

# VOLUMEN 3 GENOMA Y EPIGENÉTICA

**Coordinadores del tema:**

Lluís Montoliu José  
y Álvaro Rada Iglesias

DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030

Desafíos coordinados por:  
Jesús Marco de Lucas y M.<sup>a</sup> Victoria Moreno-Arribas

VOLUMEN 3

# GENOMA Y EPIGENÉTICA

Reservados todos los derechos por la legislación en materia de propiedad intelectual. Ni la totalidad ni parte de este libro, incluido el diseño de la cubierta, puede reproducirse, almacenarse o transmitirse en manera alguna por medio ya sea electrónico, químico, óptico, informático, de grabación o de fotocopia, sin permiso previo por escrito de la editorial.

Las noticias, los asertos y las opiniones contenidos en esta obra son de la exclusiva responsabilidad del autor o autores. La editorial, por su parte, solo se hace responsable del interés científico de sus publicaciones.

*Catálogo de publicaciones de la Administración General del Estado:*  
<https://cpage.mpr.gob.es>

EDITORIAL CSIC:  
<http://editorial.csic.es> (correo: [publ@csic.es](mailto:publ@csic.es))



© CSIC  
© de cada texto, sus autores  
© de las ilustraciones, las fuentes mencionadas

ISBN Vol. 3: 978-84-00-11009-3  
ISBN O.C.: 978-84-00-11008-6  
e-ISBN Vol. 3: 978-84-00-11011-6  
e-ISBN O.C.: 978-84-00-11010-9  
NIPO: 833-22-101-8  
e-NIPO: 833-22-102-3  
DL: M-15686-2022

Diseño y maquetación: gráfica futura

DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030

# VOLUMEN 3

# GENOMA Y EPIGENÉTICA

**Coordinadores del tema**

Lluís Montoliu José  
y Álvaro Rada Iglesias

## **DESAFÍOS CIENTÍFICOS DEL CSIC: RUMBO AL 2030**

¿Cuáles son los principales desafíos científicos de la primera mitad del siglo XXI?  
¿Podemos establecer las prioridades para el futuro? ¿Cómo debe abordarlos la comunidad científica?

Este libro presenta las reflexiones del Consejo Superior de Investigaciones Científicas (CSIC) sobre 14 temas estratégicos establecidos en función de su impacto científico e importancia social.

Se abordan cuestiones fundamentales como el origen de la vida, la exploración del universo, la inteligencia artificial, el desarrollo de energías limpias, seguras y eficientes o la comprensión del funcionamiento del cerebro. El documento identifica desafíos complejos en áreas como la salud y las ciencias sociales, y los temas estratégicos seleccionados abarcan tanto cuestiones básicas como posibles aplicaciones del conocimiento. En este análisis han participado cerca de 1100 investigadores de más de 100 centros del CSIC y otras instituciones (organismos públicos de investigación, universidades, etc.). Todos coinciden en la necesidad de adoptar un enfoque multidisciplinario y en fomentar la investigación colaborativa que permita poner en marcha proyectos ambiciosos centrados en temas concretos.

Estos 14 *libros blancos*, concebidos como marco de referencia para el desarrollo de la estrategia científica de la institución, permitirán conocer la investigación que se realiza actualmente en el CSIC y, al mismo tiempo, construir una visión global de lo que serán los principales desafíos científicos en la próxima década.

### **VOLÚMENES QUE COMPONEN LA OBRA**

- 1 *Nuevos fundamentos para una sociedad global sostenible*
- 2 *Orígenes, (co)evolución, diversidad y síntesis de la vida*
- 3 *Genoma y epigenética*
- 4 *Desafíos en el ámbito de la biomedicina y la salud*
- 5 *Cerebro, mente y comportamiento*
- 6 *Producción primaria sostenible*
- 7 *Impactos del cambio global*
- 8 *Energía limpia, segura y eficiente*
- 9 *Comprensión de los elementos básicos del universo, su estructura y evolución*
- 10 *Información digital y compleja*
- 11 *Inteligencia artificial, robótica y ciencia de los datos*
- 12 *¿Nuestro futuro? Espacio, colonización y exploración*
- 13 *Desafíos de las ciencias del mar para el 2030*
- 14 *Dynamic Earth: Explorando el pasado, preparando el futuro*

## ***Desafíos científicos del CSIC: rumbo al 2030***

### **Desafíos coordinados por:**

Jesús Marco de Lucas y M.<sup>a</sup> Victoria Moreno-Arribas

### **Volumen 3**

### ***Genoma y epigenética***

#### **Coordinadores del tema:**

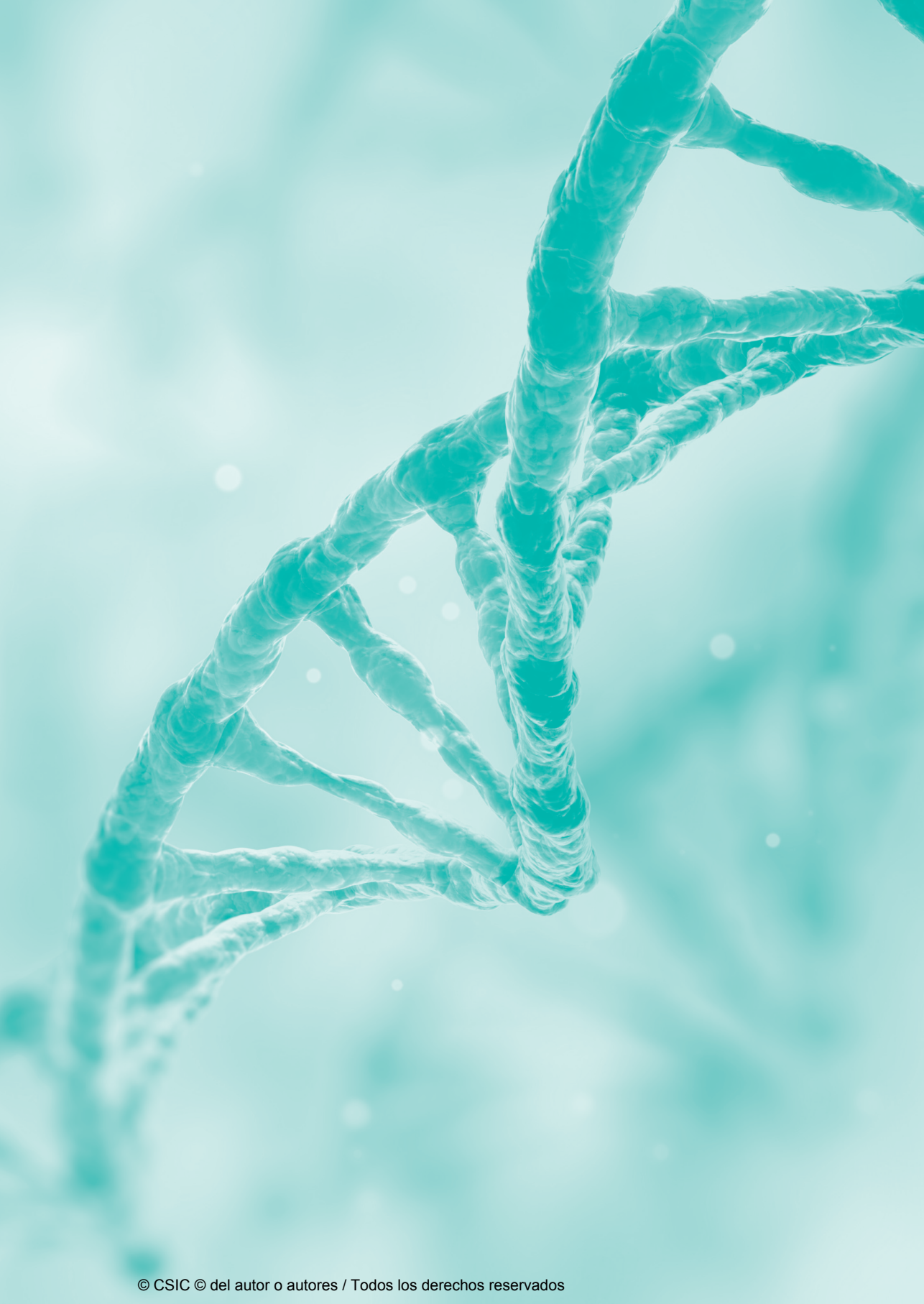
Lluís Montoliu José (CNB, CSIC) y Álvaro Rada Iglesias (IBBTEC, CSIC-UC)

#### **Coordinadores de los capítulos:**

María Domínguez (IN, CSIC-UMH); Pablo Huertas (CABIMER, CSIC-Junta de Andalucía-US-UPO); Javier De Las Rivas (IBMCC, CSIC-USAL); Ana Rojas (CABD, CSIC-Junta de Andalucía-UPO); Ferran Azorín (IBMB, CSIC); Josep Rotllant (IIM, CSIC); Crisanto Gutiérrez (CBM, CSIC-UAM); Cristina Hernández Munain (IPBLN, CSIC); Ángel Barco (IN, CSIC-UMH); María Gómez Vicentefranqueira (CBM, CSIC-UAM); Antonia Herrero (IBVF, CSIC-US); Lourdes Ramos (IQOG, CSIC); Mario Fernández Fraga (CINN, CSIC-UNIOVI-Principado de Asturias) y Sonia Ramos Rivero (ICTAN, CSIC)

#### **Centros participantes:**

Centre for Research in Agricultural Genomics (CRAG, CSIC-IRTA-UAB-UB)  
Centro Andaluz de Biología del Desarrollo (CABD, CSIC-Junta de Andalucía-UPO)  
Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER, CSIC-Junta de Andalucía-US-UPO)  
Centro de Biología Molecular Severo Ochoa (CBM, CSIC-UAM)  
Centro de Investigación en Nanomateriales y Nanotecnología (CINN, CSIC-UNIOVI-Principado de Asturias)  
Centro Nacional de Biotecnología (CNB, CSIC)  
Instituto Botánico de Barcelona (IBB, CSIC-ICUB)  
Instituto de Agroquímica y Tecnología de Alimentos (IATA, CSIC)  
Instituto de Biología Integrativa de Sistemas (I2SysBio, CSIC-UV)  
Instituto de Biología Molecular de Barcelona (IBMB, CSIC)  
Instituto de Biomecánica (IBV, CSIC)  
Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC, CSIC-UNICAN)  
Instituto de Bioquímica Vegetal y Fotosíntesis (IBVF, CSIC-US)  
Instituto de Catálisis y Petroleoquímica (ICP, CSIC)  
Instituto de Ciencia y Tecnología de Alimentos y Nutrición (ICTAN, CSIC)  
Instituto de Ciencias del Mar (ICM, CSIC)  
Instituto de Diagnóstico Ambiental y Estudios del Agua (IDAEA, CSIC)  
Instituto de Investigaciones Biomédicas Alberto Sols (IIBM, CSIC-UAM)  
Instituto de Investigaciones Marinas (IIM, CSIC)  
Instituto de Neurociencias (IN, CSIC-UMH)  
Instituto de Parasitología y Biomedicina López-Neyra (IPBLN, CSIC)  
Instituto de Química Orgánica General (IQOG, CSIC)  
Instituto Universitario de Biología Molecular y Celular del Cáncer (IBMCC, CSIC-USAL)



- 10 **CAPÍTULO 1**  
**MÉTODOS PARA ANALIZAR Y MODIFICAR EL GENOMA**  
**Coordinadores:** María Domínguez y Pablo Huertas
- 30 **CAPÍTULO 2**  
**TECNOLOGÍAS ÓMICAS Y MEDICINA DE PRECISIÓN**  
**Coordinadores:** Javier de las Rivas y Ana Rojas
- 54 **CAPÍTULO 3**  
**ARQUITECTURA TRIDIMENSIONAL DEL GENOMA**  
**Coordinadores:** Ferran Azorín y Josep Rotllant
- 84 **CAPÍTULO 4**  
**EL GENOMA NO CODIFICANTE**  
**Coordinadores:** Crisanto Gutiérrez y Cristina Hernández Munain
- 110 **CAPÍTULO 5**  
**EPIGENÉTICA FUNCIONAL Y EPITRANSCRIPTÓMICA  
Y SU PAPEL EN LA SALUD Y LA ENFERMEDAD**  
**Coordinadores:** Ángel Barco y María Gómez Vicentefranqueira
- 132 **CAPÍTULO 6**  
**GENÓMICA Y EPIGENÓMICA MEDIOAMBIENTAL**  
**Coordinadores:** Antonia Herrero y Lourdes Ramos
- 150 **CAPÍTULO 7**  
**EPIGENÓMICA Y ESTILO DE VIDA**  
**Coordinadores:** Mario Fernández Fraga y Sonia Ramos Rivero





En las últimas décadas, se han secuenciado por completo los genomas de cientos de organismos vivos diferentes. La descodificación de esta gran cantidad de información genética promete desvelar los secretos moleculares de la vida en nuestro planeta, así como la base molecular de las enfermedades humanas. Sin embargo, esto dista mucho de ser trivial, ya que en muchas especies las secuencias que codifican las proteínas (es decir, los genes) solo representan una pequeña fracción de su genoma, mientras que el resto de las secuencias no codificantes desempeñan funciones reguladoras. Además, dentro de los organismos multicelulares, el genoma se utiliza de forma exclusiva en cada tipo de célula introduciendo modificaciones epigenéticas o adoptando conformaciones tridimensionales particulares que pueden afectar a la actividad y expresión de los genes sin cambiar las secuencias de ADN subyacentes. A pesar de esta complejidad, los recientes avances en diversas tecnologías *ómicas* y de edición del genoma han mejorado notablemente nuestros actuales conocimientos sobre la función del genoma. Por consiguiente, ahora estamos en una disposición única para secuenciar, analizar y modificar los genomas y, de ese modo, mejorar no solo nuestra calidad de vida sino también la de nuestro planeta.

## VOLUMEN 3 PALABRAS CLAVE

genómica	epigenética	epigenómica	epitranscriptómica	arquitectura del genoma
edición del genoma	genoma no codificante	medicina de precisión	metagenómica	
microbiota	estilo de vida	genómica medioambiental		

Los autores de este volumen sobre genoma y epigenética dedican esta publicación a la memoria de José Luis Gómez Skarmeta (1966-2020), profesor de investigación del CSIC en el Centro Andaluz de Biología del Desarrollo (CABD).

## CAPÍTULO 1

### RESUMEN

Nuestra capacidad para conocer los sistemas biológicos se ve limitada por nuestra capacidad para identificar, manipular y controlar la información genética. Sin embargo, en los últimos años, los grandes avances técnicos han ampliado nuestras capacidades para predecir, influenciar y *comprender* la información genómica en prácticamente todos los organismos vivos. La capacidad de identificar las enfermedades y los rasgos genéticos y epigenéticos deletéreos redefinirá la investigación biológica y biomédica. Asimismo, los posibles usos de esta revolución técnica y científica y la información que esta genere tendrán consecuencias sobre el diagnóstico y tratamiento de las enfermedades y sobre el proceso de adopción de decisiones de biología reproductiva.

### PALABRAS CLAVE

edición genética edición epigenética  
tecnologías de célula única ómicas  
bioinformática

# MÉTODOS PARA ANALIZAR Y MODIFICAR EL GENOMA

**Coordinadores**

María Domínguez

(IN, Alicante, coordinadora)

Pablo Huertas

(CABIMER, Sevilla, coordinador adjunto)

**Investigadores y centros de  
investigación participantes  
(por orden alfabético)**

Marta Casado

(IBV, Valencia)

José Pascual López Atalaya

Martínez

(IN, Alicante)

Alberto M. Pendás

(CIC, Salamanca)

Lluís Montoliu José

(CNB, Madrid)

Marian Ros

(IBBTEC, Santander)

**RESUMEN EJECUTIVO**

En las dos últimas décadas, hemos asistido a importantes avances tecnológicos que han anunciado el advenimiento de la era genómica. Ha habido notables iniciativas a escala internacional que no solo nos han brindado la información genómica de muy diversas especies, sino que ahora también de diversos individuos de una misma especie. El desarrollo de las técnicas *ómicas* (genómicas, transcriptómicas, epigenómicas, proteómicas, metabolómicas, etc.) también nos aporta varias capas de información que analizar. Y, por último, se ha producido un gran avance al materializarse la posibilidad de transformar prácticamente cualquier genoma de cualquier especie a voluntad gracias a la identificación y popularización de las herramientas de modificación genómica. Por lo tanto, actualmente estamos a punto de ser capaces, por vez primera, de comprender a ciencia cierta la información genómica y manipularla para la investigación básica o aplicaciones concretas.

En este capítulo, repasaremos el estado actual de los métodos disponibles para analizar y modificar el genoma. Además, identificaremos cuáles son los principales desafíos a los que nos enfrentamos en este ámbito y que convendría acometer en los próximos años. Entre ellas, destacaremos la necesidad de implantar y mejorar herramientas para extraer y leer la información procedente de los datos *ómicos*, así como para modificar el genoma.

Por último, analizaremos los puntos fuertes y débiles del CSIC para la investigación en este ámbito. El CSIC se halla en una posición excepcional para contribuir a este propósito, gracias a su dilatada experiencia en ciencias muy diversas, desde la física hasta la biología y desde la informática hasta la química, así como en una amplia gama de modelos experimentales; y esta disposición tan multidisciplinar favorecerá el desarrollo de las herramientas mencionadas. Asimismo, nos consta que en el CSIC existen institutos y laboratorios específicos que son expertos y referentes mundiales en materias concretas que resultan imprescindibles para entender el funcionamiento de estas tecnologías y mejorarlas aún más.

## 1.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

En el presente capítulo, expondremos nuestra percepción actual y los desafíos que plantearán en el futuro las distintas herramientas de investigación, entre las que cabe destacar las siguientes:

- Las estrategias computacionales y los procedimientos bioinformáticos
- Las herramientas de modificación genética y epigenética
- Los nuevos métodos: por ejemplo, las tecnologías de célula única

Uno de los mayores desafíos que plantea la secuenciación del genoma completo reside en la exactitud del ensamblaje de genes grandes y complejos, las regiones repetitivas, y la anotación y ensamblaje de todas las variantes genéticas. Todos estos pasos resultan esenciales para analizar exhaustivamente genomas completos, transcriptomas, epigenomas y la organización tridimensional del genoma. Los métodos computacionales deben optimizarse para permitirnos acceder de forma inmediata y sencilla a todos estos datos *ómicos*. Para que podamos comprender el funcionamiento de la biología, la salud y la enfermedad, es imprescindible definir las secuencias y marcas epigenéticas funcionalmente importantes a partir de datos genómicos y epigenómicos sin procesar.

Para ser capaces de extraer información pertinente de conjuntos de datos de tan magnas proporciones e interpretarlos, es preciso que se produzcan avances determinantes y radicales en los procedimientos bioinformáticos y computacionales de los análisis genómicos funcionales de alto rendimiento, las tecnologías *ómicas* de célula única y las metodologías de modificación genómica. Al mejorar e incrementar su exactitud, las tecnologías de modificación genómica, de modificación genética y epigenética de célula única, y de obtención de imágenes de superresolución a escala de célula única y orgánica nos brindan oportunidades para cumplir el objetivo de hacer avances notables en la investigación genética en humanos y otros organismos (tanto modelos como no modelos).

El progreso en las herramientas de modificación genómica está revolucionando la investigación genética y ha supuesto un estímulo notable para el avance de la genética humana y médica. La modificación del genoma humano pronto será una realidad en la terapia clínica, sobre todo en el caso de los tratamientos *ex vivo*. Además, estas herramientas de modificación también están mejorando las anotaciones funcionales del genoma. La anotación funcional requiere conocer las funciones moleculares, celulares y orgánicas de

todos y cada uno de los genes de un genoma; así como comprender el *contexto* en el que los genes funcionan y reaccionan ante los desafíos ambientales, y la relación entre la variación genética, la epigenética y, en última instancia, la fenotípica. La anotación del genoma humano y de los organismos y microorganismos modelo y no modelo (por ejemplo, la microbiota) está ofreciendo oportunidades sin precedentes para la interpretación biológica de la actividad genómica, la evolución y la diversidad, entre otras cuestiones. Hasta hace pocos años, lo que ralentizaba la anotación de los genes funcionales era la disponibilidad de mutantes. La rapidez y la eficiencia de las tecnologías de modificación genómica están potenciando nuestra capacidad para interrogar a los genomas de cualquier organismo, lo que permite realizar estudios funcionales con una perspectiva evolutiva. Dichos estudios pudieran incluir ahora especies que antes no se podían manipular genéticamente. Asimismo, en algunos casos, pudieran permitirnos llevar a cabo este tipo de experimentos de modificación génica a gran escala. Hasta la fecha, el análisis genético de los organismos no modelo ha revestido una gran complejidad, ha llevado mucho tiempo y se ha limitado a unas cuantas especies. Sin embargo, a medida que se vayan secuenciando (así como anotando y ensamblando) los genomas de especies nuevas, las nuevas herramientas de modificación genómica serán aptas de inmediato para el análisis genético funcional.

La plétora de especies que se analizarán en un futuro a corto plazo resultará útil para **comprender varios procesos biológicos**, entre los que se incluyen la **formación de patrones** (diatomeas y Stentor), la **morfogénesis ramificada** (*Physcomitrella* y *Ashbya*), la **regeneración** (ajolotes e hidras), la **pluricelularidad** (*Volvox*), el **desarrollo y la enfermedad humanos** (organoides derivados de iPS), el **envejecimiento** (kilis y ratopines), el **cáncer** (perros, gatos y ratopines), la complejidad de los **ciclos vitales de los parásitos** (malaria), **las infecciones por coronavirus** (organoides), la **hibernación** (murciélagos), la **adaptación a la hipoxia y al frío** (marmota), y desde el **estrés salino** (cultivos) hasta el **control de las plagas** más complejo mediante las estrategias más innovadoras de manipulación genética dirigida.

Durante esta actual nueva era de la investigación biológica, surgirán nuevas herramientas de manipulación genómica —con toda probabilidad, guiadas por el emparejamiento de bases de Watson y Crick— a partir de proyectos metagenómicos, o dichas herramientas se diseñarán/mejorarán sintéticamente mediante la intervención de la inteligencia artificial (IA). Por tanto,

dada la multitud de procesos que se analizarán *de novo* por primera vez, es probable que asistamos a una explosión de hitos biológicos y avances técnico-científicos sin precedentes históricos, cuyo potencial transformador resulta difícil de prever.

El proyecto **Genome Project-write** (GP-write) es uno de estos avances. El GP-write generará ingeniería del genoma completo con estirpes celulares humanas y otros organismos de relevancia para la agricultura y la salud pública. El centro de interés del proyecto Human GP-write (HGP-write) consistirá en sintetizar genomas humanos en su totalidad o en parte, y su ámbito de trabajo se circunscribirá a las células y los organoides derivados. El objetivo principal del GP-write residirá en ampliar las herramientas de ingeniería genética disponibles y en generar información que vincule la secuencia de bases de nucleótidos del ADN con sus propiedades fisiológicas y comportamientos funcionales para estudiar sus aplicaciones en el ámbito de la sanidad, la energía, la agricultura o la biorreparación. Para poder descifrar la información codificada en el genoma y el epigenoma y manipular esta información con precisión, se requieren tecnologías que trabajen a escala de célula única y a escala genómica en células únicas. Dichas tecnologías revolucionarán la capacidad de traducir la información genómica en una medicina y una asistencia sanitaria precisas y personalizadas, así como de manipular el contenido del genoma para definir y conocer el funcionamiento de los genes, las redes genéticas y las interacciones genético-ambientales.

El uso de **nuevas herramientas de modificación genómica** —tales como CRISPR-Cas9, TALENs y las nucleasas con dedos de zinc (ZFN, del inglés *zinc finger nucleases*)— para la manipulación de genomas de célula única supondrá un desafío considerable, ya que puede aportar información trascendental sobre la capacidad por parte de una sola célula de influir sobre las células adyacentes dentro de los mismos tejidos, en otros tejidos o en el contexto de todo el organismo.

Las **tecnologías de célula única** se están convirtiendo en una herramienta esencial en los estudios biológicos. Estas técnicas nos están brindando la oportunidad de estudiar la heterogeneidad celular y de desenmascarar poblaciones celulares que antes estaban ocultas. Las tecnologías de célula única, unidas a las técnicas de obtención de imágenes en alta resolución —tales como la superresolución, la espectrometría de masas y la secuenciación profunda— permitirán analizar, identificar y dar a conocer subtipos celulares y tipos de células poco habituales para estudiar en profundidad la biología y



anatomía patológica de las células y los órganos. Estas tecnologías poseen el potencial de revelar los cambios dinámicos que se producen continuamente en el tipo/estado de las células a lo largo de procesos biológicos como la diferenciación, la respuesta inmunitaria o la extensión del cáncer. No obstante, los métodos de célula única y genómicos de célula única también plantean una serie de desafíos y limitaciones. Por ejemplo, están apareciendo una serie de metodologías de epigenómica de célula única, pero aún no se sabe con claridad cuál será todo su potencial. La elaboración de perfiles epigenómicos de célula única de células cancerosas, combinada con otros métodos de modificación génica y análisis genómico (secuenciación ATAC, CUT&Tag o scTrio-seq), revolucionará el análisis epigenético y también puede ser útil para modificar células poco habituales, como es el caso de las células madre cancerosas y las células metastásicas. Hay que desarrollar modelos animales validados junto con plataformas de modelos *in vitro* que ofrezcan la oportunidad de investigar las interacciones multicelulares y los procesos dinámicos de varios pasos.

En este sentido, la tecnología de **órganos en chips** (OOC, por las siglas en inglés de *organ-on-a-chip*) ha evolucionado a partir de una combinación de varias plataformas tecnológicas para solventar las dificultades de los modelos convencionales de evaluación de fármacos. Los cultivos de organoides supusieron un gran hito en el cultivo *in vitro* de células tumorales de pacientes y se están convirtiendo en la herramienta más atractiva como plataforma de cribado *in vitro*.

A continuación resumimos los principales desafíos que podemos prever, que desarrollaremos con mayor detalle en las siguientes secciones:

### **Desafíos de las herramientas de modificación genómica:**

- i. Los métodos tienen que ser sólidos, con una elevada eficacia y escasos efectos colaterales.
- ii. Más allá de la modificación génica mediante CRISPR: se están dedicando esfuerzos a desarrollar sistemas de modificación genómica más innovadores y seguros.
- iii. Es necesario que las metodologías permitan modificar cualquier *locus*, independientemente de su posición, estructura y secuencias adyacentes, así como en células aisladas o en el organismo en células somáticas y de la estirpe germinal.

**Implantación de nuevos métodos:**

- i.** Manipulación precisa del contenido genómico de cualquier célula, *in vivo* y en cultivo celular.
- i.** Herramientas bioinformáticas potentes para interpretar la información almacenada en los genomas, así como el efecto de la manipulación para promover la investigación sobre la biología animal y humana, el comportamiento, enfermedades como el cáncer, los trastornos metabólicos, las enfermedades degenerativas y la longevidad.
- ii.** La combinación de cultivos de organoides con desarrollos novedosos en el ámbito de la obtención de imágenes en directo, la ingeniería genética y los materiales biológicos representa una proeza que, en un futuro muy cercano, influirá en nuestra forma de estudiar el desarrollo humano y de tratar las enfermedades humanas.

## 1.2 REPERCUSIÓN SOBRE EL PANORAMA DE LA CIENCIA BÁSICA Y LAS POSIBLES APLICACIONES

La posibilidad de leer con precisión, analizar, interpretar y, por último, controlar a voluntad el genoma y el epigenoma de cualquier organismo parece factible en un futuro cercano. La adquisición y el posterior desarrollo de dichas capacidades darán lugar a una revolución técnica y científica trascendental que afectará por igual a las ciencias básicas y a las aplicadas. Asimismo, suscitarán inquietudes éticas sobre la posible utilidad de esta información genética y su manipulación para tomar decisiones terapéuticas, así como para identificar —y eliminar— rasgos genéticos deletéreos en las fases embrionarias iniciales. Esto último solo será posible tras mejorar los métodos actuales y decidir si convendría modificar las leyes que actualmente prohíben, en España y en muchos otros países, la modificación irreversible del genoma de un embrión humano.

Los investigadores básicos se beneficiarán enormemente de estos avances en la genómica y la epigenómica para adquirir indicios e información nuevos que ayudarán a conocer los procesos biológicos en mayor profundidad y con mayor exhaustividad. Seremos capaces de lo siguiente: interrogar genomas completos y, además, buscar y hallar por medios informáticos los genes o las secuencias genéticas con una mayor probabilidad de estar implicados en el control de procesos específicos o patológicos; extraer información valiosa de las variantes naturales existentes; y analizar y predecir cómo hacen el entorno y las experiencias para controlar la información genética e influir sobre

ella, tanto a escala de las secuencias de ADN como de la epigenética y de la organización tridimensional de la cromatina. Conocer y entender esta información genética nos ayudará a dilucidar cómo se traducen las variantes genéticas y epigenéticas en cambios en el funcionamiento, el rendimiento y el comportamiento de proteínas específicas, ARN no codificantes o su interacción. Es importante señalar que la posibilidad de secuenciar el genoma completo de un animal o una persona a costes cada vez más asequibles pudiera combinarse con la modificación genómica o epigenómica precisa, lo cual será fundamental para el desarrollo de una medicina personalizada y de precisión. La capacidad para demostrar la causa y el efecto también revolucionará y guiará los tratamientos personalizados. Asimismo, las técnicas de célula única y de alta resolución ampliarán el conocimiento actual de varios procesos biológicos o enfermedades, lo cual nos ayudará a entender las dinámicas y la heterogeneidad biológicas que se dan de forma natural y patológica. Por lo tanto, estos avances tecnológicos nos permitirán analizar en detalle procesos que hace tan solo unos años apenas podíamos vislumbrar.

La mejora de las tecnologías genómicas y epigenómicas y el análisis con una resolución de célula única ampliarán estas capacidades a prácticamente cualquier organismo, no solo a los modelos. De este modo, con las nuevas tecnologías pudieran reformularse y abordarse cuestiones biológicas nuevas y antiguas. Por ejemplo, la modificación genómica mediada por CRISPR/Cas9 ya se ha aplicado con éxito en muy diversos taxones de animales, de hongos, de plantas y de microorganismos (ElMounadi *et al.*, 2020; Lee *et al.*, 2020; Schuster y Kahmann, 2019; Sun *et al.*, 2017; y Zhang *et al.*, 2018).

Además, los nuevos desarrollos permiten ahora manipular las regiones reguladoras del genoma, el epigenoma y el genoma no codificante; identificar las interacciones de la cromatina; y etiquetar *loci* específicos para obtener imágenes de células vivas y analizar el genoma (Pickar-Oliver y Gersbach, 2019). La información de las interacciones físicas entre los *loci* y otras regiones genómicas localizadas distalmente, que se obtiene a través de las interacciones de la cromatina de largo alcance, también resulta esencial para entender las dinámicas de la organización tridimensional de la cromatina durante la diferenciación y en respuesta a la proliferación, a los factores de diferenciación y a las hormonas (Gutiérrez *et al.*, 2019). El mapeo sistemático de las interacciones de las proteínas y la ARN-cromatina (Sridhart *et al.*, 2017) y la facilidad para adaptar las tecnologías a las pruebas de cribado del genoma (o epigenoma) completo conformarán una herramienta potente para la investigación

de fenotipos complejos. El uso generalizado de estas novedosas metodologías también pondrá a disposición de la comunidad una enorme biblioteca de herramientas de investigación. Por ejemplo, el ya mencionado proyecto Genome Project-write (GP-write) supondrá uno de los muchos avances en esa dirección. Este proyecto de investigación internacional pretende minimizar el coste de la ingeniería genómica a gran escala utilizando tanto la modificación genómica como la síntesis a gran escala con estirpes celulares de varios organismos. Tanto la tecnología desarrollada al amparo del proyecto GP-write como las estirpes celulares constituirán un recurso muy valioso para la comunidad. Otras propuestas internacionales semejantes aportarán una sólida reserva de variantes genéticas para la investigación.

En cuanto a las posibles aplicaciones de dichas herramientas, la única limitación reside en la imaginación de los investigadores. La mejora de las metodologías de análisis del genoma nos ayudará a comprender la salud y las enfermedades humanas; la forma en que las plantas repercuten sobre el medioambiente y se adaptan a él; la interacción entre el huésped y los patógenos; la biología de la microbiota en el suelo y el intestino; y la aparición y evolución de nuevas enfermedades como el SARS, el MERS y la pandemia por la COVID-19. La facultad de predecir y controlar la información genómica y epigenómica dependerá de nuestra capacidad para modular el genoma con precisión, lo cual ofrece posibilidades prometedoras en lo referente a la reparación o el restablecimiento de la actividad genética asociada a determinadas enfermedades (Doudna, 2020; Picard-Oliver y Gersbach, 2019); la edición del microbioma (Menchaca *et al.*, 2020); la creación de cultivos mejorados mediante modificación genética (Doudna, 2020); y la mejora de los tratamientos veterinarios, lo cual repercutirá positivamente sobre el bienestar y la salud de los animales de compañía y el ganado (Menchaca *et al.*, 2020); entre otras oportunidades. Asimismo, facilitará la generación de animales mutantes para comprender los procesos biológicos y crear modelos de las enfermedades humanas. Estas aptitudes también requerirán acuerdos internacionales para actualizar las normas éticas que regulan la investigación en genómica, que deberán tener en cuenta los posibles riesgos y beneficios para la sociedad humana y el medioambiente.

En general, estamos asistiendo a la aparición de nuevas herramientas que aumentarán drásticamente nuestra capacidad para descifrar y modificar la información genética y epigenética, lo cual revolucionará indudablemente varios campos de investigación que se tratarán con mayor detalle en los siguientes capítulos.

## 1.3 PRINCIPALES PUNTOS PROBLEMÁTICOS

### 1.3.1 Estrategias computacionales y procedimientos bioinformáticos

**Hacia una definición más exacta de lo que es un gen:** desde que se acuñó el término *gen* hace más de cien años, su definición ha ido evolucionando para adaptarse a nuestros conocimientos: desde la representación inicial y abstracta de una unidad hereditaria empleada por Johannsen a principios del siglo XX, pasando por la de «un gen, un ARNm, un polipéptido» de los años 60 hasta el concepto más molecular de una secuencia localizada de nucleótidos del ADN que codifica un ARN, independientemente de que se traduzca o no en uno o varios polipéptidos. De este modo, el término *gen* se ha flexibilizado y su actual definición molecular se ha vuelto más compleja. Uno de los mayores desafíos en materia de análisis bioinformático consistirá en adquirir la capacidad de detectar y definir genes a partir de una secuencia genómica en bruto, y buscar y localizar sus elementos críticos (los puntos de activación y finalización de la transcripción, los intrones y exones, y los elementos reguladores), lo cual se complica aún más en el caso de los genes anidados y de los genes comprendidos por varios genes no codificantes y elementos genéticos móviles. Estos análisis deben posibilitar la descripción de todas las variantes de ARN producidas por cualquier gen, predecir su(s) secuencia(s) codificante(s) (en el caso de los ARNm), sus estructuras (en el caso de los ARN estructurales) o sus funciones (en el caso de los ARNm y los ARN no codificantes [ARNnc]).

**Hacia la creación de herramientas que mejoren el conocimiento sobre el control de la expresión génica:** las reacciones del sistema biológico dependen en gran medida de cómo se regulen los distintos loci genómicos de forma intrínseca y en respuesta a los estímulos ambientales. Por consiguiente, resulta esencial desarrollar nuevas herramientas que nos permitan —a partir de los datos genómicos en bruto— no solo extraer las secuencias reales de los genes, sino también predecir cómo estos pudieran regularse. La información sobre la secuencia se deberá complementar con una caracterización total de las proteínas de la cromatina y sus modificaciones. El desarrollo y la combinación de técnicas que abarcan todo el genoma como, por ejemplo, ChIP-seq, ATAC-seq y BiSseq, nos están ayudando a catalogar las proteínas de la cromatina y sus modificaciones. El desarrollo de análisis computacionales novedosos para estos datos permite segregar esta complejidad en un número discreto de estados de la cromatina que, a su vez, revelan cómo la cromatina dirige funciones tales como la transcripción y el ARN, el procesamiento y, algo que

es crucial, la contribución de la biología de la cromatina a la enfermedad. Los nuevos métodos con resolución de célula única pudieran discriminar aún más los estados transcripcionales entre células individuales (véanse los capítulos 4 y 5 para conocer más detalles sobre estos temas).

**Hacia un análisis de datos genómicos en tiempo real potenciado por herramientas bioinformáticas fáciles de usar:** el uso de herramientas bioinformáticas se está extendiendo rápidamente en la investigación biológica y biomédica. Sin embargo, su uso presenta limitaciones profundas. En primer lugar, los experimentos *ómicos* generan una cantidad ingente de información que hay que almacenar de forma segura, en un formato y un soporte que se puedan recuperar de forma rápida y sencilla. Con esta pronunciada y constante acumulación de datos, dichos experimentos requerirán una mejora del *hardware* y el *software* en los próximos años, y estos avances deberían provenir esencialmente de los ingenieros informáticos. En segundo lugar, una de las restricciones de muchas herramientas bioinformáticas consiste en que requieren una pronunciada curva de aprendizaje. Aunque algunas herramientas son fáciles de conseguir y de utilizar, muchas de ellas están en desarrollo continuo y requieren unos profundos conocimientos informáticos. En el futuro, prevemos que esos métodos y estas herramientas de *software* evolucionarán hacia opciones más eficientes y flexibles, así como a protocolos más fáciles de usar, gracias a lo cual se propagará el uso de herramientas bioinformáticas en la investigación. Estos temas se tratarán exhaustivamente en el capítulo 2.

### 1.3.2 Herramientas de modificación genómica

**Hacia métodos consistentes, eficientes y específicos:** las herramientas de modificación genómica deben ser muy eficientes y específicas, y lo ideal es que sus efectos colaterales sean escasos o nulos. Estos son los principales desafíos a los que nos enfrentamos en la actualidad en lo relativo a la modificación genómica eficiente, y se están dedicando muchos esfuerzos a la búsqueda de soluciones. La mejora de estos aspectos requiere un conocimiento más profundo de los procesos de reparación del ADN, ya que la modificación genómica se basa en engañar a la célula para alterar secuencias genómicas concretas durante tales procesos. Se ha propuesto o se está produciendo un sinfín de modificaciones de las tecnologías CRISPR/Cas para modificar el sistema con el fin de aumentar su eficacia y/o consistencia. Estas modificaciones son muy diversas, y entre ellas se incluyen las del propio sistema CRISPR/Cas9 original (Kleinstiver *et al.*, 2016), la fusión con proteínas específicas de reparación del

ADN (Charpentier *et al.*, 2018; Jayavaradhan *et al.*, 2019; y Rees *et al.*, 2019), y la manipulación temporal de los procesos de reparación (Jinek *et al.*, 2013; Wienert *et al.*, 2020; y Xu *et al.*, 2020a). Asimismo, actualmente se ha enfatizado mucho el desarrollo de nuevas herramientas, ya sea mediante la búsqueda de nuevos sistemas de modificación genómica naturales y más eficientes (véase más adelante) o mediante la alteración directa *in vitro* de las enzimas de modificación genómica, como sucede al crear derivados de CRISPR/Cas9 que no rompen el ADN (Anzalone *et al.*, 2019; y Komor *et al.*, 2016).

**Más allá de la modificación génica mediante CRISPR:** antes de la aparición de la tecnología de CRISPR/Cas9 (Jinek *et al.*, 2012), se implantaron otros métodos, como es el caso de las proteínas con dedos de cinc (ZNF) o las nucleasas TALEN, para la modificación genética (Xu *et al.*, 2020b). Todos ellos tienen en común la explotación de herramientas de biología molecular ya disponibles en la naturaleza o derivadas de nuestros conocimientos sobre el funcionamiento de algunas proteínas nucleares. Por lo tanto, es previsible que en los próximos años descubramos nuevas opciones con solo buscar en distintos organismos. De hecho, ya se utilizan alternativas a la enzima Sp-Cas9 modificada de *Streptococcus pyogenes* que se emplea de forma más generalizada (Jinek *et al.*, 2012; y Kleinstiver *et al.*, 2016). Se han detectado proteínas Cas9 más cortas, procedentes de *Staphylococcus aureus* (SaCas9), *Neisseria meningitidis* (NmCas9), *Streptococcus thermophilus* (St1Cas9) o *Brevibacillus laterosporus* (BlatCas9), que se han aplicado con éxito para la modificación génica (Xu *et al.*, 2020b). Además, a partir de otras bacterias, se han aislado ortólogos de esta familia de proteínas que también se pudieran aplicar a tal efecto. Aparte de los distintos ortólogos de Cas9, también se han descubierto otras proteínas Cas, entre las que se incluyen la Cpf1 (CRISPR de *Prevotella* y *Francisella* 1, también conocida como Cas12a), y las diversas variantes de Cas13 y Cas14 (Abudayyeh *et al.*, 2016; Konermann *et al.*, 2018; Strecker *et al.*, 2019; y Zetsche *et al.*, 2015), algunas de las cuales propician nuevas aplicaciones innovadoras como es el caso del diagnóstico genético (Gootenberg *et al.*, 2017). Además, se ha propuesto que otras enzimas no relacionadas con las proteínas Cas, como es el caso de las enzimas CasX, pudieran llevar a cabo la modificación genómica guiada por ARN (Liu *et al.*, 2019). Así pues, la incorporación de enzimas del tipo Cas9 alternativas o incluso la detección de sistemas completamente nuevos para llevar a cabo la modificación génica darán lugar en un futuro cercano a una infinidad de herramientas de modificación genómica con capacidades distintas.



**Hacia unas herramientas de modificación genómica más flexibles y generales:** las metodologías deberían posibilitar la modificación de cualquier *loci*, independientemente de su posición, estructura y secuencias adyacentes, así como en células somáticas o de la estirpe germinal de organismos enteros. Uno de los mayores desafíos para las herramientas de modificación génica ha sido ampliar su campo de aplicación más allá del cultivo de células *in vitro*, de forma que resulten útiles en aplicaciones médicas y para la creación de organismos modificados genéticamente (Doudna, 2020; Lee *et al.*, 2020; Menchaca *et al.*, 2020; Seruggia *et al.*, 2015; Sun *et al.*, 2017; Xu *et al.*, 2019; y Zhang *et al.*, 2018). En vista del desarrollo y las ventajas que aportan las tecnologías de modificación genómica, en los últimos años ha aumentado considerablemente el número de estudios en los que se emplea la modificación genómica para mejorar los cultivos hortícolas. La combinación de esta tecnología de modificación genómica en rápido avance con los cruces genéticos aumentará en gran medida la producción y la calidad de dichos cultivos (Xu *et al.*, 2019). En organismos modelo como la mosca de la fruta (*Drosophila melanogaster*), la capacidad por parte de los investigadores para diseñar modificaciones del genoma dirigidas en estudios de genes y elementos genéticos se ha visto transformada significativamente por la generación de moscas transgénicas que expresan las variantes Cas9 y Cas9 modificadas (Gratz *et al.* 2013; y Ewen-Campen *et al.* 2017) y de transgenes a fin de expresar ARN guía sintéticos (ARNgs) para la interrupción, eliminación o activación de los genes. En un futuro cercano, se dispondrá de herramientas que permitan realizar modificaciones genómicas en animales de forma cada vez más precisa (Lino *et al.*, 2018). No obstante, las herramientas de modificación genómica siguen sufriendo de un componente secuencia-contexto muy acusado que limita las secuencias que se pueden delimitar de forma eficiente. Por ello, habrá que poder relajar o romper dichas restricciones con el fin de modificar cualquier secuencia dada, independientemente del contexto genómico. En este sentido, también será pertinente el desarrollo de nuevos métodos bioinformáticos para mejorar el diseño y la selección de guías de ARN únicas y más eficientes que se emplearán en combinación con las tecnologías de CRISPR-Cas (Oliveros *et al.*, 2016; y Torres-Pérez *et al.*, 2019).

### 1.3.3 Nuevos métodos de análisis y visualización del genoma

**Cultivos tridimensionales y de organoides:** un *organoide* es una estructura tridimensional derivada de células troncales y contiene tipos celulares de órganos específicos que se autoorganiza y se asemeja a la característica fisiológica de dicho órgano (Clevers, 2016; y Yin *et al.*, 2016). La importancia de estos



cultivos de órganos *en placas* es que pueden recapitular y, con el tiempo y al ir mejorando, simular el microambiente natural de un órgano, lo que permite a los investigadores plantear preguntas más complejas sobre el funcionamiento del (epi)genoma humano durante el desarrollo o en reacción a diversos estímulos. Además, en la investigación biomédica, tienen el potencial de crear modelos más exactos de las enfermedades humanas mediante células troncales pluripotentes derivadas de pacientes (Yin *et al.*, 2016) o, en el futuro, de ser una fuente de órganos para trasplantes en la medicina regenerativa. Los organoides humanos también pueden superar las diferencias genéticas y de dosificación génica que a veces existen entre los humanos y los organismos modelo y que pueden complicar el estudio de determinados trastornos humanos. **Sin embargo, el verdadero desafío consiste en crear los propios organoides. Aunque en el laboratorio se han creado con éxito organoides de varios órganos, estos carecen de la complejidad y organización de los órganos reales.** Por lo tanto, la mejora de los organoides supone un desafío importante en el futuro cercano. Además, trabajar con los organoides continúa resultando difícil, exigiendo mucho tiempo y esfuerzo y teniendo un coste superior al de los cultivos bidimensionales tradicionales, y su estudio mediante técnicas estándar sigue resultando más complejo (por ejemplo, se complica al tratar de aplicar dispositivos de obtención de imágenes) (Jensen y Teng, 2020).

**Nuevos desarrollos en el ámbito de la obtención de imágenes en directo:** aunque la relación entre las técnicas de adquisición de imágenes y las herramientas genómicas no es obvia, existen entre ellas sinergias interesantes que conviniere explorar en el futuro. Como ya se ha mencionado, la modificación genómica puede aportar nuevas herramientas para la obtención de imágenes en directo al ayudar a etiquetar y visualizar regiones específicas del ADN, lo que, por ejemplo, puede mejorar el conocimiento actual de la organización tridimensional del genoma y superar algunas de las limitaciones de los métodos genómicos masivos (por ejemplo, la heterogeneidad y la dinámica) (revisado en Pickar-Oliver y Gersbach, 2019). Por otra parte, el desarrollo de la microscopia celulómica de alto rendimiento pudiera complementar a la perfección los datos obtenidos mediante técnicas *ómicas* de célula única. Además, resultará pertinente para racionalizar los procedimientos que vinculan los cribados del genoma completo con los resultados obtenidos mediante microscopia.

**Ómica de célula única:** la gran ventaja de efectuar análisis *ómicos* de células sueltas es que permite a los investigadores tener en cuenta la heterogeneidad natural de los sistemas biológicos, ya que, en lugar de promediar la señal de

varias células en un solo resultado, se conserva la información de cada una de las células. En la actualidad, poseemos la capacidad de interrogar a células de manera individual para diversas capas *ómicas*, ya se trate del genoma, el transcriptoma, el epigenoma o el genoma tridimensional (Chappell *et al.*, 2018). A pesar de ello, aún quedan desafíos por resolver en el futuro. Todavía cuesta efectuar algunas tecnologías *ómicas* con una resolución de célula única debido al escaso material biológico, por lo que todavía hay margen de mejora en esa dirección. Se están implantando métodos de reciente desarrollo para posibilitar un mapeo de alto rendimiento de la expresión génica a escala de célula única dentro de los tejidos (transcriptómica espacial). Además, el coste de dichas tecnologías sigue siendo elevado, y habrá que mejorar el componente computacional para mejorar el análisis. Por último, la combinación de varias de estas tecnologías en la multiómica de célula única experimentará un mayor desarrollo en un futuro cercano (Chappell *et al.*, 2018).

