientemente, el laboratorio de Núria ha desarrollado métodos de inteligencia artificial inspirados en la biología evolutiva (boostDM) que han conducido a la identificación de numerosas mutaciones responsables de la carcinogénesis humana. Estos avances han abierto nuevos caminos para el tratamiento personalizado de pacientes con cáncer. En este sentido, el grupo de Núria ha creado una herramienta informática (ver <http://www.cancergenomeinterpreter.org>), denominada Cancer Genome Interpreter (CGI) que es ampliamente utilizada para anotar las mutaciones conductoras y los biomarcadores de respuesta al tratamiento de tumores individuales con orígenes muy diferentes”*.

Para acabar, no me puedo resistir a lanzar una reflexión con la intención de iniciar con Núria López-Bigas una discusión que deseo continúe por muchos años, el mayor placer para los científicos. Discusión entendida en términos anglosajones como describe el Cambridge Dictionary: intercambiar ideas y opiniones, muy diferente de la primera acepción del diccionario de la lengua española que señala como sinónimos de discusión: disputa, controversia, litigio. Pero no se alarmen, la segunda acepción habla de *“comparación de resultados de una investigación a la luz de otros existentes o posibles”*. Esto último es de una gran belleza.

La discusión a la que me refiero es que, tras comprender mucho mejor los cambios que ocurren en el genoma asociados a la tumorigénesis gracias al trabajo de Núria y colaboradores, sabemos que además hay cambios que ocurren más allá de las mutaciones, debidos a los promotores antes mencionados y a un componente muy relevante de interacción de las células cancerosas entre sí y con el microambiente celular. En fin, una demostración más de la evolución de las células pertenecientes a una comunidad, y resultado de presiones de selección. En los últimos años se están obteniendo datos muy valiosos de miles de células únicas de cada tumor, poniendo de manifiesto una heterogeneidad y una complejidad intra-tumoral que no podíamos anticipar, y que es adicional a los cambios en el ADN. Ocurren cambios secuenciales en la transcripción y en el proteoma celular a lo largo con trayectorias complejas que determinan la evolución del tumor y con ello, el pronóstico de los pacientes. Esto se presenta como un nuevo reto, al que idealmente debemos aproximarnos no sólo estudiando sus mecanismos intrínsecos sino a través de la integración de los datos del genoma, el transcriptoma y el proteoma. En ciencia de datos, la integración y el metaanálisis es una práctica habitual, pero también un desafío. Acompañando a los múltiples estudios de transcriptomas y proteomas de células únicas ya hay intentos de integración que resultan muy informativos, pero nos queda mucho camino. ¿Podremos hacerlo? Me gustaría pensar que sí y que con la entrada de Núria López-Bigas en esta Real Academia, éste sea un tema de **discusión** frecuente que dé lugar a nuevas ideas y proyectos que podrían desarrollarse a la luz de la inmensa cantidad de datos disponibles e incipientes. Podríamos alcanzar una visión holística con un análisis de la evolución del cáncer en cuatro dimensiones, con el consiguiente beneficio para el **mejor pronóstico y tratamiento de los pacientes**. En paralelo, sigue su curso la aproximación que la nueva academia está llevando a cabo con sus colaboradores en el gran proyecto PROMINENT, en el empeño de encontrar la naturaleza de las lesiones pre-tumorales, ya existentes en el tejido aún sano, para poder en un futuro desarrollar “promolíticos” y **prevenir el desarrollo del cáncer**. Te deseamos muchos éxitos, los cuales redundarán en el prestigio de la ciencia en España y sobre todo, en el progreso para el beneficio de la sociedad.

Bienvenida a la Real Academia de Ciencias, Núria. Sabes que a partir de ahora ésta también es tu casa, aunque esté un poco más lejos. Te recibimos con cariño y admiración, deseando iniciar el camino de proyectos comunes y de disfrutar de tu contribución al buen hacer de esta casa.

Muchas gracias.