## CAPÍTULO 1 BIBLIOGRAFÍA

- Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L. *et al.* (2016). C2c2 is a single-component programmable RNAGuided RNA-targeting CRISPR effector. *Science* 353, aaf5573.
- Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. *et al.* (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149-157.
- Bastida, M.F., Pérez-Gómez, R., Trofka, A., Zhu, J., Rada-Iglesias, A., Sheth, R., Stadler, H.S., Mackem, S. y Ros, M.A. (2020). The formation of the thumb requires direct modulation of Gli3 transcription by Hoxa13. *Proc. Natl. Acad. Sci. USA*. 117, 1090-1096.
- Caburet, S., Arboleda, V.A., Llano, E., Overbeek, P.A., Barbero, J.L., Oka, K., Harrison, W., Vaiman, D., BenNeriah, Z., García-Tuñón, I. *et al.* (2014). Mutant cohesin in premature ovarian failure. *N. Engl. J. Med.* 370, 943-949.
- Chappell, L., Russell, A.J.C. y Voet, T. (2018). Single-Cell (Multi) omics Technologies. *Annu. Rev. Genomics Hum. Genet.* 19, 15-41. doi: 10.1146/annurev-genom-091416-035324.
- Charpentier, M., Khedher, A.H.Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S. *et al.* (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9, 1-11.
- Chneiweiss, H., Hirsch, F., Montoliu, L., Müller, A.M., Fenet, S., Abecassiss, M., Merchant, J., Baertschi, B., Bothol-Baum, M., Houghton, J.A. *et al.* (2017). Fostering responsible research with genome editing technologies: a European perspective. *Transgenic Res.* 26, 709-713.
- Clevers, H. (2016). Modeling Development and Disease with Organoids. *Cell* 165, 1586-1597.
- Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. *Nature* 578, 229-236.
- El-Mounadi, K., Morales-Floriano, M.L. y García-Ruiz, H. (2020). Principles, Applications, and Biosafety of Plant Genome Editing Using CRISPR-Cas9. *Front. Plant Sci.* 11, 1-16.
- Ewen-Campen, B., Yang-Zhou, D., Fernandes, V.R. *et al.* (2017). Optimized strategy for in vivo Cas9-activation in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. 114(35):9409-9414.
- Gratz, S.J., Wildonger, J., Harrison, M.M. y O'Connor-Giles, K.M. (2013). CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly* 7(4), 249-255. doi: 10.4161/fly.26566. Publicación electrónica: 2 de octubre de 2013.
- Fernández-Albert, J., Lipinski, M., López-Cascales, M.T., Rowley, M.J., Martín-González, A.M., Del Blanco, B., Corces, V.G. y Barco, A. (2019). Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat. Neurosci.* 22, 1718-1730.
- Gómez-H, L., Felipe-Medina, N., Sánchez-Martín, M., Davies, O.R., Ramos, I., García-Tuñón, I., de Rooij, D.G., Dereli, I., Tóth, A., Barbero, J.L. *et al.* (2016). C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nat. Commun.* 7, 13298.
- Gómez-H, L., Felipe-Medina, N., Condezo, Y.B., García-Valiente, R., Ramos, I., Suja, J.A., Barbero, J.L., Roig, I., Sánchez-Martín, M., de Rooij, D.G. *et al.* (2019). The PSMA8 subunit of the spermatoproteasome is essential for proper meiotic exit and mouse fertility. *PLoS Genet.* 15, e1008316.
- Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A. *et al.* (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356, 438-442.
- Gutiérrez-Pérez, I., Jordan-Rowley, J., Lyu, X., Valadez-Graham, V., Vallejo, D.M., Ballesta-Illán, E., López-Atalaya, J.P., Kremsky, I., Caparrós, E., Corces, V.G., Domínguez, M. (2019). Ecdysone-Induced 3D Chromatin Reorganization Involves Active Enhancers Bound by Pipsqueak and Polycomb. *Cell Rep.* 28, 2715-2727.e5.

- Hellmuth, S., Gutiérrez-Caballero, C., Llano, E., Pendás, A.M. y Stemmann, O. (2018). Local activation of mammalian separase in interphase promotes double-strand break repair and prevents oncogenic transformation. *EMBO J.* 37(22), e99184.
- Hirsch, F., Lemaitre, C., Chneiweiss, H. y Montoliu, L. (2019). Genome Editing: Promoting Responsible Research. *Pharmaceut. Med.* 33, 187-191.
- Inserte, J., Mollá, B., Aguilar, R., Través, P.G., Barba, I., Martín-Sanz, P., Boscá, L., Casado, M. y García-Dorado, D. (2009). Constitutive COX-2 activity in cardiomyocytes confers permanent cardioprotection Constitutive COX-2 expression and cardioprotection. *J. Mol. Cell. Cardiol.* 46, 160-168.
- Jayavaradhan, R., Pillis, D.M., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P.R. y Malik, P. (2019). CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nat. Commun.* 10, 1-13.
- Jensen, C. y Teng, Y. (2020). Is It Time to Start Transitioning From 2D to 3D Cell Culture? *Front. Mol. Biosci.* 7, 1-15.
- Jimeno, S., Mejías-Navarro, F., Prados-Carvajal, R. y Huertas, P. (2019). Controlling the balance between chromosome break repair pathways. *Advances in Protein Chemistry and Structural Biology* 4, 95-134.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. y Charpentier, E. (2012). A programmable dualRNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816-821.
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. y Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife* 2013, 1-9.
- Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. y Joung, J.K. (2016). Highfidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490-495.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. y Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420-424.
- Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N. y Hsu, P.D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* 173, 665-676.
- Lee, H., Yoon, D.E. y Kim, K. (2020). Genome editing methods in animal models. *Animal Cells Syst. (Seúl)*. 24, 8-16.
- Lino, C.A., Harper, J.C., Carney, J.P. y Timlin, J.A. (2018). Delivering crispr: A review of the challenges and approaches. *Drug Deliv.* 25, 1234-1257.
- Lipinski, M., Muñoz-Viana, R., Del Blanco, B., Márquez-Galera, A., Medrano-Relinque, J., Caramés, J.M., Szczepankiewicz, A.A., Fernández-Albert, J., Navarrón, C.M., Olivares, R. *et al.* (2020). KAT3-dependent acetylation of cell type-specific genes maintains neuronal identity in the adult mouse brain. *Nat. Commun.* 11, 2588.
- Liu, J.J., Orlova, N., Oakes, B.L., Ma, E., Spinner, H.B., Baney, K.L.M., Chuck, J., Tan, D., Knott, G.J., Harrington, L.B. *et al.* (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 566, 218-223.
- López-Saavedra, A., Gómez-Cabello, D., Domínguez-Sánchez, M.S., Mejías-Navarro, F., Fernández-Ávila, M.J., Dinant, C., Martínez-Macías, M.I., Bartek, J. y Huertas, P. (2016). A genome-wide screening uncovers the role of CCR2 as an antagonist of DNA end resection. *Nat. Commun.* 7, 12364.
- Menchaca, A., Santos-Neto, P.C., Mulet, A.P. y Crispo, M. (2020). CRISPR in livestock: From editing to printing. *Theriogenology* 150, 247-254.
- Moncayo-Arlandi, J., Guasch, E., Sanz de la Garza, M., Casado, M., García, N.A., Mont, L., Sitges, M., Knöll, R., Buyandelger, B., Campuzano, O. *et al.* (2016). Molecular disturbance underlies to arrhythmogenic cardiomyopathy induced by transgene content, age and exercise in a truncated PKP2 mouse model. *Hum. Mol. Genet.* 25, 3676-3688.
- Montoliu, L., Merchant, J., Hirsch, F., Abecassis, M., Jouannet, P., Baertschi, B., Sarrauste de Menthère, C. y Chneiweiss, H. (2018). ARRIGE Arrives: Toward the Responsible Use of Genome Editing. *Cris. J.* 1, 128-129.

- Motino, O., Francés, D.E., Casanova, N., Fuertes-Agudo, M., Cucarella, C., Flores, J.M., Vallejo-Cremades, M.T., Olmedilla, L., Pérez Peña, J., Bañares, R. *et al.* (2019). Protective Role of Hepatocyte Cyclooxygenase-2 Expression Against Liver Ischemia-Reperfusion Injury in Mice. *Hepatology* 70, 650-665.
- Oliveros, J.C., Franch, M., Tabas-Madrid, D., San León, D., Montoliu, L., Cubas, P. y Pazos, F. (2016). Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. *Nucleic Acids Res.* 44, W267-71.
- Pickar-Oliver, A. y Gersbach, C.A. (2019). The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* 20, 490-507.
- Rees, H.A., Yeh, W.H. y Liu, D.R. (2019). Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* 10, 2212.
- Remesal, L., Roger-Baynat, I., Chirivella, L., Maicas, M., Brocal-Ruiz, R., Pérez-Villalba, A., Cucarella, C., Casado, M. y Flames, N. (2020). PBX1 acts as terminal selector for olfactory bulb dopaminergic neurons. *Development* 147, dev186841.
- Saiz-López, P., Chinnaiya, K., Campa, V.M., Delgado, I., Ros, M.A. y Towers, M. (2015). An intrinsic timer specifies distal structures of the vertebrate limb. *Nat. Commun.* 6, 8108.
- Schuster, M. y Kahmann, R. (2019). CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genet. Biol.* 130, 43-53.
- Seruggia, D., Fernández, A., Cantero, M., Pelczar, P. y Montoliu, L. (2015). Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis. *Nucleic Acids Res.* 43, 4855-4867.
- Soria-Bretones, I., Cepeda-García, C., Checa-Rodríguez, C., Heyer, V., Reina-San-Martín, B., Soutoglou, E. y Huertas, P. (2017). DNA end resection requires constitutive sumoylation of CtIP by CBX4. *Nat. Commun.* 8(1), 113.
- Sridhar, B., Rivas-Astroza, M., Nguyen, T.C., Chen, W., Yan, Z., Cao, X., Hebert, L. y Zhong, Sheng (2017). Systematic Mapping of RNA-Chromatin Interactions. *Vivo. Curr. Biol.* 27 (4), 610-612.
- Sun, D., Guo, Z., Liu, Y. y Zhang, Y. (2017). Progress and prospects of CRISPR/Cas systems in insects and other arthropods. *Front. Physiol.* 8, 1-22.
- Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K.S., Koonin, E. V. y Zhang, F. (2019). Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* 10(1), 112.
- Sun, D., Guo, Z., Liu, Y. y Zhang, Y. (2017). Progress and prospects of CRISPR/Cas systems in insects and other arthropods. *Front. Physiol.* 8, 1-22.
- Torres-Pérez, R., García-Martín, J.A., Montoliu, L., Oliveros, J.C. y Pazos, F. (2019). CRISPR Tools-Live Repository of Computational Tools for Assisting CRISPR/Cas Experiments. *Bioeng. (Basilea, Suiza)*, 6(3), 63.
- Vicente García, C., Villarejo Balcels, B., Irastorza Azcárate, I., Naranjo, S., Acemel, R.D., Tena, J.J., Rigby, P.W.J., Devos, D.P., Gómez-Skarmeta, J.L. y Carvajal, J.J. (2017). Regulatory landscape fusion in rhabdomyosarcoma through interactions between the PAX3 promoter and FOXO1 regulatory elements. *Genome Biol.* 18, 106.
- Wienert, B., Nguyen, D.N., Guenther, A., Feng, S.J., Locke, M.N., Wyman, S.K., Shin, J., Kazane, K.R., Gregory, G.L., Carter, M.A.M. *et al.* (2020). Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nat. Commun.* 11, 2109.
- Xu, J., Hua, K. y Lang, Z. (2019). Genome editing for horticultural crop improvement. *Hortic. Res.* 6, 113.
- Xu, S., Kim, J., Tang, Q., Chen, Q., Liu, J., Xu, Y. y Fu, X. (2020a). CAS9 is a genome mutator by directly disrupting DNA-PK dependent DNA repair pathway. *Protein Cell* 11, 352-365.
- Xu, X., Hulshoff, M.S., Tan, X., Zeisberg, M. y Zeisberg, E.M. (2020b). Crispr/cas derivatives as novel gene modulating tools: Possibilities and in vivo applications. *Int. J. Mol. Sci.* 21(9), 3038.
- Yin, X., Mead, B.E., Safaei, H., Langer, R., Karp, J.M. y Levy, O. (2016). Engineering Stem Cell Organoids. *Cell Stem Cell* 18, 25-38.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A. *et al.* (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 759-771.
- Zhang, Z.-T., Jiménez-Bonilla, P., Seo, S.-O., Lu, T., Jin, Y.-S., Blaschek, H.P. y Wang, Y. (2018). Bacterial Genome Editing with CRISPR-Cas9: Taking *Clostridium beijerinckii* as an Example. En *Synthetic Biology: Methods and Protocols*, J.C. Braman, ed. (Nueva York, NY: Springer New York), 297-325.



## CAPÍTULO 2

### RESUMEN

Las tecnologías *ómicas* de alto rendimiento están llamadas a revolucionar el ejercicio de la medicina mediante la implantación generalizada de métodos de medicina personalizada y de precisión (MPP) para adaptar las estrategias de diagnóstico, terapéuticas y de vigilancia a cada paciente en función de sus firmas genéticas y moleculares.

### PALABRAS CLAVE

ómicas    medicina de precisión  
medicina personalizada    big data  
integración

# TECNOLOGÍAS ÓMICAS Y MEDICINA DE PRECISIÓN

**Coordinadores**

Javier de las Rivas  
(CIC-IBMCC, Salamanca, coordinador)

Ana Rojas  
(CABD-CSIC/UPO/JA, Sevilla,  
coordinadora adjunta)

**Investigadores y centros de  
investigación participantes  
(por orden alfabético)**

Lluís Montoliu José  
(CNB, Madrid)

Leticia Mora  
(IATA, Valencia)

Florencio Pazos  
(CNB, Madrid)

## RESUMEN EJECUTIVO

En la actualidad, la genómica, que se centra en el estudio de genomas completos y fue la primera disciplina *ómica* en aparecer, se aplica con bastante éxito para estratificar subtipos de enfermedades (por ejemplo, varios subtipos de tumores en el cáncer) y ofrecer un tratamiento más personalizado a cada persona. El resto de las recientes tecnologías *ómicas* de genoma completo que permiten analizar de forma masiva la metilación del ADN o la acetilación de las histonas (epigenómica), el ARNm (transcriptómica), las proteínas (proteómica), los metabolitos (metabolómica), etc., todavía se aplican poco en el ámbito clínico y, lo que es más importante: se suelen estudiar por separado con métodos distintos, lo cual genera información biológica fragmentada en lugar de un conocimiento integrado. Esta escasa integración de la información genómica en la información biológica funcional a gran escala resultante constituye una de las principales limitaciones para extender la medicina de precisión a enfermedades complejas causadas por cambios en la regulación de los genes o en otros moduladores importantes de los fenotipos y respuestas celulares, más que de mutaciones genéticas concretas. De hecho, es habitual que en las publicaciones se pasen por alto las relaciones entre estas tecnologías, que ni son independientes ni se suelen tener en cuenta en los análisis. Por ejemplo, la expresión génica se regula en gran medida por los niveles de metilación del ADN, pero también por los niveles de proteínas y metabolitos, por lo que medir cada variable por separado es engañoso y las observaciones



resultantes tienden a ser erróneas. Esto es solo un ejemplo de la importancia de incorporar datos a todos los frentes desde los genomas hasta los fenomas —y de analizar cómo se formalizan las relaciones— para materializar por completo la promesa de la medicina de precisión. El análisis integrador de todos estos datos *ómicos* en el ámbito clínico no solo pretende dar con el tratamiento adecuado en el momento oportuno para cada paciente, sino también posibilitar su uso en planes individualizados para tratar a poblaciones de alto riesgo. Este desafío global obliga a incorporar y coordinar recursos de varios campos (la biología, la bioinformática, el análisis de datos y la medicina) para superar principalmente cuatro pasos constrictivos: (1) la adquisición e integración de datos; (2) el análisis de datos masivos, incluidas técnicas computacionales de aprendizaje automático; (3) la aplicación de métodos de diagnóstico y pronóstico a los datos *ómicos* clínicos; y (4) los aspectos éticos asociados al procesamiento de datos *ómicos*.

## 2.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La genómica fue la primera disciplina *ómica* en aparecer y hasta la fecha se ha aplicado con bastante éxito en el ámbito clínico. La secuenciación del genoma mediante métodos globales amplios (como es el caso de la secuenciación del genoma completo [SGC] o de la secuenciación del exoma completo [SEC]) o dirigidos más específicos (como es el caso de la secuenciación de paneles de genes o de la secuenciación cromosómica) ya se ha introducido como tecnología esencial en muchos estudios biomédicos con aplicaciones clínicas, por lo que ha pasado a la práctica médica (Biesecker *et al.*, 2014). En el caso de la oncología, por ejemplo, es bien sabido que estas tecnologías se han aplicado para identificar firmas genéticas específicas, que luego se utilizaron para estratificar a los pacientes con cáncer a partir de la identificación de varios subtipos tumorales biomoleculares, lo que ha aportado perfiles genómicos y tratamientos personalizados para cada paciente (Shen *et al.*, 2015).

Junto con la genómica, podemos distinguir un repertorio principal del resto de las tecnologías *ómicas* que complementan aspectos esenciales de la actividad y regulación de los genes, los productos génicos y las biomoléculas que actúan en un sistema vivo. Estas tecnologías *ómicas* del genoma completo son las siguientes:

- i. la tecnología descrita que mide secuencias de genomas completos a diferentes profundidades (la genómica);
- ii. las tecnologías que miden el estado de metilación del ADN o las modificaciones epigenéticas de las histonas (metilación, acetilación, etc.) (la epigenómica);
- i. las tecnologías que miden la conformación del genoma y la arquitectura tridimensional (la genómica tridimensional);
- ii. las tecnologías que identifican las regiones reguladoras y la regulación transcripcional (la regulómica);
- iii. las tecnologías que miden la expresión global de todos los genes, ya sean ARNm codificantes o ARNnc (la transcriptómica);
- iv. las tecnologías directamente asociadas al proteoma que miden la presencia y la actividad de proteínas y péptidos y sus modificaciones postraduccionales (la proteómica);
- v. las tecnologías que miden los metabolitos a escala global en las células (la metabolómica);
- iii. las tecnologías que miden las interacciones entre las moléculas en las células, en diferentes órganos, en todo un organismo, o entre organismos distintos como un patógeno y su huésped (la interactómica).

Aparte de la genómica, muchas de estas otras técnicas que abarcan todo el genoma siguen aplicándose escasamente en el ámbito clínico y, algo que es importante: la mayor parte de ellas se suelen estudiar por separado, en lugar de hacerlo forma **integrada en las ómicas**. La escasa integración de los datos multiplexados que se derivan de las diversas tecnologías *ómicas* constituye una de sus principales limitaciones y un desafío primordial para ampliar la medicina personalizada y de precisión y, por consiguiente, para tratar de forma eficiente enfermedades complejas y actualmente incurables. De hecho, las asociaciones funcionales y los vínculos relacionales entre los datos generados por las distintas tecnologías *ómicas* no se tratan muy a menudo en las publicaciones científicas. Sin embargo, está claro que las diferentes tecnologías *ómicas*, ya estén centradas en el genoma o arraigadas en él (es decir, en el genoma humano en el caso de los estudios biomédicos), no son independientes y se las debe considerar un compendio coherente de metodologías experimentales de alto rendimiento que aportan datos complementarios para estudiar los sistemas biológicos a escala global. Por lo tanto, aunque el genoma es la base de todos los estudios *ómicos*, y la genómica y la epigenética deberían estar a la vanguardia de cualquier enfoque biomolecular integral que pretenda estudiar un sistema biológico, el análisis integrador de varios tipos de datos *ómicos* resulta esencial para desarrollar métodos adecuados de diagnóstico y pronóstico, encontrar tratamientos personalizados e implementar planes individualizados para el tratamiento de casos raros o subpoblaciones de alto riesgo.

En conclusión, el poder traslativo de las distintas tecnologías *ómicas* solo se aprovechará plenamente si se siguen estrategias integradoras, multidisciplinarias e interdisciplinarias, lo cual requiere incorporar y coordinar recursos y conocimientos teórico-prácticos de los campos de la biología general, la medicina, la bioinformática, las ciencias informáticas, la ingeniería tecnológica instrumental, la gestión y el análisis de datos masivos, la deontología y el derecho. Ahora nos centraremos en cuatro áreas de trabajo principales que suelen constituir pasos constrictivos en el campo de la *ómica*: (I) la adquisición e integración de datos; (II) la gestión y el análisis de datos masivos; (III) la aplicación de métodos de diagnóstico a la medicina de precisión; (IV) los aspectos éticos asociados al procesamiento de datos *ómicos*. Aunque estas áreas de trabajo se abordarán en su mayor parte desde una perspectiva humana y patológica, las cuestiones examinadas se pueden aplicar sin duda al uso de las tecnologías *ómicas* para el estudio de todos los demás organismos vivos y de muchos procesos biológicos (como se explica en los siguientes capítulos del presente tema).

### 2.1.1 Adquisición e integración de datos

Dado que la genómica es una disciplina que requiere un gran volumen de datos, el tratamiento informático de estos es igual de importante que los propios métodos experimentales. Aunque en otras disciplinas científicas el procesamiento de datos puede considerarse una ayuda para el diseño experimental, en la genómica (y en cualquier tecnología *ómica*) constituye un componente intrínseco, por lo que sin él estos métodos serían inviables. El procesamiento de las grandes cantidades de datos que generan estos métodos plantea una serie de desafíos tecnológicos considerables. Del mismo modo, cada vez es más habitual obtener datos de diversos métodos *ómicos* para la misma muestra o el mismo paciente (multiómica). Aunque cada tecnología *ómica* aporta un punto de vista diferente sobre el problema biológico objeto de estudio, solo a través de su integración se puede llegar a comprender la complejidad característica e intrínseca de los sistemas vivos.

La genómica plantea algunos de los desafíos computacionales más complejos a los que tendremos que enfrentarnos en la próxima década. Los requisitos de procesamiento de datos de la genómica están a la altura de otras disciplinas científicas que consumen datos con voracidad, como es el caso de la astronomía y la física de partículas, así como de las modernas tecnologías de la información y la comunicación (TIC) como, por ejemplo, YouTube o Twitter. Como sucede en estas disciplinas, el ritmo de generación de nuevos datos en la genómica está aumentando exponencialmente, lo que impone una serie de cambios computacionales (Stephens *et al.*, 2015; y Wong *et al.*, 2019).

Hace cinco años, había más de 2500 secuenciadores de alto rendimiento en 55 países de todo el mundo. La cantidad total de datos de secuencias producidos se duplica aproximadamente cada siete meses (<http://omicsmaps.com/>). Se calcula que, para 2025, se habrá secuenciado el genoma del 25 % de la población mundial, ya que muchos países están llevando a cabo proyectos de secuenciación masiva debido a todo lo que promete la genómica para la medicina personalizada. Se podría dar incluso el caso de que a muchas personas se les secuenciara más de una muestra (por ejemplo, de cáncer o de diversos tejidos) durante periodos distintos, así como otros conjuntos de datos *ómicos* basados en definitiva en la secuenciación de ácidos nucleicos (por ejemplo, la epigenómica o la transcriptómica). De este modo, incluso en el caso de una sola persona, el volumen de datos genómicos necesarios para la medicina personalizada podría ser colosal (Chen *et al.*, 2012). Por otro lado, las tecnologías de secuenciación novedosas, tales como la de nanoporos, podrían disminuir

un poco estas cifras, ya que requieren menos sobremuestreo. Es importante tener en cuenta que todas estas previsiones se basan en un supuesto de amplia adquisición de datos mediante la aplicación de las tecnologías actuales, es decir, que las tecnologías actuales se apliquen en el futuro a más personas y muestras, por lo que no tienen en cuenta las eventuales nuevas tecnologías que podrían cambiar drásticamente el panorama y aumentar aún más estas cifras. Por consiguiente, hace falta una gran capacidad de almacenamiento de datos para procesar la información genómica en bruto. Aunque los algoritmos de compresión de datos pueden aliviar esto hasta cierto punto, se demostró que el tiempo de compresión/descompresión puede ser un problema en determinadas situaciones (Loh *et al.*, 2012). En determinadas circunstancias, se puede prescindir de los datos sin procesar originales o incluso de los genomas ensamblados y almacenar, por ejemplo, solo una lista de las variantes relativas a un genoma de referencia. Pero esto no es algo generalizado, ya que, por ejemplo, los genomas cancerosos y otras muestras complejas pueden presentar grandes reordenamientos que no se pueden codificar ni almacenar de este modo. Un paso prometedor en esa dirección es el uso de gráficos genómicos para representar colecciones de genomas (Rakocevic *et al.*, 2019). Es cierto que, en el futuro se irá prescindiendo cada vez más de los datos sin procesar a medida que vayan mejorando las metodologías para deducir datos de nivel superior a partir de ellos; lo cual conllevará que, por ejemplo, en la transcriptómica se almacenen únicamente los valores de expresión, en lugar de las lecturas originales. En cualquier caso, no cabe duda de que, aun con estos paliativos, el almacenamiento eficiente de estos crecientes volúmenes de datos genómicos supondrá un desafío en el futuro.

Con respecto a la distribución de los datos, en comparación con otras disciplinas y tecnologías, el rasgo distintivo de los datos genómicos es que se solicitan y transmiten en unidades que abarcan una amplia gama de tamaños: desde unas cuantas bases o genes (por ejemplo, para comparar con una base de datos de motivos o para realizar una búsqueda de secuencias) hasta descargas masivas de repositorios centrales (por ejemplo, para realizar análisis masivos a escala local). La computación en la nube puede contribuir a mitigar los requisitos de transmisión de datos, ya que permite ejecutar los análisis en la misma máquina (remota) en la que se encuentran los datos (Marx *et al.*, 2013), por lo que solo hace falta transmitir el código para efectuar los análisis y los resultados destilados. Otra cuestión que hay que abordar a la hora de encarar la transmisión de datos genómicos es la privacidad, ya que en muchos casos los datos genómicos médicos confidenciales podrían tener que enviarse a máquinas externas

para su análisis. En este sentido, hay un método prometedor que consiste en utilizar métodos de cifrado que nos permitan manipular los datos cifrados y efectuar determinadas consultas al respecto sin tener que descifrarlos. Por ejemplo, se podrían recuperar las mutaciones o variantes en un punto concreto de un genoma cifrado sin tener que acceder a toda su secuencia. Aunque son muy costosos desde el punto de vista computacional, estos métodos podrían facilitar la adopción generalizada de la medicina genómica al aliviar los problemas asociados a la privacidad de los datos (Erlich *et al.*, 2014).

### 2.1.2 Gestión y análisis de datos masivos

El concepto de *datos masivos* lleva utilizándose desde hace mucho tiempo, está heredado de las ciencias «duras» clásicas (física, química cuántica, etc.) y se refiere a la cantidad de datos que genera un experimento en particular. Los datos masivos requieren la regla de las cuatro «V» para cumplir los requisitos: Volumen de datos, Velocidad de procesamiento de los datos, Variabilidad de las fuentes de datos y Veracidad de la calidad de los datos. Dado que la biología se ha vuelto cuantitativa desde hace muy poco tiempo, es ahora cuando estamos empezando a alcanzar una masa crítica en la disponibilidad de los datos (Stephens *et al.*, 2015), lo cual se debe en parte a la inmensa cantidad de datos genómicos producto de la secuenciación de última generación. En 2019, teníamos al menos 91 000 secuencias de especies, y las previsiones para el 2025 indican que incluso con las estimaciones más conservadoras de duplicación cada 12 meses o cada 18 meses (equivalente a la ley de Moore), deberíamos alcanzar la escala de exabases en la genómica por entrada en la próxima década (Stephens *et al.*, 2015). En 2015, el Sequence Read Archive (SRA) ya contenía más de 3,6 petabases de datos sin procesar de secuencias que reflejaban los ~32 000 genomas microbianos, los ~5000 genomas vegetales y animales, y los ~250 000 genomas humanos individuales que se habían secuenciado o estaban en curso hasta entonces. Dado que las capacidades de secuenciación se han ampliado considerablemente, de seguir al ritmo actual duplicándose cada siete meses, deberíamos alcanzar más de una exabase de secuencias al año en los próximos cinco años y acercarnos a una zetabase de secuencias para el año 2025.

Como se ha mencionado anteriormente, también deberíamos tener en cuenta que existe una gran posibilidad de que se haya secuenciado el genoma de una fracción considerable de la población humana mundial, por lo que las estimaciones se hallan entre los 100 millones y hasta los 2000 millones de genomas humanos secuenciados para el año 2025 (lo cual supone un

crecimiento de entre cuatro y cinco órdenes de magnitud al cabo de diez años) (Stephens *et al.*, 2015), lo cual requiere generar nuevos datos de secuenciación varias veces por persona para supervisar la actividad molecular. De hecho, este número podría crecer aún más, sobre todo ahora que las nuevas tecnologías de secuenciación de célula única del genoma están empezando a revelar niveles de variación que antes resultaban inimaginables (Massoni-Badosa *et al.*, 2020; y Mereu *et al.*, 2020). Además, la tecnología empleada para secuenciar el ADN a partir de los genomas se ha ampliado y desplegado de forma creativa para otras aplicaciones *ómicas*, algunas de las cuales están ligadas a muy diversos tipos de secuenciación de alto rendimiento (como es el caso de la transcriptómica, la epigenómica o la regulómica), a otras aplicaciones basadas en la secuenciación simultánea de varios genomas (como la microbiómica y la metagenómica), o a otras basadas en diferentes tipos de tecnologías de alto rendimiento (como, por ejemplo, la proteómica, la metabolómica, etc.).

Todas estas aplicaciones tecnológicas de alto rendimiento requieren una cuantificación exacta y precisa de la señal masiva (por ejemplo, miles de millones de lecturas de secuenciación) para captar la diversidad de la señal y la diversidad de las abundancias, lo que requiere millones de puntos de datos para estimar con exactitud las distribuciones subyacentes a medida que estas van cambiando con el tiempo. Estos procedimientos resultan costosos desde el punto de vista computacional. Por ejemplo, la lectura de variantes en 2000 millones de genomas al año, con 100 000 CPU en paralelo, requeriría métodos que procesaran 2 genomas por hora y CPU. Otro ejemplo ilustrativo es que la alineación del genoma completo empleada para diversos objetivos (desde la reconstrucción de la filogenia hasta la anotación del genoma mediante metodologías comparativas) es una tarea costosa. Una sola alineación del genoma completo entre el ser humano y el ratón consume unas ~100 horas de CPU. Por lo tanto, alinear todos los pares de los aproximadamente 2,5 millones de especies que se prevé que estén disponibles en 2025 asciende a 50-100 billones de esas alineaciones del genoma completo, lo que requeriría que su velocidad fuera seis órdenes de magnitud superior a la actual. Además, en el caso de las aplicaciones médicas y clínicas, no bastará con tener el genoma: con cada persona, este se tendrá que acoplar a otros conjuntos de datos *ómicos* pertinentes, algunos de los cuales se recopilan periódicamente y a partir de diversos tejidos, para comparar los estados saludables y patológicos. En este sentido, las cuestiones de integración son muy importantes.

En lo referente al procesamiento de datos, en la actualidad hay una infinidad de recursos genómicos de gran tamaño en la nube que utilizan paradigmas de computación en la nube, especialmente para respaldar los requisitos de los mayores centros de secuenciación o para sustentar las necesidades de grandes comunidades y proyectos internacionales (por ejemplo, el Atlas del Genoma del Cáncer [TCGA], el International Cancer Genome Consortium [ICGC] y el Sequence Read Archive [SRA], entre otros). Para sacarles el máximo partido a estos sistemas, se requiere un desarrollo en mayor profundidad de interfaces de programación de aplicaciones (API) sólidas para descubrir y consultar grandes conjuntos de datos en sistemas remotos. En esta línea, por ejemplo, la Global Alliance for Genomics and Health (<https://www.ga4gh.org/>) adoptó estas normas para los datos genómicos humanos, y se prevé que la sigan otras comunidades semejantes.

Por último, a la hora de compartir y distribuir datos, surgen problemas muy fundamentales, como es el caso de la autenticación, el cifrado y otras protecciones de seguridad que hay que desarrollar para garantizar en todo momento la privacidad de los datos genómicos. A menudo se utilizan varias tecnologías de ciencia de datos, entre las que se incluyen R, Mahout y otros sistemas de aprendizaje automático con tecnología de Hadoop y otros sistemas muy escalables. Estos componentes ya se están desarrollando de la mano de empresas de ciencia de datos, Google e iniciativas de código abierto, con un amplio grado de éxito. Sin embargo, la genómica plantea desafíos únicos en lo referente a la adquisición, la distribución, el almacenamiento y, sobre todo, el análisis de datos, por lo que probablemente no baste con esperar que se produzcan innovaciones de fuera de este campo (véanse los desafíos).

### **2.1.3 Aplicación de los métodos de diagnóstico y pronóstico a la medicina de precisión**

La medicina de precisión permite adaptar las intervenciones sanitarias a grupos de pacientes en función de su propensión a la enfermedad, su diagnóstico o su pronóstico, así como su reacción al tratamiento. De hecho, lo ideal es que el sistema vincule la analítica, la práctica clínica, la investigación y la ciencia de datos con el objetivo de mejorar la eficiencia y la eficacia de la prevención, el diagnóstico y el tratamiento de las enfermedades (Ginsburg, 2014; David *et al.*, 2015; y Ginsburg y Phillips, 2018).

El desarrollo de metodologías de diagnóstico y pronóstico requiere comprender el procesamiento y la codificación de información genética, epigenética y



posgenética, que no puede deducirse únicamente de la secuencia del genoma. En este sentido, por ejemplo, las proteínas como máquinas moleculares están relacionadas más directamente con los fenotipos de los distintos estados patológicos y, por lo tanto, más cerca de las vías subyacentes que causan la enfermedad. Se espera que las metodologías de diagnóstico y pronóstico basadas en la *ómica* mejoren su especificidad y factibilidad, con pruebas únicas decidir la vía terapéutica, la elección del tratamiento o incluso el riesgo de padecer varias enfermedades simultáneamente (Slade *et al.*, 2015; y Matthews *et al.*, 2016).

### **2.1.4 Aspectos éticos asociados al procesamiento de datos ómicos**

Nuestro mundo avanza hacia la medicina personalizada y de precisión (MPP), en la que a los pacientes no solo se les acabará tratando en función de sus síntomas y enfermedades asociadas, sino también teniendo en cuenta sus variantes genéticas. Las sutiles variaciones genómicas podrían tener consecuencias enormes con respecto a la conveniencia o no de prescribir un fármaco a alguien. Por supuesto, esto también repercute sobre cuestiones posteriores. Habrá que educar a los pacientes, darles la debida información y aconsejarles que acepten la posibilidad de acceder de forma regulada a sus genomas para explorar las posibles variaciones beneficiosas o perjudiciales para su salud y sus tratamientos. Por eso hace tanto tiempo que se pretende llegar a la medicina personalizada y de precisión. Habrá que desarrollar medidas para hacer valer y garantizar la privacidad de los genomas de todos estos pacientes, lo que incluye el uso de métodos de cadenas de bloques (Ozerkan *et al.*, 2018; y Kuo *et al.*, 2020).

## **2.2 REPERCUSIÓN SOBRE EL PANORAMA DE LA CIENCIA BÁSICA Y LAS POSIBLES APLICACIONES**

### **2.2.1 Adquisición e integración de datos**

Está claro que la principal repercusión con respecto a la adquisición de datos *ómicos* está ligada al desafío de lograr una cobertura adecuada de todas las tecnologías *ómicas* (es decir, de llevar a cabo estudios verdaderamente multiómicos) y lograr una integración adecuada de las distintas tecnologías *ómicas* complementarias (dentro de los 8 niveles arriba expuestos). Los estudios multiómicos, en los que se recuperan diversos conjuntos de datos *ómicos* para la misma muestra o paciente, son cada vez más populares (Noor *et al.*, 2019). Aunque estos conjuntos de datos *ómicos* se analizaron primero por separado y luego se combinaron los resultados, quedó claro que integrarlos y analizarlos de forma sinérgica bastó para poder obtener una imagen biológica clara.

Los métodos de aprendizaje automático se utilizan a menudo para analizar conjuntos de datos multiómicos debido a su versatilidad intrínseca para procesar y combinar datos diversos. En este sentido, el aprendizaje a varios niveles (Serra *et al.*, 2016) resulta especialmente apropiado para procesar datos de naturaleza diversa e incluso codificados de forma distinta. Cuando se dispone de información mecanicista aparte de los propios datos, las redes son el objeto matemático predilecto para representarla. Pongamos el caso de una representación reticular de una vía biológica conocida con los datos multiómicos dispuestos en la parte superior de los nodos (genes/proteínas) (Barabási y Oltvai, 2004). Acto seguido, se emplean métodos basados en redes para minarlas en búsqueda de conocimientos nuevos. De hecho, la interactómica (que mide las interacciones moleculares entre proteínas, entre proteínas y ADN o ARN, o entre cualquier tipo de biomolécula) solo puede estudiarse adecuadamente utilizando redes y aplicando la teoría de grafos (De las Rivas y Fontanillo, 2010; y De las Rivas y Fontanillo, 2012), lo cual ha dado lugar al nuevo campo de la biología de redes (Barabási y Oltvai, 2004).

No solo es importante la integración de los datos procedentes de las distintas *ómicas* (es decir, la integración interómica), sino también la combinación de distintas instancias o distintos conjuntos de datos del mismo tipo de datos genómicos (es decir, la integración intraómica). Por ejemplo, genomas humanos diferentes producidos con tecnologías de secuenciación distintas, o diferentes conjuntos de datos transcriptómicos de expresión producidos en laboratorios distintos con muestras de diversas poblaciones. A este respecto, deben tenerse en cuenta las cuestiones técnicas relativas a las normas de representación de los datos, los identificadores coherentes de las bases de datos para las proteínas y los genes, las ontologías para representar el conocimiento de forma normalizada, etc.

## **2.2.2 Gestión y análisis de datos masivos**

El análisis de los datos genómicos conlleva un abanico de métodos muy diverso debido a la variedad de pasos que intervienen en la lectura de una secuencia genómica y la obtención de información útil a partir de ella. En última instancia, nuestro objetivo es poder interpretar determinados aspectos de las secuencias genómicas y dar respuesta a diversas preguntas relacionadas con ellas. Se plantean una serie de cuestiones como, por ejemplo, cómo se relacionan las mutaciones del ADN, los cambios de expresión u otras mediciones moleculares con el desarrollo, el comportamiento, la evolución o la enfermedad desde un punto de vista biomédico.

La consecución de este objetivo en el marco de los datos masivos requerirá claramente un planteamiento multidisciplinario que incorpore a varios expertos del ámbito biológico y biomédico que formulen las preguntas biomédicas esenciales, así como a científicos e ingenieros informáticos capaces de gestionar y aplicar sistemas de aprendizaje automático a gran escala en infraestructuras de computación sólidas, capaces de respaldar consultas flexibles y dinámicas para buscar patrones en compilaciones muy grandes con una dimensionalidad muy alta. Además, los matemáticos y estadísticos aportarán un tercer eje fundamental de este planteamiento multidisciplinario, que resultará esencial para traducir los datos en números y expresiones cuantitativas adecuadas para lograr una verdadera ciencia de datos masivos en la próxima generación.

Por otro lado, las recientes reimplantaciones del aprendizaje automático y la analítica de datos en los datos genómicos ha generado la impresión generalizada de que dichos métodos son capaces de resolver la mayoría de los problemas sin tener que recurrir a métodos de investigación científicos convencionales. No obstante, entender el funcionamiento de la biología resulta esencial para interpretar los datos y utilizar adecuadamente estas tecnologías informáticas. De hecho, aunque el aprendizaje automático lleva existiendo incluso desde antes de que existiera la bioinformática, llegó a integrarse en el campo mediante análisis de secuencias de proteínas hace casi 30 años (Rost y Sander, 1993) para predecir las estructuras secundarias de las proteínas. De este modo, aunque estos métodos y otros afines se han utilizado profusamente en situaciones prácticas, es ahora en los subdominios de la genómica y la transcriptómica cuando se ha producido toda una explosión de los métodos de aprendizaje automático, que son independientes a la teoría.

Los análisis de datos masivos están cobrando fuerza en el ámbito clínico, por ejemplo, para predecir los factores de riesgo individuales de los distintos tipos de enfermedades, para respaldar la toma de decisiones clínicas y para poner en práctica la medicina de precisión utilizando información genómica (Alonso Betanzos y Bolón Canedo, 2018). En particular, a lo largo de la última década, la inteligencia artificial (IA) oncológica ha contribuido a resolver determinados problemas biomédicos específicos en estudios sobre el cáncer. El aprendizaje profundo (AP), un subcampo de la IA que es muy flexible, que posibilita la extracción automática de características y que se aplica cada vez más en diversos campos de investigación, tanto básica como clínica contra el cáncer (Shimizu y Nakayama, 2020). En general, los estudios de datos masivos y las nuevas

tecnologías asociadas seguirán encarrilando una investigación novedosa y apasionante que, en última instancia, mejorará la atención sanitaria y la medicina, pero también somos realistas en el sentido de que sigue habiendo dudas sobre la privacidad, la equidad, la seguridad y el beneficio para todos (Car *et al.*, 2019).

Otro campo importante en el que los datos masivos podrían ganar peso es la farmacogenómica, que ha sobrepasado a la farmacogenética, lo cual propiciará nuevas perspectivas para el diseño de fármacos y la identificación de los factores de reacción y resistencia a estos, en cuanto se superen las precauciones relativas a la recopilación, el procesamiento, el análisis y la interpretación de los datos (Barrot *et al.*, 2019). Por último, la proteogenómica (Nesvizhskii, 2014), un enfoque integrador que aúna genómica y proteómica, es un campo que podría beneficiar enormemente a la medicina personalizada y de precisión (MPP) (Nishimura y Nakamura, 2016), lo cual hace que esta integración de datos sea una novedosa fuente de información para mejorar el conocimiento genómico de las enfermedades.

### **2.2.3 Aplicación de los métodos de diagnóstico y pronóstico a la medicina de precisión**

En los últimos años, el uso de la espectrometría de masas en los métodos de diagnóstico y pronóstico se ha convertido en una herramienta indispensable, sobre todo para la búsqueda y detección de biomarcadores epigenéticos. La genómica ha permitido progresar en el mapeo genómico de las posiciones de los nucleosomas, en el conocimiento de las modificaciones del ADN y de las histonas, así como de los factores ligados a la cromatina (Hawkins *et al.*, 2010; y Zhou *et al.*, 2011). Sin embargo, sigue habiendo la necesidad de estudiar los mecanismos mediante los cuales la epigenética puede influenciar o modificar las funciones biológicas, por lo que el uso de la espectrometría de masas y los métodos proteómicos o peptidómicos tienen una repercusión importante sobre la investigación en epigenética, lo cual posibilita el desarrollo de métodos para la identificación de las proteínas características de la enfermedad y los péptidos endógenos. Además de participar en el desarrollo de métodos de diagnóstico y pronóstico para mejorar la reacción al tratamiento, el uso de la proteómica y la peptidómica permiten conocer mejor la enfermedad y utilizar de forma más eficaz los tratamientos modernos (DiMeo *et al.*, 2016). Por ejemplo, los métodos de espectrometría de masas han contribuido al desarrollo de la epigenética del cáncer mediante la identificación de biomarcadores para el diagnóstico y el pronóstico, al mejorar la comprensión de los mecanismos de regulación de la cromatina. En los últimos años, ha surgido una nueva área

denominada *proteogenómica*, que es un producto avanzado de combinar con eficiencia las *ómicas* centradas en el genoma y las *ómicas* centradas en el proteoma. Uno de los primeros estudios proteogenómicos pertinentes en el ámbito biomédico fue el que publicaron Mertins y sus colaboradores que trabajaban en el cáncer de mama (Mertins *et al.*, 2016). Estos autores demostraron que el análisis proteogenómico del cáncer de mama permite esclarecer las consecuencias funcionales de las mutaciones somáticas, delimitar las nominaciones de candidatos a genes iniciadores dentro de grandes delecciones genómicas y regiones amplificadas, e identificar los objetivos terapéuticos.

#### **2.2.4 Aspectos éticos asociados al procesamiento de datos ómicos**

El procesamiento y la utilización de los datos genómicos de los pacientes tendrán que llevarse a cabo en cumplimiento de la ley. En Europa, el reciente Reglamento General de Protección de Datos (RGPD, 2016/679) impone límites estrictos y rigurosos a la circulación y el intercambio libres de estos datos genómicos. Al mismo tiempo, el RGPD concede a los pacientes y a sus asociaciones plenos derechos para gestionar sus propios datos, incluidos los datos genómicos, que después pueden ofrecerse a los investigadores y a los médicos para su análisis en mayor detalle. Estas nuevas posibilidades plantean una serie de desafíos a los investigadores y a los médicos, que se enfrentan a la obligación de proteger la privacidad de toda la información genómica y, al mismo tiempo, dar respuesta a la petición por parte de los pacientes para acceder a sus genomas (o a una compilación de genomas de pacientes pertenecientes a una asociación determinada que los represente jurídicamente) (Schickhardt *et al.*, 2020). Estos problemas requieren soluciones urgentes, pues de lo contrario muchos conjuntos de datos genómicos útiles que se conservan en diversas instituciones y países permanecerán paralizados y no se podrán compartir, por lo que todo el mundo saldrá perdiendo. Se han presentado algunas estrategias innovadoras, entre las que se incluyen la Nube europea de investigación e innovación sanitaria (HRIC por sus siglas en inglés; Aarestrup *et al.*, 2020). Tal y como presentan estos autores, la HRIC permitirá compartir y analizar datos para la investigación sanitaria en toda la UE, de conformidad con nuestra rigurosa legislación sobre protección de datos, al tiempo que preserva la confianza total por parte de los pacientes y voluntarios participantes. Del mismo modo, a medida que comencemos a acceder a los genomas de las personas, necesitaremos una estrategia clara para tratar los resultados incidentales, en lo referente al cómo y al cuándo comunicarlos a las personas (y a sus médicos), concretamente, aquellos que afecten a las condiciones de salud actuales o futuras de los participantes en estudio o que las pongan en peligro (Ayuso *et al.*, 2015)

## 2.3 PRINCIPALES PUNTOS PROBLEMÁTICOS

### 2.3.1 Adquisición cuantitativa, eficiente y segura de datos ómicos

Las tecnologías *ómicas* se están convirtiendo en herramientas esenciales en la mayoría de los campos biológicos y biomédicos. Para satisfacer estas necesidades crecientes, hay una serie de problemas y limitaciones que habrá que superar en los próximos años, entre los que se incluyen los siguientes:

- Reducir los costes de secuenciación en al menos dos órdenes de magnitud.
- Recopilar otros datos y metadatos asociados a las secuencias.
- Desarrollar nuevos sistemas centrados en el genoma para representar los datos de secuenciación más allá de las cadenas de caracteres.
- Mejorar las tecnologías de almacenamiento de datos (memoria, disco, etc.) y de transmisión (redes).
- Mejorar los sistemas de compresión e indexación para racionalizar el almacenamiento y el acceso a los datos.
- Desarrollar la autenticación, el cifrado y otras protecciones de seguridad para garantizar la privacidad de los datos.
- Mejorar la *ómica* cuantitativa (es decir, la medición de los cambios cuantitativos) y la *ómica* longitudinal (es decir, la medición de los cambios a lo largo del tiempo).
- Mejorar la adquisición de datos y la integración de la información clínica y fenotípica en los datos biomoleculares *ómicos*.
- Desarrollar enfoques técnicos y analíticos para combinar conjuntos de datos intra e interómicos.

### 2.3.2 La interpretación: un desafío recurrente en la bioinformática

Desde que la biología empezó a volverse cuantitativa, se depositaron muchas expectativas en el aspecto computacional y en el nuevo campo de la bioinformática (Ouzounis, 2012), con la idea ingenua de que los ordenadores resolverían todos nuestros problemas. La mayoría de las promesas iniciales dependían de la interpretación automática final de los datos (Thornton, 1998), que pronto pasó de ser una expectativa a, en la actualidad, un desafío clave. Lo cierto es que la interpretación de los datos se ha convertido en un desafío recurrente, sobre todo cuando se declaró que la integración de los datos era el primer cuello de botella. Este desafío inicial lo proclamaron equipos de expertos en secuenciación del genoma y bases de datos genómicas: «managing huge

data volumes, integrating information from various discovery platforms and discerning phenotypic implications» (gestionar volúmenes de datos inmensos, integrar la información de diversas plataformas de descubrimiento y discernir las implicaciones fenotípicas) (Scherer *et al.*, 2007). En cambio, en algunos ámbitos concretos de la genómica, se ha propuesto desde hace tiempo que el cuello de botella de nuestros conocimientos es la falta de datos disponibles. Sin embargo, esto solo es cierto en algunas situaciones porque, por ejemplo, la disponibilidad de miles de genomas de la población general ha contribuido a revisar la penetrancia de determinadas enfermedades (Check Hayden, 2016) en función de la frecuencia de las mutaciones.

Luego viene la hiperexpectación. Para ilustrarlo, tenemos el informe de los llamados «genetic heroes» (héroes genéticos) o «bullet-proof genomes» (genomas a prueba de balas), 13 personas que deberían estar muertas porque se encontraron mutaciones mendelianas en sus genomas (Chen *et al.*, 2016). Esto último se refutó por completo, ya que la metodología y la interpretación eran defectuosas y exageradas (MacArthur, 2016). Este polémico ejemplo demuestra que **el análisis de datos exactos y reproducibles sigue siendo el gran desafío de los estudios ómicos a gran escala**. A este respecto, en muchos casos, el incremento del tamaño de la muestra que se suele recomendar no ayudará a resolver los problemas ligados al análisis y la interpretación de los datos, donde la asignación funcional y biológica sigue siendo un paso fundamental. Por lo tanto, este es probablemente uno de los aspectos más problemáticos en la investigación biomédica actual para lograr una verdadera medicina personalizada y de precisión. Por ejemplo, en la genómica médica, la identificación de variantes genómicas en un solo genoma y la interpretación correcta de cuál es la repercusión de una mutación en concreto siguen siendo muy problemáticas. En otras palabras, **descubrir si una mutación en particular es la verdadera causante de un estado patológico determinado o no sigue siendo una cuestión sin resolver**. Por lo tanto, aunque los casos obvios vinculados a enfermedades bien definidas se abordan con facilidad, la gran cantidad de mutaciones de importancia desconocida continúa resolviéndose incorrectamente, y la elección de nuestras herramientas computacionales predictivas y los datos introducidos afectan en gran medida a los resultados (McCarthy *et al.*, 2014). De hecho, se ha calculado aproximadamente que faltan las posibles consecuencias fenotípicas para la variación de >75 % de los ~20 000 genes anotados en el genoma humano (Posey *et al.*, 2019), lo cual plantea una situación paradójica, ya que cuantos más datos tenemos, más cuesta interpretarlos. En este sentido, se han desarrollado decenas de métodos bioinformáticos y de



biología computacional para dar respuesta a esta cuestión precisamente, lo cual indica la pertinencia del asunto. De hecho, la evaluación comparativa de los procedimientos y los anotadores se ha aplicado en entornos clínicos, donde los resultados deficientes y el escaso consenso requirieron debates finales para alcanzar unas directrices consensuadas (Amendola *et al.*, 2016). Este ejemplo expone un problema común en la genómica computacional, en el que el desarrollo o la aplicación de un método más no mejora la interpretabilidad de los resultados, lo cual se debe en parte al hecho de utilizar siempre un genoma humano de referencia, por lo que otro de los desafíos clave en la genómica médica estriba en avanzar hacia ensamblajes y análisis genómicos sin referencias, o incluso hacia el uso simultáneo de varias referencias aplicando metodologías de aprendizaje difuso y profundo.

En resumen, los datos masivos biológicos solo los pueden gestionar y analizar de forma eficiente biólogos computacionales expertos en estrecho contacto con médicos e investigadores biológicos, quienes solo pueden **lograr una transición desde los estudios de asociación a los de causalidad** trabajando juntos.

### 2.3.3 Capacidades informáticas (CPU)

Las mejoras en las capacidades de CPU podrían resultar útiles, pero las tendencias en la potencia de cómputo suelen estar orientadas a operaciones de coma flotante y no aportan necesariamente mejoras en el análisis genómico, en el que las **operaciones de cadena y la gestión de la memoria** suelen plantear los desafíos más importantes. Además, si se tiene en cuenta el desafío tecnológico, quizá el principal cuello de botella del análisis de datos masivos en el futuro no se encuentre en las capacidades de CPU, sino en la *hardware* de entrada/salida (E/S) que transporta los datos entre el almacenamiento y los procesadores, un problema que requiere la investigación de nuevos algoritmos y *hardware* de E/S paralelos que puedan utilizarlos eficazmente.

### 2.3.4 Formación adecuada en biología, informática y estadística

Otro desafío que quedó de manifiesto a lo largo del desarrollo del campo fue el enfrentamiento entre la «automation utopia» (utopía de la automatización) y la «people paradox» (paradoja de las personas): la constatación de que la aplicación de las ciencias informáticas y el aprendizaje automático a los resultados biológicos provoca un aumento de la demanda de personas bien cualificadas (Miller y Attwood, 2003). Esta constatación se ha vuelto aún más evidente con las competencias necesarias para analizar los datos que tenemos entre manos. En el caso particular de la genómica, plantea desafíos únicos en



lo referente a la adquisición, la distribución, el almacenamiento y los análisis de datos. Debemos afrontar estos desafíos nosotros mismos, empezando por integrar la ciencia de los datos en los planes de estudios de grado, licenciatura y bachillerato para formar a las próximas generaciones de biólogos cuantitativos, bioinformáticos, científicos informáticos y bioingenieros. Este desafío concreto quedará muy bien ilustrado en los siguientes capítulos de este tema, ya que en todas las áreas destacadas de la investigación (epi)genómica existe una necesidad apremiante no solo de infraestructura informática, sino también de biólogos informáticos.

Además, la falta de pensamiento estadístico es una norma en el campo biomédico y biológico, y se hace evidente en lo referente al diseño experimental y las elecciones de los procedimientos metodológicos. Por ejemplo, **los valores de p se emplean profusamente, pero no se suelen entender bien**, por lo que se abusa demasiado de ellos en muchas publicaciones de investigación biomédica (Leek y Peng, 2015). Además, en muchos estudios *ómicos* se suele presentar un escrutinio y un debate muy escaso del diseño experimental necesario para lograr una significación estadística determinada.

### 2.3.5 Necesitamos la teoría cuando aplicamos la inteligencia artificial a los datos biológicos

Uno de los principales desafíos de la investigación biológica y biomédica es la heterogeneidad de los datos, ya que todo sistema biológico u organismo está compuesto por decenas de miles de componentes, y necesitamos conocer las relaciones entre estos componentes. Además, existe una concepción generalizada de que la inteligencia artificial (IA) solventará la mayoría de los problemas sin ninguna base de la lógica científica, como si confiar en patrones ocultos nos brindara los resultados que deseamos. Esta concepción errónea se ve acelerada por la facilidad de obtener datos y de aplicar muchos métodos como cajas negras. Independientemente de la profundidad y la sofisticación de los métodos basados en datos —como es el caso de las redes neuronales de aprendizaje profundo—, al final tan solo se utilizan para generar unas sencillas curvas de ajuste de los resultados de los datos existentes. No obstante, estos métodos requieren cantidades de datos mucho mayores que las previstas por los aficionados a los datos masivos para producir resultados estadísticamente fiables. Es más, en muchos casos se necesita y aún no se puede sustituir la teoría con respecto a los métodos, y lo mismo sucede con los conocimientos sobre los procesos para los que queremos respuestas (Coveney *et al.*, 2016).

### 2.3.6 Biomarcadores reproducibles para uso clínico

Aparte de las dificultades para integrar los datos de las diferentes *ómicas*, también surgen complicaciones en el descubrimiento y la aplicación de biomarcadores estables y reproducibles debido a las dificultades para manipular muestras heterogéneas, el reconocimiento adecuado de los errores estadísticos en las identificaciones (es decir, el cálculo de las tasas de falsos positivos y falsos negativos), la posible reactividad cruzada o los factores ocultos en las técnicas, la identificación molecular incompleta, o el análisis e interpretación final incorrecta de los datos. De este modo, la estandarización de las metodologías *ómicas* mediante instrumentos de última generación para evitar la variabilidad de los procedimientos es una cuestión difícil de superar a fin de implantar con éxito el uso de los biomarcadores recién descubiertos en las aplicaciones clínicas.

### 2.3.7 Directrices éticas en la era de la *ómica* y la medicina de precisión

La mejora continua de la tecnología *ómica* en términos de calidad y rendimiento de los datos generados está revolucionando todos los ámbitos de la investigación biomédica. Sin embargo, estos rápidos avances también han planteado cuestiones éticas que exigen respuestas, directrices y regulación:

- i. ¿Qué hacemos con la información no procesable de una posible enfermedad identificada en un cribado? ¿Debemos avisar a los pacientes?
- ii. ¿Cómo se anonimizará esta información? Hoy en día, resulta muy fácil llegar a un nombre o a un código postal utilizando datos genómicos, lo cual tendrá consecuencias en lo referente a la ascendencia, las enfermedades, etc. ¿Cómo vamos a encarar esta situación como sociedad?
- iii. La medicina de precisión debe respetar las preocupaciones éticas y morales de todos los grupos y culturas, así como garantizar la seguridad de la información en un entorno en el que los sistemas informáticos de los hospitales pudieran estar en el punto de mira de ciberataques.
- iv. Los investigadores y las instituciones, que pueden ser públicas o privadas, deben proteger la integridad de sus datos y sus estudios, así como procurar que los resultados sean reproducibles. Se debe fomentar el uso de organismos independientes de verificación y validación. Deben implantarse políticas para liberar los datos o regularlos.

Junto con los métodos *ómicos*, los recientes avances en las tecnologías de modificación genómica harán que la medicina de precisión y personalizada sea una realidad en un futuro próximo. Por ejemplo, pronto se podrá modificar el

genoma de cualquier organismo vivo, incluidos los seres humanos, gracias a herramientas genómicas populares, asequibles y eficientes, como es el caso de CRISPR-Cas, que ya se han presentado en el capítulo anterior (véase el capítulo 3.1). Tanto si la estrategia perseguida se dirige a las células del paciente *ex vivo* o *in vivo*, tendremos que evaluar los aspectos éticos y los posibles daños colaterales no deseados que pudieran interferir en el genoma de las personas. Los efectos colaterales y el mosaicismo son problemas habituales no deseados ligados a todo método actual de modificación genómica. Las consecuencias difieren a la hora de tratar con especies animales o vegetales, o con seres humanos. Eliminar las mutaciones no deseadas mediante esquemas de cultivo, seleccionando los individuos más adecuados, es algo aceptable para las especies animales y vegetales, mientras que estrategias semejantes resultan éticamente inaceptables para los pacientes. Por lo tanto, es imprescindible reflexionar sobre el uso responsable en el futuro de las tecnologías de modificación genómica, y esta es una cuestión que habrá que abordar.

Los aspectos éticos de nuestra capacidad actual para manipular los genomas van más allá de los ejemplos clínicos. En las plantas, la posibilidad de aplicar mutaciones simples asociadas a rasgos de interés agroeconómico se enfrenta a problemas éticos y legales en muchos países, incluida la Unión Europea. La sentencia del Tribunal de Justicia de la Unión Europea de julio de 2018, en la que se falló que los cultivos modificados genómicamente no estarán exentos de cumplir con la Directiva sobre OMG de la UE (2001/18), supuso una brecha clara para el desarrollo de nuevas plantas con el genoma modificado, que experimentó una acusada penalización por esta decisión legal. En contraste, en otras partes del mundo no se dan estas inquietudes y, por consiguiente, sus posibilidades de comerciar y expandir sus mercados repercutirán negativamente sobre nuestro futuro inmediato en Europa (Hundleby y Harwood, 2019).

## CAPÍTULO 2 BIBLIOGRAFÍA

- Aarestrup, F.M., Albeyatti, A., Armitage, W.J. *et al.* (2020). Towards a European health research and innovation cloud (HRIC). *Genome Med.* 12(1), 18. doi: 10.1186/s13073-020-0713-z.
- Alonso Betanzos, A. y Bolón Canedo, V. **Big-Data Analysis, Cluster Analysis, and Machine-Learning Approaches.** En: Kerkhof, P. y Miller, V. (eds.) **Sex-Specific Analysis of Cardiovascular Function** (2018). *Adv. Exp. Med. Biol.* 1065, 607-626. Springer, Cham. doi: 10.1007/978-3-319-77932-4\_37.
- Amendola, L.M., Jarvik, G.P., Leo, M.C. *et al.* (2016). Performance of ACMG-AMP Variant-Interpretation Guidelines Among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am. J. Hum. Genet.* 98(6), 1067-1076. doi: 10.1016/j.ajhg.2016.03.024.
- Ayuso, C., Millán, J.M. y Dal-Re, R. (2015). Management and return of incidental genomic findings in clinical trials. *Pharmacogenomics J.* 15(1), 1-5. doi: 10.1038/tpj.2014.62.
- Barabási, A.L. y Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5(2), 101-113. doi: 10.1038/nrg1272.
- Barrot, C.C., Woillard, J.B. y Picard, N. (2019). Big data in pharmacogenomics: current applications, perspectives and pitfalls. *Pharmacogenomics*, 20(8), 609-620. doi: 10.2217/pgs-2018-0184.
- Biesecker, L.G. y Green, R.C. (2014). Diagnostic clinical genome and exome sequencing. *N. Engl. J. Med.* 370(25), 2418-2425. doi:10.1056/NEJMr1312543.
- Car, J., Sheikh, A., Wicks, P. y Williams, M.S. (2019). Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Med.* 17(1), 143. doi:10.1186/s12916-019-1382-x.
- Casacuberta, J.M. y Puigdomènech, P. (2018). Proportionate and scientifically sound risk assessment of gene-edited plants. *EMBO Rep.* 2018a, 19(10), e46907. doi: 10.15252/embr.201846907.
- Casacuberta, J.M. y Puigdomènech, P. (2018). European politicians must put greater trust in plant scientists. *Nature* 561(7721), 33. doi: 10.1038/d41586-018-06129-2.
- Chakravorty, S. y Hegde, M. (2017). Gene and Variant Annotation for Mendelian Disorders in the Era of Advanced Sequencing Technologies. *Annu. Rev. Genomics Hum. Genet.* 18, 229-256. doi:10.1146/annurev-genom083115-022545.
- Chen, R., Mias, G.I., Li-Pook-Than, J. *et al.* (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 148(6), 1293-1307. doi:10.1016/j.cell.2012.02.009.
- Chen, R., Shi, L., Hakenberg, J. *et al.* (2016). Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat. Biotechnol.* 34(5), 531-538. doi: 10.1038/nbt.3514.
- Coveney, P.V., Dougherty, E.R. y Highfield, R.R. (2016). Big data need big theory too. *Philos Trans A Math. Phys. Eng. Sci.* 374(2080), 20160153. doi:10.1098/rsta.2016.0153.
- Check Hayden, E. (2016). A radical revision of human genetics. *Nature.* 538(7624), 154-157. doi: 10.1038/538154a.
- David, S.P., Johnson, S.G., Berger, A.C. *et al.* (2015). Making Personalized Health Care Even More Personalized: Insights from Activities of the IOM Genomics Roundtable. *Ann. Fam. Med.*, 13(4), 373-380. doi: 10.1370/afm.1772.
- De Las Rivas, J. y Fontanillo, C. (2010). Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* 6(6), e1000807. doi: 10.1371/journal.pcbi.1000807.
- De Las Rivas, J. y Fontanillo, C. (2012). Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell. *BriefFunct. Genomics.* 11(6), 489-496. doi: 10.1093/bfpg/els036.
- Di Meo, A., Pasic, M.D. y Yousef, G.M. (2016). Proteomics and peptidomics: moving toward precision medicine in urological malignancies. *Oncotarget.* 7(32), 52460-52474. doi: 10.18632/oncotarget.8931.
- Dopazo, J., Amadoz, A., Bleda, M. *et al.* (2016). 267 Spanish Exomes Reveal Population-Specific Differences in Disease Related Genetic Variation. *Mol. Biol. Evol.* 33(5), 1205-1218. doi: 10.1093/molbev/msw005.
- Erlich, Y. y Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet.* 15(6), 409421. doi: 10.1038/nrg3723.

- Gelabert, P., Ferrando Bernal, M., de Dios, T. et al. (2019).** Genome-wide data from the Bubi of Bioko Island clarifies the Atlantic fringe of the Bantu dispersal. *BMC Genomics*. 20(1), 179. doi: 10.1186/s12864-019-5529-0.
- Ginsburg, G. (2014).** Medical genomics: Gather and use genetic data in health care. *Nature*. 508(7497), 451-453. doi: 10.1038/508451a.
- Ginsburg, G.S. y Phillips, K.A. (2018).** Precision Medicine: From Science to Value. *Health Aff. (Millwood)*. 37(5), 694701. doi: 10.1377/hlthaff.2017.1624.
- Hawkins, R.D., Hon, G.C. y Ren, B. (2010).** Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11(7), 476-486. doi: 10.1038/nrg2795.
- Hirsch, F., Lemaitre, C., Chneiweiss, H. y Montoliu, L. (2019).** Genome Editing: promoting responsible research. *Pharmaceut. Med.* 33(3), 187-191. doi: 10.1007/s40290-019-00276-1.
- Hundleby, P.A.C. y Harwood, W.A. (2019).** Impacts of the EU GMO regulatory framework for plant genome editing. *Food Energy Secur.* 8(2), e00161. doi: 10.1002/fes3.161.
- Iriart, J.A.B. (2019).** Precision medicine/ personalized medicine: a critical analysis of movements in the transformation of biomedicine in the early 21st century. *Cad Saude Publica* 35(3), e00153118. doi: 10.1590/0102311X00153118.
- Kuo, T.T., Kim, J. y Gabriel, R.A. (2020).** Privacy-preserving model learning on a blockchain network-of-networks. *J. Am. Med. Inform. Assoc.* 27(3), 343-354. doi: 10.1093/jamia/ocz214.
- Leek, J. y Peng, R. (2015).** Statistics: p-values are just the tip of the iceberg. *Nature* 520(7549), 612. doi: 10.1038/520612a.
- Loh, P.R., Baym, M. y Berger, B. (2012).** Compressive genomics. *Nat. Biotechnol.* 30(7), 627-630. doi: 10.1038/nbt.2241.
- Lorente Galdós, B., Lao, O., Serra Vidal, G. et al. (2019).** Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. *Genome Biol.* 20(1), 77. doi: 10.1186/s13059-019-1684-5.
- MacArthur, D. (2016).** Superheroes of disease resistance. *Nat. Biotechnol.* 34, 512-513. doi: 10.1038/nbt.3555.
- Marx, V. (2013).** Drilling into big cancer-genome data. *Nat. Methods*. 10(4): 293-297. doi: 10.1038/nmeth.2410.
- Massoni-Badosa, R., Iacono, G., Moutinho, C. et al. (2020).** Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol.* 21(1), 112. doi: 10.1186/s13059-020-02032-0.
- Matthews, H., Hanison, J. y Nirmalan, N. (2016).** "Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. *Proteomes*. 4(3), 28. doi: 10.3390/proteomes4030028.
- McCarthy, D.J., Humburg, P., Kanapin, A. et al. (2014).** Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* 6, 26 doi: 10.1186/gm543.
- Mereu, E., Lafzi, A., Moutinho, C. et al. (2020).** Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* 38(6), 747-755. doi: 10.1038/s41587-020-0469-4.
- Mertins, P., Mani, D.R., Ruggles, K.V. et al. (2016).** Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, 534(7605), 55-62. doi:10.1038/nature18003.
- Miller, C.J. y Attwood, T.K. (2003).** Bioinformatics goes back to the future. *Nat. Rev. Mol. Cell Biol.* 4(2), 157-162. doi:10.1038/nrm1013.
- Montoliu, L., Merchant, J., Hirsch, F. et al. (2018).** ARRIGE Arrives: toward the responsible use of genome editing. *CRISPR J.* 1(2), 128-129. doi:10.1089/crispr.2018.29012.mon.
- Nesvizhskii, A. (2014).** Proteogenomics: concepts, applications and computational strategies. *Nat. Methods*. 11, 1114-1125. doi: 10.1038/nmeth.3144.
- Nishimura, T. y Nakamura, H. (2016).** Developments for Personalized Medicine of Lung Cancer Subtypes: Mass Spectrometry-Based Clinical Proteogenomic Analysis of Oncogenic Mutations. *Adv. Exp. Med. Biol.* 926, 115-137. doi: 10.1007/978-3-319-42316-6\_8.
- Noor, E., Cherkaoui, S. y Sauer, U. (2019).** Biological insights through omics data integration. *Curr. Opin. Syst. Biol.* 15, 39-47. doi: 10.1016/j.coisb.2019.03.007.

- Oliveros, J.C., Franch, M., Tabas Madrid, D. *et al.* (2016). Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. *Nucleic Acids Res.* 44(W1), W267-W271. doi: 10.1093/nar/gkw407.
- Ouzounis, C. (2012). Rise and demise of bioinformatics? *Promise and Progress. Plos Comp. Biol.* 8(4), e1002487. doi: 10.1371/journal.pcbi.1002487.
- Ozercan, H.I., Ileri, A.M., Ayday, E. y Alkan, C. (2018). Realizing the potential of blockchain technologies in genomics. *Genome Res.* 28(9), 1255-1263. doi: 10.1101/gr.207464.116.
- Rakocevic, G., Semenyuk, V., Lee, W.P. *et al.* (2019). Fast and accurate genomic analyses using genome graphs. *Nat. Genet.* 51(2), 354-362. doi:10.1038/s41588-018-0316-4.
- Rost, B. y Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA* 90(16), 7558-7562. doi:10.1073/pnas.90.16.7558.
- Pazos, F., García Moreno, A., Oliveros, J.C. (2018). Automatic detection of genomic regions with informative epigenetic patterns. *BMC Genomics* 19(1), 847. doi: 10.1186/s12864-018-5286-5.
- Posey, J.E., O'Donnell-Luria, A.H., Chong, J.X. *et al.* (2019). Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med.* 21(4), 798-812. doi: 10.1038/s41436-018-0408-7.
- Scherer, S.W., Lee, C., Birney, E. *et al.* (2007). Challenges and standards in integrating surveys of structural variation. *Nat. Genet.* 39, S7-15. doi: 10.1038/ng2093.
- Schickhardt, C., Fleischer, H. y Winkler, E.C. (2020). Do patients and research subjects have a right to receive their genomic raw data? An ethical and legal analysis. *BMC Med. Ethics.* 21(1), 7. doi: 10.1186/s12910-020-0446-y.
- Serra, A., Fratello, M., Greco, D. y Tagliaferri, R. (2016). Data integration in genomics and systems biology. *2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC*, 1272-1279, doi: 10.1109/CEC.2016.7743934.
- Shen, T., Pajaro-Van de Stadt, S.H., Yeat, N.C. y Lin, J.C. (2015). Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. *Front Genet.* 6, 215. doi: 10.3389/fgene.2015.00215.
- Shimizu, H. y Nakayama, K.I. (2020). Artificial intelligence in oncology. *Cancer Sci.* 111(5), 1452-1460. doi: 10.1111/cas.14377.
- Slade, I., Riddell, D., Turnbull, C., Hanson, H. y Rahman, N. (2015). MCG Programme. Development of cancer genetic services in the UK: A national consultation. *Genome Med.* 7(1), 18. doi: 10.1186/s13073-015-0128-4.
- Stephens, Z.D., Lee, S.Y., Faghri, F. *et al.* (2015). Big Data: Astronomical or Genomical? *PLoS Biol.* 13(7), e1002195. doi: 10.1371/journal.pbio.1002195.
- Thornton, J.M. (1998). The future of bioinformatics. *Trends Guide to Bioinformatics* 16(supl. 1), 30-31.
- Torres-Pérez, R., García-Martín, J.A., Montoliu, L., Oliveros, J.C. y Pazos, F. (2019). WeReview: CRISPR Tools-Live Repository of Computational Tools for Assisting CRISPR/Cas Experiments. *Bioengineering* 6(3), 63. doi: 10.3390/bioengineering6030063.
- Wong, K.C. (2019). Big data challenges in genome informatics. *Biophys. Rev.* 11(1), 51-54. doi: 10.1007/s12551018-0493-5.
- Yanai, I. y Chmielnicki, E. (2017). Computational biologists: moving to the driver's seat. *Genome Biol.* 18(1), 223. doi: 10.1186/s13059-017-1357-1.
- Zhou, V.W., Goren, A. y Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet.* 12(1), 7-18. doi: 10.1038/nrg2905.

### RESUMEN

Justo ahora comienza a comprenderse la organización tridimensional (3D) del genoma, si bien, a pesar de los enormes avances registrados, aún quedan muchas preguntas esenciales sin resolver. Es preciso superar importantes desafíos metodológicos y conceptuales para alcanzar una visión completa del impacto fisiológico de la arquitectura tridimensional del genoma en la salud y la enfermedad. El CSIC se encuentra en una posición privilegiada para abordarlos.

### PALABRAS CLAVE

tridimensional	arquitectura del genoma		
TAD	compartimento	3C	imagenología

# ARQUITECTURA TRIDIMENSIONAL DEL GENOMA

## Coordinadores

Ferran Azorín  
(IBMB, Barcelona, coordinador)

Josep Rotllant (IIM, Vigo,  
coordinador adjunto)

## Investigadores y centros de investigación participantes (por orden alfabético)

Albert Jordán  
(IBMB, Barcelona)

Ignacio Maeso  
(CABD, Sevilla)

Miguel Manzanares  
(CBMSO, Madrid)

Álvaro Rada Iglesias  
(IBBTec, Santander)

Joaquim Roca  
(IBMB, Barcelona)

Guillermo Vicent  
(IBMB, Barcelona)

## RESUMEN EJECUTIVO

Nuestros conocimientos sobre la regulación de las funciones genómicas se han beneficiado enormemente del esclarecimiento de la organización estructural del genoma en diferentes niveles. En primer lugar, el descubrimiento de la estructura del ADN revolucionó la biología y la medicina de muchas maneras. Asimismo, la identificación del nucleosoma como subunidad estructural básica de la cromatina sentó las bases para la comprensión de los mecanismos epigenéticos que resultan fundamentales para la regulación de las funciones del genoma, desde la transcripción del ARN y la replicación del ADN, su recombinación y reparación, hasta la segregación de los cromosomas y la estabilidad del genoma. En años más recientes, se ha consolidado la organización jerárquica de la estructura de la cromatina dentro del núcleo celular. Los agrupamientos de nucleosomas, los bucles de cromatina, los dominios asociados topológicamente (TAD, por sus siglas en inglés), los compartimentos y los territorios cromosómicos aparecen como niveles de organización que, en última instancia, conforman la organización tridimensional (3D) del genoma en el núcleo. Descifrar la arquitectura tridimensional y la dinámica conformacional del genoma con una resolución a nanoescala tendrá un impacto importante en nuestra comprensión de cómo se establecen y mantienen los programas de desarrollo y diferenciación en la célula, la respuesta adaptativa de los sistemas biológicos a los cambios ambientales y la etiología de las disfunciones del



genoma, en particular, las que derivan en las enfermedades. Sin embargo, aún quedan muchas preguntas por esclarecer. Los actuales enfoques experimentales solo permiten abordar unos pocos aspectos de la organización tridimensional de la cromatina y, en su mayor parte, solo se aplican a tipos celulares abundantes, lo que conduce a una visión aún incompleta, estática y con toda probabilidad distorsionada de la estructura tridimensional del genoma. También falta un marco teórico para comprender las fuerzas y los principios que rigen el plegado de la cromatina. Por otra parte, la caracterización de la maquinaria macromolecular y de los mecanismos implicados en el establecimiento y mantenimiento de la organización tridimensional de la cromatina sigue siendo esquiva. Es más, los pocos trabajos comparativos realizados hasta ahora están revelando importantes diferencias sobre cómo las distintas especies eucariotas estructuran su cromatina en el espacio nuclear. Y por último, pero no por ello menos importante, las consecuencias funcionales y evolutivas de la organización tridimensional del genoma y sus efectos causales en los procesos basados en plantillas de ADN son todavía poco conocidos. Para hacer frente a estos desafíos se requiere un planteamiento totalmente interdisciplinario que implique el uso de enfoques moleculares, celulares, genómicos, genéticos, evolutivos, biofísicos, físicos, de imagen y computacionales. Nuestro objetivo final se centra en **la construcción de modelos mecanicistas precisos y predictivos que describan la arquitectura tridimensional del genoma y descubran su relevancia funcional en la salud y la enfermedad**. Muchos equipos del CSIC demuestran una amplia experiencia en estos campos, lo que hace que este organismo se encuentre en una buena disposición para abordar con éxito el desafío.

### 3.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

Existe una relación íntima entre la estructura y la función del genoma que aún no se comprende del todo. En su más alta resolución, el nucleosoma representa la subunidad estructural básica de la cromatina y su descubrimiento fue esencial para el hallazgo y posterior comprensión de los mecanismos epigenéticos que ahora sabemos que son fundamentales para todos los aspectos de la función del genoma (es decir, la transcripción del ARN, la replicación del ADN, la recombinación, la reparación, la segregación de los cromosomas y la estabilidad del genoma). Más recientemente, los avances tecnológicos nos han proporcionado una visión más completa y global de la organización tridimensional de la cromatina dentro del núcleo celular. Según esta imagen emergente, la necesidad de empaquetar grandes genomas eucariotas dentro de la dimensión limitada del núcleo se cumple a través de la organización jerárquica del ADN en estructuras de creciente compactación y complejidad. La repetición monótona de nucleosomas a lo largo de la molécula de ADN constituye la organización estructural básica de la fibra de cromatina. Las fibras de cromatina se pliegan con posterioridad en niveles de organización estructural de orden superior que actúan a escalas genómicas crecientes. El análisis de microscopía de superresolución ha demostrado que la interacción con las histonas de enlace H1 da lugar a la formación de aglomeraciones de nucleosomas de tamaño variable (Ricci *et al.*, 2015). Por otra parte, la captura de la conformación del cromosoma (3C) y las aproximaciones de imagen han revelado dos niveles adicionales de plegado de la fibra de cromatina, a saber, los dominios asociados topológicamente (TAD) y los compartimentos. Los TAD son regiones genómicas aisladas que interactúan entre sí en la escala de unas pocas kilobases (kb) a megabases (Mb) (Dixon *et al.*, 2012; Nora *et al.*, 2012), mientras que los compartimentos reflejan la agrupación espacial de grandes regiones genómicas, de decenas a cientos de Mb, que comparten estados epigenéticos similares (Bonev *et al.*, 2017; Lieberman-Aiden *et al.*, 2009; Rao *et al.*, 2014). Además, el uso de sondas fluorescentes de pintado de cromosomas completos desveló que los cromosomas ocupan zonas discretas dentro del espacio nuclear (Kempfer y Pombo, 2020). Estas estructuras, conocidas como territorios cromosómicos, mostraron una mezcla limitada entre ellas y parecen distribuirse en posiciones preferentes dentro del núcleo. Los cromosomas grandes, pobres en genes, se localizan preferentemente en la periferia del núcleo —cerca de la lámina nuclear—, mientras que los cromosomas pequeños, ricos en genes, se suelen localizar en la parte central del núcleo.

**Cabe la posibilidad de que la arquitectura tridimensional del genoma desempeñe importantes funciones reguladoras.** Aunque los papeles causales directos en la función biológica siguen siendo escasos, la estructura tridimensional guarda una estrecha correlación con la función del genoma. En particular, se han descrito dos tipos principales de compartimentos: el *compartimento A* que corresponde a las regiones eucromáticas transcripcionalmente activas marcadas con modificaciones de histonas activas, y el *compartimento B* formado por la heterocromatina transcripcionalmente silenciosa decorada con modificaciones de histonas represivas. En este sentido, se ha planteado que los TAD representan unidades reguladoras fundamentales en las que tienen lugar la mayoría de las interacciones entre los potenciadores y sus genes diana. Además, durante el desarrollo del embrión de *Drosophila*, no se detectan signos de plegado tridimensional de la cromatina hasta que el genoma cigótico se vuelve transcripcionalmente activo en la transición materno-cigótica (Hug *et al.*, 2017). No obstante, las pruebas que relacionan la estructura tridimensional y la transcripción están lejos de ser concluyentes. Varios estudios han analizado los efectos de la inhibición de la transcripción sobre la organización tridimensional de la cromatina. En las bacterias *C. crescentus* y *B. subtilis*, la inhibición de la transcripción provoca una pérdida de dominios de contacto (Le *et al.*, 2013; Wang *et al.*, 2017). Sin embargo, experimentos similares realizados en *Drosophila* sostienen que la transcripción en sí misma no es necesaria para la organización en TAD y compartimentos (Hug *et al.*, 2017). Por este motivo, **se sigue debatiendo si el plegado tridimensional es una causa o una consecuencia de la función.**

La conservación evolutiva de la organización tridimensional de la cromatina también se está sometiendo a una intensa investigación. Se han observado dominios de autointeracción similares a los TAD en una gran variedad de organismos, desde mamíferos y otras especies animales —es decir, *Drosophila* (Rowley *et al.*, 2017; Wang *et al.*, 2018b)—, hasta plantas y hongos (Dong *et al.*, 2017; Hsieh *et al.*, 2015; Mizuguchi *et al.*, 2014; Ricci *et al.*, 2019; Wang *et al.*, 2018a; Winter *et al.*, 2018) e incluso en bacterias (Dame *et al.*, 2020; Le *et al.*, 2013). En este sentido, parece que se ha conservado a través de la evolución un mecanismo primordial que pliega las grandes moléculas de ADN de los genomas en dominios enlazados. Este mecanismo ancestral se basa en la actividad de los complejos SMC (es decir, cohesina, condensina) y de las topoisomerasas del ADN de tipo II, que siguen estando presentes en bacterias, arqueas y eucariotas. No obstante, muchas características de la organización tridimensional de la cromatina están lejos de conservarse universalmente.

Una mejor comprensión de los factores y mecanismos implicados en el establecimiento o mantenimiento de la organización tridimensional de la cromatina debe ayudar a aclarar estas cuestiones.

A pesar de los enormes avances en el estudio de la arquitectura tridimensional del genoma que hemos presenciado en los últimos años, muchas preguntas esenciales siguen sin respuesta. Los métodos relacionados con la captura de la conformación de los cromosomas (3C), concretamente el método Hi-C, han mejorado en gran medida nuestra comprensión actual de la organización tridimensional de la cromatina, en particular en el ámbito descriptivo. Sin embargo, el Hi-C (y otros métodos relacionados con la 3C) tiene limitaciones intrínsecas, ya que proporciona mediciones estáticas, de promedio poblacional (en masa) y emparejadas de las interacciones físicas entre pares de secuencias en todo el genoma, lo que en principio pudiera enmascarar o distorsionar principios fundamentales de la organización del genoma. A este respecto, las características estructurales reales de los TAD siguen siendo objeto de debate. Por tanto, existe una gran necesidad de desarrollar metodologías novedosas que nos proporcionen una visión más dinámica, unicelular y multidireccional de la arquitectura tridimensional del genoma sin perder las propiedades de genoma completo y alta resolución de los datos de Hi-C.

**También falta un marco teórico para comprender las fuerzas y los principios que rigen el plegado de la cromatina.** Por otra parte, la caracterización de la maquinaria macromolecular y de los mecanismos implicados en el establecimiento y mantenimiento de la organización tridimensional de la cromatina sigue siendo esquiva. Además, los pocos trabajos comparativos realizados hasta ahora están revelando importantes diferencias sobre cómo las distintas especies eucariotas estructuran su cromatina en el espacio nuclear, pero no se conocen bien las bases moleculares de estas diferencias. Y por último, pero no por ello menos importante, las consecuencias funcionales de la organización tridimensional del genoma y sus efectos causales en los procesos basados en plantillas de ADN son todavía poco conocidos. A continuación, repasamos nuestros actuales conocimientos sobre la organización tridimensional del genoma y esbozamos los principales desafíos metodológicos, conceptuales y traslacionales a los que se enfrenta este campo.

## 3.2 IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS POSIBLES APLICACIONES

### 3.2.1 Desafíos técnicos y metodológicos

Los métodos actuales para el estudio de la estructura tridimensional del cromosoma y del genoma se basan en unos pocos enfoques experimentales, principalmente la microscopia de superresolución y Hi-C. Estas metodologías presentan limitaciones, sesgos y dificultades técnicas que pueden ensombrecer y desdibujar los resultados obtenidos. De hecho, los métodos basados en Hi-C se fundamentan en la fijación y ligadura de regiones genómicas en estrecha proximidad tridimensional. Pero la necesidad de ligadura también ha limitado la resolución y ha impedido que las regiones que se encuentran fuera del rango de ligadura se incorporen a la estructura general. Además, la fijación depende de los procesos químicos y, por tanto, puede revelar solo un subconjunto de interacciones. Las interacciones metafísicas o transitorias pueden pasar desapercibidas o estar infrarrepresentadas. Por otra parte, las matrices Hi-C, a partir de las cuales se infieren los TAD, no representan los contactos de cromatina presentes en una sola célula, sino que reflejan conjuntos de eventos de coligación generados durante múltiples procesos dinámicos (es decir, extrusión de bucles, transcripción, remodelación, etc.), que se encuentran en diferentes posiciones en distintas células. Los métodos de microscopia de superresolución superan algunas de estas limitaciones ya que, en principio, permiten la visualización dinámica de células individuales. Sin embargo, la detección de agentes específicos (epítomos y etiquetas proteicas, secuencias de ADN, etc.) está limitada y condicionada por el número de sondas disponibles (colorantes, fluoróforos, anticuerpos, oligonucleótidos,...) y su capacidad para producir señales sostenidas y cuantitativas. Y lo que es más importante, solo hay unos pocos métodos de imagen de células vivas que puedan utilizarse para analizar las distancias físicas a lo largo del tiempo dentro de las células individuales. Por último, las actuales tecnologías Hi-C y de superresolución no pueden trazar aún el recorrido de las moléculas individuales de ADN ni la arquitectura interna de los conjuntos macromoleculares *in vivo*, que exigen una verdadera resolución a escala nanométrica. Los esfuerzos futuros deben centrarse en implementar o combinar métodos que nos proporcionen una visión de todo el genoma, de alta resolución, multidireccional, unicelular y dinámica de la organización tridimensional del genoma. A continuación, analizamos los desarrollos metodológicos para abordar estos desafíos.

## I. Técnicas de imagen

La microscopia electrónica y de superresolución se han utilizado para diseccionar las características genómicas, incluidas las técnicas de imagen de superresolución específica de la secuencia de los dominios de contacto (Boettiger *et al.*, 2016; Ou *et al.*, 2017). Estos estudios de superresolución han proporcionado los primeros atisbos de la naturaleza física de los dominios de contacto, como sus volúmenes y formas. Es más, los métodos de microscopia de superresolución que utilizan oligonucleótidos derivados de matrices, como OligoSTORM y OligoDNA-PAINT, pueden emplearse para obtener una visión de alta resolución, de célula única y multidireccional del plegado del ADN en cualquier tipo de célula o tejido de interés (Bintu *et al.*, 2018; Mateo *et al.*, 2019). Asimismo, estos novedosos métodos de imagen pueden combinarse con el de RNA-FISH para analizar al mismo tiempo la transcripción en curso. No obstante, estos métodos siguen ofreciendo una cobertura genómica limitada, dado el acotado número de oligos que pueden utilizarse, el tamaño restringido y el número de regiones que pueden analizarse al mismo tiempo, así como el número finito de células que pueden controlarse. Las mejoras en la obtención de imágenes multicolor de células vivas y la implementación de nuevos enfoques que emplean técnicas de imagen basadas en CRISPR-Cas9 (Chen *et al.*, 2013), así como estrategias de código de barras basadas en colorantes (Beliveau *et al.*, 2015) también pueden aumentar en gran medida nuestra comprensión de la organización tridimensional del genoma en células individuales. Otra limitación importante de los métodos descritos con anterioridad hace referencia a que proporcionan una visión estática de la estructura tridimensional. Sin embargo, teniendo en cuenta que las interacciones ADN-proteína y proteína-proteína que controlan el plegado del ADN son muy dinámicas, lo más lógico es que la arquitectura del genoma también lo sea. Los métodos basados en el etiquetado de *loci* de interés (por ejemplo, el sistema de etiquetado MS2) pueden utilizarse para analizar la dinámica de la organización tridimensional del genoma en células vitales individuales (Alexander *et al.*, 2019; Fukaya *et al.*, 2016). Lamentablemente, estos métodos adolecen de una cobertura y resolución genómica limitadas.

## II. Hi-C

Hasta la fecha, la mayoría de los datos de Hi-C generados se basan en pares de contactos. Las nuevas tecnologías debieran permitir el análisis de los multi-contactos, así como de las interacciones intercromosómicas, que hoy en día creemos que son menos frecuentes, pero que bien podrían ser indetectables con las tecnologías actuales. Recientemente se han establecido métodos basados en la secuenciación, ortogonales a Hi-C (por ejemplo, GAM o SPRITE),

que pueden proporcionar perfiles del genoma completo, de alta resolución y multidireccionales de la organización tridimensional del genoma (Beagrie *et al.*, 2017; Quinodoz *et al.*, 2018). Además, estos nuevos métodos parecen especialmente adecuados para detectar contactos intercromosómicos de muy largo alcance. Por desgracia, al igual que el Hi-C, estos métodos se basan en mediciones del promedio poblacional en cientos o miles de células. Por otro lado, aunque los métodos Hi-C de célula única ya se han implementado, siguen adolecendo de una baja resolución y de mediciones por pares de las interacciones físicas (Nagano *et al.*, 2013). Las mejoras en cualquiera de estos aspectos de la metodología tendrán importantes repercusiones en nuestra comprensión de la arquitectura tridimensional del genoma.

### **III. Modelado**

Otro desafío consiste en reconstruir modelos tridimensionales de alta resolución de grandes genomas a partir de datos de Hi-C, lo que resulta necesario para estudiar las interacciones detalladas entre genes y elementos reguladores. La enorme complejidad temporal y la escasez de datos asociados al modelado de alta resolución resultan limitaciones importantes. A pesar de la mejora en los enfoques de modelado de estructuras tridimensionales, la falta de una estructura real con la que contrastar estos modelos sigue suponiendo un desafío. En particular, hoy día resulta difícil confirmar la verdadera capacidad de modelado de los métodos tridimensionales del genoma. Además, dada la posible conexión de la alteración de la estructura tridimensional del genoma con las enfermedades, es importante que científicos biomédicos puedan utilizar fácilmente en sus investigaciones los métodos de modelado tridimensional del genoma.

### **IV. Criomicroscopía electrónica (crioME)**

Actualmente, se está produciendo una revolución en la caracterización de grandes complejos macromoleculares basada en la criomicroscopía electrónica (Bai *et al.*, 2015). No queda claro si estos métodos pueden ampliarse para estudiar grandes dominios cromosómicos dentro de las células. Sin embargo, cabe la posibilidad de que los métodos de crioME mejorados y adaptados sean esenciales para desentrañar la estructura de los complejos proteicos que resultan fundamentales para la organización tridimensional del genoma.

### **V. Separación de fases**

Uno de los desafíos consistirá en conciliar la existencia de los dominios de cromatina y los TAD mapeados a partir de los datos de Hi-C con la presencia de condensados u orgánulos nucleares sin membrana formados a partir de

diferentes tipos de procesos de separación de fases. Hay que esforzarse en el desarrollo de técnicas que permitan la caracterización molecular de estos condensados *in vivo*, con lo que se conseguiría su visualización en el interior de las células, el aislamiento de su contenido y el mapeo preciso de sus interacciones tridimensionales. Sin duda, esto constituye un gran desafío debido a la naturaleza transitoria, heterogénea y dinámica de estas estructuras y, por lo tanto, requiere un enfoque multidisciplinario y una estrecha colaboración entre biólogos celulares, físicos y químicos, lo cual es fundamental en el desarrollo de nuevas tecnologías que permitan abordar estas cuestiones.

## **VI. Topología del ADN**

Las tecnologías actuales de Hi-C y de imagen no pueden trazar aún la trayectoria de las moléculas individuales de ADN. Por ese motivo, se requieren metodologías novedosas basadas en el análisis topológico del ADN intracelular para inferir cómo se deforma la trayectoria de las moléculas de ADN (es decir, la torsión helicoidal del ADN y la flexión axial) mediante elementos individuales de la cromatina *in vivo*. Las fuerzas que actúan sobre las moléculas de ADN afectan al equilibrio conformacional e impulsan las transiciones de la arquitectura de la cromatina. Este es el caso de la superhelicidad (tensión doble helicoidal) del ADN generada durante las transacciones del genoma (transcripción, replicación), que se propaga e interactúa de manera profunda con la estructura y la función de la cromatina. La forma de evaluar la tensión superhelicoidal del ADN intracelular ha supuesto un desafío desde hace mucho tiempo.

## **VII. Estudios *in vitro***

Otro desafío tiene que ver con la reconstrucción *in vitro* de los conjuntos macromoleculares que dirigen/reflejan los estados de plegado de la cromatina (complejos SMC, remodeladores de la cromatina, agentes de condensación y transición de fase), y el posterior análisis de sus propiedades fisicoquímicas y mecanicistas (pinzas ópticas y magnéticas, AFM, criome, topología del ADN, bioquímica de las enzimas, etc.).

### **3.2.2 Factores y mecanismos que rigen la arquitectura tridimensional del genoma**

La mayor parte de lo que sabemos sobre la organización tridimensional del genoma y sus mecanismos moleculares subyacentes procede del estudio de los vertebrados, especialmente los mamíferos, y de un puñado de especies modelo adicionales, en particular la mosca de la fruta *Drosophila melanogaster*. En esta sección, revisaremos brevemente los mecanismos que se conocen para



controlar los diferentes niveles de plegado tridimensional del genoma en los vertebrados y su posible conservación a través de la evolución. Y lo que es más importante, también pondremos de relieve importantes cuestiones mecanicistas que siguen sin resolverse. Aunque cabe la posibilidad de que la arquitectura tridimensional del genoma afecte a todos los procesos basados en el ADN, aquí nos centraremos preferentemente en la transcripción durante la interfase.

### ***I. TAD y compartimentos***

A una gran escala genómica (de decenas a cientos de Mb), los estudios de Hi-C revelaron que los cromosomas se organizan en compartimentos que resultan de la agrupación espacial de regiones genómicas con un estado de la cromatina y transcripcional similar (Bonev *et al.*, 2017; Le *et al.*, 2013; Rao *et al.*, 2014).

Se han descrito dos tipos principales de compartimentos: el *compartimento A* incluye *loci* genómicos ricos en genes, transcripcionalmente activos y marcados con modificaciones activas de las histonas (por ejemplo, H3K27ac o H3K4me2/3); y el *compartimento B* está formado por *loci* pobres en genes, transcripcionalmente silenciosos y marcados con modificaciones represivas de las histonas (por ejemplo, H3K9me2/3 o H3K27me3). El *compartimento B* puede subdividirse a su vez de esta forma: (i) *loci* de heterocromatina constitutiva, que se suelen encontrar en la periferia del núcleo (como parte de los dominios asociados a la lámina [LAD]) y en los nucléolos, marcados con H3K9me2/3 y con ADN hipometilado; (ii) *loci* de heterocromatina facultativa, que están unidos por complejos de proteínas PcG, con ADN hipometilado y localizados en el interior del núcleo.

En la submegabase (de decenas a cientos de kb), los compartimentos pueden subdividirse en TAD (Dixon *et al.*, 2012; Nora *et al.*, 2012). Los TAD están separados unos de otros por sus límites o fronteras, que suelen coincidir con los sitios de unión de CTCF y, en menor medida, con los genes de mantenimiento, los genes de ARNt y las repeticiones de SINE (Dixon *et al.*, 2012). Es importante destacar la propuesta de que los TAD representan unidades reguladoras fundamentales en las que tienen lugar la mayoría de las interacciones entre los potenciadores y sus genes diana. Por consiguiente, los TAD podrían: (i) facilitar las interacciones entre los potenciadores y sus genes diana; y (ii) aislar a los genes del establecimiento de interacciones ectópicas con los potenciadores equivocados (Spielmann *et al.*, 2018).

Tras el hallazgo de los TAD y los compartimentos, se han dedicado grandes esfuerzos a dilucidar los mecanismos implicados en su formación. Estudios

recientes han demostrado de forma concluyente que la formación de los TAD depende tanto de CTCF como del complejo de cohesina (Nora *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017; Zhang *et al.*, 2019), ya que la depleción de cualquiera de estas dos proteínas conduce a una pérdida casi completa de todos los TAD. Además, la formación de los TAD parece explicarse por un mecanismo de *extrusión de bucle*, por el que la cohesina se carga al principio en los potenciadores y promotores para luego formar bucles cada vez más grandes que finalmente se estancan en los límites de TAD formados por motivos convergentes de CTCF (Fudenberg *et al.*, 2016; Rao *et al.*, 2014). Actualmente se están dilucidando los detalles mecanicistas de cómo funciona en realidad este modelo de *extrusión de bucle*. Por ejemplo, ahora se ha demostrado que la cohesina humana puede extruir bucles de ADN de forma simétrica, rápida y dependiente del ATP (Davidson *et al.*, 2019; Kim *et al.*, 2019; Vian *et al.*, 2018). Del mismo modo, estudios estructurales recientes explican cómo la interfaz de interacción entre la cohesina y CTCF conduce a la aparición preferente de límites de TAD en sitios convergentes de CTCF (Li *et al.*, 2020). A pesar de estos importantes avances, **la relevancia funcional e incluso la existencia de los TAD es objeto de un intenso debate científico**. Se ha sugerido que los TAD pudieran representar una propiedad emergente del promedio poblacional celular medido mediante Hi-C, así como un artefacto computacional carente de significado biológico (Rowley *et al.*, 2017). Sin embargo, trabajos recientes basados en el Hi-C de célula única y en la microscopia de superresolución han confirmado que las estructuras tipo TAD existen realmente dentro de las células individuales (Bintu *et al.*, 2018; Nagano *et al.*, 2013). Además, aunque los límites de TAD mostraron variación entre las células individuales, todavía se superponen con frecuencia con los sitios de CTCF (Bintu *et al.*, 2018). Curiosamente, en ausencia de cohesina, aún se observaron estructuras similares a los TAD dentro de las células individuales, aunque sus límites se volvieron muy variables y no coincidieron con los sitios de CTCF (Bintu *et al.*, 2018).

Inicialmente, los TAD y los compartimentos se consideraron como dos capas relacionadas jerárquicamente de la organización tridimensional del genoma que actúan en diferentes escalas genómicas. Sin embargo, estudios recientes en los que la cohesina se eliminó de manera inducible sugirieron que, en realidad, podrían representar dos modos independientes de organización de la cromatina (Rao *et al.*, 2017; Schwarzer *et al.*, 2017). Tras la pérdida de cohesina, los TAD desaparecieron por completo, mientras que la segregación de compartimentos incluso se reforzó y se hizo evidente en un nivel más preciso de submegabase que reflejaba el estado subyacente de la cromatina. Todo ello

llevó a sugerir que la *extrusión de bucle* dependiente de la cohesina y las interacciones homotípicas de la cromatina representan dos mecanismos independientes y opuestos de organización tridimensional (Schwarzer *et al.*, 2017). Por otro lado, trabajos recientes también han aclarado algunos de los mecanismos por los que las interacciones homotípicas de la cromatina pueden conducir a la segregación de *loci* distales en compartimentos específicos. Es importante destacar que estos mecanismos parecen ser específicos para cada tipo de compartimento y actualmente se conocen mejor para el compartimento B. En concreto, la segregación espacial de los *loci* marcados con heterocromatina constitutiva parece depender de la unión a la lámina nuclear o a los nucléolos y de la separación de fases impulsada por HP1 (Larson *et al.*, 2017; Strom *et al.*, 2017; van Steensel y Belmont, 2017). Por otra parte, la agrupación espacial de los *loci* unidos por PcG está mediada por subunidades del complejo canónico PRC1, incluidas la separación de fases impulsada por CBX2 y la polimerización debido a los dominios SAM presentes en las subunidades PHC1/2 (Isono *et al.*, 2013; Plys *et al.*, 2019; Tatavosian *et al.*, 2019).

El hallazgo de los TAD y los compartimentos ha cambiado significativamente la forma de pensar en la arquitectura del genoma. Ahora bien, todavía hay importantes cuestiones relativas a las fuerzas y mecanismos moleculares que controlan estas dos características estructurales que deben resolverse y que pudieran representar importantes áreas de investigación futura.

## ***II. Interacciones genómicas de corto alcance***

La resolución relativamente baja de los estudios de Hi-C iniciales impidió una visión precisa de la arquitectura dentro de los TAD, pero esto ha cambiado debido al aumento de la resolución de los estudios de Hi-C y al uso de enfoques dirigidos basados en 3C (4C-seq, Hi-C de captura, ChIA-PET, HiChIP) que proporcionan una menor cobertura genómica pero con una mayor resolución (Hughes *et al.*, 2014; Mumbach *et al.*, 2016; Tang *et al.*, 2015; van de Werken *et al.*, 2012). Como resultado, está empezando a surgir una compleja organización tridimensional dentro de los TAD, que incluye entidades topológicas múltiples y a veces superpuestas, como sub-TAD, dominios de bucle o dominios aislados (Dowen *et al.*, 2014; Phillips-Cremins *et al.*, 2013; Rao *et al.*, 2014). Los mecanismos que controlan estas capas de organización tridimensional dentro de los TAD parecen ser similares a los descritos para los TAD y, por tanto, incluyen la extrusión de bucle dependiente de la cohesina y las interacciones entre CTCF y la cohesina. Sin embargo, estas estructuras intra-TAD son más variables y específicas del tipo de célula, y su significado funcional aún no se ha estudiado en su totalidad.

Otro grupo importante de interacciones intra-TAD son las que ponen en proximidad física a los potenciadores y a sus genes diana. Con el descubrimiento de los TAD y la confirmación de la extrusión de bucle dependiente de la cohesina como mecanismo implicado en la formación de TAD, se pensó inicialmente que la mayoría de las interacciones potenciador-gen dependerían de la cohesina y la extrusión de bucle (Dixon *et al.*, 2012; Kagey *et al.*, 2010; Phillips-Cremins *et al.*, 2013). Del mismo modo, se ha planteado que muchas interacciones entre potenciadores y genes dependen de la dimerización de las moléculas de CTCF unidas a potenciadores y promotores de genes (Guo *et al.*, 2015). Sin embargo, cada vez hay más evidencias que indican que CTCF y cohesina solo están implicados en un número limitado de contactos potenciador-gen y que otros mecanismos alternativos podrían ser más frecuentes (Rao *et al.*, 2017; Schwarzer *et al.*, 2017). Por ejemplo, de forma análoga a CTCF, YY1 puede unirse tanto a potenciadores como a promotores y formar dímeros que facilitan las interacciones entre estos elementos reguladores (Beagan *et al.*, 2017). La dimerización es también el mecanismo a través del cual la proteína coactivadora LDB1 puede facilitar los contactos potenciador-gen de largo alcance (Deng *et al.*, 2014). Por otro lado, el proceso biofísico de separación de fases líquido-líquido también pudiera estar implicado en las interacciones potenciador-gen (Hnisz *et al.*, 2017). En este modelo, los potenciadores y promotores sirven como plataformas de reclutamiento para múltiples proteínas (por ejemplo, FT, ARN pol II, mediadoras, modificaciones de histonas) y moléculas de ARN que luego se involucran en interacciones débiles pero multivalentes y cooperativas. Como resultado, se pueden formar orgánulos sin membrana o condensados con propiedades similares a las de un gel, dentro de los cuales se facilitan y establecen los contactos potenciador-gen (Hnisz *et al.*, 2017). Un componente central del modelo de separación de fases son las regiones intrínsecamente desordenadas (IDR, por sus siglas en inglés) que se encuentran en múltiples factores de transcripción y coactivadores, como ciertas subunidades del complejo mediador (Boija *et al.*, 2018; Sabari *et al.*, 2018). Estas IDR facilitan la separación de fases al servir como plataformas flexibles y pleiotrópicas para las interacciones proteína-proteína. Sin embargo, aunque se ha confirmado la existencia de condensados transcripcionales *in vivo*, la mayoría de los conocimientos sobre los mecanismos implicados en su formación proceden de experimentos *in vitro*. Por ejemplo, MED1-IDR puede formar gotas separadas por fases *in vitro*, pero la pérdida completa del complejo mediador no tiene casi ningún impacto en la organización tridimensional del genoma ni en las interacciones entre los genes y los potenciadores (El Khattabi *et al.*, 2019; Sabari *et al.*, 2018).

Los potenciadores pueden mostrar estados capacitados o iniciados, que se supone que facilitan la futura activación de sus genes diana (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011). Se ha informado de que tanto los potenciadores capacitados como los iniciados interactúan físicamente con sus genes diana antes de pasar a un estado activo (Cruz-Molina *et al.*, 2017; Ghavi-Helm *et al.*, 2014). En el caso de los potenciadores capacitados, estos contactos preformados parecen estar mediados por complejos PcG unidos tanto a los potenciadores como a los promotores de genes diana. Estos contactos dependientes de PcG están mediados por la capacidad de polimerización de los dominios SAM presentes en las proteínas PHC1/2, que son componentes clave del complejo PRC1 canónico (Isono *et al.*, 2013). En el caso de los potenciadores iniciados, las interacciones preformadas con los genes diana pudieran requerir la presencia de las proteínas H3K4me1 y MLL3/4 (Yan *et al.*, 2018). Se ha sugerido que la H3K4me1 pudiera facilitar el reclutamiento de cohesina a los potenciadores, lo que puede llevar a los genes y a los potenciadores a una proximidad física (Yan *et al.*, 2018). Por último, pero no menos importante, los silenciadores representan otra clase importante de elementos reguladores que pueden influir negativamente en la expresión de sus genes diana. Históricamente, los silenciadores han resultado difíciles de identificar y caracterizar y, por lo tanto, sus características topológicas siguen siendo en gran parte desconocidas. Sin embargo, estudios recientes indican que algunos silenciadores pudieran contactar físicamente con sus genes diana inactivos (Ngan *et al.*, 2020; Pang y Snyder, 2020).

En resumen, los contactos intra-TAD son muy complejos, con la participación de múltiples y diversas estructuras tridimensionales. En consecuencia, aún quedan muchas cuestiones mecanicistas por abordarse en los próximos años.

### ***III. ¿Son los TAD unidades universales de organización del genoma?***

Se han observado dominios de autointeracción similares de manera superficial a los TAD de mamíferos en una gran variedad de organismos, desde otras especies animales —a saber, *Drosophila* (Rowley *et al.*, 2017; Wang *et al.*, 2018b)—, hasta diferentes linajes eucariotas como plantas y hongos (Dong *et al.*, 2017; Hsieh *et al.*, 2015; Mizuguchi *et al.*, 2014; Ricci *et al.*, 2019; Wang *et al.*, 2018a; Winter *et al.*, 2018) y bacterias (Dame *et al.*, 2020; Le *et al.*, 2013). Ahora bien, muchas de las características de estas estructuras similares a los TAD distan de estar conservadas de manera universal, con diferencias en su tamaño, la nitidez de sus límites, la intensidad de los contactos y el contenido genómico (es decir, regiones ricas y pobres en genes, abundancia de elementos transponibles y

repetitivos, etc.). Aun así, parece surgir un tema común al comparar diferentes especies de plantas, hongos, apicomplejos y animales (Rowley *et al.*, 2017). En todos estos casos, la actividad transcripcional y las características de la cromatina asociadas a ella crean límites entre los dominios de contacto, lo que parte el genoma en regiones genómicas alternas transcripcionalmente activas e inactivas (Rowley *et al.*, 2017). No obstante, este principio organizativo compartido puede dar lugar a patrones de contacto muy diferentes, según el grado de compacidad del genoma de cada especie y de la distribución y densidad de los genes transcritos. Por ejemplo, en *Drosophila*, los genes transcripcionalmente activos se encuentran con frecuencia en grupos de diferentes tamaños que separan grandes TAD correspondientes a regiones reprimidas (Polycomb reprimido enriquecido en H3K27me3, así como heterocromatina clásica enriquecida en H3K9me2) (Rowley *et al.*, 2017; Wang *et al.*, 2018b). Por el contrario, en los hongos filamentosos *Epichloë festucae*, hay una alternancia entre dominios que contienen bloques de ADN muy ricos en repeticiones y regiones ricas en genes que están casi libres de repeticiones (Winter *et al.*, 2018).

#### ***IV. Evolución de las interacciones reguladoras de largo alcance***

Además de estas diferencias genómicas, varios linajes han desarrollado mecanismos moleculares adicionales además de la compartimentación preexistente encontrada en todos los eucariotas estudiados, como la presencia de proteínas aislantes/arquitectónicas. En los vertebrados, existe una capa adicional de organización y las interacciones homotípicas de la cromatina pueden sustituirse, al menos parcialmente, por la presencia de sitios de unión de CTCF invertidos. Como hemos mencionado anteriormente, CTCF junto con otras proteínas como YY1, permiten la formación de bucles de cromatina de largo alcance entre elementos reguladores y los promotores de sus genes diana. Dado que YY1 y CTCF se encuentran de forma exclusiva en animales y animales bilaterales respectivamente (Heger *et al.*, 2012; Irimia y Maeso, 2019), el origen de estas proteínas arquitectónicas pudiera explicar la evolución de la regulación de largo alcance y los TAD a gran escala que se suelen encontrar en torno a los genes del desarrollo animal (Harmston *et al.*, 2017). Esta hipótesis estaría respaldada por el hecho de que algunas especies de nematodos, como *Caenorhabditis elegans*, han perdido CTCF e YY1 y estas pérdidas parecen ser concomitantes con el desmantelamiento de la mayor parte de la regulación ancestral de largo alcance en estos linajes, los cuales se caracterizan por sus genomas altamente compactos y densos en genes (Crane *et al.*, 2015; Heger *et al.*, 2012; Heger *et al.*, 2009; Jabbari *et al.*, 2018). Sin embargo, la situación resulta más compleja de lo que se indica en este supuesto. Además, la comprensión de la relación entre la

evolución de las interacciones reguladoras de largo alcance y las proteínas arquitectónicas requerirá una cantidad significativa de trabajo en una variedad mucho más amplia de especies animales y no animales. Por ejemplo, en *Drosophila*, CTCF no se requiere para el establecimiento de bucles de cromatina (Rowley *et al.*, 2017). La evolución de múltiples proteínas arquitectónicas nuevas en los linajes de insectos y moscas (Heger *et al.*, 2013; Pauli *et al.*, 2016) pudiera haber contribuido a la pérdida del papel putativamente ancestral de CTCF, pero esta hipótesis no se ha confirmado por completo.

Además, aunque la regulación de largo alcance es ancestral en todos los animales (Gaiti *et al.*, 2017a; Gaiti *et al.*, 2017b; Grau-Bove *et al.*, 2017; Irimia *et al.*, 2013; Irimia *et al.*, 2012; Schwaiger *et al.*, 2014; Sebe-Pedros *et al.*, 2016), CTCF no lo es, ya que se originó después de la divergencia de los animales bilaterales (Heger *et al.*, 2012). De esta manera, la regulación distal en los ancestros de los metazoos era independiente de CTCF y, aunque aún no se ha probado de manera experimental, también pudiera ser el caso en linajes existentes de no bilaterales, como las esponjas y los cnidarios (Irimia y Maeso, 2019).

Por último, estudios recientes a través de diferentes especies de plantas han demostrado que los elementos reguladores en *cis* distales están muy extendidos entre las angiospermas y, de forma similar a lo que ocurre en ciertos linajes animales como los nematodos, solo aquellas especies con genomas extremadamente compactos y densos, como la especie modelo *Arabidopsis thaliana*, carecen en gran medida de interacciones reguladoras de largo alcance (Lu *et al.*, 2019; Ricci *et al.*, 2019). Cómo se mantienen los bucles reguladores en *cis* de largo alcance en las plantas es una cuestión completamente abierta y aún hoy se desconoce si estos bucles son consecuencia de la segregación compartimental, si se establecen mediante proteínas arquitectónicas específicas de la secuencia (hasta ahora no descritas en las plantas) o por la acción combinada de ambos tipos de mecanismos moleculares (Ricci *et al.*, 2019).

En resumen, parece que hay muchos tipos diferentes de estructuras tipo TAD en los distintos organismos, que pueden formarse por diferentes mecanismos moleculares, algunos de los cuales comparten todos los eucariotas mientras que otros son específicos de cada linaje. Así, aunque estos mecanismos diversos pueden conducir a la formación de dominios de interacción similares en superficie, actualmente no está claro hasta qué punto estos dominios tridimensionales representan características estructurales ancestrales o han evolucionado de manera convergente como resultado de propiedades funcionales compartidas (Szabo *et al.*, 2019).



### 3.2.3 Impacto fisiológico de la estructura tridimensional del genoma

Las estructuras de cromatina de orden superior surgen como supuestos bloques de construcción del genoma que se supone que tienen un papel funcional. Sin embargo, hallar este papel se ha vuelto más complejo de lo previsto. Ha resultado difícil encontrar papeles causales directos en la función biológica y, en otros casos, parece que la estructura pudiera ser un resultado secundario de otras funciones del genoma. En esta sección, abordamos lo que se sabe sobre el papel de la estructura tridimensional de la cromatina en otras funciones básicas del genoma, así como en diferentes contextos fisiológicos.

#### Interacción de la estructura tridimensional con la función del genoma

##### *I. Replicación*

Varios estudios han abordado la relación entre la replicación del ADN y la estructura tridimensional del genoma (revisado en Marchal *et al.*, 2019). Los dominios temporales de replicación temprana y tardía corresponden de manera aproximada a los compartimentos A y B respectivamente, mientras que, a menor escala genómica, existe una buena correlación entre los TAD y los dominios de replicación. La reciente descripción de los elementos de control de replicación temprana (ERCE) sugiere un vínculo causal en el que la replicación tiene un papel instructivo en la estructura tridimensional del genoma (Sima *et al.*, 2019). Otras pruebas de que la replicación no depende de la estructura tridimensional provienen de estudios en etapas tempranas del desarrollo, donde la replicación tiene lugar mientras la estructura tridimensional aún no se ha establecido (revisado en Hug y Vaquerizas, 2018).

##### *II. Transcripción*

Durante mucho tiempo se ha asumido que el efecto principal y la consecuencia directa de la organización tridimensional de la cromatina pudiera ser la transcripción diferencial de los genes localizados en diferentes dominios estructurales. Esto resulta más evidente para los compartimentos A y B, que corresponden a la eucromatina y la heterocromatina respectivamente y están enriquecidos en genes activos e inactivos a su vez. Los genes de las regiones que cambian de compartimento A a B durante la diferenciación reducen su expresión y viceversa en el caso de los cambios de B a A (Dixon *et al.*, 2015). En cuanto a la relevancia funcional de los TAD como unidades reguladoras fundamentales de la expresión génica, existen varios estudios contradictorios. Por un lado, las variantes estructurales (deleciones, inversiones, duplicaciones) que alteran la organización del TAD pueden conducir a una pérdida de las interacciones endógenas



entre el potenciador y el gen (*desconexión del potenciador*) o a una ganancia de interacciones ectópicas entre el potenciador y el gen (*adopción del potenciador*) que derivan en el silenciamiento o la activación patológica del gen respectivamente (Laugsch *et al.*, 2019; Lupiáñez *et al.*, 2015; Smol *et al.*, 2020; Spielmann *et al.*, 2018). Por otro lado, la alteración estructural de ciertos TAD no tiene efectos importantes en la expresión génica (Ghavi-Helm *et al.*, 2019; Laugsch *et al.*, 2019), si bien la alteración global de la organización de los TAD debido a la pérdida de CTCF o cohesina se traduce en cambios moderados en dicha expresión (Nora *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017). En lo que respecta a los bucles o dominios de autointeracción, estas estructuras acercarían los potenciadores y otros elementos reguladores en *cis* a sus genes diana, lo que facilita su correcta regulación (revisado en Schoenfelder y Fraser, 2019). Sin embargo, el análisis cuidadoso de casos paradigmáticos de elementos potenciadores de largo alcance, como los de los *loci* Shh y Sox2, indica que la estructura de los TAD o el contacto físico potenciador-promotor mediado por los bucles de cromatina no son un requisito previo para la actividad (Alexander *et al.*, 2019; Benabdallah *et al.*, 2019; Williamson *et al.*, 2019).

En este sentido, **existen pruebas contradictorias sobre el papel de la organización tridimensional del genoma en la regulación de la expresión génica**, y tal vez tenga un papel facilitador más que instructivo. Además, la actividad transcripcional influye en la estructura de la cromatina. Por lo tanto, existe una conversación cruzada y dinámica entre la transcripción y la organización del genoma, donde cada una puede modular la actividad de la otra (revisado en van Steensel y Furlong, 2019).

### **3.2.4 Impacto fisiológico de la estructura tridimensional del genoma**

#### ***I. Desarrollo y evolución***

El desarrollo normal depende, no solo de la secuencia lineal del genoma que incorpora millones de CRE sino también de la organización tridimensional de la cromatina. La cromatina organiza las interacciones entre los CRE y sus genes diana y, de esta manera, modula procesos biológicos cruciales para la diferenciación y el desarrollo celular. Los TAD pueden definirse simplemente como unidades genómicas funcionales de regulación génica, en las que los CRE interactúan con sus genes diana. Los TAD son muy estables, desde el punto de vista de su posicionamiento, entre tipos de células y tejidos, independientemente del estado transcripcional (Dixon *et al.*, 2015). Además, la actividad de los

promotores y potenciadores parece estar muy descoordinada dentro del mismo TAD. Esto ha sugerido la existencia de una topología muy estable y preformada que establece la proximidad física entre los potenciadores y sus genes diana, aunque se han planteado dudas sobre su papel real en los procesos de regulación específicos dentro de las células y los tejidos (Lupiáñez *et al.*, 2015). Las alteraciones específicas del linaje de la organización tridimensional del genoma suelen producirse dentro de los TAD o sub-TAD. Por consiguiente, para determinar la influencia de la organización tridimensional de la cromatina en la regulación genética específica del linaje, es fundamental generar mapas de interacciones de la cromatina con una resolución lo bastante alta para distinguir los elementos reguladores individuales y para una amplia variedad de tejidos y células durante el desarrollo normal. Actualmente, se están utilizando Capture-C/Hi-C y PLAC-seq/HiChIP (Hui y Wei, 2019) para generar dichos mapas de alta resolución. De este modo, parece crucial seguir desarrollando estas tecnologías. El estudio de las posiciones de los CRE y de los genes que regulan en relación con los TAD ofrece numerosas oportunidades para el estudio de la variación de la expresión génica durante la evolución. Hasta la fecha, las escasas pruebas disponibles se han obtenido a partir de un estudio de elementos evolutivos no codificantes conservados (CNE) (Gómez-Marín *et al.*, 2015). Estos elementos se organizan en localizaciones sinténicas, principalmente alrededor de genes clave para el desarrollo (Gómez-Marín *et al.*, 2015). El estudio de dichos elementos ha permitido concluir que, al menos en lo que respecta a los genes del desarrollo, los TAD son estructuras evolutivamente conservadas que pueden desempeñar alguna función en el mantenimiento de la correlación entre los CRE y sus genes diana. Otro punto importante en la conservación evolutiva de los TAD estriba en que, durante la evolución, los genomas animales han experimentado una profunda reorganización, lo que ha cambiado el orden relativo de los TAD y esto, a su vez, ha generado importantes modificaciones en la expresión génica que han derivado en la aparición de novedades evolutivas. Ahora bien, lo que queda por describir son los mecanismos moleculares que subyacen a esta reorganización de los TAD durante la evolución. Además, todavía no sabemos con exactitud cómo interactúan los CRE con sus genes diana. Entender los mecanismos que facilitan las interacciones funcionales tanto dentro de los TAD como entre ellos resulta fundamental para comprender el control de la expresión génica durante el desarrollo desde un punto de vista evolutivo.

## II. Enfermedad

La descripción de la organización tridimensional del genoma llevó rápidamente a plantearse si esta pudiera estar implicada en diferentes aspectos de

las enfermedades humanas. Esta relación puede ser doble. En primer lugar, pudiera estar estableciéndose una relación directa en la que la alteración de la estructura tridimensional de la cromatina condujera a un estado patológico. Por otro lado, el análisis de la estructura del genoma puede aportar nuevos conocimientos sobre los mecanismos de las enfermedades.

Las variaciones estructurales del genoma, como las deleciones o las inversiones, pueden conducir a la reubicación o a la pérdida de los límites de los TAD, lo que da lugar a una reordenación de las interacciones potenciador-promotor que puede causar patologías debido a una expresión génica incorrecta. Estas alteraciones se han denominado *TADopatías* y pueden explicar la base molecular de la leucodistrofia autosómica dominante de inicio en el adulto (ADLD, por sus siglas en inglés) y de algunas malformaciones congénitas de las extremidades (Matharu y Ahituv, 2015). Aunque parece que los reordenamientos de TAD no son forzosamente los causantes de la enfermedad (Smol *et al.*, 2020), hay que tener en cuenta la alteración de los TAD al analizar el efecto patológico de las variaciones estructurales vinculadas a la enfermedad. Además, hay que destacar que su efecto pudiera estar produciéndose en genes que no están incluidos en la eliminación o inversión patológica. Por ejemplo, la eliminación de un límite de TAD puede dar lugar a una expresión errónea de los genes situados fuera del segmento eliminado (Spielmann *et al.*, 2018).

Es frecuente encontrar mutaciones de proteínas de arquitectura cromosómica, como CTCF o el complejo de cohesina, en diferentes tipos de cáncer. Además, se ha demostrado que los cambios en la unión de CTCF al ADN son causantes de enfermedades. En un ejemplo sorprendente descrito en gliomas, la hipermetilación de un sitio CTCF, causada por una mutación de ganancia de función de IDH que produce un metabolito que permite la disminución de la actividad de las desmetilasas TET, lleva a la desregulación de un elemento aislante y a la activación ectópica de PDGFRA, que codifica un potente conductor oncogénico (Krijger y de Laat, 2016). Por supuesto hay más ejemplos relacionados con la metilación aberrante de los sitios CTCF a la espera de ser descubiertos, lo que demuestra una vez más que el conocimiento de la estructura tridimensional de la cromatina puede revelar genes inesperados como causantes de una enfermedad o circunstancia particular. También debemos ser conscientes de que la naturaleza no aleatoria y localmente heterogénea de la estructura del genoma tiene un impacto directo en la distribución y naturaleza de las mutaciones, como las que vemos en el cáncer. Los diferentes

estados de la cromatina imponen restricciones a la secuencia, al igual que la disponibilidad física de los nucleótidos a los agentes mutagénicos, lo que tiene un impacto directo en su mutabilidad (Akdemir *et al.*, 2018; Schuster-Böckler y Lehner, 2012).

Por último, el estudio de la estructura tridimensional del genoma proporciona una herramienta de gran valor para relacionar las variantes de la enfermedad identificadas mediante los estudios de asociación del genoma completo (GWAS, por sus siglas en inglés) con los genes causales subyacentes. Una proporción muy elevada de variantes asociadas a enfermedades o rasgos se encuentra en regiones intergénicas no codificantes del genoma, que muy probablemente contienen elementos reguladores. Estos elementos controlan la transcripción de los genes, pero a menudo la asignación del gen correcto a un supuesto elemento regulador no resulta evidente. Aunque la opción predeterminada consiste en asignar el gen más cercano a las variantes del GWAS, esta puede no ser la correcta. El conocimiento de la estructura genómica tridimensional de la región, como la distribución de TAD o bucles, puede ayudar a identificar correctamente los genes diana de las variantes de la enfermedad (Krijger y de Laat, 2016).

### 3.3 PUNTOS CLAVE DEL DESAFÍO

#### 3.3.1 Hacia una visión dinámica, de célula única y multidireccional de la organización tridimensional del genoma

Nuestra actual comprensión de la organización tridimensional del genoma carece de una visión más dinámica, de célula única y multidireccional. Por lo tanto, existe una gran necesidad de desarrollar nuevas metodologías y técnicas que nos proporcionen una visión estructural dinámica de alta resolución del genoma completo a nivel de células individuales. Lo ideal sería que estos métodos se combinaran con el análisis de otros procesos basados en plantillas de ADN (transcripción, replicación, reparación del ADN, etc.) para responder a algunas cuestiones importantes sobre la relevancia biológica y la base mecanicista de la arquitectura tridimensional del genoma.

#### 3.3.2 Hacia una visión evolutiva de la organización tridimensional del genoma

Hoy en día cuesta saber cuánto de lo que hemos aprendido sobre los mamíferos y sobre algunos organismos modelo comunes puede extrapolarse a otras especies y diferenciar los rasgos ancestrales de los específicos de un linaje, lo

que dificulta nuestra capacidad para sacar conclusiones generales sobre los mecanismos responsables de la organización tridimensional del genoma y a qué funciones biológicas contribuye en los distintos organismos. De hecho, las pocas especies no vertebradas y no animales estudiadas hasta ahora han revelado importantes diferencias entre linajes. Así pues, parece imperativo caracterizar la organización tridimensional de la cromatina en una variedad mucho más amplia de especies que las que se están analizando actualmente, incluidos representantes de diversos linajes eucariotas.

### **3.3.3 Hacia un marco mecanicista y teórico unificado de la organización tridimensional del genoma**

Falta una visión unificada de los mecanismos que rigen la organización tridimensional del genoma. La segregación de los *loci* genómicos en compartimentos guarda una estrecha correlación con los panoramas epigenéticos y de cromatina subyacentes. En cambio, los mecanismos que permiten la interacción preferente entre *loci* con un estado similar de la cromatina (es decir, las interacciones homotípicas de la cromatina) siguen siendo hasta ahora bastante indeterminados. La separación de fases, ya sea líquida o basada en polímeros, pudiera proporcionar una base biofísica general para el plegado de la cromatina. Por otra parte, la formación de condensados transcripcionales separados por fases emerge como un mecanismo potencial para corregular la expresión génica. Ahora bien, los condensados separados por fases son difíciles de estudiar *in vivo*. Los modelos actuales proponen interacciones multivalentes entre muchos componentes de proteínas y ARN, lo que sugiere un sistema altamente redundante que pudiera ser resistente a la alteración de componentes individuales (por ejemplo, Mediator). En definitiva, ¿qué tipo de aproximaciones experimentales debieran diseñarse para diseccionar los mecanismos y las fuerzas moleculares que controlan la separación de fases? ¿Cuál es la dinámica de la formación de condensados transcripcionales y qué mecanismos controlan su desmantelamiento? ¿Cómo podemos integrar los condensados transcripcionales basados en la separación de fases y los TAD basados en la extrusión de bucle? ¿Cuántos genes y potenciadores pueden encontrarse y corregularse dentro de un mismo condensado separado por fases? ¿Existen condensados especializados dentro de las células individuales en los que solo se transcriban determinados genes y potenciadores, y cuál es la base de esta especificidad? Los genomas de los vertebrados contienen miles de potenciadores, que están unidos a un gran número de factores de transcripción y coactivadores muy diversos. Teniendo esto en cuenta, los diferentes tipos de potenciadores pudieran utilizar mecanismos distintos (por ejemplo, la

polimerización) para comunicarse con sus genes diana y eso debiera diseccionarse sistemáticamente, así como compararse la dinámica de estas interacciones con la que se producen en los condensados. Estos son solo algunos de los muchos interrogantes mecanicistas que esperan una respuesta.

### **3.3.4 Hacia modelos predictivos de la organización tridimensional del genoma y sus consecuencias funcionales**

Nuestros actuales conocimientos aún son insuficientes para modelar el plegado tridimensional del genoma directamente a partir de los datos genómicos y lo que es más, **no comprendemos del todo las implicaciones funcionales del plegado de la cromatina**. Por ejemplo, la alteración de algunos TAD únicamente tiene consecuencias medibles en la expresión génica, pero seguimos sin entender las características genéticas o epigenéticas que distinguen los TAD *funcionales* de los *no funcionales*. Del mismo modo, desconocemos si todos los TAD y los límites de estos son equivalentes desde un punto de vista mecanicista y funcional. Se requiere un gran esfuerzo para reunir suficiente información sobre una amplia variedad de tipos celulares y durante el desarrollo y la diferenciación con el fin de poder construir modelos de plegado tridimensional de la cromatina y predecir los resultados funcionales y las consecuencias patológicas de las variantes estructurales. Al mismo tiempo, hay que generar instrumentos computacionales para que el análisis del plegado tridimensional del genoma sea asequible para los no especialistas, en particular para los del ámbito biomédico.

## CAPÍTULO 3 BIBLIOGRAFÍA

- Akdemir, K.C., Le, V.T., Killcoyne, S., King, D.A., Li, Y.-P., Tian, Y., Inoue, A., Amin, S., Robinson, F.S., Herrera, R.E. *et al.* (2018). Process-specific somatic mutation distributions vary with three-dimensional genome structure. *bioRxiv* 426080. doi: <https://doi.org/10.1101/426080>
- Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S. y Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* 8, e41769.
- Bai, X.C., McMullan, G. y Scheres, S.H. (2015). How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sc.* 40, 49-57.
- Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R. y Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* 27, 1139-1152.
- Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R. *et al.* (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519-524.
- Beliveau, B.J., Boettiger, A.N., Avendano, M.S., Jungmann, R., McCole, R.B., Joyce, E.F., Kim-Kiselak, C., Bantignies, F., Fonseca, C.Y., Erceg, J. *et al.* (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* 6, 7147.
- Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P. y Bickmore, W.A. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* 76(3), 473-484.e477.
- Bintu, B., Mateo, L.J., Su, J.H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N. y Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783.
- Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T. y Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418-422.
- Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M. *et al.* (2018). Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842-1855.e1816.
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A. *et al.* (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell* 171, 557572.e524.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. *et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479-1491.
- Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J. y Meyer, B.J. (2015). Condensin-driven remodelling of x chromosome topology during dosage compensation. *Nature* 523, 240-244.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931-21936.
- Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H. *et al.* (2017). PRC2 facilitates the regulatory topology required for poised enhancer function during pluripotent stem cell differentiation. *Cell Stem Cell* 20, 689-705.e689.
- Dame, R.T., Rashid, F.M. y Grainger, D.C. (2020). Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* 21, 227-242.
- Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G. y Peters, J.M. (2019). DNA loop extrusion by human cohesin. *Science* 366, 1338-1345.



- Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A. *et al.* (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 158, 849-860.
- Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., A.Y., L., Ye, Z., Kim, A., Rajagopal, N., Xie, W. *et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331-336.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. y Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.
- Dong, P., Tu, X., Chu, P.Y., Lu, P., Zhu, N., Grierson, D., Du, B., Li, P. y Zhong, S. (2017). 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant*, 1497-1509.
- Downen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K. *et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374-387.
- El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., K.R., Park, S., Wang, X. *et al.* (2019). A pliable Mediator acts as a functional rather than an architectural bridge between promoters and enhancers. *Cell* 178, 1145-1158.e1120.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. y Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038-2049.
- Fukaya, T., Lim, B. y Levine, M. (2016). Enhancer control of transcriptional bursting. *Cell* 166, 358-368.
- Gaiti, F., Calcino, A.D., Tanurdzic, M. y Degnan, B.M. (2017a). Origin and evolution of the metazoan noncoding regulatory genome. *Dev. Biol.* 427, 193-202.
- Gaiti, F., Jindrich, K., Fernandez-Valverde, S.L., Roper, K.E., Degnan, B.M. y Tanurdzic, M. (2017b). Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *Elife* 6, e22194.
- Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S. *et al.* (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* 176, 377-390.e319.
- Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korbel, J.O. y Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* 51, 1272-1282.
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. y Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96-100.
- Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E. *et al.* (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. USA* 112, 7542-7547.
- Grau-Bove, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., Richards, T.A. y Ruiz-Trillo, I. (2017). Dynamics of genomic innovation in the unicellular ancestry of animals. *Elife* 6, e26036.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900-910.
- Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merckenschlager, M. y Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* 8, 441.
- Heger, P., George, R. y Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. *Evolution* 67, 2945-2956.
- Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. y Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* 109, 17507-17512.
- Heger, P., Marin, B. y Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. *BMC Mol. Biol.* 10, 84.



- Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. y Sharp, P.A. (2017). A phase separation model for transcriptional control. *Cell* 169, 13-23.
- Hsieh, T.H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N. y Rando, O.J. (2015). Mapping nucleosome resolution chromosome folding in Yeast by Micro-C. *Cell* 162, 108-119.
- Hug, C.B., Grimaldi, A.G., Kruse, K. y Vaquerizas, J.M. (2017). Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* 169, 216-228.e219.
- Hug, C.B. y Vaquerizas, J.M. (2018). The birth of the 3D genome during early embryonic development. *Trends Genet.* 34, 903-914.
- Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. y Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46, 205-212.
- Irimia, M. y Maeso, I. (2019). Boosting Macroevolution: Genomic Changes Triggering Qualitative Expansions of Regulatory Potential. En *Old Questions and Young Approaches to Animal Evolution, Fascinating Life Sciences*, M.-D. J. y V. B. (eds.). (Springer, Cham).
- Irimia, M., Maeso, I., Roy, S.W. y Fraser, H.B. (2013). Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet.* 29, 521-528.
- Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L. y Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* 22, 2356-2367.
- Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E. y Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell.* 26, 565-577.
- Jabbari, K., Heger, P., Sharma, R. y Wiehe, T. (2018). The Diverging Routes of BORIS and CTCF: An Interatomic and Phylogenomic Analysis. *Life (Basel)* 8(1), 4.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S. *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430-435.
- Kempfer, R. y Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207226.
- Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. y Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. *Science* 366, 1345-1349.
- Krijger, P.H. y de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. *Nat. Rev. Mol. Cell Biol.* 17, 771-782.
- Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S. y Narlikar, G.J. (2017). Liquid droplet formation by HP1a suggests a role for phase separation in heterochromatin. *Nature* 547, 236-240.
- Laugsch, M., Bartusel, M., Rehim, R., Alirzayeva, H., Karaolidou, A., Crispatzu, G., Zentis, P., Nikolic, M., Bleckwehl, T., Kolovos, P. *et al.* (2019). Modeling the pathological long-range regulatory effects of human structural variation with patient-specific hiPSCs. *Cell Stem Cell* 24, 736-752.e712.
- Le, T.B., Imakaev, M.V., Mirny, L.A. y Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 342, 731-734.
- Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D. *et al.* (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature* 578, 472-476.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.
- Lu, Z., Marand, A.P., Ricci, W.A., Ethridge, C.L., Zhang, X. y Schmitz, R.J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* 5, 1250-1259.

- Lupiañez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R. *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012-1025.
- Marchal, C., Sima, J. y Gilbert, D.M. (2019). Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* 20, 721-737.
- Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A. y Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49-54.
- Matharu, N. y Ahituv, N. (2015). Minor loops in major Ffolds: enhancer-promoter Looping, chromatin restructuring, and their association with transcriptional regulation and disease. *PLoS Genet* 11, e1005640.
- Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J. *et al.* (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432-435.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J. y Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 13, 919-922.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A. y Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59-64.
- Ngan, C.Y., Wong, C.H., Tjong, H., Wang, W., Goldfeder, R.L., Choi, C., He, H., Gong, L., Lin, J., Urban, B. *et al.* (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.* 52, 264-272.
- Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. y Bruneau, B.G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930-944.e922.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J. *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381-385.
- Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H. y O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357(6349), eaag0025.
- Pang, B. y Snyder, M.P. (2020). Systematic identification of silencers in human cells. *Nat. Genet.* 52, 254-263.
- Pauli, T., Vedder, L., Dowling, D., Petersen, M., Meusemann, K., Donath, A., Peters, R.S., Podsiadlowski, L., Mayer, C., Liu, S. *et al.* (2016). Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects: Insect insulator proteins. *BMC Genomics* 17, 861.
- Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y. *et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281-1295.
- Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K. y Kingston, R.E. (2019). Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes. Dev.* 33, 799-813.
- Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y. *et al.* (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* 174, 744-757.e724.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A. y Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279-283.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.

- Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D. *et al.* (2017). Cohesin loss eliminates all loop domains. *Cell* 171, 305-320.e324.
- Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M. y Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* 160, 1145.
- Ricci, W.A., Lu, Z., Ji, L., Marand, A.P., Ethridge, C.L., Murphy, N.G., Noshay, J.M., Galli, M., Mejía-Guerra, M.K., Colomé-Tatché, M. *et al.* (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* 5, 1237-1249.
- Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y. y Corces, V.G. (2017). Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* 67, 837852 e837.
- Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C. *et al.* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361(6400), eaar3958.
- Schoenfelder, S. y Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437-455.
- Schuster-Böckler, B. y Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488, 504-507.
- Schwaiger, M., Schonauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D. y Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* 24, 639-650.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L. *et al.* (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51-56.
- Sebe-Pedros, A., Ballare, C., Parra-Acero, H., Chiva, C., Tena, J.J., Sabido, E., Gómez-Skarmeta, J.L., Di Croce, L. y Ruiz-Trillo, I. (2016). The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* 165, 1224-1237.
- Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K.N., Holcomb, N.P., Turner, J.L., Paulsen, M.T., Rivera-Mulia, J.C., Trevilla-Garcia, C. *et al.* (2019). Identifying cis elements for spatiotemporal control of mammalian DNA replication. *Cell* 176, 816-830.e818.
- Smol, T., Sigé, J., Thuillier, C., Frénois, F., Brunelle, P., Rama, M., Roche-Lestienne, C., Manouvrier-Hanu, S., Petit, F. y Ghomid, J. (2020). Lessons from the analysis of TAD boundary deletions in normal population. *bioRxiv*. doi: 10.1101/2020.04.01.021188
- Spielmann, M., Lupiáñez, D.G. y Mundlos, S. (2018). Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453-467.
- Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X. y Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. *Nature* 547, 241-245.
- Szabo, Q., Bantignies, F. y Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5(4), eaaw1668. doi: 10.1126/sciadv.aaw1668.
- Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Włodarczyk, J., Ruszczycski, B. *et al.* (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 163, 1611-1627.
- Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H.N., Huynh, T.N., Zhen, C.Y., Ma, B., Wang, H. y Ren, X. (2019). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J. Biol. Chem.* 294, 1451-1463.
- van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., ValdesQuezada, C., Oz, Y., Bouwman, B.A. *et al.* (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* 9, 969-972.
- van Steensel, B. y Belmont, A.S. (2017). Lamina-Associated Domains: Links with chromosome architecture, heterochromatin, and gene repression. *Cell* 169, 780-791.

- van Steensel, B. y Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* 20, 327-337.
- Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El Khattabi, L., Dose, M., Pruett, N. *et al.* (2018). The energetics and physiological impact of cohesin extrusion. *Cell* 175, 292-294.
- Wang, M., Wang, P., Lin, M., Ye, Z., Li, G., Tu, L., Shen, C., Li, J., Yang, Q. y Zhang, X. (2018a). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* 4, 90-97.
- Wang, Q., Sun, Q., Czajkowsky, D.M. y Shao, Z. (2018b). Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* 9, 188.
- Wang, X., Brandao, H.B., Le, T.B., Laub, M.T. y Rudner, D.Z. (2017). *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* 355, 524-527.
- Williamson, I., Kane, L., Devenney, P.S., Flyamer, I.M., Anderson, E., Kilanowski, F., Hill, R.E., Bickmore, W.A. y Lettice, L.A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development* 146, pii: dev179523.
- Winter, D.J., Ganley, A.R.D., Young, C.A., Liachko, I., Schardl, C.L., Dupont, P.Y., Berry, D., Ram, A., Scott, B. y Cox, M.P. (2018). Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLoS Genet.* 14, e100746.
- Yan, J., Chen, S.A., Local, A., Liu, T., Qiu, Y., Dorigi, K.M., Preissl, S., Rivera, C.M., Wang, C., Ye, Z. *et al.* (2018). Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res.* 28, 204-220.
- Zhang, H., Emerson, D.J., Gilgenast, T.G., Titus, K.R., Lan, Y., Huang, P., Zhang, D., Wang, H., Keller, C.A., Giardine, B. *et al.* (2019). Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* 576, 158-162.

## CAPÍTULO 4

### RESUMEN

El genoma no codificante contribuye a establecer una sofisticada red de regulación que resulta fundamental para la función celular, con un impacto directo en el crecimiento, desarrollo, evolución y salud de los organismos. Actualmente se desconocen la mayoría de sus características estructurales, los factores de interacción y los mecanismos de acción. Descifrar estos aspectos junto con las consecuencias que su función pudiera tener a múltiples niveles son desafíos importantes para futuras investigaciones.

### PALABRAS CLAVE

genoma no codificante	repeticiones
secuencias reguladoras	ARN no codificante

# EL GENOMA NO CODIFICANTE

## Coordinadores

Crisanto Gutiérrez  
(CBMSO, Madrid, coordinador)

Cristina Hernández Munain  
(IPBLN, Granada, coordinadora adjunta)

## Investigadores y centros de investigación participantes (por orden alfabético)

Carlos Estella  
(CBMSO, Madrid)

Sònia García  
(IBB, Barcelona)

José Carlos Reyes  
(CABIMER, Sevilla)

Esther Serrano  
(CBMSO, Madrid)

## RESUMEN EJECUTIVO

El estudio de la función de cada uno de los genes que codifican proteínas en el genoma ha concentrado muchos esfuerzos para comprender el funcionamiento de las células y los organismos y lo seguirá haciendo en el futuro. Ahora bien, el genoma codificante constituye una fracción muy pequeña del genoma completo, en particular del genoma de los organismos multicelulares. La presencia de cantidades muy grandes de ADN genómico no codificante (GENOMAnc) es una característica del genoma de una gran diversidad de especies, incluidos todos los linajes evolutivos analizados, que abarca animales, plantas y levaduras, y que también está presente en procariotas e incluso en virus con grandes genomas. La idea hasta ahora generalmente aceptada de que esta enorme fracción de los genomas eucariotas está formada por ADN no codificante ha cambiado de manera significativa en los últimos años, lo que ha transformado el concepto de genoma de *materia oscura* en una de las estrellas de la biología moderna. Este cambio conceptual se basa en gran medida en la enorme diversidad de elementos y productos derivados del GENOMAnc, y en la identificación de su papel decisivo tanto en condiciones normales como con enfermedades, como se ha informado en algunos ejemplos concretos. En conjunto, los diferentes elementos y productos del GENOMAnc contribuyen a establecer una sofisticada red reguladora que resulta esencial para el funcionamiento normal del genoma codificante. La mayoría de las características estructurales, los factores de interacción, los mecanismos de acción, las

dianas y las interacciones de bucle del GENOMAnc se consideran prácticamente desconocidos. Descifrar estos diversos aspectos del GENOMAnc, así como identificar las consecuencias que su función pudiera tener a múltiples niveles se consideran desafíos importantes para las futuras investigaciones. Una lista no exhaustiva de las posibles funciones del GENOMAnc incluye la transcripción, el empalme y el procesamiento del ARN mensajero, el transporte, la traslación y la descomposición, la interferencia del ARN, el sellado, las modificaciones epigenéticas, la remodelación de la cromatina, el ensamblaje de los orgánulos subnucleares mediante procesos de separación de fases líquido-líquido y la organización nuclear tridimensional. Un aspecto clave de la función de varios componentes del GENOMAnc estriba en que sus ARN no codificantes derivados se asocian con una gran cantidad de proteínas, otras moléculas de ARN y el ADN para lograr su función. En este sentido, el estudio de las características estructurales tanto de la fracción de ARN como de las proteínas con las que interactúa resulta de una importancia capital para comprender la función del GENOMAnc.

En resumen, el GENOMAnc es mucho mayor que el genoma codificante y sabemos muy poco sobre él. ¿Por qué es crucial abordar el enorme desafío de comprender el papel que tiene en los próximos años? La respuesta a esta pregunta resulta bastante sencilla: el GENOMAnc tiene un impacto directo en el crecimiento, el desarrollo y la evolución de todos los organismos multicelulares, incluidos los animales y las plantas, así como en su salud. Basándonos en el estado actual de la técnica en este campo, resumimos una serie de cuestiones que surgen y que creemos que debieran constituir el centro de atención en el futuro.

## 4.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La presencia de grandes cantidades de ADN genómico no codificante (GENOMAnc) es una característica del genoma que está presente en todos los organismos analizados. Dentro de este término incluimos el ADN no codificante y los ARN derivados del mismo (ARNnc). La importancia de estos transcritos se puso de manifiesto en 2012 con el proyecto ENCODE, que estableció que el 75 % del genoma humano se transcribe, pero solo el 2 % de los transcritos se traducen en proteínas (Consortio ENCODE, 2012). La observación de estos abundantes transcritos se ha confirmado en muchos organismos diferentes.

El GENOMAnc se compone de diversos elementos que pueden clasificarse en tres grandes grupos: secuencias que dan lugar a transcritos con función de mantenimiento (GENOMAmán), secuencias repetidas (GENOMArepet) y regiones que contienen secuencias reguladoras o conducen a la génesis de transcritos reguladores (GENOMArepet).

La transcripción del **GENOMAmán** da lugar a la generación de ARN de transferencia (ARNt), ARN ribosómico (ARNr) y ARN pequeño nuclear (ARNpn) con funciones constitutivas.

El **GENOMArepet** incluye no solo las regiones teloméricas y centroméricas sino también todas las diferentes clases de ADN repetitivo que en conjunto constituyen una gran fracción del GENOMAnc. Las secuencias de GENOMArepet pueden transcribirse, como es el caso de los ARN que contienen repeticiones teloméricas (TERRA), implicados en la protección de los telómeros. También incluye repeticiones que pueden estar dispuestas en tándem o dispersas, cuyos principales constituyentes son los elementos transponibles (ET), conocidos no solo por dar lugar a reordenamientos cromosómicos sino también por modular la expresión de genes cercanos y de ellos mismos mediante la generación de ARN asociados a Piwi (ARNpi).

El **GENOMArepet** está formado por un grupo muy diverso de secuencias genómicas. Entre ellos se encuentran los elementos de acción en *cis* que conforman secuencias proximales y distales que regulan la transcripción, como los promotores, los silenciadores y los potenciadores que contienen motivos para mediar la unión de factores celulares. El GENOMArepet también incluye las regiones intrónicas e intergénicas que conducen a la generación de ARNnc de diversos tamaños y configuraciones, como los ARNnc largos (ARNncl), los micro-ARN (miARN) y los ARN potenciadores (ARNp), que actúan en *cis* o *trans*



con diversas funciones reguladoras. Por último, existen transcritos lineales sentido y antisentido y ARNnc circulares (ARNcirc) derivados de regiones exónicas e intrónicas que también desempeñan un papel regulador. Dado que los ARNcirc son, al menos en parte, una consecuencia de la transcripción de exones, los hemos clasificado como transcritos derivados del genoma codificante, en lugar del GENOMAnc.

Los elementos que componen el GENOMAmán, el GENOMAreple y el GENOMAreple están regulados entre sí y también se entrecruzan funcional y estructuralmente con el genoma codificador y las proteínas de unión al ARN/ADN para regular el desarrollo, el crecimiento, la enfermedad y la evolución.

Durante el desarrollo embrionario, numerosos procesos deben estar debidamente organizados en el espacio y el tiempo mediante la regulación de la transcripción de genes. Para lograrlo, es preciso integrar las aportaciones reguladoras de muchos elementos del GENOMAnc, un lenguaje cifrado de enorme complejidad. Es fundamental comprender cómo el GENOMAnc puede afectar al crecimiento y al bienestar y, por tanto, a la productividad de las plantas y los animales utilizados por el ser humano, al tiempo que se mantiene una productividad sostenible. Las mutaciones o alteraciones epigenéticas que afectan al GENOMAnc tienen un gran impacto en el crecimiento de los organismos, incluidas las enfermedades. Los estudios de asociación del genoma completo (GWAS) han puesto de manifiesto que la mayor parte (70-90 %) de la variación genética asociada a las enfermedades (polimorfismos de nucleótido simple o SNP, por sus siglas en inglés), y también en animales y plantas, se encuentra fuera de los cuerpos de los genes, en el GENOMAnc. En cambio, la gran mayoría de estos SNP no están caracterizados de manera funcional. Responder a estas y otras preguntas relacionadas representa un gran desafío con importantes implicaciones científicas, sociales y económicas, especialmente si se tiene en cuenta el constante aumento de la población humana que necesita alimentarse y mantener una buena salud.

## 4.2 IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS POSIBLES APLICACIONES

### 4.2.1 El GENOMAmán

El GENOMAmán genera transcritos con funciones constitutivas que desempeñan funciones esenciales en una gran variedad de procesos biológicos y causan múltiples patologías.

Los **ARNt** son componentes básicos de la maquinaria de traducción que actúan como descodificadores durante la traducción de los ARNm en proteínas (Tuorto y Parlato, 2019). Son relativamente pequeños en comparación con otras especies de ARN (70-90 nt), con una estructura tridimensional (3D) en forma de hoja de trébol, una innovación evolutiva que permitió la estandarización del código genético de tres nucleótidos. Los ARNt también desempeñan papeles importantes en otras funciones celulares, por ejemplo, la modulación de la proliferación celular y la respuesta al estrés (Thompson *et al.*, 2008), así como en la regulación de la expresión génica (Raina e Ibba, 2014). En particular, la actividad de fragmentos específicos de ARNt (ARNt escindido) es responsable de funciones aún desconocidas en su mayoría. Estos fragmentos, que antes se consideraban intermediarios de degradación no funcionales, se reconocen ahora como especies principales de ARN cuyas funciones reguladoras se están empezando a comprender, incluidos la respuesta al estrés, la apoptosis y el cáncer (Hanada *et al.*, 2013). Otros ARNt participan en otros procesos bioquímicos, como la formación de la pared celular, el etiquetado de proteínas para su degradación y la biosíntesis de antibióticos (Kanai, 2014).

Los **ARNr** y los genes de ARNr (ADNr) son parte integral de los ribosomas con funciones estructurales y catalíticas en la síntesis de proteínas. La mayor parte del ARN de una célula (aproximadamente el 80 %) es ARNr, y hay varias moléculas de ARNr diferentes: tres en procariotas (5S, 16S y 23S) y cuatro en eucariotas (5S, 5,8S, 18S y 28/26S). El ARNr 5S eucariótico es el más enigmático y sus funciones aún no están resueltas; curiosamente, mientras que el 5,8S eucariótico se corresponde con el 5S procariota, ninguna molécula procariota se corresponde con el 5S en eucariotas, y además se produce en una localización diferente (el nucléolo) (Ha y Bhagavan, 2011). Al igual que el ARNt, el ARNr también media en las condiciones de estrés celular, por lo que la síntesis de ambas especies de ARN es interdependiente.

#### ***ARNnp: ARNpno, ARNcca y ARNnp-U***

Los ARNnp son transcritos cortos (60-300 nt de longitud) presentes en todos los organismos eucariotas y su número aumenta con la complejidad del organismo. Además, la organización genómica de los genes que codifican ARNnc sigue un árbol evolutivo. En *Sacharomyces cerevisiae*, se organizan en genes independientes, mientras que en los humanos se localizan predominantemente dentro de los intrones, donde se liberan mediante empalme. Los ARNnp tienen importantes funciones estructurales y funcionales, por lo que son

decisivos en la formación de los orgánulos subnucleares conocidos como *cuerpos nucleares*. Estos compartimentos sin membrana están formados por altas concentraciones locales de moléculas que promueven la formación de enlaces débiles no covalentes, muy probablemente impulsados por la separación de fases líquido-líquido (Khosraviani *et al.*, 2019). Entre estas estructuras dinámicas se encuentran los nucléolos, las motas, los *paraspeckles*, los cuerpos de leucemia promielocítica, los cuerpos de Cajal (CB, por sus siglas en inglés), los cuerpos de locus de histonas (HLB, por sus siglas en inglés) y los cuerpos de Polycomb. Los cuerpos nucleares ocupan el espacio de la intercromatina y se encuentran muy enriquecidos con factores nucleares específicos y, en muchos casos, con ARNnc estructurales. También se han implicado otros ARNnc de manera estructural en su formación, por ejemplo: los ARNr en la formación del nucléolo y los ARNncl estructurales en la formación de los *paraspeckles*. Los ARNpno y los ARNcca se ensamblan con las proteínas en las ribonucleoproteínas nucleares pequeñas (RNPnp), responsables de las modificaciones postranscripcionales de los ARN, como la pseudouridilación y la 2'-O-metilación de otros ARN. Las RNP formadas con ARNpno (RNPpno) son responsables de la modificación de los ARNr en los nucléolos, mientras que las formadas con ARNcca (RNPcca) median la modificación de los ARNnp-U en los *speckles* y los CB. Los ARNnp-U también se ensamblan en RNP (RNPnp-U) que funcionan en el procesamiento de pre-ARNm. Datos recientes indican que los ARNpno y los ARNcca tienen especificidades y localización superpuestas en el núcleo, lo que sugiere que los nucléolos y los CB pueden tener funciones intercambiables (Deryusheva y Gall, 2019).

Además de su papel en la modificación y maduración del ARN, los cuerpos nucleares participan en el guiado del plegado de los cromosomas con el fin de proporcionar una plataforma para la organización espacial de los *loci* genómicos, lo que afecta a su expresión. De esta manera, los genes transcritos de manera activa se asocian a la periferia de los *speckles*, los HLB participan en la organización tridimensional de los grupos de genes de histonas y los dominios de heterocromatina se localizan principalmente cerca de la periferia nuclear o del nucléolo (Khosraviani *et al.*, 2019). Una hipótesis provocadora sostiene que la organización agrupada de los genes dentro de los cromosomas y la capacidad de los ARNnc para agregarse en los cuerpos nucleares han evolucionado conjuntamente para facilitar el ensamblaje impulsado por el ARN de los núcleos nucleares (Smith *et al.*, 2020). Por lo tanto, los ARNnp tienen el potencial de modular las estructuras subnucleares con un papel inexplorado en la conformación del genoma para así promover una expresión génica adecuada.

#### 4.2.2 El GENOMAre<sub>p</sub>

Los genomas eucariotas están compuestos en gran parte por secuencias de ADN repetitivas que constituyen el GENOMAre<sub>p</sub>. Estas secuencias pueden clasificarse como repeticiones en tándem o intercaladas. Las repeticiones en tándem incluyen los centrómeros y los telómeros (compuestos por ADN satélite), mientras que las repeticiones intercaladas incluyen los ET. Las repeticiones son extremadamente variables tanto en su abundancia como en la longitud de la secuencia, presentándose hasta en millones de copias por genoma, con tamaños que van desde unos pocos pares de bases hasta muchos miles (Heslop-Harrison y Schwarzscher, 2011). Estas regiones generan transcritos con funciones importantes, como los generados a partir de los telómeros, conocidos como TERRA, que son esenciales para la protección de los telómeros, y los ARNpi, generados a partir de los ET, que participan en la regulación de su propia transcripción.

##### *Telómeros y TERRA*

Los telómeros son complejos nucleoproteicos situados en el extremo de los cromosomas que los protegen de los daños durante el proceso de replicación y mantienen la homeostasis y el envejecimiento de los cromosomas. Su parte repetitiva consta de un minisatélite de secuencia de 6-8 nt que está muy conservado en grandes grupos de organismos. La longitud de los telómeros es específica de cada especie, pero puede modularse mediante un equilibrio entre las señales de acortamiento y elongación durante la vida de un organismo. La telomerasa, la enzima responsable de alargar los telómeros, y los TERRA, que regulan la actividad de la telomerasa y mantienen el estado heterocromático en los extremos de los cromosomas, junto con los telómeros conforman la tríada interdependiente del «telomere world» (el mundo de los telómeros) (Mensà *et al.*, 2019).

##### *Centrómeros*

Los centrómeros y las regiones pericentroméricas, que son específicos de cada especie, están compuestos por repeticiones satélite. Los centrómeros organizan los movimientos de los cromosomas desde la profase hasta la anafase, mediante la interacción con los microtúbulos del aparato del huso y la promoción de la segregación fiel de los cromosomas. La composición de las regiones centroméricas se conoce para muy pocos taxones, en su mayoría organismos modelo (Hirsch y Jiang, 2013). Los centrómeros suelen ir acompañados de ET en las regiones pericentroméricas y la interacción entre ambos componentes genómicos apenas empieza a comprenderse. La combinación

de la secuenciación de última generación (NGS, por sus siglas en inglés) con la inmunoprecipitación de la cromatina (ChIP-seq) permite analizar la composición de los centrómero en la mayoría de las especies.

### ***Elementos transponibles (ET) y ARNpi***

La fracción más importante del ADN repetitivo está compuesta por repeticiones dispersas, en su mayoría elementos transponibles (ET). Anteriormente conocidos como *genes saltarines*, los ET pueden cambiar el paisaje del ADN y ser fuente de innovación genética al alterar la expresión génica o promover reordenamientos cromosómicos (Pantartzzi *et al.*, 2018). Los ET ocupan la mayor parte de los genomas eucariotas, por ejemplo, casi el 50 % del genoma humano o hasta el 85 % del maíz. La mayoría de los ET en el genoma han perdido su movilidad, siendo los restos de elementos móviles antes activos —hace más de 200 millones de años— y las mutaciones aleatorias los que los han hecho irreconocibles. Las características intrínsecas de los ET, con sus rasgos de secuencia repetitiva pero también no conservada, los han convertido en un objeto de estudio difícil según los planteamientos tradicionales. De nuevo, la llegada de la NGS ha hecho posible su análisis. Existen dos grandes categorías, los ADN-ET, que utilizan un mecanismo de transposición de cortar y pegar, y los retroelementos, los más abundantes en la mayoría de los genomas eucariotas que se mueven y amplifican a través de una (retro)transposición mediada por ARN. En el genoma humano, las secuencias Alu, un tipo de elementos nucleares intercalados cortos (SINE, por sus siglas en inglés), clasificados como retroelementos, son las más abundantes. Otra fracción importante del genoma humano (en torno al 8 %) está formada por retrovirus endógenos (ERV, por sus siglas en inglés), un tipo de retroelementos que pueden tener su origen en antiguas inserciones de retrovirus. Aunque históricamente se han considerado *elementos genéticos egoístas*, los ET contribuyen a una amplia gama de funciones reguladoras. Además, los últimos descubrimientos respaldan un papel relevante de los ET en procesos relacionados con la especiación pero también en la salud y la enfermedad (Reilly *et al.*, 2013; Serrato-Capuchina y Matute, 2018).

El **ARNpi** es una nueva clase de ARN no codificante (perteneciente al grupo de los ARN de interferencia) del que hace poco se ha descubierto que desempeña una función en el silenciamiento epigenético y postranscripcional de los elementos transponibles, impidiendo por lo general su expansión. Sin embargo, la amplia variación de las secuencias de ARNpi entre las especies dificulta el establecimiento de su funcionalidad.

### 4.2.3 El GENOMAre<sub>g</sub>

El GENOMAre<sub>g</sub> incluye no solo los elementos reguladores que actúan en *cis* y los límites de los dominios asociados topológicamente (TAD) o aisladores en la cromatina sino también ARNnc de diversos tamaños y configuraciones que se originan en regiones intrónicas y exónicas. La importancia de los TAD y de sus límites en la organización tridimensional del genoma y en los controles de la expresión génica se ha descrito a fondo en el capítulo 3.3, por lo que aquí nos centraremos en los elementos reguladores en *cis* y en los ARNnc.

#### *Regiones reguladoras que actúan en cis*

Los potenciadores y promotores son secuencias que tienen la capacidad de activar la transcripción de los genes en *cis*. Mientras que los promotores actúan a una corta distancia del sitio de inicio transcripcional, los potenciadores y silenciadores lo hacen a largas distancias, desde unos cientos de pares de bases hasta varios cientos de kilobases, de manera independiente a su orientación (Parker *et al.*, 2013; Hnisz *et al.*, 2013). La organización de la cromatina en diferentes TAD y sub-TAD con paisajes reguladores específicos depende de las interacciones entre estas regiones reguladoras en *cis* por medio del reclutamiento de factores de transcripción (FT) (Franke y Gómez-Skarmeta, 2018; Schoenfelder y Fraser, 2019).

Los potenciadores han merecido una gran atención debido al papel esencial que desempeñan en el control de los programas de desarrollo. Los potenciadores contienen numerosos sitios de unión al ADN para FT, específicos para tipos de células o etapas de desarrollo diferentes (Heinz *et al.*, 2015). La unión del FT sirve para reclutar complejos coactivadores, como el complejo mediador o las enzimas modificadoras de las histonas, que en conjunto determinan la actividad del potenciador. En algunos casos, densos grupos de potenciadores se unen en estrecha proximidad tridimensional y colaboran para actuar como una sola unidad reguladora (superpotenciadores) que puede impulsar altos niveles de transcripción (Sabari *et al.*, 2018; Nair *et al.*, 2019).

Los ARNe corresponden a un sorprendente descubrimiento realizado en 2010, que relacionó la presencia de estos transcritos derivados de potenciadores con la de los ARNm de los genes diana (De Santa *et al.*, 2010; Kim *et al.*, 2010). Los ARNe son diversos, con estructura, longitud y modificaciones postranscripcionales heterogéneas, y presentan especificidad tisular y de linaje. Aunque en un principio se consideraron como un posible subproducto de la transcripción, cada vez hay más pruebas de que constituyen biomoléculas funcionales

que promueven la transcripción de los genes mediante la formación de bucles entre los potenciadores y los promotores y la modificación de la cromatina (Arnold *et al.*, 2020). Ahora bien, estos transcritos suelen ser muy inestables y carecen de una buena conservación de la secuencia, lo que sugiere que algunos ARNe pudieran tener funciones más bien sutiles o incluso representar subproductos de la transcripción. La transcripción de los superpotenciadores da lugar a grandes cantidades de ARNe que interactúan con complejos, como el mediador, la cohesina, p300/CBP y BRD4, para regular la transcripción tanto de forma local como a distancia.

### **ARNnc**

Los ARNnc, transcritos intergénicos e intrónicos/exónicos que no se traducen, han surgido como actores principales del GENOMareg durante la evolución, el desarrollo y la enfermedad. Aunque el término ARNnc incluye también los transcritos generados a partir del GENOMaman y el GENOMarep, la descripción en esta sección se limita a los ARNnc derivados del GENOMareg. Estos ARNnc están implicados en numerosos procesos celulares, como la transcripción, el empalme y la traducción de proteínas. Estos transcritos pueden regular la expresión génica en *cis* (el locus genético desde el que se ha generado) mediante la unión al ADN, o en *trans* (en otro lugar del genoma), como componentes estructurales de los orgánulos nucleares o funcionando como señuelos moleculares para valorar proteínas y otros ARN. Se pueden clasificar en función de su tamaño en ARNnc largos (ARNncl) con más de 200 pb y cortos (ARNnce) con menos de 200 pb y, en función de su configuración, en ARN lineales o circulares.

Los ARNnc incluyen miARN con aproximadamente 22 nucleótidos. Las vías de biogénesis y maduración de los miARN, así como su modo de acción se han definido adecuadamente (Pu *et al.*, 2019). Estos transcritos cortos actúan como reguladores postranscripcionales de la expresión génica a través de la unión directa a los ARNm, lo que regula diversos procesos biológicos en todos los tejidos y estadios celulares.

Los ARNcirc están formados por bucles cerrados de manera covalente a través del retroempalme y pueden ser exónicos, intrónicos y exón-intrónicos. Los ARNcirc derivan del genoma codificante, pero se consideran ARNnc porque no se traducen, aunque algunos pueden generar pequeños péptidos (Pamadurti *et al.*, 2017). Al igual que los ARNm, los ARNcirc son transcritos abundantes y bien conservados que se expresan de forma regulada por el tejido, el



estadio celular y el tiempo, y están alterados en varias enfermedades (Wang *et al.*, 2017). Los ARNcirc participan en varias actividades biológicas esponjando a los miARN y a las proteínas de unión al ARN. Lo curioso es que son más estables que los ARN lineales, por lo que tienen una vida media cinco veces mayor que la de los ARNm (Jeyaraman *et al.*, 2020). Según estas características, se han propuesto como nuevos e interesantes biomarcadores y agentes terapéuticos para el cáncer.

Se sabe poco sobre la biología de los ARNncl, que muestran un alto nivel de diversidad y se originan en una parte importante del GENOMAnc. A diferencia de los 2000 miARN diferentes identificados en humanos, se han detectado aproximadamente 50 000 ARNncl distintos, que representan la mayor parte del transcriptoma en animales y plantas, y se observan en casi todas las especies, incluidos levaduras, procariotas e incluso virus (Alessio *et al.*, 2020). Estos transcritos se transcriben mediante el ARN polimerasa II, por lo que normalmente se les añade una estructura llamada caperuzy (cap) y se poliadenuilan, y podrían o no procesarse con la maquinaria de empalme (Quinn y Chang, 2017). En función de sus modos de acción, pueden clasificarse en *cis* o *trans*. Un ejemplo de modo de acción en *cis* incluye cuando un ARNncl funciona como puente entre el ADN y la proteína, lo que sirve como andamio para llevar complejos modificadores de histonas a *loci* específicos, como es el caso de los ARNe. Un ejemplo de modo de acción en *trans* incluye un ARNncl que actúa como señuelo molecular para atraer proteínas o miARN (Grüll y Masé, 2019). Sin embargo, debido a su baja abundancia, existe un debate sobre si pueden agotar de forma eficaz los miARN (Ulitsky, 2018). De hecho, la existencia de una red de interacciones entre los ARNncl y los miARN resulta actualmente un tema importante de investigación.

El papel que desempeñan los ARNncl en muchos procesos biológicos, como la transcripción, el sellado, el empalme y la traslación, con consecuencias funcionales en el ciclo celular, la apoptosis, la pluripotencia y la reprogramación, la respuesta al choque térmico y las enfermedades, refleja la versatilidad del propio ARN, capaz de plegarse en una variedad de estructuras secundarias y de unirse a un gran número de sustratos. El interactoma de los ARNncl incluye proteínas, otros ARN y ADN. Los ARNncl muestran una conservación muy pobre en comparación con otros ARN (Necsulea y Kaessmann, 2014). Aunque no conservan su secuencia primaria, su función se mantiene en diferentes especies, lo que indica que su conservación se basa en rasgos estructurales más que en rasgos de secuencia. En este sentido, su estructura secundaria/terciaria desempeña



un papel fundamental en la función de los ARNncl. Curiosamente, estudios recientes indican que alrededor del 23 % del interactoma de ARNncl está compuesto por péptidos cortos que están codificados dentro de los ARNncl (Matsumoto y Nakayama, 2018). A pesar del entusiasmo actual por los ARNncl y sus posibles funciones biológicas, cabe mencionar que cuando se han evaluado sus funciones *in vivo* mediante la supresión individual de un grupo seleccionado de 24 de 727 de estos transcritos en el pez cebra, lo que corresponde al 3,3 % del total de ARNncl presentes en este organismo, no dio lugar a ningún defecto fenotípico evidente (Goudarzi *et al.*, 2019). Si bien este estudio sugiere que los ARNncl podrían no tener ningún papel funcional o tenerlo más bien sutil. En cambio, otros estudios han demostrado que la regulación errónea o la eliminación de ARNncl tienen consecuencias fenotípicas significativas, como COOLAIR en el tiempo de floración en las plantas (Csorba *et al.*, 2014), y los transcritos sentido/antisentido originados en los genes receptores de antígenos en el control de la recombinación V(D)J (Abarrategui y Krangel, 2006; Giallorakis *et al.*, 2010) o Xist en la regulación de la compensación de dosis de los cromosomas X entre machos y hembras (Loda y Heard, 2019) en animales. Por lo tanto, debieran realizarse más análisis sistemáticos de supresión, incluida la supresión simultánea de más de un transcrito para descartar efectos de compensación entre transcritos diferentes, con el fin de establecer claramente su función *in vivo* en este y otros organismos modelo.

#### 4.2.4 El GENOMAnc en desarrollo

Todos los tipos de células de un organismo multicelular comparten una secuencia de ADN casi idéntica, pero desempeñan funciones muy diferentes durante el desarrollo y la edad adulta. Esta gran diversidad de tipos celulares se consigue gracias a la regulación precisa de la expresión génica en el tiempo y el espacio por parte del GENOMAre a través de la compartimentación de la cromatina, los elementos reguladores en *cis* y los mecanismos postranscripcionales. Las mutaciones que afectan al GENOMAre son una fuente común de divergencia fenotípica y cambio evolutivo, además de una causa frecuente de enfermedades humanas.

#### *Cómo descifrar el código transcripcional espacial y temporal durante el desarrollo*

La accesibilidad del ADN en la cromatina es dinámica durante el desarrollo y diferente entre los tipos de células (Klemm *et al.*, 2019). Los potenciadores activos suelen estar desprovistos de nucleosomas, y los FT pioneros son importantes para la apertura de sitios de cromatina cerrados y la posterior unión de

FT adicionales (Iwafuchi-Doi *et al.*, 2016). La actividad potenciadora se correlaciona con ciertas propiedades de la cromatina, por ejemplo, la reducción de los nucleosomas y las modificaciones postraduccionales, principalmente H3K4me1 y H3K27ac (Yáñez-Cuna *et al.*, 2013; Shlyueva *et al.*, 2014). Además, un mismo gen del desarrollo emplea varios potenciadores o FT distintos para activar con precisión su expresión en diferentes células, en niveles variables y en múltiples momentos del desarrollo. Por otra parte, se ha demostrado que los ARNe pudieran regular el estado de la cromatina, la dinámica de unión de los FT o la estabilización del bucle de cromatina, lo que proporciona otra capa de regulación transcripcional (Lewis *et al.*, 2020). Desde la constatación de la función de los potenciadores en las primeras etapas del desarrollo (Martínez-Salas *et al.*, 1989), la comprensión de la biología de los potenciadores se ha convertido en un área de gran interés, ya que se aprecia cada vez más su importancia en el desarrollo, la evolución y las enfermedades (Corradin y Scacheri, 2014).

### ***Mecanismo postranscripcional en la regulación génica a través del desarrollo***

Los reguladores postranscripcionales, como los ARNnc1, miARN y ARNcirc, son actores clave en la regulación de la expresión génica durante el desarrollo animal y vegetal, incluidos casi todos los pasos del desarrollo, desde la regulación de la pluripotencia de las células madre embrionarias hasta el proceso de diferenciación, la determinación del sexo, la adquisición y el mantenimiento de la identidad celular, el control del tiempo de desarrollo y la herencia epigenética transgeneracional y el sellado genómico. Los ARNnc actúan en redes jerárquicas; algunos regulan los eventos iniciales del establecimiento de los patrones embrionarios, y su pérdida o mala regulación generan fenotipos letales o fuertes, mientras que otros realizan funciones más sutiles pero esenciales (Davidson *et al.*, 2003; Alberti y Cochella, 2017). Los ARNnc están representados en todas las posiciones de la jerarquía, pero sus bajos niveles de expresión y sus funciones redundantes han dificultado su estudio. El papel de los miARN como reguladores del tiempo de desarrollo en *C. elegans* y los mecanismos implicados en la pubescencia de los mamíferos muestran sorprendentes similitudes, lo que destapa la cuestión de la prevalencia de los elementos de los miARN en el control del tiempo de desarrollo. Tras el hallazgo de los ARNnc1 Xist y HOTAIR, se considera que los ARNnc1 funcionan de manera preferente en contextos celulares, tipos de células, etapas de desarrollo y enfermedades específicas. Su estudio ha aportado un conocimiento muy valioso de los mecanismos de acción de los ARNnc1 en cuanto al direccionamiento de los modificadores de la cromatina en *cis* y *trans*. No

obstante, aunque los estudios de los ARNnci están todavía en sus inicios, los ARNnci se han convertido en un paradigma para entender los mecanismos de largo alcance de la regulación génica y la compensación de dosis.

**Los ARNnc se han convertido en los candidatos modernos para conciliar la noción de herencia epigenética adaptativa** (Heard y Martienssen, 2014). La herencia epigenética transgeneracional, es decir, la transmisión de características epigenéticas de una generación a la siguiente a través de la línea germinal y con persistencia en las generaciones posteriores, está bien documentada en numerosos organismos, incluidos plantas, nematodos, moscas de la fruta y mamíferos (Tyebji *et al.*, 2020). Un obstáculo importante de la herencia epigenética transgeneracional tiene que ver con la reprogramación de la línea germinal, mediante la cual se restablece la metilación del ADN, las variantes de las histonas y sus modificaciones (Tang *et al.*, 2015). Las moléculas de ARN son excelentes candidatas para transportar la información epigenética entre generaciones debido a su especificidad y larga vida, menos afectada por la reprogramación, si bien aún no se han resuelto en gran parte los mecanismos. Las señales de ARN pequeño son muy móviles y median el silenciamiento transcripcional heredable a través de las generaciones, como se ha demostrado en la línea germinal de *C. elegans*. Asimismo, los ARN pequeños también pueden viajar a través de la vasculatura y los plasmodesmos en las plantas, así como de los exosomas e incluso el suero en los mamíferos (Chen *et al.*, 2016). Siguen descubriéndose componentes de este proceso, tanto básicos como específicos de cada especie.

La impronta genómica es un proceso complejo y muy regulado que consiste en el silenciamiento monoalélico de ciertos genes. Los ARNnci regulan la estructura de la cromatina y la expresión génica de los genes con impronta genómica a través de interacciones con las proteínas modificadoras de las histonas, la formación de bucles y la promoción de la compartimentación de la cromatina intracromosómica (Kanduri, 2016). El perfil de metilación del ADN reveló que, aunque el grueso del genoma (incluidos los *loci* con impronta genómica) se desmetila en las células germinales primordiales, una serie de *loci*, predominantemente asociados a secuencias repetitivas, escapan a la desmetilación. Los motivos aún no están claros, pero estos pudieran representar los principales candidatos para una posible herencia transgeneracional en los mamíferos. En el capítulo 3.7 se ofrecen más detalles sobre este interesante, aunque poco conocido, tipo de herencia y sus implicaciones para la salud y la enfermedad humanas.

#### 4.2.5 Relevancia patológica del GENOMAnc

El GENOMAnc tiene un profundo impacto en la enfermedad. Los GWAS han demostrado que la mayoría (70-90 %) de los SNP asociados a enfermedades se encuentran en el GENOMAnc (ENCODE Consortium, 2012; Maurano *et al.*, 2012). Mientras que algunos de estos SNP son determinantes para el desarrollo de la enfermedad, otros contribuyen de forma combinada. Además de las variaciones genéticas heredadas, las mutaciones somáticas en el GENOMAnc también son responsables de enfermedades, como el cáncer. Asimismo, se han empezado a analizar los estudios de asociación basados en rasgos epigenéticos (EWAS, del inglés *Epigenome-Wide Association Studies*), principalmente la metilación del ADN. También en este caso, e incluso más que en los estudios GWAS, la mayoría de las variaciones epigenéticas se concentran en regiones que residen en el GENOMAnc (Rakyan *et al.*, 2011). Para ilustrar cómo los diferentes elementos del GENOMAnc están implicados en la enfermedad, se incluyen algunos ejemplos.

##### **ARNr, ARNt y ARNpn**

La desregulación en la síntesis del ARNr puede dar lugar a trastornos, como el alzhéimer u otras enfermedades neurodegenerativas (Tuorto y Parlato, 2019). Y lo que es más, el cáncer, el envejecimiento prematuro y el deterioro neurológico en la ataxia telangiectasia y el síndrome de Bloom, entre otros, se relacionan con el aumento de la inestabilidad celular del ADNr, pero su papel es aún desconocido (Warmerdam y Wolthuis, 2019). En particular, el conjunto de ADNr se ha resistido a las tecnologías de secuenciación y ensamblaje debido a su naturaleza repetitiva y, como consecuencia, esta porción del genoma sigue siendo un área no cubierta en los ensamblajes del genoma, incluso en organismos modelo bien estudiados (Wang y Lemos, 2019).

La alteración en los pasos de maduración de cualquier ARNt o ARNr y sus enzimas relacionadas puede afectar a la homeostasis celular en varios niveles y ser la causa de diferentes trastornos, como el cáncer, las infecciones, muchas enfermedades neurodegenerativas y otras condiciones patológicas (Tuorto y Parlato, 2019). Más allá de comprender el papel de los ARNt en las enfermedades humanas, en algunos casos estos fragmentos de ARNt pueden servir como biomarcadores útiles (Anderson e Ivanov, 2014).

La expresión anómala de los ARNpn se ha asociado con el cáncer, las enfermedades cardiovasculares y varios trastornos neurológicos y neuromusculares. En consecuencia, estos ARNnc se han propuesto como atractivos agentes

terapéuticos capaces de activar o inhibir el empalme de ARNm y nuevos biomarcadores, debido a su expresión específica en tejidos particulares y a su circulación estable en fluidos biológicos (Isakova y Quake, 2018).

### ***Telómeros y centrómeros***

El acortamiento de los telómeros está muy relacionado con las enfermedades asociadas a la edad. Las mutaciones en los genes de mantenimiento de los telómeros provocan su erosión, lo que da lugar a diferentes síndromes de disfunción telomérica, como el síndrome de Hoyeraal-Hreidarsson, la disqueratosis congénita y la anemia aplásica, todos ellos caracterizados por un envejecimiento prematuro (Martínez y Blasco, 2017). El acortamiento de los telómeros también provoca reordenamientos cromosómicos que contribuyen a la aparición o la progresión de los tumores. Además, la expresión ectópica de la telomerasa suele asociarse a la transformación de células malignas.

Los centrómeros, probablemente debido a la alta densidad de secuencias repetitivas, son regiones frágiles y propensas a la rotura de cromosomas y los reordenamientos (Barra y Fachinetti, 2018). Por consiguiente, son una fuente potencial de inestabilidad del genoma que puede relacionarse con las enfermedades humanas. De hecho, algunos tipos de tumores se caracterizan por una alta frecuencia de reordenamientos cromosómicos que afectan a la región pericentromérica. Además, las alteraciones centroméricas se encuentran en enfermedades congénitas humanas que incluyen algunos tipos de inmunodeficiencias, la inestabilidad de la región centromérica y el síndrome de anomalías faciales. En este síndrome, las mutaciones en los genes DNMT3B, ZBTB24, CDCA7 y HELLS provocan la pérdida de metilación del ADN y la alteración de la estructura de la heterocromatina de las regiones centroméricas, lo que produce roturas pericentroméricas y reordenamientos cromosómicos. Todavía no se comprende bien cómo, en qué momento y en qué medida las disfunciones de los centrómeros pueden intervenir en estas patologías, por lo que se necesitan nuevos modelos y tecnologías. Concretamente, la investigación de los centrómeros estaba limitada por la ausencia de tecnologías de secuenciación sensibles para el ADN repetitivo. Por ese motivo, serán fundamentales nuevos avances en secuenciación y bioinformática con el fin de revelar los procesos que reportan inestabilidad en estas regiones de las enfermedades humanas.

### ***Potenciadores***

El genoma humano contiene más de 1 000 000 de potenciadores parcialmente caracterizados que constituyen una parte importante del GENOMAnc.

Además, en ellos se encuentra una gran fracción de todos los SNP asociados a enfermedades humanas (Maurano *et al.*, 2012). Las enfermedades relacionadas con los potenciadores se denominan *potenciopatías* e incluyen la polidactilia, causada por mutaciones en uno de los potenciadores del gen SHH o algunas  $\beta$ -talasemias causadas por translocaciones en las regiones de control del locus de globina, un conocido superpotenciador (Smith y Shilatifard, 2014). Las mutaciones en los genes que codifican las proteínas encargadas de regular la función del potenciador también se consideran *potenciopatías*. Un ejemplo es el síndrome de Cornelia de Lange que está provocado por mutaciones en los genes que codifican NIPBL, otras subunidades del complejo de cohesina y el lector epigenético BRD4. Las translocaciones y los reordenamientos cromosómicos también causan *potenciopatías* al cambiar la posición tridimensional de las regiones reguladoras con respecto a sus genes diana, lo que provoca la expresión ectópica de nuevos genes diana. Así, el linfoma de Burkitt se origina por la translocación del potenciador IGH en proximidad al gen MYC (Gillies *et al.*, 1983). La reprogramación epigenética de los potenciadores también está implicada en la predisposición al cáncer (Aran y Hellman, 2013) y la metástasis (Roe *et al.*, 2017). Estos son solo algunos ejemplos, sin embargo, la mayoría de las enfermedades y susceptibilidades causadas por alteraciones de los potenciadores no están caracterizadas y su estudio constituye uno de los principales desafíos de la genética humana.

### ***miARN***

El papel de los miARN en el cáncer ha suscitado una atención especial. Los perfiles de expresión de los miARN son específicos de cada tumor y difieren entre los distintos tipos de tumores. De hecho, los perfiles de miARN se utilizan para la estratificación de tumores y para la predicción del pronóstico. Los miARN específicos actúan como oncogenes (miARN<sub>onco</sub>), supresores de tumores, o ambos, según el contexto (Di Leva *et al.*, 2014). Además, la detección precoz del cáncer en biopsias líquidas mediante la determinación de miARN circulantes es una estrategia prometedora (Toiyama *et al.*, 2017), que con el tiempo pudiera servir también para otras enfermedades.

### ***ARNncl***

Los ARNncl desempeñan importantes papeles reguladores en casi todas las vías de señalización que afectan a la proliferación y diferenciación celular y, en consecuencia, están directamente implicados en la transformación maligna (Huar-te, 2015). Por ejemplo, la sobreexpresión del ARNncl HOTAIR provoca metástasis en las células de cáncer de mama y promueve el silenciamiento del grupo

HOXD, entre otros genes, lo que desencadena la desdiferenciación y aumenta la invasividad de las células cancerosas y la metástasis. Otro ejemplo incluye el ARNnc antisentido CDKN2B-AS1, cuya sobreexpresión provoca el silenciamiento del locus supresor de tumores INK. Los ARNnc también pueden desempeñar funciones supresoras de tumores regulando negativamente la expresión de oncogenes, como se ha demostrado para PVT1, que controla la expresión del oncogén contiguo MYC (Cho *et al.*, 2018). Además del cáncer, los ARNnc también se asocian a otras enfermedades como las cardiovasculares (Liu *et al.*, 2014), las del neurodesarrollo (Ang *et al.*, 2019) y las celíacas (Castellanos-Rubio *et al.*, 2016). El número de ARNnc implicados en enfermedades está creciendo rápidamente y es probable que, en los próximos años, miles de ellos guarden relación con diferentes patologías. Sin embargo, debido a las preocupaciones planteadas por un informe reciente (Goudarzi *et al.*, 2019), también será importante validar su relevancia funcional y patológica *in vivo* y con métodos adecuados y ortogonales.

## 4.3 PUNTOS CLAVE DEL DESAFÍO

La gran abundancia, la diversidad y nuestra todavía escasa comprensión del GENOMAnc originan muchas cuestiones complejas que deben resolverse en futuras investigaciones y que pueden organizarse en cuatro grupos diferentes. Esta lista de preguntas que resolver no es completa, pero debe incluir las siguientes cuestiones:

### 4.3.1 Conocimiento básico de la función y estructura del GENOMAnc

- ¿Por qué hay tantos tipos de elementos en el GENOMAnc?
- ¿Cuál es el mapa completo de las interacciones entre los GENOMAnc y sus dianas en animales y plantas?
- ¿Cómo influyen las estructuras de los ARNnc en la separación de fases líquido-líquido y qué desencadena la organización de dichas estructuras?
- ¿Impulsan los ARNe el bucle potenciador-promotor y desencadenan la transcripción de genes mediante un proceso mediado por la separación de fases líquido-líquido?
- ¿Cómo modulan los ARNnc la estructura subnuclear y la organización tridimensional del genoma?
- ¿Cómo afectan las modificaciones epigenéticas a la función de los ARNnc?
- ¿Cómo se organiza la interacción reguladora combinatoria entre los diferentes componentes del GENOMAnc?



Otros desafíos relacionados con la función de los potenciadores incluyen la identificación de los determinantes moleculares que definen no solo los potenciadores más complejos sino también el conjunto mínimo de elementos que les confieren actividad, así como la comprensión de varias características de los potenciadores, incluidos la función de la presencia de modificaciones postraduccionales específicas de las histonas, la unión de una variedad de proteínas, como CTCF y cohesinas, y la función del ARNe.

Otros desafíos relacionados con los ARNnc incluyen entender con claridad cómo los ARNnc ejercen sus funciones y en qué medida dependen de sus características estructurales, su interacción con otras moléculas o incluso su localización intranuclear en el núcleo 3D. También hay que dedicar muchos esfuerzos a determinar las estructuras secundarias y terciarias de los ARNnc con el fin de identificar los principales motivos para la función. Otra línea importante de investigación futura será la determinación de los socios de unión de estos transcritos. Por último, pero no por ello menos importante, el estudio de las modificaciones epigenéticas en el contexto de los ARNnc puede revelar mecanismos desconocidos y nuevas capas de complejidad en las redes de regulación capaces de repercutir en la salud y la enfermedad.

#### **4.3.2 Relevancia del GENOMAnc en el desarrollo y la enfermedad**

- ¿Por qué los genomas de los eucariotas son tan ricos en ADN no codificante?
- ¿Cuál es la importancia de los diferentes elementos del GENOMAnc en el desarrollo, la organogénesis, la regeneración, el envejecimiento y la enfermedad?
- ¿Son todos los potenciadores similares desde un punto de vista mecanicista o hay tipos de potenciadores fundamentalmente diferentes?
- ¿Cómo se definen por sus características moleculares las especificidades del potenciador-promotor durante el desarrollo?
- ¿Cuál es la dinámica de la interacción potenciador-promotor durante el desarrollo?
- ¿Por qué hay varios potenciadores que controlan los genes en el mismo contexto celular (redundancia de potenciadores)?
- ¿Qué identidad tienen los FT pioneros y cómo actúan en la regulación de la función potenciadora durante el desarrollo?
- ¿Cómo podemos explotar el GENOMAnc para generar elementos reguladores sintéticos que con el tiempo se utilicen como agentes terapéuticos?
- ¿Qué impacto tienen las modificaciones epigenéticas de los ARNnc en el desarrollo y la enfermedad?



Se necesita un enfoque sistemático de alto rendimiento en todo el genoma para validar la actividad potenciadora *in vivo*. Los estudios sistemáticos a gran escala de las secuencias e interacciones de potenciadores y promotores, combinados con los análisis computacionales, resultan clave para descifrar las reglas que subyacen a la compleja red de interacciones entre potenciadores y promotores durante el desarrollo. Por otra parte, las estrategias basadas en CRISPR dirigidas a editar potenciadores específicos mutados o a modificar su estado epigenético descubrirán patrones regulados por el desarrollo y proporcionarán una nueva herramienta terapéutica para tratar a los pacientes. Además, de los miles de ARNnc anotados, solo una pequeña fracción se ha interrogado de manera funcional. Por este motivo, sigue suponiendo un desafío identificar y caracterizar de forma sistemática todos los ARNnc funcionales de un organismo y su papel específico en el tipo de célula durante el desarrollo, además de cómo se controlan epigenéticamente. Dado que las moléculas de ARN están formadas por secuencias específicas, resulta realista predecir que servirán para diseñar objetivos farmacológicos que modulen la actividad de los ARNnc (miARN, ARNnc y ARNcirc) mediante la modulación de su función.

#### 4.3.3 Relevancia del GENOMAnc en la evolución

- ¿Cuáles son las diferencias estructurales y funcionales entre los linajes evolutivos (animales y plantas) en los distintos elementos del GENOMAnc y sus transcritos?
- ¿Qué puede decirnos la genómica comparativa sobre los mecanismos universales y específicos de cada grupo?
- ¿Cuál es la contribución del GENOMAnc a la especiación, la evolución y las interacciones ecofisiológicas?
- ¿Cuál es el papel de los ET en la reorganización, transferencia y dinámica del genoma en el árbol de la vida?
- ¿Qué impacto tiene el GENOMAnc en la dinámica del genoma, incluidas, entre otras, la poliploidía o la aneuploidía?

Conviene destacar que al ensamblar los genomas, las regiones más difíciles de resolver están constituidas por las regiones repetitivas y no codificantes. Por ese motivo, las nuevas técnicas de NGS capaces de tratar el ADN repetitivo revelarán en el futuro los misterios del GENOMArepetitivo, los modos evolutivos basados en secuencias repetitivas y el origen de organizaciones alternativas. Los futuros análisis globales de los cambios en los paisajes reguladores entre diferentes organismos también ayudarán a dilucidar los mecanismos genéticos de la evolución.

## Desarrollo técnico de métodos para el estudio del GENOMAnc

- ¿Qué desafíos tecnológicos surgirán para avanzar en las estrategias de análisis del GENOMAnc más allá de las actuales Hi-C, ChIP-seq, ATAC-seq, DNase-seq o microscopia de alta resolución, entre otras?
- ¿Cuáles son los desafíos técnicos para avanzar en nuestra capacidad de identificar los elementos del GENOMAnc y comprender sus características estructurales y funcionales?

Como se describe con más detalle en el capítulo 3.1, la mejora de las tecnologías de células individuales, las potentes herramientas de NGS, las técnicas de edición de genes y la microscopia de superresolución pudieran tener un impacto en nuestra comprensión del GENOMAnc. Por otro lado, las nuevas herramientas informáticas y bioinformáticas (véase el capítulo 3.2) serán esenciales para interpretar esta enorme cantidad de datos que, al final, nos proporcionarán interesantes conocimientos sobre la regulación temporal y espacial del GENOMAnc, ayudándonos a identificar nuevos mecanismos que subyacen a los rasgos de desarrollo, enfermedad y evolución. Además, es necesario desarrollar nuevos métodos de edición del ADN para comprender el impacto del GENOMAnc en el desarrollo y la enfermedad del organismo.

## CAPÍTULO 4 BIBLIOGRAFÍA

- Abarrategui, I., Krangel, M.S. (2006).** Regulation of T-cell receptor  $\alpha$  gene recombination by transcription. *Nat. Immunol.* 7, 1109-1115.
- Alberti, C., Cochella, L. (2017).** A framework for understanding the roles of miRNAs in animal development. *Development.* 144, 2548-2559.
- Alessio, E., Bonadio, R.S., Buson, L., Chemello, F., Cagnin, S. (2020).** A single cell but many different transcripts: A journey into the world of long non-coding RNAs. *Int. J. Mol. Sci.* 21, 302.
- Anderson, P., Ivanov, P. (2014).** tRNA fragments in human health and disease. *FEBS Lett.* 588, 4297-4304.
- Ang, C.E., Ma, Q., Wapinski, O.L., Fan, S., Flynn, R.A., Lee, Q.Y., Coe, B., Onoguchi, M. et al. (2019).** The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *eLife* 8, e41770.
- Arnold, P.R., Wells, A.D., Li, X.C. (2020).** Diversity and emerging roles of enhancer RNA in regulation of gene expression and cell fate. *Front. Cell. Dev. Biol.* 7, 377.
- Barra, V., Fachinetti, D. (2018).** The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nat. Commun.* 9, 1-17.
- Castellanos-Rubio, A., Fernández-Jiménez, N., Kratchmarov, R., Luo, X., Bhagat, G., Green, P.H., Schneider, R., Kiledjian et al. (2016).** A long noncoding RNA associated with susceptibility to celiac disease. *Science* 352, 91-95.
- Chen, Q., Yan, W., Duan, E. (2016).** Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat. Rev. Genet.* 17, 733-743.
- Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J. et al. (2018).** Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. *Cell* 173, 1398-1412.
- Corradin, O., Scacheri, P.C. (2014).** Enhancer variants: evaluating functions in common disease. *Genome Med.* 6, 85.
- Csorba, T., Questa, J.I., Sun, Q., Dean, C. (2014).** Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. USA* 111, 16160-16165.
- Davidson, E.H., McClay, D.R., Hood, L. (2003).** Regulatory gene networks and the properties of the developmental process. *Proc. Natl. Acad. Sci. USA* 100, 1475-1480.
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J. et al. (2010).** A large fraction of extragenic RNA pol II enhancers. *PLoS Biol.* 8, e1000384.
- Deryusheva, S., Gall, J.G. (2019).** scaRNAs y snoRNAs: Are they limited to specific classes of substrate RNAs? *RNA* 25, 17-22.
- Di Leva, G., Garofalo, M., Croce, C.M. (2014).** MicroRNAs in cancer. *Annu. Rev. Pathol.* 9, 287-314.
- ENCODE Project Consortium (2012).** An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.
- Franke, M., Gómez-Skarmeta, J.L. (2018).** An evolutionary perspective of regulatory landscape dynamics in development and disease. *Curr. Opin. Cell. Biol.* 55, 24-29.
- Gillies, S.D., Morrison, S.L., Oi, V.T., Tonegawa, S. (1983).** A tissue-specific enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, 717-728.
- Giallourakis, C.C., Franklin, A., Guo, C., Cheng, H.L., Yoon, H.S., Gallagher, M., Perlot, T., Andzelm, M. et al. (2010).** Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineagespecific V(D)J recombination. *Proc. Natl. Acad. Sci. USA* 107, 22207-22212.
- Grüll, M.P., Massé, E. (2019).** Mimicry, deception and competition: the life of competing endogenous RNAs. *Wiley Interdiscip. Rev. RNA* 10, e1525.
- Goudarzi, M., Berg, K., Pieper, L.M., Schier, A.F. (2019).** Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *eLife*, 8, e40815.

- Ha, C.E., Bhagavan, N.V. (2011). *Essentials of medical biochemistry: with clinical cases*. Academic Press.
- Hanada, T., Weitzer, S., Mair, B., Bernreuther, C., Wainger, B.J., Ichida, J., Hanada, R., Orthofer, M. *et al.* (2013). CLP1 links tRNA metabolism to progressive motor-neuron loss. *Nature* 495, 474-480.
- Heard, E., Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157, 95109.
- Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* 16, 144-154.
- Heslop-Harrison, J.S.P., Schwarzacher, T. (2011). Organization of the plant genome in chromosomes. *Plant J.* 66, 1833.
- Hirsch, C.D., Jiang, J. (2013). Centromeres: Sequences, structure, and biology, 59-70. En: Wendel, J., Greilhuber, J., Dolezel, J. y Leitch, I.J. (eds.), *Plant genome diversity, vol. 1, Plant genomes, their residents, and their evolutionary dynamics*. Vienna: Springer.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., Young, R.A. (2013). Superenhancers in the control of cell identity and disease. *Cell* 155, 934-947.
- Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253-1261.
- Isakova, A., Quake, S.R. (2018). A mouse tissue atlas of small nuclear non-coding RNA. *bioRxiv* 430561.
- Iwafuchi-Doi, M., Zaret, K.S. (2016). Cell fate control by pioneer transcription factors. *Development* 143, 1833-1837.
- Jeyaraman, S., Hanif, E.A.M., Mutalib, N.S.A., Jamal, R., Abu, N. (2020). Circular RNAs: Potential regulators of treatment resistance in human cancers. *Front. Genet.* 10, 1369.
- Kanai, A. (2014). Welcome to the new tRNA world! *Front. Genet.* 5, 336.
- Kanduri, C. (2016). Long noncoding RNAs: Lessons from genomic imprinting. *Biochim. Biophys. Acta* 1859, 102111.
- Khrosraviani, N., Ostrowski, L.A., Mekhail, K. (2019). Roles for non-coding RNAs in spatial genome organization. *Front. Cell. Dev. Biol.* 7, 336.
- Klemm, S.L., Shipony, Z., Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207-220.
- Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptevich, M. *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182-187.
- Lewis, M.W., Li, S., Franco, H.L. (2019). Transcriptional control by enhancers and enhancer RNAs. *Transcription* 10, 171-186.
- Liu, J.Y., Yao, J., Li, X.M., Song, Y.C., Wang, X.Q., Li, Y.J., Yan, B., Jiang, Q. (2014). Pathogenic role of lncRNAMALAT1 in endothelial cell dysfunction in diabetes mellitus. *Cell Death Dis.* 5, e1506.
- Martínez, P., Blasco, M.A. (2017). Telomere-driven diseases and telomere-targeting therapies. *J. Cell. Biol.* 216, 875887.
- Loda, A., Heard, E. (2019). Xist RNA in action: past, present and future. *PLoS Genet.* 15, e1008333.
- Martinez-Salas, E., Linney, E., Hassel, J., De Pamphilis, M.L. (1989). The need for enhancers in gene expression first appears during mouse development with formation of zygotic nucleus. *Genes Dev.* 3, 1493-1506.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R. *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190-1195.
- Matsumoto, A., Nakayama, A.I. (2018). Hidden peptides encoded by putative non-coding RNAs. *Cell Struct. Func.* 43, 75-83.
- Mensà, E., Latini, S., Ramini, D., Storci, G., Bonafè, M., Olivieri, F. (2019). The telomere world and aging: analytical challenges and future perspectives. *Ageing Res. Rev.* 50, 27-42.
- Nair, S.J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M.J., Wang, S., Suter, T. *et al.* (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat. Struct. Mol. Biol.* 26, 1932013.
- Necsulea, A., Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* 15, 734-748.

- Pamadurti, N.R., Bartok, O., Jens, M., Ashwall-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E. *et al.* (2017). Translation of circRNAs. *Mol. Cell* 66, 9-21.
- Pantartzzi, C.N., Pergner, J., Kozmik, Z. (2018). The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. *Sci. Rep* 8, 1-11.
- Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S. *et al.* (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* 110, 17921-17926.
- Pu, M., Chen, J., Tao, Z., Miao, L., Qui, X., Wang, Y., Ren, J. (2019). Regulatory network of miRNA on its target: Coordination between transcriptional and post-transcriptional regulation of gene expression. *Cell. Mol. Life Sci.* 76, 441-451.
- Raina, M., Ibba, M. (2014). tRNAs as regulators of biological processes. *Front. Genet.* 5, 171.
- Rakyan, V.K., Down, T.A., Balding, D.J., Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12, 529-541.
- Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I., Gage, F.H. (2013). The role of transposable elements in health and diseases of the central nervous system. *J. Neurosci.* 33, 17577-17586.
- Roe, J.S., Hwang, C.I., Somerville, T.D.D., Milazzo, J.P., Lee, E.J., Da Silva, B., Maiorino, L., Tiriach, H. *et al.* (2017). Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* 170, 875-888.
- Quinn, J.J., Chang, H.Y. (2017). Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17, 47-62.
- Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannet, M.N. *et al.* (2018). Coactivator condensation at super-enhancers links phase-separation and gene control. *Science* 361, eaar3958.
- Schoenfelder, S., Fraser, S. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20, 437-455.
- Serrato-Capuchina, A., Matute, D.R. (2018). The role of transposable elements in speciation. *Genes* 9, 254.
- Shlyueva, D., Stampfel, G., Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272-286.
- Smith, K.P., Hall, L.L., Lawrence, J.B. (2020). Nuclear hubs built on RNAs and clustered organization of the genome. *Curr. Opin. Cell Biol.* 64, 67-76.
- Smith, E., Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* 21, 210-219.
- Tang, W.W., Dietmann, S, Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F. *et al.* (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell* 161, 14531467.
- Thompson, D.M., Lu, C., Green, P.J., Parker, R. (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* 14, 2095-2103.
- Toiyama, Y., Okugawa, Y., Tanaka, K., Araki, T., Uchida, K., Hishida, A., Uchino, M., Ikeuchi, H. *et al.* (2017). A panel of methylated microRNA biomarkers for identifying high-risk patients with ulcerative colitis-associated colorectal cancer. *Gastroenterology* 153, 1634-1646.
- Tuorto, F., Parlato, R. (2019). rRNA and tRNA bridges to neuronal homeostasis in health and disease. *J. Mol. Biol.* 431, 1763-1779.
- Tyebji, S., Hannan, A.J., Tonkin, C.J. (2020). Pathogenic Infection in Male Mice Changes Sperm Small RNA Profiles and Transgenerationally Alters Offspring Behavior. *Cell Rep.* 31, 107573.
- Ulitsky, I. (2018). Interactions between short and long noncoding RNAs. *FEBS Lett.* 592, 2874-2883.
- Wang, M., Lemos, B. (2019). Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res.* 29, 325-333.
- Wang, Y., Mo, Y., Gong, Z., Yang, X., Yang, M., Zhang, S., Xiong, F., Xiang, B. *et al.* (2017). Circular RNAs in human cancer. *Mol. Cancer* 16, 1-8.
- Warmerdam, D.O., Wolthuis, R.M. (2019). Keeping ribosomal DNA intact: a repeating challenge. *Chromosome Res.* 27, 57-72.



### RESUMEN

En los últimos años, se han realizado grandes esfuerzos para caracterizar el epigenoma y el epitranscriptoma de diferentes tipos de células y organismos. Sin embargo, la mayoría de esos estudios son descriptivos, por lo que seguimos ignorando el papel de muchas epimodificaciones. Debátimos la necesidad de obtener una visión funcional y mecanicista y el gran impacto que este conocimiento tendrá en la comprensión y el tratamiento de numerosas enfermedades.

### PALABRAS CLAVE

epigenética	epitranscriptómica
cromatina	enfermedades poco frecuentes
cáncer	degenerativo
enfermedades autoinmunes e infecciosas	
metabólico	
trastornos neurológicos y psiquiátricos	

# EPIGENÉTICA FUNCIONAL Y EPITRANSCRIPTÓMICA Y SU PAPEL EN LA SALUD Y LA ENFERMEDAD

## Coordinadores

Ángel Barco  
(IN, Alicante, coordinador)

María Gómez Vicentefranqueira  
(CBMSO, Madrid, coordinadora adjunta)

## Investigadores y centros de investigación participantes (por orden alfabético)

Sandra Blanco Benavente  
(IBMCC, Salamanca)

Elena Gómez Díaz  
(IPBLN, Granada)

Javier Martín  
(IPBLN, Granada)

Marián Martínez Balbás  
(IBMB, Barcelona)

Jordi Pérez-Tur  
(IBV, Valencia)

Isidro Sánchez García  
(IBMCC, Salamanca)

Carles Suñé  
(IPBLN, Granada)

Mario Vallejo  
(IIBM, Madrid)

## RESUMEN EJECUTIVO

Durante los últimos años, se han realizado grandes esfuerzos internacionales para caracterizar el epigenoma de diferentes tipos de células y organismos. Hasta la fecha se han identificado más de 100 modificaciones covalentes distintas de la cromatina que afectan tanto al ADN como a las proteínas histónicas. La aparición de la epitranscriptómica es más actual y estudios recientes están desvelando nuevas e inesperadas capas de regulación de la expresión génica. A pesar de estos avances, seguimos ignorando la función específica —si es que existe— de la mayoría de estas modificaciones epigenéticas y epitranscriptómicas y su impacto en la transcripción, la traducción y la biología celular.

En este capítulo, analizaremos primero la importancia de caracterizar de forma funcional y mecanicista el epigenoma y el epitranscriptoma en diferentes contextos celulares. Seguidamente, haremos hincapié en la relevancia de estos estudios para la comprensión y el tratamiento de numerosas afecciones, como las enfermedades poco frecuentes, el cáncer, las enfermedades degenerativas, autoinmunes e infecciosas, y los trastornos metabólicos, neurológicos y psiquiátricos. La investigación de los mecanismos epigenéticos en todas estas condiciones ya ha contribuido a una mejor



comprensión de su etiopatología. A continuación, se analizan los desafíos a los que todavía se enfrenta el sector. El constante desarrollo tecnológico está desvelando nuevos mecanismos y eventos que esculpen el epigenoma y el epitranscriptoma, lo que conduce a la producción de una enorme cantidad de información genómica y transcriptómica que debe procesarse e integrarse para comprender en su totalidad la función de estos epicambios y su implicación en las enfermedades. Aunque se ha informado de cambios en la cromatina en muchos trastornos y estos cambios suelen correlacionarse con la progresión de la enfermedad, sigue sin saberse si estas alteraciones epigenéticas son una causa o una consecuencia de la patología. Las tecnologías innovadoras para la manipulación precisa del epigenoma y el epitranscriptoma deberían permitirnos abordar el enigma de la causalidad en un futuro próximo. Además, deben identificarse compuestos con el potencial de restablecer el paisaje epigenético y epitranscriptómico normal y evaluarse en ensayos clínicos. Estos fármacos epigenéticos pueden abrir nuevas vías para el tratamiento de una gran cantidad de trastornos.

El CSIC cuenta con numerosos y excelentes grupos de investigación que trabajan en epigenética desde distintos puntos de vista. En la parte final de este capítulo, se analizarán los recursos que el CSIC dedica a hacer frente a estos desafíos y las acciones que se podrían poner en marcha en la institución para mejorar su posición en este campo de investigación esencial y de rápida evolución.

## 5.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

### 5.1.1 Epigenoma

Aunque todas las células somáticas de un organismo pluricelular tienen el mismo genoma (es decir, idéntica secuencia de ADN), los distintos tipos de células tienen transcriptomas diferentes (conjunto de todas las moléculas de ARN expresadas), proteomas diferentes (conjunto de todas las proteínas) y, por tanto, funciones diferentes. Esto se consigue en gran medida mediante modificaciones en la cromatina. La unidad estructural básica de la cromatina eucariótica es el nucleosoma, en el que aproximadamente 150 pb de ADN se enrollan alrededor de un núcleo proteico básico integrado por dos copias de histonas H2A, H2B, H3 y H4. Las modificaciones de la cromatina que afectan tanto al ADN como a las histonas permiten la propagación de los estados de actividad activa y silenciosa de las células madre a las hijas dentro de un linaje celular determinado. Los cambios arquitectónicos y de conformación en los nucleosomas, y los bucles de retroalimentación reguladores de los genes también contribuyen a modificar la expresión de la información genética sin alterar la propia información. Estos mecanismos se denominan de manera colectiva *epigenoma*.

Uno de los procesos epigenéticos más investigados es la metilación del ADN, que se produce con diferentes patrones en el genoma de microorganismos, plantas y filos animales, al tiempo que desempeña sistemáticamente funciones de protección y regulación. Otros sistemas epigenéticos importantes en los organismos eucariotas son los complejos de heterocromatina (HP1 y H3K-9me3), Polycomb (PRC1 y PRC2) y Trithorax. La transmisión de las marcas epigenéticas a través de la división celular requiere que sobrevivan a la replicación del ADN y a la mitosis, lo que es bastante relevante para las modificaciones de las histonas, ya que los nucleosomas no tienen un sistema de duplicación basado en plantillas de ADN. Se cree que los complejos mencionados perpetúan las respuestas funcionales uniendo modificaciones específicas de las histonas y modificando otras proteínas histónicas en la parte central del nucleosoma para transmitir una herencia estable. La herencia epigenética suele implicar la cooperación de señales parcialmente superpuestas, con lo que cada una de ellas añade un grado de estabilidad, si bien siguen siendo reversibles, lo que permite la plasticidad en presencia de señales reguladoras (Cavalli y Heard, 2019). Los genomas de los organismos unicelulares también son portadores de información epigenética. Análisis recientes del metiloma han demostrado que la metilación del ADN está muy extendida en los

genomas de bacterias y arqueas, incluidos los pequeños genomas de ciertos parásitos estrictos. Estos estudios hacen tambalear algunas viejas concepciones y demuestran que los linajes epigenéticos permiten la adaptación de las poblaciones bacterianas a entornos difíciles o cambiantes y modulan la interacción de los patógenos con sus huéspedes eucariotas. Además de su función en la herencia epigenética, las modificaciones de la cromatina también desempeñan un papel fundamental en la plasticidad celular. Gracias a los cambios en su epigenoma, las células que no se dividen en los órganos, como el músculo o el cerebro, pueden adaptarse o responder de forma diferente a los cambios en su entorno en función de su historial de activación anterior.

La biología molecular, la genómica y la proteómica basada en la espectrometría de masas han identificado más de un centenar de modificaciones postraduccionales (PTM) de residuos específicos de las histonas que pueden actuar de forma combinada generando miles de patrones, muchas de cuyas funciones están siendo objeto de intensa investigación. El ADN también puede modificarse de múltiples maneras, desde la metilación abundante de citosinas en la cromatina de mamíferos y plantas hasta cambios menos frecuentes tanto en las citosinas como en las demás bases. Con todo, aún se debate si son meros intermediarios metabólicos o si desempeñan un papel funcional en la biología del ADN. La investigación minuciosa sobre la cromatina mediante enfoques químicos, celulares y de biología molecular ha proporcionado valiosos conocimientos sobre la función molecular de los reguladores epigenéticos. Estos conocimientos han puesto de manifiesto la naturaleza altamente dinámica del epigenoma y han procurado el fundamento molecular que nos permite centrarnos en estas proteínas con fines terapéuticos. El resultado de 25 años de intensa investigación en el campo de la epigenética y la epigenómica ha puesto de manifiesto un escenario extraordinariamente complejo con cientos de proteínas que introducen (escritores), eliminan (borradores) o se unen (lectores) a modificaciones específicas de la cromatina. Esta complejidad recalca la contribución esencial de los mecanismos epigenéticos en el desarrollo, el envejecimiento, la regeneración de tejidos y la plasticidad celular, así como la implicación de la desregulación epigenética en una gran cantidad de enfermedades.

### **5.1.2 Epitranscriptoma**

Al igual que el ADN, el ARN puede modificarse de formas diversas y complejas. La deposición química puede ocurrir en todos los nucleótidos y posiciones (carbono, nitrógeno u oxígeno en las bases o ribosa) y abarca más de 170 modificaciones, incluidas la metilación, la hidroximetilación, la

hidroxilación, la acetilación, la pseudouridilación y la glicosilación (Boccalitto *et al.*, 2018). El procesamiento postraducciona del ARN no solo abarca la deposición de grupos químicos sino también la edición de secuencias, el recorte y empalme de secuencias innecesarias, la adición de ribonucleótidos no incluidos en el código genético y la unión a proteínas para bloquear su estructura tridimensional o ganar propiedades catalíticas. En conjunto, estas modificaciones del ARN constituyen el *epitranscriptoma*, un término acuñado hace pocos años. Dos importantes avances desencadenaron la aparición de esta nueva área de investigación. En primer lugar, el hallazgo de que algunos genes vinculados a la enfermedad codifican para enzimas modificadoras del ARN (por ejemplo, FTO que codifica una desmetilasa de ARN 6-metiladenosina [m6A]) (Jia *et al.*, 2011). En segundo lugar, el avance de sólidos métodos de detección para el mapeo de m6A y otras modificaciones del ARN ha permitido el desarrollo de métodos de secuenciación masiva para estudiar el epitranscriptoma (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). La investigación pionera sobre la función fisiológica de estas modificaciones ha dejado al descubierto funciones reguladoras en una variedad de procesos celulares, incluidos la autorrenovación y diferenciación de las células madre, la proliferación, el desarrollo, las respuestas a las señales ambientales, la migración, la supervivencia al estrés, la respuesta inmune, la función mitocondrial y el reloj circadiano (Frye *et al.*, 2018). La comprensión del papel de estas modificaciones postraduccionales dinámicas del ARN representa una nueva frontera en la investigación, a menudo denominada *epitranscriptómica*.

La intensa investigación de los últimos años, junto con el desarrollo de novedosas tecnologías y herramientas, ha desvelado un complejo conjunto de nuevos mecanismos reguladores de la expresión génica que afectan a la localización, estabilidad y eficiencia de traducción de las moléculas de ARN (Davalos *et al.*, 2018). Además de los ARN mensajeros (ARNm), también se producen modificaciones del ARN en el genoma no codificante, mucho menos explorado, incluidos los ARN no codificantes pequeños y largos y los elementos transponibles, cuyas funciones en el desarrollo normal y en los procesos patológicos pudieran ser de una importancia fundamental. Es más, al igual que con las enzimas epigenéticas, estamos descubriendo un gran número de enzimas modificadoras del ARN que depositan (escritores) o eliminan (borradores) modificaciones específicas en diferentes posiciones de nucleótidos, dominios de ARN o tipos de ARN. Por otra parte, se han identificado grupos de proteínas que se unen específicamente a los nucleótidos modificados (lectores), lo que afecta al destino del ARN. También, como ocurre con las histonas y el ADN,

las proteínas de unión al ARN (RBP, por sus siglas en inglés) pueden modificarse como parte del repertorio del epitranscriptoma. Las mutaciones en todas estas proteínas (escritores, borradores y lectores) se han asociado a diversas patologías, del cáncer a la disfunción neuronal, la fertilidad o el metabolismo (Harries, 2019), lo que convierte a las enzimas epitranscriptómicas y a sus sustratos en prometedores objetivos terapéuticos.

## 5.2 IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS POSIBLES APLICACIONES

### 5.2.1 Epigenética funcional y epitranscriptómica

El estudio de la epigenética y la epitranscriptómica ha experimentado un fuerte impulso en los últimos años. Esto se debe principalmente a la llegada de avances tecnológicos como la capacidad de analizar todo el epigenoma o todo el epitranscriptoma en un solo experimento y la posibilidad de estudiar una mayor diversidad de epimarcas. Los últimos avances han demostrado que la 5-metilcitosina y su contraparte hidroxilada son solo dos de una serie de modificaciones que sufre el ADN para modular la regulación de los genes. La constatación de que las moléculas de ARN están sometidas a una gran cantidad de modificaciones químicas añade nuevas capas de complejidad al sofisticado panel de control de la expresión génica en una célula. Estas técnicas incluyen: la secuenciación de bisulfitos del genoma completo para investigar la metilación del ADN y el ARN; ChIP-seq para mapear las modificaciones de las histonas y la unión de factores de transcripción; ATAC-seq para explorar la accesibilidad y la ocupación de la cromatina; Hi-C para dilucidar la arquitectura de la cromatina; iCLIP-seq para identificar los sitios de unión de RBP en el ARN; miCLIP, m6A-seq o m1A-seq para explorar la metilación del ARN en las adeninas del ARN; RiboMeth-seq y Nm-seq para detectar la metilación de la ribosa en el ARN; aza-IP para detectar la metilación de la citosina en el ARN; o X-seq y CeU-seq para detectar la pseudouridina, entre otras. Además, se ha descubierto que las modificaciones de la cromatina y el ARN pueden ser específicas del tipo de célula o del estado celular, así como del organismo, lo que aumenta aún más la complejidad de la regulación epigenética y epitranscriptómica.

Las marcas epigenéticas y epitranscriptómicas son de naturaleza dinámica y tienen la capacidad de aparecer o desaparecer en respuesta a estímulos externos e influencias ambientales, como los nutrientes o el estrés. También cambian desde el desarrollo hasta los organismos envejecidos. Cada vez hay más evidencias que demuestran las funciones reguladoras que permiten una

rápida adaptación de las células a los cambios del medioambiente. Los cambios del epigenoma y el epitranscriptoma pueden mostrar tanto el pasado medioambiental como la futura susceptibilidad a la enfermedad. Miles de estudios han demostrado la asociación de la deposición aberrante de epimodificaciones con afecciones que van desde las enfermedades poco frecuentes a las comunes, y desde las metabólicas a las autoinmunes, psiquiátricas o cancerosas, como analizaremos en detalle en la siguiente sección. Más allá de la salud humana, los estudios epigenéticos y epitranscriptómicos proporcionan pistas para la mejora de los cultivos y la producción animal, para manipular las interacciones entre el huésped y los microbios y para mejorar los productos biotecnológicos pertinentes y la industria alimentaria (véase el capítulo 3.6 para conocer más detalles). Los contaminantes suelen ser precursores químicos de los modificadores del ADN, las histonas y el ARN. Como se describe con más detalle en el capítulo 3.6, la comprensión de cómo la exposición a diferentes situaciones ambientales altera el paisaje epigenético servirá para predecir lo que puede ocurrir en el futuro, así como para identificar las exposiciones que se produjeron en el pasado, lo que contribuye a describir la etiología de las enfermedades ambientales, así como a permitir intervenciones preventivas. Además, cada vez hay más pruebas que demuestran la transmisión de información epigenética a través de la línea germinal (es decir, la llamada *herencia epigenética transgeneracional*) (Horsthemke, 2018). Las modificaciones residuales de ADN e histonas en las células germinales y las moléculas de ARN de larga duración se han postulado como posibles portadores de información epigenética a través de las generaciones. Estos mecanismos transgeneracionales y su importancia para la salud y la enfermedad se tratarán con más detalle en el capítulo 3.7.

### 5.2.2 Mecanismos epigenéticos en la enfermedad

La investigación multidisciplinaria y colaborativa de genetistas, bioquímicos, químicos médicos, biólogos celulares, médicos y bioinformáticos ha aportado una enorme cantidad de conocimientos sobre el papel fundamental de los reguladores epigenéticos en la etiopatología. Estos avances, y los que aún están en desarrollo, han permitido identificar varios objetivos terapéuticos prometedores que actualmente se están explorando en su mayoría en trastornos cerebrales, autoinmunes o relacionados con el cáncer. Sin embargo, descifrar la función mecanicista de los cambios epigenéticos en el origen y la progresión de las enfermedades supone todo un desafío y requerirá el desarrollo de estudios funcionales en organismos modelo o sistemas de cultivo celular. La investigación sobre estos modelos ha demostrado que las epimarcas pueden

modificarse mediante intervenciones farmacológicas o cambios en el entorno, lo que abre nuevas e insospechadas vías de tratamiento. La reproducción de estos esfuerzos de colaboración en el campo emergente de la epitranscriptómica representa un nicho de oportunidad para futuras investigaciones. En los próximos párrafos subrayaremos la importancia de este tipo de estudios en diferentes áreas de la biomedicina.

**Etiología epigenética de enfermedades poco frecuentes:** Los últimos planteamientos sobre el genoma completo han permitido identificar un número cada vez mayor de enfermedades hereditarias poco frecuentes causadas por mutaciones en factores que actúan en la cromatina, como las metiltransferasas del ADN, las enzimas modificadoras de las histonas, los factores de remodelación de la cromatina y las proteínas lectoras (Bjornsson, 2015; Velasco y Francastel, 2019). Aunque las enfermedades poco frecuentes afectan por separado a menos de 1 de cada 2000 personas, se estima que, solo en la UE, más de 30 millones de personas se ven afectadas por enfermedades poco frecuentes (The Lancet Diabetes, 2019). La mayoría de estas enfermedades no tienen un tratamiento aprobado y representan una enorme carga para los pacientes, las familias y la sociedad. La etiología de estas afecciones implica defectos en el establecimiento de marcas epigenéticas en las primeras etapas del desarrollo o en la perpetuación de dichas marcas en etapas posteriores. En muchos casos, estas alteraciones epigenéticas están asociadas a neuropatías, trastornos del neurodesarrollo, discapacidad intelectual e inmunodeficiencia. En todos los casos se altera el paisaje epigenómico y se observan cambios en los perfiles de transcripción. Curiosamente, la deficiencia de diferentes factores epigenéticos suele generar fenotipos y síntomas comunes, lo que sugiere que sus acciones convergen en vías y genes comunes. Todavía entendemos mal la naturaleza de estos nodos y los mecanismos que vinculan los cambios epigenéticos con las manifestaciones clínicas. Existe una gran necesidad de identificar los mecanismos epigenéticos que rigen el resultado de estas afecciones con el fin de identificar biomarcadores fiables para mejorar el diagnóstico y los tratamientos que disminuyan o acaben con los síntomas más graves.

**Epigenética del cáncer:** A pesar de la enorme cantidad de datos recopilados en las últimas cuatro décadas sobre la biología de las células tumorales, nuestra capacidad para comprender y controlar el desarrollo del cáncer sigue siendo, por desgracia, limitada. Todavía no sabemos cómo prevenir la conversión de una célula precancerosa en un tumor, sobre todo debido al hecho de que los primeros acontecimientos que desencadenan la adhesión a un nuevo

linaje canceroso siguen siendo en gran parte desconocidos. Un punto crucial en la historia de un tumor es la transición de una célula normal a un estado maligno. Las pruebas recientes de tumores hematopoyéticos y epiteliales revelaron que la contribución de los oncogenes al desarrollo del cáncer está mediada principalmente a través del iniciado epigenético de las células iniciadoras del cáncer, lo que sugiere que las lesiones genéticas que inician el proceso del cáncer podrían ser prescindibles para la posterior progresión y mantenimiento del tumor (Vicente-Dueñas *et al.*, 2018). En las fases iniciales del desarrollo del cáncer, una célula normal se va a convertir en una célula precancerosa por la acción de una determinada transformación oncogénica o por la exposición a un factor ambiental (como el humo del tabaco en el adenocarcinoma de pulmón); ambos pueden *restablecer* el estado epigenético o del transcriptoma y reprogramar el epigenoma para dar lugar a una célula precancerosa. Una vez realizada su función en la reprogramación oncogénica, ya no es necesaria la transformación inicial para la progresión del tumor. El iniciado epigenético maligno puede tener lugar en los primeros años de vida y permanecer en silencio hasta que segundos acontecimientos específicos desencadenen la aparición del cáncer. Estas segundas transformaciones pueden suceder al azar o pueden desencadenarse por factores ambientales o por el envejecimiento (Sen *et al.*, 2016; Tomasetti y Vogelstein, 2015). Por ello, una comprensión detallada del recableado epigenético es un requisito previo para el desarrollo de cualquier posible tratamiento contra el cáncer dirigido al epigenoma de las células precancerosas. Además, aunque todavía no sabemos si el iniciado epigenético puede ser el único impulsor del cáncer, la aparición de nuevas herramientas para detectar el evento de iniciado antes del desarrollo de un tumor completo está empezando a arrojar luz sobre este mecanismo. Los avances en este campo y el desarrollo de nuevas herramientas terapéuticas son muy prometedores para revertir los cambios epigenéticos relacionados con el cáncer. Al igual que la epigenética, el epitranscriptoma representa una nueva capa de complejidad en la biología del cáncer. Los investigadores han ido recopilando datos que implican las modificaciones postraduccionales del ARN con funciones supresoras o promotoras de tumores (Barbieri y Kouzarides, 2020).

**Epigenética de los trastornos metabólicos:** Desde el punto de vista epidemiológico, las dos enfermedades metabólicas más importantes son la obesidad y la diabetes de tipo 2. Ambos trastornos han alcanzado proporciones epidémicas en todo el mundo y se cree que tienen un importante componente epigenético. Su mayor prevalencia se ha relacionado con la mejora del nivel de vida junto con



el aumento del sedentarismo y el fácil acceso a abundante comida rápida y de alto contenido energético. Varios estudios han demostrado una estrecha correlación entre las firmas epigenéticas y los rasgos clínicos asociados a la obesidad o a la distribución del tejido adiposo, que a menudo afectan a genes relacionados con el metabolismo de la insulina/glucosa, el metabolismo de los lípidos y la adipogénesis, o la regulación de la ingesta de alimentos. Sin embargo, no está del todo claro si todos estos cambios son la causa o son secundarios a la disfunción metabólica, especialmente en los estudios realizados con muestras de sangre en lugar de tejido adiposo (Wahl *et al.*, 2017). Las regiones que presentan una metilación anómala del ADN suelen estar asociadas a genes que se sabe que regulan el metabolismo y muestran una expresión génica diferencial, lo que vincula los mecanismos epigenéticos con la disfunción de los islotes pancreáticos (Volkov *et al.*, 2017). Sin embargo, la contribución individual de estos genes o *loci* a la diabetes es bastante pequeña, en consonancia con la compleja naturaleza poligénica y multifactorial de la enfermedad. Varios estudios también han demostrado que las dietas altas en grasas y la ingesta excesiva de ácidos grasos saturados inducen cambios epigenéticos y de expresión génica en el músculo, el tejido adiposo y los islotes pancreáticos. Además, las modificaciones epigenéticas durante el desarrollo intrauterino en madres con malas condiciones nutricionales o de salud pueden conllevar la aparición de trastornos metabólicos en su progenie años después del nacimiento, como la obesidad y la diabetes, e incluso en generaciones posteriores (Sales *et al.*, 2017). Por lo tanto, un aspecto importante para futuros estudios será comprender los mecanismos por los que los hábitos alimentarios afectan al epigenoma, especialmente en individuos predispuestos (véase también el capítulo 3.7). La modificación postranscripcional del ARN m6A también desempeña un papel importante en el metabolismo de la glucosa y los lípidos, y algunos reguladores de m6A pueden estar implicados en vías hepáticas críticas relacionadas con la obesidad y el síndrome metabólico. Las investigaciones posteriores allanarán el camino hacia intervenciones dirigidas a vías epigenéticas y epitranscriptómicas específicas para el tratamiento.

**Epigenética de las enfermedades degenerativas:** Los trastornos degenerativos son un grupo heterogéneo de afecciones que pueden afectar a tejidos, órganos o a todo el organismo. Aunque algunas de estas afecciones tienen un origen genético, otras son provocadas por factores ambientales o surgen con el envejecimiento. El papel de la epigenética en el envejecimiento y las enfermedades relacionadas con la edad está bien documentado en la bibliografía. Las alteraciones en el ADN, las modificaciones y la composición de las histonas y la

regulación de los ARN no codificantes forman parte del proceso de envejecimiento y se cree que contribuyen a las enfermedades neurodegenerativas relacionadas con la edad. La incidencia de estas enfermedades, incluida la enfermedad de Alzheimer y otras demencias, la enfermedad de Parkinson y otros trastornos del movimiento o neuromusculares, casi se duplica con cada década de edad y por eso representan uno de los principales problemas de salud de la población que envejece. Al igual que ocurre con el envejecimiento, hay una reducción general de la 5-metilcitosina en varias áreas anatómicas, junto con un aumento/disminución de genes específicos como el SNCA en el párkinson o el IL-1 en el alzhéimer. También se observan alteraciones en los niveles de modificación de las histonas o, incluso, en la composición de estas, así como desregulación o expresión diferencial de algunos micro-ARN. El análisis de las modificaciones epigenéticas, tanto en el genoma como en los locus específicos, descubrirá el impacto genómico del proceso de envejecimiento y señalará enfoques de intervención que pudieran aliviar o prevenir la degeneración relacionada con la edad y otros efectos no deseados del envejecimiento.

**Epigenética de los trastornos neurológicos y psiquiátricos:** Nuestro conocimiento de la etiología de muchos de los trastornos neurológicos y psiquiátricos sigue siendo limitado, en parte debido a su origen complejo. Aunque algunas afecciones neurológicas se producen por mutaciones en genes únicos, la mayoría de los trastornos mentales tienen un origen poligénico. Por lo general, estas afecciones tienen un componente hereditario, pero la contribución de los factores ambientales oscila entre menos del 25 % en afecciones como la esquizofrenia, el trastorno bipolar, los trastornos del espectro autista o el trastorno por déficit de atención con hiperactividad (TDAH), y más del 60 % en los trastornos de ansiedad, el trastorno obsesivo-compulsivo (TOC), el trastorno por estrés postraumático (TEPT) y los trastornos depresivos mayores. La investigación de los mecanismos epigenéticos en la función cerebral ha contribuido a una mejor comprensión de este componente no heredable de la enfermedad mental. Esta investigación ha demostrado que la regulación epigenética de la expresión génica no se limita a los procesos de desarrollo sino que también desempeña un papel fundamental en las neuronas maduras, lo que influye en una gran variedad de mecanismos básicos del cerebro adulto y proporciona nuevas pistas sobre cómo el entorno y las experiencias pueden interactuar con nuestro genoma. Del mismo modo, las modificaciones del ARN, como la m6A, pueden impulsar redes reguladoras postranscripcionales específicas de la región en el cerebro y contribuir a las enfermedades cerebrales (Chang *et al.*, 2017). La desregulación epigenética y epitranscriptómica parece ser una

característica de numerosos trastornos neurológicos y psiquiátricos que puede contribuir de forma importante a su etiología. Este es el caso de numerosos trastornos del neurodesarrollo asociados a la discapacidad intelectual y el autismo, enfermedades neurodegenerativas como el alzhéimer o la enfermedad de Huntington, y trastornos psiquiátricos como la drogadicción y la esquizofrenia. El difícil acceso al tejido cerebral y su extrema complejidad, tanto en términos de diversidad como de número de células, provoca desafíos específicos para la investigación en este ámbito, ya que la mayoría de los métodos actuales para detectar los epicambios se basan en laboriosos procedimientos bioquímicos o inmunológicos que requieren una cantidad significativa de células homogéneas.

**Epigenética de las enfermedades autoinmunes:** En los últimos años se ha observado un creciente interés y aprecio por el papel de la regulación epigenética en el sistema inmunitario sano y en la autoinmunidad. El desarrollo de la tecnología basada en matrices de metilación del ADN en el genoma completo y de la secuenciación de alto rendimiento ha permitido la evaluación de marcas epigenéticas específicas en todo el genoma en pacientes con una serie de enfermedades autoinmunes, así como la identificación y caracterización de regiones específicas dentro del genoma que están alteradas epigenéticamente en comparación con los controles sanos. Sin embargo, los conocimientos existentes no pueden explicar en su totalidad si las alteraciones epigenéticas causan o siguen al aumento de la activación inmunitaria, lo que hace que su caracterización precisa sea un requisito para una comprensión global de los mecanismos patogénéticos que complementen los estudios genéticos y clínicos. La integración de los datos de los estados de metilación del ADN específicos de la enfermedad y de las células, de las modificaciones de las histonas y de la actividad de los ARN no codificantes, además de los datos de la genómica y la transcriptómica, y la aplicación de metodologías novedosas como el RNA-seq de célula única o la edición génica proporcionarán una mejor imagen del papel de la epigenética en la etiología, el pronóstico y el tratamiento de las enfermedades autoinmunes.

**Epigenética de las enfermedades infecciosas:** Las enfermedades infecciosas humanas causadas por bacterias, virus, parásitos y hongos son las segundas más frecuentes y representan el 20 % de todas las enfermedades humanas. Las interacciones entre el patógeno y el huésped requieren una rápida adaptación y evolución. La supervivencia de los parásitos depende de la evasión del sistema inmunitario del huésped para asegurar su persistencia. A su

vez, el huésped debe desarrollar mecanismos de defensa para evitar la invasión y eliminar el microorganismo invasor (resistencia), o para limitar el daño causado por la infección (tolerancia). Los procesos epigenéticos, tanto en el huésped como en el patógeno, desempeñan un papel fundamental en la regulación de las interacciones entre el huésped y el patógeno durante la infección, así como en la capacidad de evolución y la rápida adaptación de los agentes infecciosos (véase también el capítulo 3.6). Una posibilidad intrigante es que la maquinaria epigenética de los patógenos intracelulares pueda alterar directamente el genoma del huésped (Sánchez-Romero y Casadesus, 2020). Una mejor comprensión de los mecanismos reguladores que controlan las variantes de los fenotipos de infección es esencial para prevenir la aparición y reaparición de enfermedades infecciosas y el fracaso de las intervenciones de control/erradicación existentes. Además, aunque el epitranscriptoma se ha investigado sobre todo en organismos eucariotas, las modificaciones del ARN también están presentes en el genoma de numerosos microorganismos. Por ejemplo, los genomas de ARN de muchos virus contienen numerosas modificaciones de ARN que influyen en su crecimiento e infectividad (Netzband y Pager, 2020). El reciente descubrimiento de las enzimas encargadas de añadir y eliminar estas modificaciones postranscripcionales, así como los grandes avances tecnológicos en la determinación y análisis del epitranscriptoma mediante inmunoprecipitación y secuenciación masiva, han permitido comenzar a conocer el efecto de estas modificaciones postranscripcionales del ARN en la patogénesis viral. La reciente prevalencia de las infecciones por virus de ARN aboga firmemente por la integración de los estudios epitranscriptómicos en este ámbito de investigación.

## 5.3 PUNTOS CLAVE DEL DESAFÍO

### 5.3.1 Desarrollo de métodos y herramientas para la detección rápida y cuantitativa de epicambios e interacciones ácido nucleico-proteína con precisión de un solo nucleótido

Los científicos ya disponen de un potente arsenal de técnicas para cartografiar las modificaciones en el genoma y el transcriptoma. La reducción de los costes y los ajustes que permiten el escalado descendente de estas técnicas a números de células bajos, en el rango de una a unos pocos miles de células, están posibilitando un progreso muy rápido en nuestra comprensión de los paisajes reguladores. Sin embargo, aún estamos lejos de completar el mapa de las marcas epigenéticas y, sobre todo, epitranscriptómicas que utilizan los distintos tipos de células de un organismo determinado. El progreso futuro requerirá técnicas

sofisticadas de alto rendimiento para ampliar y complementar las existentes. Las nuevas tecnologías de secuenciación de moléculas individuales directas y el desarrollo de anticuerpos específicos para diversas modificaciones del ARN pudieran permitir el trazado de marcas epitranscriptómicas específicas de los transcritos en todos los tipos de células. Además, para desvelar el *código epitranscriptómico* y su importancia para la salud y la enfermedad humanas, es necesario descubrir la interacción reguladora combinatoria entre las diferentes modificaciones del ARN. Asimismo, será de suma importancia el desarrollo de metodologías que puedan permitir la determinación inequívoca de las epimodificaciones en un tipo celular específico, con independencia del tejido de origen. El objetivo principal debería ser alcanzar la capacidad de detectar una modificación concreta, en una célula específica dentro de la resolución del gen/locus o incluso del nucleótido. Igualmente importante es el desarrollo de métodos innovadores y tecnologías de secuenciación capaces de detectar a la vez diferentes marcas epigenéticas en un mismo locus o marcas epitranscriptómicas dentro de un mismo transcrito. Entre estos avances, la descripción detallada de las alteraciones del paisaje de la cromatina y de la estructura global tridimensional de la cromatina parece esencial para comprender la contribución de la cromatina a la regulación de los genes y al establecimiento de enfermedades. La traducción de las firmas epigenéticas específicas de los tejidos a la arquitectura tridimensional de la cromatina representa una nueva y prometedora línea de investigación (véase también el capítulo 3.3 para obtener más detalles). Se necesitan mapas completos de alta resolución para ordenar las interacciones entre los perfiles epigenéticos y los *loci* de rasgos cuantitativos o las variantes genéticas asociadas a la enfermedad, ya que algunas de ellas están relacionadas con polimorfismos de un solo nucleótido que afectan a los sitios de metilación CpG.

### **5.3.2 Comprensión de las modificaciones y la dinámica del ADN y el ARN en todo el sistema**

Las nuevas técnicas descritas anteriormente tienen el potencial de generar una cantidad de datos masiva y difícil de manejar. Así pues, nos enfrentamos al enorme desafío de integrar los nuevos datos epigenómicos y epitranscriptómicos con los resultados genéticos, transcriptómicos y proteómicos, las asociaciones genómicas y las interacciones tridimensionales de la cromatina en el contexto específico del tipo de célula, del estímulo y de la localización con el fin de proporcionar una mejor imagen de sus implicaciones funcionales. La disponibilidad de perfiles genómicos y epitranscriptómicos para el repertorio completo de proteínas escritoras, lectoras y borradoras y sus sustratos

debiera conducir a una comprensión exhaustiva de la regulación y la dinámica de la expresión génica. Para desvelar la interrelación entre las modificaciones de las histonas, el ADN y el ARN se necesitarán soluciones informáticas estandarizadas que permitan identificar, analizar e integrar las modificaciones del ARN de alto rendimiento con los datos epigenómicos. Será preciso desarrollar nuevos análisis bioinformáticos integradores de datos con una mayor potencia informática.

### **5.3.3 Mapeo del epigenoma y del epitranscriptoma en cuatro dimensiones**

Tanto el epigenoma como el epitranscriptoma son dinámicos por naturaleza y, por tanto, tendremos que comprender los mecanismos que impulsan su cambio durante el desarrollo, la experiencia vital, la interacción con el entorno y el envejecimiento. Es fundamental determinar las alteraciones transcripcionales y epigenéticas que se producen durante el desarrollo normal porque esta información nos ayudará a determinar la ventana espacio-temporal en la que la deficiencia de factores epigenéticos o epitranscriptómicos contribuye más al desarrollo de la enfermedad, lo que facilita las intervenciones terapéuticas prematuras. Esto debería dictar un diagnóstico basado no solo en los síntomas sino también en las características moleculares, incluidos los patrones alterados de metilación del ADN, las histonas o las modificaciones del ARN. Estos conocimientos y la ampliación de los cribados genómicos —actualmente limitados en gran medida al exoma— al genoma no codificante debieran ayudar a dilucidar los mecanismos moleculares subyacentes a un gran número de enfermedades poco frecuentes no diagnosticadas. Es más, una comprensión precisa de la dinámica de las epimodificaciones debiera aclarar también el tipo de rasgos e información que pueden transmitirse a la descendencia y afectar a las generaciones futuras.

### **5.3.4 Resolución de los mecanismos epigenéticos y epitranscriptómicos de la etiopatología**

Existe una gran necesidad de elaborar un catálogo completo de variaciones epigenómicas y epitranscriptómicas asociadas a las enfermedades humanas en general. Estas pudieran servir como biomarcadores de la actividad de la enfermedad o de su evolución. Hasta la fecha, la mayoría de los estudios de asociación del epigenoma completo se han basado en el uso de *microarrays* para identificar los sitios de metilación CpG y, en este sentido, una proporción muy grande del epigenoma no ha sido aún descubierta. Además, muchos de los posibles mecanismos epigenéticos permanecen inexplorados y su investigación puede

llevarnos en direcciones inesperadas. Por su parte, los recientes avances en la investigación epitranscriptómica asocian de forma clara las alteraciones de las enzimas modificadoras del ARN con la aparición de enfermedades como el cáncer, lo que demuestra que la inhibición de esas enzimas puede tener un enorme potencial terapéutico. Pero los estudios de asociación de todo el epitranscriptoma siguen sin descubrirse y cientos de posibles mecanismos epitranscriptómicos sin explorarse. Una vez que completemos el catálogo de epimodificaciones vinculadas a las enfermedades, todavía tendremos que determinar su contribución específica al estado de la enfermedad, así como la diafonía entre la genética y la epigenética, y entre la transcriptómica y la epitranscriptómica. Este enorme desafío debe llevarse a cabo, preferentemente, en los estadios iniciales de la enfermedad, ya que cualquier alteración epigenética observada en un tejido autopsiado puede reflejar (en concreto en enfermedades de larga duración como las neurodegenerativas) cambios relacionados con el avance de la enfermedad, lo que hace bastante difícil diferenciar entre causa y consecuencia. En el campo de la epigenética del cáncer, todavía no sabemos si la decisión de iniciar el cáncer tiene lugar durante la diferenciación del tumor o si está compuesta por una serie de decisiones consecutivas. Los futuros avances en la prevención del cáncer pueden depender de la identificación de exposiciones específicas como desencadenantes del iniciado epigenético y de la protección de los individuos susceptibles de exponerse a factores ambientales que pueden desencadenar el cáncer (por ejemplo, las infecciones en niños portadores de genes susceptibles). También habría que dedicar esfuerzos a estudiar los cambios en el epigenoma y el epitranscriptoma de la célula huésped inducidos por la infección del patógeno para influir en las respuestas del huésped y contribuir a otras formas de enfermedad humana.

### 5.3.5 Comprensión de los trastornos a nivel unicelular

La llegada de las estrategias de célula única ofrece una oportunidad óptima para obtener información sobre los mecanismos que subyacen a la identidad celular, el fenotipo y la respuesta a los estímulos, los factores de estrés y los patógenos (Avraham *et al.*, 2015). La naturaleza individual de los cambios epigenéticos, el hecho de que afecten de forma diferencial a las células de un tejido, es uno de los principales obstáculos a la hora de estudiar la influencia o el papel de los cambios epigenéticos en la etiopatología, sobre todo en el caso de las enfermedades cerebrales y otros tejidos complejos. Mientras que en el análisis genético no es imprescindible el uso de tejidos o células afectadas por determinadas enfermedades, ya que la mayor parte de la variación genética causante de una enfermedad se encuentra en todas las células, esto no puede aplicarse a las



enfermedades epigenéticas. **Las marcas epigenéticas cambian según el tipo de célula y, por tanto, su estudio requiere el análisis específico de esas células.** El continuo perfeccionamiento y el nuevo desarrollo de técnicas que abarcan el genoma completo para explorar el epigenoma, el transcriptoma y el epitranscriptoma a nivel unicelular debieran resolver los problemas causados por la diversidad celular y reducir la cantidad de tejido necesaria (véase también el capítulo 3.1 para obtener más detalles sobre los métodos de célula única). Estas técnicas permitirán analizar los cambios epigenéticos, epitranscriptómicos y de expresión génica en poblaciones celulares restringidas. Esto puede iniciar una era de nuevos descubrimientos que tienen el potencial de cambiar radicalmente nuestra comprensión de las enfermedades, en particular las que afectan a tejidos complejos y heterogéneos como el cerebro y los tejidos sometidos a una infección patógena. Este conocimiento es fundamental, por ejemplo, para comprender las estrategias de variación de fase y de apuesta diversificada implicadas en la resistencia a los antibióticos y la evasión inmunológica, y para anticipar la rápida adaptación de los patógenos a los nuevos medicamentos y vacunas (Chattopadhyay *et al.*, 2018). Esto también debiera aclarar la sorprendente especificidad del tipo de neurona de la mayoría de las afecciones neurodegenerativas y las diferencias de éxito en el tratamiento de los distintos tipos de cáncer. La tecnología de análisis de células individuales también será un buen complemento al arsenal de herramientas de investigación para estudiar los trastornos metabólicos.

### 5.3.6 Inferencia de la causalidad de las epimodificaciones

En la última década, hemos podido dibujar mapas epigenómicos y epitranscriptómicos de gran alcance y hemos aprendido cómo unos pocos actores (principalmente escritores, lectores y borradores) perturban estos sistemas. Aunque proporcionan información posicional útil y pistas sobre los mecanismos moleculares en juego, es necesario invertir mucho esfuerzo en el desarrollo de ensayos funcionales que incorporen los datos recogidos mediante los métodos globales mencionados para comprender los efectos de la compleja interacción de estas modificaciones. En muchos trastornos se han notificado cambios en la metilación del ADN, en las modificaciones postraduccionales de las histonas (HPTM) y otros cambios en la cromatina, que a menudo se correlacionan con la progresión de la enfermedad. Sin embargo, se desconoce si estas alteraciones epigenéticas son causa o consecuencia, incluso indirecta, de la patología. La capacidad de editar específicamente el epigenoma y el epitranscriptoma promete mejorar nuestra comprensión del funcionamiento de las epimodificaciones y permitir la manipulación del fenotipo celular con fines terapéuticos. La



reciente revolución en las tecnologías de ingeniería genómica ha permitido el uso de herramientas altamente específicas de ADN y ARN para incorporar con precisión cambios epigenéticos o editar secuencias de ARN de forma específica para cada locus, lo que crea diversas plataformas de edición del epigenoma y el epitranscriptoma. Por ejemplo, la tecnología CRISPR (del inglés *Clustered Regularly Interspaced Short Palindromic Repeats*)/Cas9 se ha adaptado a la edición epigenética mediante la creación de proteínas quiméricas entre una nucleasa Cas9 muerta (dCas9) con dominios catalíticos responsables de la modificación de la cromatina. Dado que los ARN guía son fáciles de diseñar y pueden dirigir la actividad catalítica a prácticamente cualquier región del genoma, uno puede, en principio, alterar de forma local el perfil epigenético en cualquier locus de diferentes maneras según la actividad enzimática acoplada a dCas9. Es importante destacar que estas nuevas herramientas genéticas (a diferencia de la edición del genoma mediante el sistema CRISPR/Cas9 convencional) son igual de eficaces en las células en división y en las que no lo están, y proporcionan medios sin precedentes para aumentar o disminuir la expresión de cualquier gen de interés en tipos de células potencialmente resistentes a los enfoques de edición génica. Se trata de un área que aún está en desarrollo y en la que cada vez hay más herramientas novedosas y sofisticadas para la edición. Un mayor desarrollo de las tecnologías actuales basadas en el sistema CRISPR/dCas9 para manipular con precisión el epigenoma y la innovación de tecnologías pioneras para la manipulación del epitranscriptoma deberían permitirnos abordar el enigma de la causalidad para la mayoría de las epimodificaciones en un futuro próximo (Voigt y Reinberg, 2013).

### 5.3.7 Nuevas epiterapias

En la última década, los químicos médicos han producido un conjunto sin precedentes de pequeñas moléculas que se dirigen a las proteínas responsables de la escritura, lectura o borrado de las marcas epigenéticas en la cromatina. Varias de estas terapias epigenéticas ya han llegado a los hospitales para combatir el cáncer, y muchas otras han avanzado hasta las primeras fases de los ensayos clínicos en una gran cantidad de enfermedades. El uso de fármacos epigenéticos en combinación con otras terapias ha abierto nuevas vías muy prometedoras para combatir la enfermedad (Michalak *et al.*, 2019). Por ejemplo, la existencia de un mecanismo de iniciación tumoral impulsado por la epigenética abre nuevas posibilidades para prevenir o suprimir el cáncer, ya que las modificaciones epigenéticas, a diferencia de los cambios genéticos, pueden borrarse, manipularse y reiniciarse incluso antes de que una célula precancerosa pueda evolucionar hacia el cáncer.

Una mayor investigación de los compuestos dirigidos a los modificadores del ARN relacionados con la enfermedad y la interacción entre el epitranscriptoma y el epigenoma identificará sin duda nuevos objetivos terapéuticos. La comprensión de las maquinarias y factores que introducen, eliminan y leen las modificaciones de la cromatina y el ARN permitirá su modulación mediante el desarrollo de nuevos fármacos con valor farmacéutico. También es urgente ampliar los métodos de administración de estos epifármacos; las estrategias basadas en la nanotecnología facilitarán esta tarea. Además de los enfoques farmacológicos, también existe un gran interés en explorar la posibilidad de corregir directamente las alteraciones epigenéticas mediante los métodos de edición mencionados con anterioridad (Hilton *et al.*, 2015; Kwon *et al.*, 2017). Aunque el uso de esta tecnología incipiente en órganos complejos, como el cerebro, se enfrenta a los destacados desafíos asociados a la terapia génica en estos órganos (es decir, bioseguridad, especificidad celular, accesibilidad al tejido enfermo, etc.), sigue representando un área importante de desarrollo que puede permitir enfoques terapéuticos personalizados para corregir alteraciones de la cromatina.

### 5.3.8 Epigenómica social y epidemiología

Resulta urgente emprender estudios poblacionales de gran envergadura para caracterizar el paisaje epigenético y epitranscriptómico en situaciones de partida, y cómo evolucionan estos paisajes con el proceso de envejecimiento y la influencia del entorno (véanse también los capítulos 3.6 y 3.7). También es muy necesario comprender cómo afectan las diferentes patologías a estos paisajes e incorporar la heterogeneidad de la enfermedad en el diseño del estudio. Las marcas epigenéticas pueden prever la aparición de enfermedades. Esto debiera revelar importantes indicios sobre la susceptibilidad a las enfermedades en las distintas poblaciones y la mejora del entorno, lo que permitiría mejorar la salud pública y la equidad social. La identificación de epibiomarcadores que pudieran predecir la evolución de la enfermedad y la respuesta al tratamiento es más difícil de conseguir que las mutaciones genéticas clásicas. En primer lugar, porque pueden ser específicos de un tejido; en segundo lugar, porque incluso con aproximaciones por PCR de alta sensibilidad, es todavía un reto técnico identificar cambios en el perfil epigenético de un locus determinado o en el epitranscriptoma derivado de unas pocas células. La herencia transgeneracional en mamíferos se entiende poco todavía, pero si las investigaciones futuras demuestran un impacto más amplio que el anticipado, tal conocimiento sería importante para proteger a las generaciones sucesivas.

## CAPÍTULO 5 BIBLIOGRAFÍA

- Avraham, R., Haseley, N., Brown, D., Penaranda, C. *et al.* (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell* 162, 1309-1321.
- Barbieri, I. y Kouzarides, T. (2020). Role of RNA modifications in cancer. *Nat. Rev. Cancer* 20(6), 303-322.
- Bjornsson, H.T. (2015). The Mendelian disorders of the epigenetic machinery. *Genome Res.* 25, 1473-1481.
- Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P. *et al.* (2018). MODOMICS: a database of RNA modification pathways. Actualización de 2017. *Nucleic Acids Res.* 46, D303-D307.
- Cavalli, G. y Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature* 571, 489-499.
- Chang, M., Lv, H., Zhang, W., Ma, C. *et al.* (2017). Region-specific RNA m(6)A methylation represents a new layer of control in the gene regulatory network in the mouse brain. *Open Biol.* 7(9), 170166.
- Chattopadhyay, P.K., Roederer, M. y Bolton, D.L. (2018). A deadly dance: the choreography of host-pathogen interactions, as revealed by single-cell technologies. *Nat. Commun.* 9, 4638.
- Davalos, V., Blanco, S. y Esteller, M. (2018). SnapShot: Messenger RNA Modifications. *Cell* 174, 498-498.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M. *et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201-206.
- Frye, M., Harada, B.T., Behm, M. y He, C. (2018). RNA modifications modulate gene expression during development. *Science* 361(6409), 1346-1349.
- Harries, L.W. (2019). RNA Biology Provides New Therapeutic Targets for Human Disease. *Front Genet.* 10, 205.
- Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I. *et al.* (2015). Epigenome editing by a CRISPR-Cas9based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510-517.
- Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in humans. *Nat. Commun.* 9, 2973.
- Jia, G., Fu, Y., Zhao, X., Dai, Q. *et al.* (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885-887.
- Kwon, D.Y., Zhao, Y.T., Lamonica, J.M. y Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* 8, 15315.
- Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O. *et al.* (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635-1646.
- Michalak, E.M., Burr, M.L., Bannister, A.J. y Dawson, M.A. (2019). The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat. Rev. Mol. Cell Biol.* 20, 573-589.
- Netzbund, R. y Pager, C.T. (2020). Epitranscriptomic marks: Emerging modulators of RNA virus gene expression. *Wiley Interdiscip. Rev. RNA* 11, e1576.
- Sales, V.M., Ferguson-Smith, A.C. y Patti, M.E. (2017). Epigenetic Mechanisms of Transmission of Metabolic Disease across Generations. *Cell Metab.* 25, 559-571.
- Sánchez-Romero, M.A. y Casadesus, J. (2020). The bacterial epigenome. *Nat. Rev. Microbiol.* 18, 7-20.
- Sen, P., Shah, P.P., Nativio, R. y Berger, S.L. (2016). Epigenetic Mechanisms of Longevity and Aging. *Cell* 166, 822-839.
- The Lancet Diabetes Endocrinology (2019). Spotlight on rare diseases. *Lancet Diabetes Endocrinol.* 7, 75.
- Tomasetti, C. y Vogelstein, B. (2015). Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347(6217), 78-81.
- Velasco, G. y Francastel, C. (2019). Genetics meets DNA methylation in rare diseases. *Clin. Genet.* 95, 210-220.
- Vicente Dueñas, C., Hauer, J., Cobaleda, C., Borkhardt, A. *et al.* (2018). Epigenetic Priming in Cancer Initiation. *Trends Cancer* 4, 408-417.
- Voigt, P. y Reinberg, D. (2013). Epigenome editing. *Nat. Biotechnol.* 31, 1097-1099.

Volkov, P., Bacos, K., Ofori, J.K., Esguerra, J.L. *et al.* (2017). Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. *Diabetes* 66, 1074-1085.

Wahl, D., Coogan, S.C., Solon-Biet, S.M., de Cabo, R. *et al.* (2017). Cognitive and behavioral evaluation of nutritional interventions in rodent models of brain aging and dementia. *Clin. Interv. Aging* 12, 1419-1428.

### RESUMEN

La contaminación ambiental y el cambio climático están influyendo mucho en todas las formas de vida de nuestro planeta. Resulta necesario comprender cómo la exposición a los contaminantes altera el genoma, el epigenoma y la microbiota en los seres humanos, los animales y las plantas, y cómo las poblaciones microbianas presentes en la naturaleza responden a las fluctuaciones medioambientales. La genómica, la epigenética y la metagenómica integradas pueden contribuir al desarrollo de soluciones agrícolas y ganaderas respetuosas con el medioambiente, así como a nuevos usos biotecnológicos basados en los microbios.

### PALABRAS CLAVE

medioambiente	contaminantes	
microbiota	metagenómica	agricultura
ganado	exposoma	

# GENÓMICA Y EPIGENÓMICA MEDIOAMBIENTAL

## Coordinadores

Antonia Herrero  
(IBVF, Sevilla, coordinadora)  
Lourdes Ramos  
(IQOG, Madrid, coordinadora adjunta)

## Investigadores y centros de investigación participantes (por orden alfabético)

Miguel A. Bañares  
(ICP, Madrid)  
Myriam Calonje  
(IBVF, Sevilla)  
M.<sup>a</sup> Carmen Collado  
(IATA, Valencia)  
Gustavo Gómez  
(I2SysBio, Valencia)  
Joan Grimalt  
(IDAEA, Barcelona)  
Paloma Mas  
(CRAG, Barcelona)  
José M. Pardo  
(IBVF, Sevilla)  
Laia Ribas  
(ICM, Barcelona)  
Federico Valverde  
(IBVF, Sevilla)

## RESUMEN EJECUTIVO

La contaminación ambiental y el cambio climático se han convertido en una de las amenazas más graves para los seres humanos y otras formas de vida del planeta. Es importante entender cómo estos factores influirán en la vida de nuestro planeta. El objetivo de este desafío consiste en desentrañar cómo el medioambiente interactúa con el genoma y el epigenoma para dar forma a la fisiología, el desarrollo y la patología de los seres humanos, los animales y las plantas, y cómo los cambios medioambientales afectan a la evolución de las comunidades microbianas en la naturaleza.

Una necesidad crucial estriba en comprender cómo la exposición a los contaminantes (sustancias químicas y nanomateriales) y los cambios medioambientales influyen en el genoma y el epigenoma durante toda la vida de un organismo. Asimismo, resulta importante aclarar cómo la firma genética, incluido el epigenoma, determina la susceptibilidad de cada individuo, así como la herencia epigenética transgeneracional, en respuesta a estas amenazas. Entender cómo las fluctuaciones de las condiciones medioambientales se traducen en cambios genómicos y epigenómicos mejorará nuestros conocimientos sobre

los mecanismos reguladores que controlan la fisiología y el desarrollo humano, animal y vegetal. También hay una necesidad emergente de encontrar nuevas soluciones para una agricultura y una ganadería respetuosas con el medioambiente que debería abordarse con una mejor comprensión de la epigenética como fuente de variabilidad fenotípica y adaptabilidad, y la genómica integradora para el descubrimiento de nuevos rasgos de interés agronómico. Las poblaciones microbianas desempeñan un papel fundamental en la lucha contra la contaminación ambiental y en la regulación del clima mundial. La metagenómica puede proporcionar un conocimiento profundo de las adaptaciones de las comunidades microbianas a los entornos cambiantes, incluidas las perturbaciones impuestas por la actividad humana, lo que permitirá prevenir efectos desastrosos y diseñar estrategias de desarrollo sostenible. Hay que prestar especial atención a los microorganismos oceánicos que contribuyen de manera global a cerca de la mitad de la producción primaria total terrestre y tienen un gran impacto en los ciclos geoquímicos globales y en el clima de la Tierra. La metagenómica en entornos naturales debe incluir el complemento viral, que se ha dado a conocer como un factor decisivo en la dinámica de las poblaciones microbianas. Por otra parte, la comprensión de la interrelación entre la microbiota y el genoma y epigenoma del huésped contribuirá a fomentar las interacciones beneficiosas entre especies y a proteger a los seres humanos, los animales y las plantas contra las enfermedades. Finalmente, la metagenómica funcional y la epigenómica debieran proporcionar un repertorio mundial de capacidades metabólicas y biomarcadores, así como guiar el diseño de aplicaciones médicas y biotecnológicas. Las relaciones emergentes entre microbios, seres humanos, animales y plantas que se derivan de los intensos movimientos geográficos en un mundo conectado de manera global constituyen un importante desafío, como se ha observado con motivo de las últimas pandemias.

## 6.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

La contaminación ambiental y el cambio climático amenazan todas las formas de vida de nuestro planeta, incluida la nuestra. Por ello, resulta fundamental estudiar y comprender la influencia que tienen estos y otros factores ambientales en los organismos vivos con el fin de minimizar y prevenir efectos perjudiciales. Más concretamente, aquí nos centraremos en las interacciones que los factores ambientales suelen tener con el genoma y el epigenoma de plantas y animales, incluidos los seres humanos, así como con las poblaciones microbianas de la naturaleza.

Los cambios ambientales en general y la exposición a los contaminantes en particular pueden dañar el genoma o el epigenoma de los organismos vivos. Es más, diferentes individuos pertenecientes a la misma especie pueden manifestar distintas susceptibilidades a los daños tras la exposición a los mismos factores ambientales en función de su información genética o epigenética. En este sentido, resulta primordial entender las interacciones mutuas que se establecen entre el medioambiente y el (epi)genoma, ya que esto puede tener un gran impacto en nuestra actual comprensión del desarrollo, la fisiología y la enfermedad.

El cambio climático exige formas de agricultura y ganadería nuevas y respetuosas con el medioambiente para satisfacer las necesidades de una población humana que no deja de aumentar. Para lograr estos objetivos será importante utilizar enfoques genómicos integradores con los que descubrir y comprender nuevos rasgos de interés agronómico. Del mismo modo, se podrán aprovechar los mecanismos epigenéticos como fuente de variabilidad fenotípica y adaptabilidad con capacidad de ajuste y orientación. Con el fin de reducir la contaminación ambiental y el cambio climático, es imprescindible conocer a fondo las poblaciones microbianas. Una de las principales estrategias para lograrlo se centra en la metagenómica, que puede utilizarse para identificar nuevos organismos, así como la variabilidad de las comunidades microbianas en respuesta a las perturbaciones ambientales, en particular las causadas por el hombre. La aplicación de estos enfoques metagenómicos al estudio de los microorganismos oceánicos será particularmente significativa, ya que siguen estando bastante inexplorados a pesar de su contribución clave a la producción primaria total y a la regulación del clima en nuestro planeta.

Por último, las aproximaciones genómicas y epigenómicas resultarán esenciales para comprender con todo detalle las interacciones que se establecen de manera constante entre los microorganismos y sus huéspedes (plantas y



animales) y que pueden tener consecuencias beneficiosas o patológicas. Esto tiene especial repercusión si tenemos en cuenta el mundo conectado de manera global en el que vivimos y que puede dar lugar a nuevas interacciones no deseadas entre los microbios y otros organismos, como por desgracia ponen de manifiesto las últimas pandemias.

## 6.2 IMPACTO EN EL PANORAMA DE LA CIENCIA BÁSICA Y EN LAS POSIBLES APLICACIONES

### 6.2.1 Contaminantes ambientales y herencia transgeneracional epigenética

La contaminación ambiental solo podría contrarrestarse basándose en un conocimiento detallado de las consecuencias de las perturbaciones del hábitat en la vida de los seres humanos y otros seres vivos. El número de productos químicos y nanomateriales sintetizados por el ser humano durante el último siglo es enorme. Muchos de ellos han contribuido a mejorar nuestra calidad de vida lo que ha reducido el número de enfermedades o facilitado el desarrollo de nuevas tecnologías y procesos industriales. Sin embargo, también se ha demostrado que un número importante de estos productos son nocivos para el medioambiente y los seres humanos debido a su naturaleza recalcitrante y tóxica. La exposición accidental a altas concentraciones de compuestos y nanomateriales especialmente tóxicos (por ejemplo, di-benzo-p-dioxinas policloradas, furanos, ZnO, CeO<sub>2</sub>) provocó graves daños en la población y evidenció efectos transgeneracionales debido a su carácter teratogénico y mutagénico. Además, la exposición crónica a compuestos con actividad de disrupción endocrina puede alterar la función o funciones del sistema endocrino y, en consecuencia, provocar efectos adversos para la salud en un organismo intacto, su progenie o (sub)poblaciones. Estas evidencias condujeron a la evaluación toxicológica de estos y otros compuestos relacionados químicamente, incluidos sus usos y niveles máximos permitidos en una variedad de matrices (por ejemplo, aire, suelo, agua o alimentos). Aunque la legislación y los acuerdos nacionales e internacionales han regulado el uso de algunos de estos compuestos con el fin de proteger la salud humana y el medioambiente, estas medidas son sin duda insuficientes para prevenir las vías de resultados metabólicos adversos generados por la exposición química. La Comisión Europea está realizando un gran esfuerzo para evaluar los efectos adversos de los nanomateriales y establecer la gobernanza adecuada. A pesar del indudable valor de estas normativas y de sus programas de control asociados, las consecuencias de la exposición

crónica a niveles residuales de estas y otras muchas sustancias químicas y nanomateriales aún no caracterizados en uso siguen siendo básicamente desconocidas. Ahora bien, estudios recientes con animales de laboratorio han señalado una posible relación entre la exposición a sustancias tóxicas ambientales, como pesticidas y componentes plásticos, y fenotipos reproductivos o metabólicos anómalos que se transmiten transgeneracionalmente (Ost *et al.*, 2014; Nilsson *et al.*, 2012). Estos nuevos conocimientos plantean un llamamiento urgente a la investigación sobre cómo la exposición a contaminantes ambientales y otros factores de estrés puede inducir cambios epigenómicos, tanto en los seres humanos como en los animales de granja, que se transmiten a las siguientes generaciones y que se asocian con fenotipos de enfermedades (véase también el capítulo 3.7 para conocer otros ejemplos de herencia transgeneracional epigenética).

### 6.2.2 Caracterización genómica de las comunidades microbianas

Los microbios de la naturaleza, y en particular de los océanos, a menudo pasan desapercibidos en nuestra visión antropocéntrica del mundo y desempeñan un papel crucial para hacer de nuestro planeta un lugar habitable (Falkowski *et al.*, 1998). Las picocianobacterias oceánicas como *Prochlorococcus* y *Synechococcus* son los organismos fotosintéticos más abundantes y los principales productores primarios de nuestro planeta (Farrant *et al.*, 2016), mientras que el *Pelagibacter* heterótrofo (SAR11) es el microorganismo más abundante en los océanos y, probablemente, en el planeta. Algunos microbios fototróficos del fitoplancton pueden fijar el nitrógeno atmosférico —que vive libremente o en simbiosis—, como el formado por diatomeas como huéspedes y cianobacterias como simbiontes (asociaciones diazotróficas de diatomeas, DDA por sus siglas en inglés) (Foster *et al.*, 2011; Karl *et al.*, 2016). Se distribuyen por todo el mundo y realizan la tarea crucial de reponer el nitrógeno en la biosfera de una forma que otros seres vivos puedan utilizar, lo que facilita la fijación del CO<sub>2</sub>. Así, estos microorganismos se consideran actores principales en los ciclos biogeoquímicos del carbono, nitrógeno, oxígeno, fósforo y hierro en nuestro planeta, por lo que amplias alteraciones en la dinámica de sus poblaciones tendrían un grave impacto en las cadenas tróficas de los ecosistemas marinos, con consecuencias para todos los seres vivos del planeta.

La actividad humana y los cambios ambientales asociados pueden tener un profundo impacto en las comunidades microbianas de la naturaleza, así como en sus actividades esenciales. La metagenómica, que consiste en la secuenciación

masiva del ADN de muestras naturales y la posterior reconstrucción cromosómica, puede proporcionar una visión completa de la estructura y la dinámica de la población microbiana a pesar de nuestra capacidad de cultivo de sus miembros. Este es un punto crucial si tenemos en cuenta las estimaciones de que solo se ha cultivado alrededor del 1 % de los microbios que habitan nuestro planeta. Desde la contribución pionera de Venter y colaboradores (Venter *et al.*, 2004) basada en la secuenciación rápida de genomas completos en muestras de agua del mar de los Sargazos, diversas iniciativas de colaboración, como la de TARA Oceans (<https://oceans.taraexpeditions.org/en/>) (Bork *et al.*, 2015) y la expedición Malaspina (<http://www.expedicionmalaspina.es/>), se han dedicado al estudio de la biodiversidad marina y de los efectos que el cambio climático mundial está teniendo sobre esta diversidad. Estos esfuerzos han proporcionado una gran cantidad de datos metagenómicos que han respaldado un análisis sin precedentes de la estructura de las comunidades microbianas de zonas oceánicas específicas. Asimismo, un número cada vez mayor de proyectos se ha dirigido al análisis de la estructura de las comunidades microbianas en otros nichos ambientales específicos y, en algunos casos, a la detección de vías metabólicas concretas o actividades de productos genéticos (por ejemplo, Dietrich *et al.*, 2019). Conocer en profundidad las comunidades microbianas y las características fisicoquímicas de su entorno ayudaría a predecir los efectos de las perturbaciones impuestas por la actividad humana, a prevenir impactos desastrosos y a diseñar comunidades controladas para usos biotecnológicos rentables (Duarte *et al.*, 2020).

### 6.2.3 Interacciones entre el microbioma y el (epi)genoma del huésped

Las comunidades microbianas habitan en casi todos los nichos terrestres, desempeñan funciones cruciales para contrarrestar la contaminación ambiental y pueden influir de un modo general en la salud y la enfermedad humanas. El huésped y las comunidades microbianas mantienen una relación íntima clave que les beneficia mutuamente. De ahí que las comunidades microbianas reciban una fuente continua de nutrientes, mientras que el huésped obtiene una gran variedad de metabolitos procedentes de la digestión bacteriana, la protección contra patógenos y virus y la educación del sistema inmunitario, entre otras funciones beneficiosas. Por eso, se considera a los interactores huésped-microbiota como una única unidad evolutiva y biológica, el holobionte, que representa un campo de gran relevancia en la biología y las ciencias médicas (Simon *et al.*, 2019). Además, la composición de la comunidad es más similar dentro de un mismo ambiente que entre ambientes distintos, y la diferencia interpersonal

dentro de los hábitats es mayor que la variabilidad intraindividual con el paso del tiempo (Näpflin *et al.*, 2019). La complejidad de la comunidad microbiana depende del hábitat concreto, y solo los microorganismos seleccionados podrán sobrevivir y colonizar en las condiciones características de cada hábitat. Lo curioso es que, aunque la microbiota asociada al huésped puede adquirirse del entorno circundante, la composición de las comunidades microbianas varía mucho de la de los microorganismos comunes que viven libremente. Además, la composición microbiana la conforman diferentes factores como la genética del huésped y el medioambiente, incluidos la temperatura, la contaminación atmosférica, los xenobióticos y los recursos nutritivos. De hecho, el efecto del exposoma sigue siendo en gran medida inexplorado y la interacción bidireccional entre la exposición a sustancias químicas y nanomateriales, las comunidades microbianas y sus huéspedes apenas empieza a considerarse un factor relevante que, entre otras cosas, puede configurar la salud y la enfermedad humanas.

Los principales objetivos de los estudios sobre la microbiota se centran en discernir cómo la composición, la diversidad y las funciones de los microorganismos que la componen influyen y regulan la fisiología del huésped y su asociación con la salud y la enfermedad. Al entender esto, podremos aprender a manipular la composición del microbioma y las actividades metabólicas de este, al tiempo que maximizamos los beneficios para la salud del huésped. Los avances en esta área de investigación han sido posibles gracias al desarrollo de diferentes tecnologías genómicas independientes de los cultivos, basadas en la secuenciación masiva, en combinación con innovadores enfoques bioinformáticos y de biología de sistemas. Por otra parte, la comprensión de la interrelación entre la microbiota y el genoma y el epigenoma de sus huéspedes contribuirá a fomentar las interacciones beneficiosas entre especies y a proteger a los seres humanos, los animales y las plantas contra las enfermedades. La evidencia acumulada demuestra que la microbiota regula el epigenoma del huésped a través de señales microbianas específicas que incluyen metabolitos, ácidos biliares y otros compuestos (Sironi *et al.*, 2015). La comprensión de las complejas interacciones entre la microbiota, los factores ambientales y el epigenoma del huésped, incluida la metilación del ADN, las modificaciones de las histonas y los ARN no codificantes, supone un reto científico muy importante.

#### **6.2.4 La genómica y la epigenómica en el desarrollo de nuevos enfoques agrícolas y ganaderos**

Los seres humanos dependen de la agricultura y la ganadería para su aporte calórico diario, mientras que las plantas y los animales se ven cada vez más

amenazados por el entorno. A medida que aumenta la población mundial, la seguridad alimentaria se ve cuestionada por las limitadas y siempre deterioradas superficies de tierra cultivable, que afectan a la productividad y a la calidad de los productos. Esta situación se ve agravada por el aumento del uso de tierras de cultivo para alimentar a los animales debido a la creciente demanda de productos ganaderos, efectos que posiblemente se verán multiplicados por las devastadoras consecuencias del calentamiento global. En este sentido, la obtención de cosechas y animales resilientes capaces de adaptarse a entornos extremos y el desarrollo de enfoques sostenibles para aumentar el rendimiento se han convertido en un desafío urgente en el que la genómica y la epigenómica desempeñarán un papel fundamental. Los conocimientos derivados de estas disciplinas fomentarán los programas de cría y desvelarán objetivos clave de manipulación para explotar cosechas y cultivos animales modificados a la carta.

## 6.3 PUNTOS CLAVE DEL DESAFÍO

### 6.3.1. Exposoma: efecto en la salud humana de la exposición a factores ambientales nocivos

Desde finales del siglo pasado, la incidencia de ciertas enfermedades no infecciosas ha aumentado en las poblaciones humanas de todo el mundo. Algunas de estas enfermedades, como la obesidad, el síndrome de ovario poliquístico (SOP) o la infertilidad masculina, se han asociado en modelos animales con la exposición a factores estresantes ambientales específicos (en su mayoría contaminantes tóxicos) y se han reconocido como heredadas transgeneracionalmente (Guerrero-Bosagna y Jensen, 2015). Es muy probable que esta herencia transgeneracional lleve asociados mecanismos epigenéticos que aún no se conocen bien y que deberían investigarse a fondo en los próximos años (véase también el capítulo 3.7). Estos hallazgos han suscitado la preocupación por los posibles efectos para los seres humanos —y en particular para las generaciones futuras— de la exposición crónica a niveles residuales de mezclas complejas de contaminantes ambientales. Factores como la agricultura intensiva y el desarrollo industrial, pero también la globalización y el cambio climático, contribuyen a un aumento constante del número de nanomateriales y sustancias químicas en uso. Incluso hoy en día, y a pesar de la gran capacidad de las técnicas analíticas más modernas, desentrañar la composición de las complejas mezclas de contaminantes orgánicos presentes en la mayoría de las muestras ambientales y humanas es una tarea difícil que solo pueden abordar con éxito equipos de investigación experimentados y

con instrumentación avanzada. Se aplican consideraciones similares a la evaluación de la toxicidad y el destino de los nanomateriales cada vez más utilizados. Uno de los principales objetivos de la investigación para el futuro es comprender los mecanismos moleculares por los que las sustancias químicas y los nanomateriales pueden introducir modificaciones (genéticas o epigenéticas) en el ADN, lo que también es fundamental para evaluar mejor su seguridad y perfeccionar su diseño. Este conocimiento molecular puede ayudar a modelizar la toxicidad de las sustancias químicas y los nanomateriales, lo que resultará decisivo para agruparlos y permitir una capacidad de extrapolación, prediciendo así los efectos adversos y el mecanismo. Por último, la correcta identificación de las alteraciones (epi)genéticas inducidas por el medioambiente y la comprensión de su relación etiológica con los fenotipos de las enfermedades debería permitir el diseño de estrategias eficaces para proteger la salud humana y reducir la incidencia de las enfermedades no transmisibles en las generaciones futuras.

### **6.3.2 El microbioma y las interacciones huésped-patógeno**

A pesar de los grandes avances en las tecnologías ómicas, la caracterización del microbioma es un campo en desarrollo y aún por explorar en su totalidad. Los enfoques de secuenciación del genoma completo permiten reconstruir los clones, caracterizar las variantes y los clones conocidos, y buscar genes de virulencia o resistencia. El cultivo y la secuenciación del amplicón 16S aumentaron los conocimientos, pero aún se necesita una mayor determinación. Los estudios de las interacciones entre el huésped y el patógeno han pasado del estudio de genes individuales a enfoques del genoma completo, que incluyen tanto el genoma del huésped como el microbiano. La metagenómica puede proporcionar un conocimiento profundo de las adaptaciones de las comunidades microbianas a los entornos cambiantes y permitir la identificación de cepas microbianas específicas. Los genomas ensamblados con metagenomas (MAG, por sus siglas en inglés) hacen referencia a un método introducido hace poco que permite el ensamblaje de genomas microbianos a partir de datos metagenómicos y que está proporcionando nuevos conocimientos sobre la diversidad microbiana, así como las interacciones huésped-patógeno (Quince *et al.*, 2017). Es preciso realizar estudios con muestreos longitudinales y múltiples perspectivas moleculares para descifrar la dinámica subyacente y aportar información nueva sobre la interacción huésped-patógeno en contextos ambientales específicos. Para investigar la dinámica del ecosistema holobionte, se requieren enfoques multiómicos que incluyan la genómica, la transcriptómica, la proteómica y

la metabolómica. Estos enfoques proporcionarían una descripción molecular detallada y una nueva visión mecanicista de la composición y el metabolismo de la microbiota, así como la regulación del fenotipo del huésped a través de las interacciones de la microbiota con el transcriptoma del huésped, las marcas epigenéticas y las vías metabólicas (Miro-Blanch y Yanes, 2019). Además, debido a la dificultad para clasificar la información *ómica* de la relación huésped-microbioma y para explorar la diversidad microbiana a nivel de cepas individuales, la bioinformática y la biología computacional se consideran algunos de los desafíos más apremiantes en este campo. Los aspectos bioinformáticos se han desarrollado en el capítulo 3.2.

### **6.3.3 Genómica integradora para acelerar la detección de rasgos**

La genómica lleva décadas contribuyendo a los avances en el desarrollo de cosechas y cultivos animales. La llegada de las plataformas de secuenciación de próxima generación (NGS, por sus siglas en inglés) cambió el impacto de la secuenciación en nuestro conocimiento de los genomas y la regulación de los genes. Una vez que se dispone de la secuencia del genoma, se pueden identificar todos los genes y variantes genéticas que contribuyen a los rasgos agronómicos. Además, los cambios realizados durante el proceso de mejora pueden evaluarse en el genotipo. La secuenciación del genoma se ha convertido en un paso inicial para conocer el genoma y la evolución, mientras que los pasos posteriores de resecuenciación permiten dilucidar la variabilidad genética entre individuos. No obstante, todavía existen importantes limitaciones en la NGS, que da lugar a ensamblajes genómicos muy fragmentados que complican el análisis. Estos problemas pueden resolverse ahora con la aparición de las tecnologías de secuenciación de tercera generación (TGS, del inglés *Third-Generation Sequencing*) que permiten la generación de lecturas largas, así como la producción de ensamblajes genómicos más precisos y contiguos (Jiao y Schneeberger, 2017; Korlach *et al.*, 2017), lo que a su vez facilita los estudios genómicos.

Se han aplicado varias aproximaciones genómicas para acelerar la detección de asociaciones entre genes y rasgos. Como las plantas y los animales evolucionan en entornos complejos, con lo que adquieren de forma gradual la capacidad de hacer frente a diferentes condiciones ambientales, un elevado número de rasgos o fenotipos deseables se definen como rasgos cuantitativos complejos. Los estudios de *loci* de rasgos cuantitativos (QTL, por sus siglas en inglés) se han utilizado para identificar regiones del genoma que co-segregan con un rasgo determinado (Hu *et al.*, 2018). Con todo, el mapeo de



QTL tiene dos limitaciones fundamentales: su restringida resolución y el hecho de que solo se pueden realizar ensayos de la diversidad alélica que se segrega entre los progenitores de una población segregante (Korte y Farlow, 2013). En los últimos años, los estudios de asociación del genoma completo (GWAS) conectados con estrategias de secuenciación del genoma completo han evolucionado hasta convertirse en una poderosa herramienta para volver a conectar los rasgos con su genética subyacente, con lo que se superan las limitaciones de los QTL (Korte y Farlow, 2013). Estas nuevas estrategias de GWAS proporcionan una mayor resolución para identificar múltiples eventos de recombinación y pueden emplearse para localizar mutaciones genómicas vinculadas a rasgos en poblaciones diversas y no relacionadas, lo que permite explorar las variaciones naturales asociadas a las diferencias fenotípicas. La integración de los GWAS en múltiples estudios y la combinación de sus datos con otras técnicas de alto rendimiento, como la transcriptómica, la proteómica y la metabolómica, acelerarán la detección de asociaciones sólidas entre genes y rasgos y ayudarán a comprender rasgos fenotípicos complejos. Por otro lado, la paleogenómica tiene como objetivo reconstruir los genomas antiguos mediante la secuenciación directa de material fósil o de los ancestros de cultivos actuales, ayudando así a comprender la domesticación de los cultivos y a predecir cómo evolucionarán las poblaciones futuras en respuesta al calentamiento global (Pont *et al.*, 2019). También permite introducir en los programas de cría modernos rasgos antiguos perdidos durante la historia de la cría natural, como la resistencia al estrés, el color o el sabor.

#### **6.3.4 La epigenética como nueva fuente de mayor variabilidad y adaptabilidad**

Los programas de cría intensiva han reducido la diversidad genética en los cultivos y el ganado. La reciente finalización de la secuenciación de numerosos genomas ha aumentado la eficiencia de la cría al proporcionar nuevas herramientas que han ayudado a identificar una proporción sustancial de la variación fenotípica heredable basada en la secuencia. Ahora bien, la variabilidad de la secuencia de los genes que controlan los rasgos agronómicos no puede explicar todo el espectro de la diversidad fenotípica observada en las plantas y los animales, y sigue habiendo una proporción significativa de heredabilidad no explicada. Pruebas recientes indican que la variación epigenética puede explicar la heredabilidad de rasgos complejos (Cortijo *et al.*, 2014). Además, los estudios en organismos modelo han proporcionado una gran cantidad de información sobre la implicación de los mecanismos



epigenéticos en la regulación del desarrollo y en la respuesta a diferentes situaciones de estrés abiótico y biótico, lo que desempeña un papel importante en la configuración de la plasticidad fenotípica.

La capacidad de un mismo genotipo para expresar múltiples fenotipos en respuesta a diferentes entornos, ya sean externos o internos, está muy extendida tanto en animales como en plantas. En general, se considera que la plasticidad es adaptativa o ventajosa para los organismos sésiles que tienen que adaptarse en el lugar a las condiciones ambientales, y puede aprovecharse para obtener cosechas más resilientes. Por otra parte, la inclusión de la plasticidad en los modelos de cría de animales (por ejemplo, la plasticidad sexual en las especies de peces de cultivo) será importante para criar animales más fuertes o en los programas de cría (por ejemplo, animales de cría) que producen material genético para una serie de entornos de producción. No obstante, nuestra capacidad para aprovechar la plasticidad de animales y plantas requiere una mejor comprensión de los mecanismos epigenéticos subyacentes y definir los estados epigenómicos.

La información epigenética está mediada en su mayor parte por la metilación del ADN y las modificaciones postraduccionales de las histonas, las llamadas marcas de la cromatina que alteran la lectura y la escritura del genoma, lo que da lugar a la regulación de la arquitectura de la cromatina, la actividad y la expresión de los genes y todo ello sin que ocurran cambios en la secuencia del ADN. Aunque los análisis de la regulación epigenética de la expresión génica se remontan a los años 80, los métodos para analizar las modificaciones epigenéticas en el genoma no se desarrollaron hasta principios de la década del 2000. Los avances en la tecnología de secuenciación del ADN, el desarrollo de métodos como la secuenciación por bisulfito y la secuenciación por inmunoprecipitación de la cromatina, y la generación de anticuerpos muy específicos contra las histonas modificadas postraduccionalmente han supuesto una oportunidad para crear mapas epigenómicos (Gallusci *et al.*, 2017). Sin embargo, es importante tener en cuenta que cada genoma puede dar lugar a un gran número de *epigenomas*, que suelen ser ajustables y muy dinámicos. Por lo tanto, se debe describir el *estado básico* epigenómico en diferentes etapas de desarrollo en distintas especies de animales y plantas, y cómo este estado cambia en respuesta al entorno. Asimismo, resulta conveniente determinar si las variantes asociadas a los rasgos están enriquecidas en firmas epigenéticas específicas de los tejidos. La generación de mapas del genoma de las marcas de histonas y del metiloma

dentro de células, tejidos y órganos específicos en condiciones ambientales variables allanará el camino para descubrir la implicación de determinados tipos de células y tejidos en rasgos específicos. Además, sería crucial buscar nuevas marcas epigenómicas para comprender plenamente la diversidad de modificaciones epigenómicas que pudiera ser necesario considerar. En este sentido, el objetivo general ahora es llevar a cabo investigaciones básicas y aplicadas sobre cómo los procesos epigenéticos/epigenómicos contribuyen al desarrollo y la respuesta a las condiciones ambientales. Los datos obtenidos a partir de estos estudios deberían incluirse en bases de datos epigenómicas integradoras. Esto permitirá llevar a cabo estudios de asociación del epigenoma (EWAS, por sus siglas en inglés), como se hace con los seres humanos (Verma, 2012), utilizando diferentes genotipos, tejidos, tipos de células o condiciones ambientales, lo que puede aportar información valiosa para el desarrollo de epimarcas que pueden emplearse en la mejora de los cultivos y la cría de animales.

### **6.3.5 Comprensión de la memoria epigenética**

Las respuestas al estrés en los animales y las plantas pueden producirse y alimentarse por episodios previos de estrés (Chang *et al.* 2020). La base molecular de esta memoria del estrés se considera en gran medida epigenética y con frecuencia transgeneracional; es decir, la tolerancia adquirida puede transmitirse a la progenie. Los desafíos futuros para aprovechar la regulación epigenética de la expresión génica con el fin de dar lugar a cosechas y cultivos animales más resistentes incluirán descifrar el código del paisaje epigenético inducido por el estrés, cómo las señales ambientales se traducen en epimarcas específicas y aprender a preservar las modificaciones epigenéticas deseables a lo largo de generaciones posteriores o a eliminar las no deseadas. Responder a estas preguntas resultará fundamental para entender la estabilidad, reversibilidad y heredabilidad de los epialelos, y el uso de la ingeniería epigenética para mejorar la resistencia y la productividad de las plantas de interés agronómico. La identificación de los componentes moleculares que conectan los cambios ambientales con el epigenoma proporcionará nuevos objetivos susceptibles de ingeniería epigenética para una mejor resiliencia. Evitar el estrés que se produce en los animales y plantas sensibilizados suele conllevar una reducción de la tasa de crecimiento y de las tallas, rasgos que pueden transferirse epigenéticamente a la descendencia. Hay que esforzarse por comprender cómo los cambios epigenéticos transgeneracionales restringen todo el potencial de rendimiento de la agricultura y la ganadería de forma que se evite el síndrome de evitación del estrés (Maggio *et al.*, 2018).

### **6.3.6 Genómica microbiana e implicaciones medioambientales**

Un impacto importante de las actividades humanas en los entornos naturales tiene que ver con el calentamiento global de los océanos, que es probable que condicione la idoneidad de nuestro planeta para la vida a medio y largo plazo (Cavicchioli *et al.*, 2019). El aumento previsto de las zonas templadas de los océanos afectará a la distribución de los microorganismos, al favorecer a los adaptados a las zonas intertropicales oligotróficas en detrimento de otros menos tolerantes a las altas temperaturas. De forma simultánea al aumento de la temperatura, la acidificación de los océanos, como consecuencia de una mayor concentración de CO<sub>2</sub> en la atmósfera, puede alterar en gran medida la composición del fitoplancton marino —por ejemplo, los efectos sobre muchas algas que incluyen estructuras de carbonato cálcico— e impactar en parámetros esenciales para la productividad oceánica, como la luz incidente o el reciclaje de nutrientes. Los desarrollos clave serían los siguientes: la implementación de la capacidad de analizar las comunidades microbianas mediante metagenómica en numerosos nichos, que deberían incluir los océanos abiertos pero también los entornos costeros y terrestres; la mejora de la capacidad de detección de todas las variedades de microorganismos en las muestras; y la capacidad de atribuir la participación de cada miembro de la comunidad a las actividades biológicas detectadas en cada nicho medioambiental. En conjunto, la aplicación de estas prácticas proporcionaría una capacidad inestimable para evaluar el alcance y la dirección de las respuestas microbianas a las actividades humanas. Por último, en los ecosistemas marinos, los virus capaces de infectar bacterias y otros microorganismos superan en un gran número a los de sus huéspedes y es posible que afecten enormemente a la dinámica de las poblaciones de productores primarios. Se trata de un efecto poco caracterizado que merece más atención y que también puede investigarse mediante aproximaciones metagenómicas.

### **6.3.7 Desarrollos biotecnológicos mediante la genómica microbiana**

Los microbios ya se utilizan para un amplio repertorio de usos biotecnológicos que incluyen la producción de sustancias químicas con gran valor añadido, la desintoxicación de entornos contaminados y el tratamiento de aguas. Como ejemplo, la empresa española PharmaMar se dedica a la investigación de recursos marinos para la industria farmacéutica. La metagenómica podría permitir la reconstrucción de las vías metabólicas estructuradas por la

comunidad, así como la identificación de nuevas enzimas y reacciones. La disección exhaustiva de las actividades microbianas representa una oportunidad sin precedentes para la identificación de nuevas vías metabólicas, nuevas actividades enzimáticas que pueden ser usadas para el diseño de reacciones y redes biotecnológicamente relevantes, y productos naturales de interés para la humanidad. En este sentido, se plantea un reto para el desarrollo de procedimientos para generar bibliotecas de expresión de tipo *shot-gun* derivadas de información metagenómica, así como reconstruir vías para analizar actividades. En conjunto, un profundo conocimiento de las comunidades microbianas en la naturaleza y sus respuestas a las perturbaciones humanas será esencial para la mitigación de impactos perniciosos y para la utilización sostenible de los recursos oceánicos.

## CAPÍTULO 6 BIBLIOGRAFÍA

- Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science*. 348, 873-873.
- Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis M. *et al.* (2019). Scientists' warning to humanity: microorganisms and climate change. *Nature Rev. Microbiol.* 17, 569-586.
- Chang, Y.N., Zhu, C., Jiang, J., Zhang, H., Zhu, J.K., Duan, C.G. (2020). Epigenetic regulation in plant abiotic stress responses. *J. Integr. Plant Biol.* 62, 563-580.
- Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K. *et al.* (2014). Mapping the epigenetic basis of complex traits. *Science* 343, 1145-1148.
- Dietrich, J., Schmidt, E., Kotze, D.J., Hornung, E., Setälä, H., Yesilonis, I. *et al.* (2019). Metagenomics reveals Bacterial and Archaeal adaptation to urban land-use: N catabolism, methanogenesis, and nutrient acquisition. *Front. Microbiol.* 10, 2330.
- Duarte, C.M., Agusti, S., Barbier, E., Britten, G.L., Castilla, J.C., Gattuso, J.P. (2020). Rebuilding marine life. *Nature*. 580, 39-51. doi: 10.1186/s13148-014-0043-3. eCollection 2015.
- Falkowski, P.G., Barber, R.T., Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*. 281, 200-206.
- Farrant, G.K., Dore, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M. *et al.* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci. USA*. 113, E3365-3374.
- Foster, R.A., Kuypers, M.M.M., Vagner, T., Paerl, R.W., Musat, N. y Zehr J.P. (2011). Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *ISME J.* 5, 1484-1493.
- Gallusci, P., Dai, Z., Génard, M., Gauffretau, A., Leblanc-Fournier, N., Richard-Molard, C., Vile, D., Brunel-Muguet, S. (2017). Epigenetics for plant improvement: Current knowledge and modeling avenues. *Trends Plant Sci.* 22, 610623.
- Guerrero-Bosagna, C., Jensen, P. (2015). Globalization, climate change, and transgenerational epigenetic inheritance: will our descendants be at risk? *Clinical Epigenetics*. 7(1), 8.
- Hu, H., Scheben, A., Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture*. 8, 75.
- Jiao, W-B., Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Curr. Opin. Plant Biol.* 36: 64-70.
- Karl, D.M., Church, M.J., Diore, J.E., Letelier, R.M., Mahaffey, C. (2016). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Pro. Natl. Acad. Sci. USA* 109, 1842-1849.
- Korlach, J., Gedman, G., Kingan, S.B., Chin, C., Howard, J.T. (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience* 6, 1-16.
- Korte, A., Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 29.
- Maggio, A., Bressan, R.A., Zhao, Y., Park, J., Yun, D.J. (2018). It's hard to avoid avoidance: Uncoupling the evolutionary connection between plant growth, productivity and stress «Tolerance». *Int. J. Mol. Sci.* 19(11), 3671.
- Miro-Blanch, J., Yanes, O. (2019). Epigenetic regulation at the interplay between gut microbiota and host metabolism. *Front. Genet.* 10, 638. doi: 10.3389/fgene.2019.00638.
- Näpflin, K., O'Connor, E.A., Becks, L., Bensch, S., Ellis, V.A., Hafer-Hahmann, N., Harding, K.C. *et al.* (2019). Genomics of host-pathogen interactions: challenges and opportunities across ecological and spatiotemporal scales. *Peer J.* 201, e8013. doi: 10.7717/peerj.8013. eCollection 2019.
- Nilsson, E., Larsen, G., Manikkan, M., Guerrero-Bosagna, C., Savenkova, M.I., Skinner, M.K. (2012). Environmentally induced epigenetic transgenerational inheritance of ovarian disease. *PLoS ONE* 7, e36129.

- Ost, A., Lempradt, A., Casas, E., Weigert, M., Tiko, T., Deniz, M. *et al.* (2014). Peternal diet defines offspring chromatin state and intergenerational obesity. *Cell* 159, 1352-64.
- Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., Salse, J. (2019). Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biology* 20, 29.
- Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833-844.
- Simon, J., Marchesi, J.R., Mougél, C. *et al.* (2019). Host-microbiota interactions: from holobiont theory to analysis. *Microbiome* 7, 5.
- Sironi, M., Cagliani, R., Forni, D., Clerici, M. (2015). Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat. Rev. Genet.* 16, 224-36. doi: 10.1038/nrg3905.
- Verma, M. Epigenome-wide association studies (EWAS) in cancer. (2012). *Curr. Genomics* 13, 308-313.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74.

**RESUMEN**

Es bien sabido que los cambios epigenéticos a lo largo de la vida (es decir, los que tienen una finalidad biológica, la deriva epigenética y los relojes epigenéticos) dependen de una compleja mezcla de factores, así como el hecho de que, en función de los loci, se pueden modular mediante factores genéticos y/o externos, de entre los cuales el estilo de vida es uno de los más importantes. Sin embargo, el mecanismo molecular subyacente sigue siendo en gran medida desconocido, por lo que encarar esta cuestión constituirá un desafío importante para el CSIC en este campo.

**PALABRAS CLAVE**

estilo de vida    envejecimiento  
relojes epigenéticos  
ejercicio físico    nutrigenómica  
herencia transgeneracional

# EPIGENÓMICA Y ESTILO DE VIDA

## Coordinadores

Mario Fernández Fraga  
(CINN, Oviedo, coordinador)

Sonia Ramos Rivero  
(ICTAN, Madrid, coordinadora adjunta)

## Investigadores y centros de investigación participantes (por orden alfabético)

Laura Bravo Clemente  
(ICTAN, Madrid)

Carmen Collado  
(IATA, Valencia)

Raúl Fernández Pérez  
(CINN, Oviedo)

Belén Gomara  
(IQOG, Madrid)

Rocío González Urdingio  
(CINN, Oviedo)

María Ángeles Martín Arribas  
(ICTAN, Madrid)

## RESUMEN EJECUTIVO

Las dos últimas décadas han servido para consolidar el concepto de que las modificaciones epigenéticas cambian a lo largo de la vida. De hecho, en numerosas obras científicas se han caracterizado varias firmas epigenéticas asociadas al envejecimiento, y lo destacable es que recientemente se han desarrollado *relojes* epigenéticos que pueden predecir la edad biológica midiendo los niveles de marcas epigenéticas tales como la metilación del ADN. Estos relojes también actúan como biomarcadores de la fisiología porque se alteran en estados patológicos. Sin embargo, aunque estos predictores de la edad cronológica se han caracterizado adecuadamente, todavía se desconocen en gran medida los mecanismos moleculares subyacentes. Por lo tanto, urge especificar estos mecanismos moleculares y determinar si los relojes epigenéticos son una causa o una consecuencia del envejecimiento.

En los últimos años, se ha propuesto en numerosas obras (principalmente descriptivas) que los cambios epigenéticos a lo largo de la vida se pueden modular, al menos en parte, mediante estímulos externos. También se ha propuesto que el estilo de vida durante el embarazo puede programar el epigenoma durante el desarrollo embrionario y determinar una serie de fenotipos patológicos concretos en la vida adulta. De ser cierto, esto implicaría que el estilo de vida podría moldear los fenotipos de la salud y la enfermedad a través de mecanismos epigenéticos. Además, en la actualidad no queda claro si



la herencia epigenética transgeneracional o las exposiciones intrauterinas repercuten sobre la salud y la propensión a enfermedades de la descendencia. Por lo tanto, es preciso explorar el papel de la epigenética en la herencia transgeneracional.

La dieta es uno de los factores más importantes en el estilo de vida y tiene el potencial de modular los programas de expresión génica a través de un mecanismo epigenético, por lo que repercute sobre la salud individual y la esperanza de vida. De cara al futuro, uno de los desafíos consistirá en determinar tanto la influencia de los nutrientes y los componentes bioactivos de los alimentos sobre la actividad epigenética como los mecanismos moleculares subyacentes afectados, lo cual será importante para definir nuevos métodos preventivos y terapéuticos eficaces basados en la nutrición. De este modo, la epigenética nutricional puede ser decisiva para el desarrollo de programas personalizados que contribuyan a reducir el riesgo de padecer enfermedades y a mejorar la salud. Además, aparte de nutrientes beneficiosos, los alimentos también pueden contener cantidades variables de elementos contaminantes que presentan diversos mecanismos de toxicidad y bioactividad. Por lo tanto, en el futuro, los estudios nutrigenómicos deberán tener en cuenta no solo los efectos beneficiosos de los nutrientes, sino también los posibles efectos perjudiciales de los contaminantes presentes en los alimentos.

Otro componente fundamental del estilo de vida es la actividad física. De hecho, es bien sabido que el ejercicio asiduo influye sobre la salud y previene trastornos tales como las enfermedades cardiovasculares y metabólicas o el cáncer. Al igual que con la dieta, también se ha propuesto que los mecanismos epigenéticos constituyen el vínculo molecular entre estos beneficios para la salud y la actividad física. Sin embargo, dado que la mayoría de los estudios se centran en genes candidatos específicos en los tejidos objetivo, es necesario llevar a cabo en el futuro un extenso análisis de las modificaciones epigenéticas en todo el genoma. También habrá que identificar firmas epigenómicas específicas asociadas a diversos tipos de actividad física (por ejemplo, ejercicios aeróbicos suaves, actividades de fuerza o resistencia, etc.).

Aparte de la dieta y la actividad física, también deben tenerse en cuenta otros aspectos importantes del estilo de vida, como es el caso del consumo de alcohol y otras drogas recreativas, los tratamientos farmacológicos, etc. Para todos estos factores, la investigación en el futuro no solo debe limitarse a explicar los cambios epigenéticos en reacción a una determinada señal ambiental, sino que también debe tratar de identificar las consecuencias funcionales y fisiológicas de estos cambios.

Por último, dado que los cambios epigenéticos son en principio reversibles, ofrecen una vía para el desarrollo de tratamientos encaminados a contrarrestar procesos complejos, como es el caso del envejecimiento. Entre los desafíos para el futuro se hallan la identificación de las alteraciones implicadas; la separación de los cambios biológicamente pertinentes del resto del ruido ocasionado por el medioambiente; el diseño de las herramientas biológicas para reprogramar las alteraciones epigenéticas; y, algo que es importante, el desarrollo de instrumentos tales como los relojes epigenéticos que nos permitan determinar si nuestras intervenciones afectan de algún modo a nuestra esperanza de vida con un envejecimiento saludable.

## 7.1 INTRODUCCIÓN Y DESCRIPCIÓN GENERAL

### 7.1.1 Cambios epigenéticos a lo largo de la vida (deriva epigenética frente a relojes epigenéticos)

El envejecimiento es un fenómeno universal mediante el cual las capacidades funcionales biológicas disminuyen gradualmente y que, al final, conduce a la muerte. A escala celular, estos cambios influyen sobre una amplia gama de vías moleculares e incluyen alteraciones tanto genéticas como epigenéticas (López Otín *et al.*, 2013). De hecho, los mecanismos epigenéticos pueden, al menos en parte, intervenir en características del envejecimiento tales como la inestabilidad del genoma o el ruido transcripcional (Pal y Tyler, 2016). Estas alteraciones se han definido a todos los niveles de la regulación epigenética, de entre los cuales los más estudiados han sido la metilación del ADN y la modificación de las histonas, tanto en humanos como en organismos modelo (Pal y Tyler, 2016). En concreto, la metilación del ADN se ha investigado en los seres humanos, lo cual ha suscitado una serie de nociones aceptadas en lo referente a los cambios graduales observados a lo largo de la vida: una pérdida global de metilación del ADN en las regiones intergénicas y repetitivas, incrementos locales en las regiones con gran concentración de CpG y, en general, un aumento de la variabilidad intra e interindividual en los patrones de metilación del ADN (Huidobro *et al.*, 2013; y Jones *et al.*, 2015). Esta última observación, a la que se suele denominar *deriva epigenética*, subraya la idea de que los cambios epigenéticos observados durante el proceso de envejecimiento constan de una combinación de alteraciones funcionales y estocásticas que, además, pueden ser relevantes o no para la regulación de la expresión génica (Tejedor y Fraga, 2017). Dado que las alteraciones epigenéticas pueden originarse de un sinfín de causas, entre las que se incluyen elementos ambientales como los factores del estilo de vida, es vital que podamos distinguir las variaciones específicas de las no específicas (Feil y Fraga, 2012).

En este contexto, debido al desarrollo de las tecnologías a escala genómica de micromatrices y de secuenciación de última generación (NGS, por sus siglas en inglés), recientemente ha surgido en este campo el concepto de *reloj epigenético*. Los relojes epigenéticos constan de modelos matemáticos que utilizan la información de la metilación del ADN para predecir la edad cronológica con un nivel de precisión sin precedentes (Horvath y Raj, 2018). Además, el comportamiento uniforme a lo largo de la vida de los focos de CpG implicados los distingue de los que participan en la deriva epigenética relacionada con la edad (Jones *et al.*, 2015). Tanto los cambios epigenéticos asociados al envejecimiento como los relojes epigenéticos recién caracterizados pueden contribuir a explicar los mecanismos moleculares implicados en la definición de la variabilidad fenotípica observada entre personas tanto a nivel fisiológico como patológico. Sin embargo, para ello hay que desligarlos eficazmente de la variación epigenética estocástica.

### 7.1.2 Nutrigenómica y nutrieipigenómica

Se considera que la nutrición es uno de los factores del estilo de vida capaces de afectar al genoma y al epigenoma más influyentes. La nutrigenómica define la interacción entre la nutrición y los genes para entender cómo pueden los componentes específicos de los alimentos o los regímenes alimenticios afectar a la salud humana. La novedosa disciplina de la nutrieipigenética integra los conocimientos de la nutrigenómica y el efecto de la dieta o los compuestos alimentarios en los programas de expresión génica a través de mecanismos epigenéticos. La dieta, los alimentos y sus componentes afectan al genoma y al epigenoma, y tienen el potencial de modular vías metabólicas críticas que repercuten sobre la salud individual y la esperanza de vida. Se ha propuesto la existencia de un vínculo entre determinados patrones dietéticos y diversas enfermedades no transmisibles (cáncer, enfermedades cardiovasculares, obesidad y diabetes). Estas asociaciones constituyen un punto de partida sobre la capacidad de las dietas y, en última instancia, los alimentos para modificar el genoma y el epigenoma y, por consiguiente, el proteoma y el metaboloma. Por lo tanto, cada patrón alimenticio podría proporcionar distintas firmas epigenéticas y genéticas, las cuales podrían estar asociadas a un estado saludable o a la propensión a la enfermedad. No obstante, la epigenómica nutricional es un área de investigación bastante reciente, y los datos actuales sobre los efectos precisos de las dietas, los alimentos o los componentes alimenticios sobre el epigenoma son muy escasos. En la actualidad, la mayoría de estos estudios son descriptivos, y tanto las modificaciones epigenéticas como los mecanismos moleculares subyacentes siguen siendo en gran medida desconocidos. Además, la mayoría de los estudios de epigenética

nutricional se han centrado en la metilación del ADN, mientras que las modificaciones postraduccionales de las histonas y la expresión del miARN se analizan aún menos.

### **7.1.3 Contaminantes presentes en los alimentos**

Se considera que la inhalación de aire y polvo y la ingestión de alimentos constituyen las principales vías de exposición a compuestos químicos tóxicos para la población general. En el caso de los compuestos lipófilos, la ingestión de alimentos es la fuente de exposición más importante. Muchos contaminantes ambientales tóxicos con propiedades de alteración endocrina son lipófilos. Por consiguiente, se bioconcentran en los organismos vivos y se bioacumulan a través de las redes tróficas, lo cual provoca que los seres humanos sean los organismos más afectados. Además, los productos agroquímicos utilizados para el control de plagas, los fármacos veterinarios empleados en la ganadería, los compuestos migratorios relacionados con los materiales en contacto con los alimentos, los aditivos alimentarios (que van desde los colorantes, los conservantes y los agentes estabilizadores de los alimentos hasta los compuestos bioactivos) y los contaminantes introducidos o formados durante el almacenamiento de los alimentos (como, por ejemplo, las micotoxinas y las aminas biógenas) y su procesamiento (como la acrilamida) llegan a los seres humanos a través de la alimentación. Se ha demostrado que muchos de estos compuestos son tóxicos y ejercen una modulación epigenética.

### **7.1.4 La nutrición de precisión en la salud humana y la prevención de enfermedades**

La nutrición desempeña un papel fundamental para la prevención de muchas patologías crónicas, y la alimentación constituye un factor modificable clave que puede influir sobre la incidencia de trastornos metabólicos muy prevalentes, tanto monogénicos (por ejemplo, la celiaquía y la intolerancia a la lactosa) como poligénicos (por ejemplo, la diabetes de tipo 2, la obesidad, el síndrome metabólico y las enfermedades cardiovasculares) o determinados tipos de cáncer. Para complicar aún más las cosas, algunas enfermedades pueden estar ligadas a factores de riesgo tanto monogénicos como poligénicos. Por ejemplo, la obesidad, cuya incidencia aumenta constantemente y está asociada a muchas enfermedades concomitantes (diabetes, dislipidemia, hipertensión, inflamación, etc.), puede representar un síntoma de hasta 40 enfermedades monogénicas y anomalías cromosómicas, pero la obesidad también puede depender de numerosas variantes genéticas, ya que en estudios de asociación del genoma completo (GWAS, por sus siglas en inglés) se asociaron más de 600 genes y

regiones de ADN a la obesidad humana, lo que demuestra la complejidad de estudiar rasgos poligénicos relacionados con numerosas vías fisiológicas.

Además, el posible papel terapéutico de la nutrición en la prevención de las enfermedades degenerativas crónicas es múltiple y complejo. No solo los nutrientes de los alimentos pueden tener un efecto acusado sobre la salud, sino, en especial, sus componentes bioactivos (polifenoles, carotenoides, fitoesteroles, isotiocianatos, glucosinolatos,  $\omega$ -3/ $\omega$ -6, etc.). Muchos de estos compuestos bioactivos actúan a varios niveles moleculares, desde la expresión del ADN hasta las modificaciones pretranscripcionales, la afectación de la funcionalidad de las proteínas, los procesos metabólicos, etc. Estas consecuencias nutri(epi)genómicas de los componentes de los alimentos también vienen acompañadas de factores nutri(epi)genéticos que afectan a aspectos cruciales como las preferencias alimentarias, el proceso de ADME (absorción, distribución, metabolismo y excreción), las vías metabólicas y, al final, el fenotipo y la reacción clínica de cada persona. Además, la microbiota intestinal puede verse afectada por estos componentes alimenticios y modificar su actividad a través de modificaciones catabólicas de las moléculas ingeridas.

Todo esto provoca una gran variabilidad en la reacción de cada persona a las dietas y a los alimentos, así como la necesidad de una estrategia nutricional molecular y *ómica* integradora. Al igual que la medicina de precisión, la nutrición de precisión también requiere una estrategia individualizada basada en los principios de las 4 P (personalizada, predictiva, preventiva y participativa), para lo cual la genética resulta esencial.

### **7.1.5 El impacto de la alimentación/microbiota y las interacciones sobre la nutrición de precisión**

La nutrición desempeña un papel destacado en la salud humana. La investigación nutricional molecular se ha definido como «la ciencia que estudia los efectos de los nutrientes, los alimentos y sus componentes sobre la fisiología en su conjunto y el estado de buena salud a escala molecular y celular». En el futuro, la investigación nutricional deberá ir más allá de la dieta única para todos hacia el estudio de la reacción personalizada del huésped a la dieta. Este concepto debe integrar tanto las variables biológicas (el microbioma, el epigenoma, el metaboloma, el genoma, etc.) como las ambientales (la alimentación, la actividad física, el consumo de fármacos, la xenobiótica, las infecciones, el estrés, etc.) para obtener predicciones detalladas de las reacciones de cada persona a nutrientes específicos y a otros compuestos bioactivos de su alimentación.

### 7.1.6 Actividad física

El efecto beneficioso de la actividad física sobre la salud es de sobra conocido (Fiuza Luces *et al.*, 2013; y Neufer *et al.*, 2015). Un estilo de vida físicamente activo y el ejercicio asiduo contribuyen a regular la presión arterial y el metabolismo, modulan la homeostasia y, en general, contribuyen a mejorar la salud y a prevenir enfermedades (Booth *et al.*, 2012). La falta de ejercicio físico influye sobre la salud a lo largo de la vida, y está asociada a mayores tasas de mortalidad prematura, cardiopatías coronarias, diabetes de tipo 2, cáncer de colon y de mama, así como de obesidad (Booth *et al.*, 2012; Nieman *et al.*, 2019; y Fernández Sanlés *et al.*, 2020). También se conocen los beneficios del ejercicio sobre la memoria y las funciones intelectuales (Hillman *et al.*, 2008; y Fernandes *et al.*, 2017), lo cual pone de relieve su importancia en el mantenimiento de la salud física y mental a lo largo de la vida.

Aunque no se conocen del todo los mecanismos biológicos que regulan los efectos beneficiosos del ejercicio sobre la salud, se ha propuesto que las marcas epigenéticas desempeñan un papel importante, ya que la epigenética puede representar el vínculo molecular que explica cómo afecta el entorno a nuestros genes (Feil y Fraga, 2012). Los cambios epigenéticos asociados a la actividad física se han estudiado a diversos niveles en lo referente al tipo de tejido y al tipo de marca epigenética analizada, aunque sobre todo en glóbulos sanguíneos, en tejidos musculares y adiposos, y en tejidos cerebrales de modelos murinos (Elsner *et al.*, 2011; Seaborne *et al.*, 2018; y Nielsen *et al.*, 2010). En este contexto, la metilación del ADN es la marca epigenética más estudiada, tanto a escala global como específica de los genes (Fernández Sanlés *et al.*, 2020; Seaborne *et al.*, 2018a; Ronn *et al.*, 2013; y Schenk *et al.*, 2019), seguida de los miARN y de las marcas postraduccionales de las histonas (Elsner *et al.*, 2011; Nielsen *et al.*, 2010; Pandorf *et al.*, 2009; y Melo *et al.*, 2015).

La aparición de las tecnologías ómicas en general, y de la epigenómica en particular, abre nuevas vías para estudiar en mayor detalle los efectos del ejercicio sobre el epigenoma y, en especial, las posibles consecuencias funcionales que estos cambios epigenéticos pueden tener sobre la salud.

### 7.1.7 Alcohol y otras drogas adictivas

Al igual que nuestra dieta, a lo largo de nuestra vida nos vemos expuestos a muchos otros tipos de compuestos. El Instituto Nacional de Estadística de España estimó en 2017 que más del 20 % de las personas de 15 años o más fumaban a diario, y una proporción similar de ciudadanos consumía alcohol

semanalmente (Instituto Nacional de Estadística, 2017). El alcohol y el tabaco tienen consecuencias de sobra conocidas y de amplio espectro sobre nuestra salud, por lo que no debería sorprender que su consumo se haya relacionado con cambios epigenéticos. Debido a su situación legal y a su prevalencia en la población, la mayor parte de los estudios se han llevado a cabo con estas dos sustancias. En cuanto al tabaco, hay que tener en cuenta que los miles de compuestos peligrosos distintos presentes en el humo pueden afectar de un sinnúmero de formas (Talhout *et al.*, 2011), mientras que los efectos del alcohol se corresponden con los del metabolismo del etanol.

## 7.2 REPERCUSIÓN SOBRE EL PANORAMA DE LA CIENCIA BÁSICA Y LAS POSIBLES APLICACIONES

### 7.2.1 Cambios epigenéticos a lo largo de la vida (deriva epigenética frente a relojes epigenéticos)

La caracterización de los cambios epigenéticos que conlleva el envejecimiento en el contexto de la deriva epigenética tan solo está empezando a abordarse. El marcado descenso del coste de las tecnologías de micromatrices ha posibilitado el desarrollo de estudios a gran escala, que permiten detectar mejor los cambios epigenéticos más sutiles en el contexto de la ruidosa variabilidad interindividual. En esta misma línea, se han llevado a cabo estudios con grandes cohortes en los que se ha revelado que la variabilidad en los cambios de metilación del ADN está ligada a las vías moleculares asociadas al envejecimiento (BIOS Consortium *et al.*, 2016), lo cual aporta una explicación de los procesos implicados en el envejecimiento y que, además, han servido para identificar loci genéticos particulares que podrían abrir vías esenciales para la intervención contra el envejecimiento (McCartney *et al.*, 2020).

Además, los relojes epigenéticos también presentan un excelente potencial como herramientas en la investigación del envejecimiento y las enfermedades relacionadas con este. Se ha demostrado que la edad epigenética se acelera en relación con una gran variedad de enfermedades, una observación importante que sirve para demostrar que los relojes se pueden emplear como biomarcadores de las enfermedades (Horvath y Raj, 2018). El hecho de que personas de la misma edad manifiesten edades epigenéticas distintas apunta hacia la noción de que estos modelos están detectando al menos en parte el «envejecimiento biológico», el cual se puede concebir como el estado fisiológico general con respecto al riesgo de morir y las enfermedades concomitantes (Bell *et al.*, 2019). Por ello, ya se está desarrollando una nueva

generación de relojes centrados en la edad biológica para detectar mejor las relaciones de las enfermedades (Lu *et al.*, 2019; y Levine *et al.*, 2018). Es posible que la más prometedora de sus aplicaciones sea el hecho de que podrían utilizarse para evaluar el éxito o el fracaso de las intervenciones contra el envejecimiento.

### 7.2.2 Nutrigenómica y nutriepigenómica

El trabajo futuro en el campo de la nutrición y la epigenética tiene el potencial de proporcionar un beneficio considerable a la salud pública. Descifrar las firmas epigenéticas desencadenadas por los componentes bioactivos de los alimentos podría propiciar intervenciones nutricionales personalizadas que tuvieran en cuenta la información genética/epigenética. La epigenética nutricional representa un potencial seguro y una estrategia innovadora para la prevención o el tratamiento de muchas enfermedades crónicas prevalentes que están íntimamente relacionadas con las modificaciones epigenéticas.

### 7.2.3 Contaminantes presentes en los alimentos

Las autoridades sanitarias controlan sistemáticamente la presencia de contaminantes en los alimentos para garantizar su seguridad. Sin embargo, los estudios sobre los efectos de estas sustancias a nivel epigenético y genómico siguen siendo bastante escasos, a pesar de ser obligatorios para elaborar una legislación y regulación alimentarias correctas. Una dificultad adicional asociada a la presencia de contaminantes industriales y agrícolas y de fármacos veterinarios es que, cuando se determina la toxicidad de estas sustancias químicas, se las sustituye en seguida por productos alternativos, lo que hace que los controles reglamentados y datos de toxicidad anteriores queden obsoletos, y obliga a la comunidad científica y a las autoridades alimentarias a desarrollar nuevas metodologías y estrategias analíticas para identificarlas, determinarlas, evaluar sus riesgos y someterlas a controles periódicos.

Por lo tanto, el futuro de la nutrigenómica pasa por tener en cuenta no solo los efectos beneficiosos de los nutrientes, sino también los posibles efectos perjudiciales que se derivan de los contaminantes, ya sean estos naturales a antropógenos, presentes en los alimentos. Se trata de un desafío complejo, ya que muchos contaminantes conocidos se han asociado a una elevada incidencia y prevalencia de diversos trastornos relacionados con el sistema endocrino de los seres humanos, pero también por la constante introducción en la red trófica de nuevas sustancias que también pueden afectar al (epi)genoma.



### **7.2.4 La nutrición de precisión en la salud humana y la prevención de enfermedades**

Existe un creciente interés en la nutrición de precisión por su potencial para la prevención y el tratamiento de enfermedades crónicas no transmisibles, tanto monogénicas como poligénicas. En los GWAS y otros estudios genéticos se han identificado más de 15 000 SNP asociados a numerosas patologías y rasgos. Aparte de los polimorfismos genéticos, las modificaciones epigenéticas, que son específicas de cada tejido, se ven muy afectadas por los factores ambientales y del estilo de vida (incluida la alimentación) y son reversibles, pero también heredables transgeneracionalmente, complican la propensión y sensibilidad diferenciales de las personas a las enfermedades relacionadas con la alimentación y a las intervenciones dietéticas. Junto a estos factores, los ARNnc también desempeñan un papel importante, ya que los ARNnc y, sobre todo, los miARN producidos en las vesículas extracelulares pueden tener efectos paracrinos o endocrinos en diversos tejidos y órganos objetivo. En consonancia con esto, también se debería considerar el papel de los ARNnc ingeridos con los alimentos como modificadores epigenéticos externos potencialmente importantes, ya que, en general, todavía se sabe poco sobre la repercusión de los ARNnc sobre las vías metabólicas y las redes de regulación.

Si tenemos esto en cuenta, identificar las variantes genéticas o marcas epigenéticas que predisponen a las personas a padecer determinadas enfermedades relacionadas con el metabolismo resulta vital tanto para la medicina como para la nutrición de precisión, a fin de estimar el riesgo de sufrir enfermedades y de diseñar estrategias preventivas. Del mismo modo, determinar las interacciones genético-dietéticas pertinentes permitirá identificar cuáles son las personas que reaccionan a intervenciones dietéticas específicas y las que no, lo cual a su vez permitiría diseñar recomendaciones personalizadas para aumentar al máximo los efectos beneficiosos de las intervenciones nutricionales. Esta nutrición orientada al genotipo resultará útil no solo para ofrecer consejos alimentarios personalizados, sino que también mejorará las recomendaciones de salud pública y el diseño de soluciones nutricionales, incluidos los alimentos funcionales y los productos nutracéuticos.

### **7.2.5 El impacto de la alimentación/microbiota y las interacciones sobre la nutrición de precisión**

Nuestra alimentación y nuestro estilo de vida, así como otros hábitos (por ejemplo, los patrones de sueño) y exposiciones (por ejemplo, el estrés, los contaminantes, la luz azul, las sustancias químicas de los alimentos y los

plásticos), han cambiado drásticamente en las últimas décadas. La exposición a alimentos poco saludables y a patrones alimentarios inadecuados como, por ejemplo, el exceso o el déficit de macro y micronutrientes, puede aumentar el riesgo de padecer enfermedades no transmisibles (ENT). Esta exposición puede actuar a través de varios mecanismos posibles, entre los que se incluyen la modulación del microbioma, pero también la afectación del metaboloma y de la regulación epigenética, de las vías celulares y fisiológicas y del sistema inmunitario, lo cual repercute sobre la sensibilidad y la salud del huésped, por lo que aumenta el riesgo de sufrir ENT. La mayoría de estos mecanismos siguen sin conocerse con claridad, por lo que la necesidad de generar pruebas científicas mediante estudios en humanos es apremiante.

Se ha observado una fuerte relación entre la disbiosis de la microbiota y el riesgo de padecer ENT como, por ejemplo, alergias, obesidad, diabetes, problemas relacionados con el sistema inmunitario y enfermedades cardiovasculares, lo cual ha provocado que el peso de estas enfermedades sea cada vez mayor y requiera tomar medidas urgentes. Según la Organización Mundial de la Salud (OMS), las ENT son la causa de más de 41 millones de muertes al año (lo que supone el 71 % de todas las muertes en el mundo), por lo que sus consecuencias para la sociedad y la economía resultan alarmantes.

Curiosamente, se considera que el periodo comprendido entre la gestación y los primeros años de vida es el más crítico en lo referente al riesgo de desarrollar ENT. La repercusión de la nutrición perinatal sobre el desarrollo del microbioma de los lactantes es una cuestión que ha cobrado un interés considerable, pero la información de la que se dispone sobre el microbioma de las mujeres y la nutrición prenatal y durante la gestación es escasa. Las exposiciones ambientales de la madre, incluidas la alimentación y los microbios, pueden propiciar cambios duraderos o incluso provocar cambios permanentes en la fisiología del feto, lo cual repercute sobre el riesgo de sufrir enfermedades en etapas posteriores de la vida. La nutrición también ocasiona efectos a corto y largo plazo sobre la salud humana mediante la programación del desarrollo inmunológico, metabólico y microbiológico. Además, la interacción entre los nutrientes, la microbiota y el huésped hacia la regulación epigenética podría repercutir sobre el desarrollo del feto, así como sobre las consecuencias para la salud del lactante y la madre.

La microbiota materna representa la principal fuente microbiana para los lactantes, y la transferencia microbiana específica de las madres a sus hijos se produce en el momento del parto y durante la lactancia.

La disbiosis microbiana materna durante el embarazo se transfiere al neonato, lo que provoca un desarrollo inadecuado del inóculo microbiano, así como efectos inmunológicos y metabólicos con consecuencias desfavorables para la salud en el futuro. En la actualidad no queda claro si la herencia epigenética transgeneracional o las exposiciones intrauterinas repercuten sobre la salud y la propensión a enfermedades de la descendencia. La alimentación desempeña un papel fundamental en la formación de la microbiota intestinal, mientras que las interacciones entre los nutrientes y la microbiota repercuten sobre las consecuencias para la salud del huésped, por lo que presentan consecuencias críticas para la salud; de hecho, como se suele decir, «somos lo que comemos». La microbiota interviene en la digestión de los alimentos, la absorción de nutrientes, el entrenamiento del sistema inmunitario y la protección contra patógenos y toxinas, así como la producción de compuestos específicos (AGCC, vitaminas, hormonas, neurotransmisores, etc.). La microbiota interactúa con el metabolismo de los carbohidratos de los alimentos, las proteínas, los polifenoles vegetales, los ácidos biliares y las vitaminas. Se necesitan más estudios para identificar cuáles son los alimentos, los macro y micronutrientes y los compuestos alimenticios específicos que influyen sobre la microbiota. Algunos nutrientes concretos de los alimentos, como es el caso de la fibra, los donantes de metilos (la betaina, la metionina y la colina), el folato y otras vitaminas del grupo B (B2, B6 y B12), están relacionados con la microbiota. Los seres humanos necesitan estos nutrientes y, aparte de las fuentes alimentarias, hay pruebas de que la microbiota intestinal también es una fuente de nutrientes esenciales para el huésped (por ejemplo, el folato y otras vitaminas B). Lo llamativo es que la mayoría de estos nutrientes desempeñan un papel crucial en la regulación epigenética y pueden repercutir sobre el epigenoma humano.

### 7.2.6 Actividad física

La aplicación de nuevas tecnologías de alto rendimiento en la era de la *ómica* no solo permitirá identificar los cambios moleculares producidos por el ejercicio en el epigenoma, sino que también ayudará a facilitar nuestra comprensión de los mecanismos que contribuyen a mejorar la salud. La utilidad de analizar en detalle los cambios o las alteraciones epigenéticas producidas por el ejercicio es incuestionable. Nos ayudará a comprender los efectos beneficiosos del ejercicio para la prevención y el tratamiento de enfermedades y, además, nos ofrecerá en el futuro nuevos objetivos terapéuticos. Aportará pistas sobre cómo combatir enfermedades cognitivas asociadas al envejecimiento como, por ejemplo, la enfermedad de Alzheimer, ya que los resultados obtenidos hasta ahora indican el efecto beneficioso del ejercicio sobre la memoria y las funciones

intelectuales (Hillman *et al.*, 2008; y Fernandes *et al.*, 2017). Además, conocer estas alteraciones en mayor profundidad generará información sobre los marcadores moleculares, indicadores del rendimiento deportivo óptimo que resultan de gran interés para los atletas profesionales. Y por último, pero no por ello menos importante, nos ayudará a aclarar cómo se mantienen a lo largo del tiempo los cambios epigenéticos producidos por el ejercicio (memoria epigenética) (Sharples *et al.*, 2016; y Seaborne *et al.*, 2018b), e incluso si estos cambios epigenéticos se heredan y, por lo tanto, presentan efectos beneficiosos para la descendencia (Segabizani *et al.*, 2019; y Spinder *et al.*, 2019).

### 7.2.7 Alcohol y otras drogas adictivas

En la actualidad, existen cuantiosos datos probatorios de que el tabaquismo provoca cambios en la metilación del ADN en el ser humano (Lee y Pausova, 2013). Los estudios de asociación por micromatrices han permitido identificar marcadores epigenéticos recurrentes tales como el gen F2RL3, relacionado con las funciones vasculares, o el gen AhR del metabolismo de sustancias xenobióticas (Breitling *et al.*, 2011; Shenker *et al.*, 2013; y Sun *et al.*, 2013). Estas y otras observaciones apuntan a que las alteraciones epigenéticas implicadas podrían ser importantes para los mecanismos moleculares asociados a los efectos adversos del tabaco. Por otro lado, los efectos epigenéticos del alcohol se han estudiado de una manera más mecanicista, principalmente porque se sabe que el etanol interrumpe el metabolismo de monocarbono y, por lo tanto, puede influir en el uso de grupos metilos por parte de las metiltransferasas del ADN (Pérez *et al.*, 2019; y Ron y Messing, 2011). Al igual que en el caso del tabaco, los estudios más recientes están empezando a aplicar tecnologías que abarcan todo el genoma para cribar loci genéticos asociados al alcohol (Zhang y Gelernter, 2017). En general, los estudios de asociación a escala genómica tienen el potencial de identificar los biomarcadores del uso de compuestos y las posibles vías implicadas en la etiología de enfermedades relacionadas con fármacos, mientras que los estudios mecanicistas resultan especialmente pertinentes para esta última aplicación.

## 7.3 PRINCIPALES PUNTOS PROBLEMÁTICOS

### 7.3.1 Entender el funcionamiento del reloj epigenético

¿Cuál es la importancia biológica de la existencia de focos de CpG cuyo estado de metilación refleja el envejecimiento cronológico o biológico? Los loci que componen los relojes se han asociado a características genómicas concretas, como es el caso de los focos asociados a policombos (Raj y Horvath, 2020), pero

todavía no hay vínculos claros con ningún proceso biológico específico. Queda por aclarar si los relojes epigenéticos son «impulsores» del envejecimiento o meros «pasajeros» que reflejan las huellas de otros procesos, una observación que será pertinente para el diseño de intervenciones antienvejecimiento.

### **7.3.2 Medir e identificar la variabilidad**

La caracterización de la variabilidad epigenética es clave para identificar la deriva epigenética y distinguirla de otros cambios funcionales o asociados a la enfermedad. Los estudios a gran escala facilitados por la integración de conjuntos de datos públicos arrojarán luz sobre esta cuestión. Sin embargo, en el campo coexisten distintas interpretaciones y mediciones de la variabilidad (BIOS Consortium *et al.*, 2016; y Gentilini *et al.*, 2015), por lo que habrá que definir mejor su importancia biológica.

### **7.3.3 Delimitar el envejecimiento cronológico y el biológico**

Los avances médicos que favorecen que se prolongue la esperanza de vida carecerán por completo de sentido si no van acompañados de la correspondiente prolongación de la vida saludable. Para poder encarar este desafío, es preciso caracterizar el envejecimiento biológico y sus biomarcadores de forma tanto general como específica; y, en este contexto, indudablemente resultará crucial contar con unos relojes epigenéticos más elaborados (Bell *et al.*, 2019; y Partridge *et al.*, 2018).

### **7.3.4 Investigar la herencia transgeneracional epigenética en los mamíferos**

Se ha propuesto que la propensión del epigenoma para adaptarse a los factores del estilo de vida —entre los que se incluye la nutrición— varía a lo largo de la vida de un organismo, y su mayor sensibilidad a los cambios se da en las etapas iniciales (periodo pre y neonatal) (Kanherkar *et al.*, 2014). Además, las modificaciones epigenómicas parecen ser reversibles, aunque se ha postulado que estas alteraciones pueden transmitirse de generación en generación. A este respecto, los datos experimentales en humanos han demostrado que los trastornos metabólicos (la desnutrición y la obesidad materna) durante los periodos iniciales del desarrollo (el embarazo) generan un entorno de desarrollo anómalo que podría modificar el epigenoma y predisponer a la descendencia a padecer enfermedades metabólicas en fases posteriores de su vida (Tobi *et al.*, 2014; y de Rooij *et al.*, 2006a,b). Esto se ha explicado mediante la llamada *hipótesis de los orígenes del desarrollo de la salud y la enfermedad*, en la que se propone una herencia epigenética transgeneracional (Gluckman y Hanson, 2004). Dicha hipótesis también se centra en los efectos de la nutrición sobre ambos progenitores

antes de la fecundación y durante esta, y que apuntarían a la posibilidad de presentar un estado nutricional favorable del desarrollo para obtener cambios epigenéticos beneficiosos (Fleming *et al.*, 2018). No obstante, en la actualidad sigue habiendo dudas sobre si la herencia epigenética transgeneracional o las exposiciones intrauterinas repercuten sobre la salud y la propensión a enfermedades de la descendencia. Estos desafíos merecen futuras investigaciones y obligarán a poner a varias generaciones en observación y seguimiento.

### **7.3.5 Analizar al detalle el papel de las modificaciones epigenéticas provocadas por la alimentación sobre la salud y la enfermedad humanas**

La nutrición también puede afectar a la salud y predisponer a padecer enfermedades en edades posteriores. Las pruebas obtenidas en los estudios de intervención dietética, así como en las investigaciones en las que se ha analizado el efecto de los componentes de los alimentos en modelos experimentales humanos y animales, han apuntado a que los componentes de la alimentación (los nutrientes y los compuestos bioactivos) ejercen diversas actividades biológicas que podrían generar efectos protectores contra distintas enfermedades no transmisibles y propiciar un envejecimiento más saludable. Sin embargo, resulta complejo dilucidar sus mecanismos moleculares y las modificaciones epigenéticas asociadas. Curiosamente, este método de evaluación de los posibles beneficios de un nutriente o componente alimentario podría ayudar a identificar cambios epigenéticos y a definir biomarcadores precoces de enfermedades, lo cual también tendrá un papel relevante a la hora de diseñar a partir de la nutrición nuevas estrategias preventivas y terapéuticas eficaces, las cuales contribuirán a mejorar la salud y a reducir el riesgo de sufrir enfermedades. No obstante, los estudios nutrigenómicos son complejos y resulta complicado demostrar una causalidad a partir de una asociación. Identificar cuáles son los componentes del fenotipo de una enfermedad que están relacionados con la nutrición plantea todo un desafío, excepto en el caso de las enfermedades causadas por un único defecto genético. Además, cuesta entender cómo reaccionan los seres humanos a dietas o nutrientes específicos, así como determinar cuáles son los componentes de los alimentos responsables de una acción concreta. De hecho, la alimentación presenta muchos componentes, y la interacción de estas sustancias provoca varios cambios metabólicos, que pueden incluso ser distintos en función de cómo se ingiera un mismo alimento (Nicodemus-Johnson y Sinnott, 2017). Además, hay que tener en cuenta que el genoma y el epigenoma pueden interactuar, es decir, que entender cómo pueden las modificaciones epigenómicas alterar la expresión

génica y regular las consecuencias de la nutrición también constituye un enorme desafío que permitirá dar los primeros pasos hacia la nutrición de precisión en la prevención de enfermedades.

### **7.3.6 Entender el impacto de los miARN exógenos**

Se han identificado miARN en líquidos biológicos (sangre, leche materna humana y leche de otras especies). Sin embargo, la influencia de estos miARN exógenos no se ha estudiado en profundidad a pesar de que estos elementos epigenéticos se detectan en la mayoría de los alimentos y se suelen conservar de una especie a otra (Xia *et al.*, 2011.; Ledda *et al.*, 2020; y Mal *et al.*, 2018). Resulta vital comprender el posible impacto de los miARN exógenos sobre la epigenética humana y su posible influencia sobre la salud y propensión a enfermedades. De hecho, la demostración de un efecto beneficioso o perjudicial podría conllevar cambios en la producción, el procesamiento o la cocción de los alimentos. Además, debido a su presencia generalizada en los alimentos, se ha propuesto que los miARN pudieran constituir biomarcadores y/o elementos de comunicación (Benmoussa y Provost 2019). Todo ello podría posibilitar una nueva estrategia terapéutica contra las enfermedades no transmisibles o favorecer un envejecimiento más saludable.

### **7.3.7 Descubrir las consecuencias genéticas y epigenéticas de las sustancias químicas de los alimentos**

Es bien sabido que en los alimentos puede haber miles de sustancias tóxicas como consecuencia de las tareas de producción, transporte, procesamiento, envasado y almacenamiento, pero también debido al impacto natural de la contaminación ambiental residual. El número de sustancias químicas en las que se detecta toxicidad epigenética aumenta continuamente (Marczylo *et al.*, 2016). Por otra parte, se prevé una ampliación considerable de nuestros conocimientos sobre los efectos epigenéticos y genéticos de las sustancias químicas antiguas y nuevas, gracias a los recientes avances en las técnicas y metodologías analíticas. Además, se ha demostrado en modelos animales que hay diversas enfermedades que se transmiten transgeneracionalmente (Guerrero-Bosagna y Jensen, 2015). Todos estos nuevos conocimientos conducen la investigación hacia estudios centrados tanto en los cambios epigenéticos que se observan actualmente como en las consecuencias transgeneracionales de la actual exposición humana a sustancias químicas tóxicas relacionadas con la ingestión de alimentos. Para alcanzar este objetivo, resulta esencial desarrollar metodologías de determinación apropiadas mediante instrumental de última generación y transferirlas desde la investigación hasta los laboratorios sistemáticos oficiales para determinar de forma

exacta, rápida y sostenible las sustancias químicas reguladas y conocidas de los alimentos, pero también los compuestos tóxicos nuevos y de reciente aparición que podrían introducirse en cualquiera de los eslabones de la cadena alimentaria. Para lograr este último objetivo, se deben implantar enfoques inespecíficos en los controles sistemáticos, lo cual obliga a adaptar las metodologías analíticas utilizadas a fin de cumplir las exigencias actuales en lo referente a la selectividad y la sensibilidad. Estos enfoques permitirán obtener datos experimentales exhaustivos sobre la exposición alimentaria que permitirán una correlación adecuada con los cambios epigenéticos y genéticos observados en la población.

### 7.3.8 Avanzar hacia la nutrición de precisión y personalizada

El objetivo final de la nutrición de precisión estriba en poder ofrecer un asesoramiento dietético personalizado teniendo en cuenta las reacciones individuales para preservar la salud y prevenir las enfermedades. Para ello, es preciso avanzar en las siguientes cuestiones:

- Se necesitan amplios estudios poblacionales basados en poblaciones caracterizadas de forma adecuada y minuciosa (anamnesis clínica, sexo, edad, estilo de vida, [epi]genética y microbiota) para avanzar en la identificación de los determinantes de la variabilidad individual en la reacción ante intervenciones dietéticas específicas a los diferentes aspectos relacionados con la alimentación. Ensayos clínicos y estudios de cohortes.
- Mejorar el conocimiento de los determinantes genéticos del proceso de ADME (de los nutrientes y los componentes bioactivos de los alimentos) y de las preferencias alimentarias, así como de los requisitos alimentarios influidos por los SNP. Lista completa de SNPs; base de datos de ADN-nutrientes.
- Abordar el impacto de los ARNnc, tanto el de cada persona como el derivado de la alimentación, en los órganos diana y su repercusión sobre diversas enfermedades (por ejemplo, la obesidad o el cáncer). Estudiar el impacto de las intervenciones dietéticas sobre los miARN y su potencial como biomarcadores o herramientas terapéuticas para la nutrición de precisión en enfermedades específicas.
- Biología de sistemas/bioinformática: La integración de los datos clínicos, los antecedentes genéticos, la microbiota y otros datos *ómicos* múltiples (epigenómicos, transcriptómicos, proteómicos, metabolómicos, metagenómicos...) en ensayos clínicos y estudios de cohortes requiere potentes recursos bioinformáticos, entre los que se incluyen algoritmos de aprendizaje automático capaces de predecir la reacción en función de la integración de los datos.



### 7.3.9 Reconfigurar la microbiota mediante la nutrición

La reconfiguración de las interacciones entre el huésped y la microbiota a través de la nutrición personalizada supondría una nueva herramienta para mejorar la salud de cara al control y la prevención de enfermedades. Conviene conocer la reacción del huésped en función del perfil microbiótico en una intervención dietética específica (por ejemplo, los que reaccionan y los que no). El cuándo (patrones de alimentación circadianos y ayuno intermitente) y el cómo (procesos de cocción) se consumen los alimentos o la ingestión alimenticia específica pueden ejercer una repercusión distinta sobre la fisiología y la microbiota del huésped, así como sobre la interacción entre el huésped y el microbioma. Hasta la fecha, solo se han llevado a cabo una cantidad limitada de estudios para determinar si los factores nutricionales provocan cambios en la microbiota e impulsan las interacciones entre el huésped y la microbiota, principalmente en los periodos críticos de la vida como la primera infancia y la ancianidad. Estas observaciones explicarían el enorme interés que suscitan las intervenciones perinatales y el posible uso de probióticos, prebióticos y simbióticos para promover una «microbiota adecuada» y, por lo tanto, influir beneficiosamente en la salud. Al mismo tiempo, existe un interés considerable por la investigación relacionada con la microbiota, cuyo objetivo es identificar los microorganismos, moléculas microbianas y metabolitos específicos que favorecen la fisiología, los metabolismos y la salud del huésped. Entender cómo reacciona el microbioma a los componentes de la alimentación y la consiguiente repercusión biológica, así como las consecuencias clínicas, puede utilizarse para el desarrollo de intervenciones dietéticas elaboradas con precisión.

- Investigar cómo benefician al organismo nutrientes específicos de la alimentación más allá del aporte nutritivo, al contribuir a mejorar el bienestar general o a reducir el riesgo de sufrir enfermedades.
- Comprender las interacciones entre el huésped, el microbioma y la alimentación, así como los mecanismos subyacentes.
- Identificar los microorganismos, moléculas microbianas y/o metabolitos concretos que contribuyen a la fisiología, los metabolismos y la salud del huésped.
- Conocer la reacción del huésped en función del perfil microbiótico en diversas intervenciones dietéticas (los que reaccionan y los que no).
- Determinar hasta qué punto modulan la salud humana el cuándo (los patrones de alimentación circadianos y ayuno intermitente) y el cómo (procesos de cocción) se consumen los nutrientes de la alimentación.

- Desarrollar modelos mecanicistas y predictivos del efecto de los componentes de la alimentación y los productos basados en la microbiota sobre la salud.

### **7.3.10 Hacia una comprensión más global y mecanicista de las consecuencias epigenéticas del ejercicio físico**

Hay una serie de cuestiones importantes que habrá que abordar en los próximos años en lo referente a la forma en que las alteraciones epigenéticas debidas a la actividad física pueden repercutir sobre la salud y la enfermedad humanas:

- El diseño de modelos animales para estudiar los mecanismos moleculares que hay detrás de los efectos beneficiosos o perjudiciales del ejercicio y que puedan trasladarse a los seres humanos (por ejemplo, en deportistas de élite).
- La implantación de tecnologías epigenéticas de última generación para estudiar las marcas epigenéticas a escala genómica.
- La integración de los datos epigenómicos en los obtenidos de otras *ómicas* (es decir, la transcriptómica y la proteómica) para identificar los cambios epigenéticos con efectos funcionales.
- El estudio del efecto del ejercicio sobre el epigenoma de diversos tipos de células y tejidos para identificar cambios comunes y específicos.

### **7.3.11 Definición del papel etiológico de las alteraciones epigenéticas producidas por el tabaco y el alcohol en las enfermedades humanas**

La mayoría de los actuales hallazgos de alteraciones epigenéticas son el producto de estudios de asociación. Sin embargo, este enfoque tiene una capacidad limitada para aclarar si estos cambios son causas o consecuencias de la enfermedad asociada al alcohol o al tabaco. Las vías de señalización y los mecanismos moleculares implicados aún están por definir mediante el desarrollo de estudios mecanicistas. Además, debido a su accesibilidad, la mayoría de los cambios epigenéticos se han referido en la sangre periférica, que no es el principal tejido objetivo del alcohol ni del tabaco. Es probable que, aparte de los biomarcadores, se detecten más asociaciones funcionales de las alteraciones epigenéticas y la expresión genética mediante el examen de otros tejidos que estén relacionados más directamente con estos compuestos.

## CAPÍTULO 7 BIBLIOGRAFÍA

- Bell, C.G., Lowe, R., Adams, P.D., Baccarelli, A.A., Beck, S., Bell, J.T., Christensen, B.C., Gladyshev, V.N., Heijmans, B.T., Horvath, S. *et al.* (2019). DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* 20, 249.
- Benmoussa, A. y Provost, P. (2019). Milk microRNAs in health and disease. *Compr. Rev. Food Sci. Food Safety* 18, 703-722.
- Biesiekierski, J.R., Jalanka, J. y Staudacher, H.M. (2019). Can Gut Microbiota Composition Predict Response to Dietary Treatments? *Nutrients* 11(5), 1134. doi:10.3390/nul1051134.
- BIOS Consortium, Sliker, R.C., van Itersson, M., Luijk, R., Beekman, M., Zhernakova, D.V., Moed, M.H., Mei, H., van Galen, M., Deelen, P. *et al.* (2016). Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol.* 17, 191.
- Booth, F.W., Roberts, C.K. y Laye, M.J. (2012). Lack of exercise is a major cause of chronic diseases. *Compr. Physiol.* 2, 1143-1211.
- Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. y Brenner, H. (2011). Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication. *The American Journal of Human Genetics* 88, 450-457.
- De Rooij, S. R. *et al.* (2006a). Impaired insulin resistance secretion after prenatal exposure to the Dutch famine. *Diabetes Care* 29, 1897-1901.
- De Rooij, S. R. *et al.* (2006b). Glucose tolerance at age 58 and the decline of glucose tolerance in comparison with age 50 in people prenatally exposed to the Dutch famine. *Diabetologia* 49, 637-643.
- Elsner, V.R., Lovatel, G.A., Bertoldi, K., Vanzella, C., Santos, F.M., Spindler, C., de Almeida, E.F., Nardin, P. y Siqueira, I.R. (2011). Effect of different exercise protocols on histone acetyltransferases and histone deacetylases activities in rat hippocampus. *Neuroscience* 192, 580-587.
- Feil, R. y Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13, 97-109.
- Fernandes, J., Arida, R.M. y Gómez-Pinilla, F. (2017). Physical exercise as an epigenetic modulator of brain plasticity and cognition. *Neurosci. Biobehav. Rev.* 80, 443-456.
- Fernández Sanlés, A., Sayols-Baixeras, S., Castro, D.E.M.M., Esteller, M., Subirana, I., Torres Cuevas, S., Pérez Fernández, S., Aslibekyan, S., Marrugat, J. y Elosua, R. (2020). Physical Activity and Genome-wide DNA Methylation: The REGISTRE Gironi del COR Study. *Med. Sci. Sports Exerc.* 52, 589-597.
- Fiuza Luces, C., Garatachea, N., Berger, N.A. y Lucia, A. (2013). Exercise is the real polypill. *Physiology (Bethesda)* 28, 330-358.
- Fleming, T.P. *et al.* (2018). Origins of lifetime health around the time of conception causes and consequences. *Lancet* 391, 1842-1852.
- Gentilini, D., Garagnani, P., Pisoni, S., Bacalini, M.G., Calzari, L., Mari, D., Vitale, G., Franceschi, C. y Di Blasio, A.M. (2015). Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with x chromosome inactivation skewing in females. *Aging* 7, 568-578.
- Gluckman, P.D. y Hanson, M.A. (2004). Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. *Ped. Res.* 56, 311-317.
- Guerrero-Bosagna, C. y Jensen, P. (2015). Globalization, climate change, and transgenerational epigenetic inheritance: will our descendants be at risk? *Clinical Epigenetics* 7, 8. doi 10.1186/s13148-014-0043-3.
- Hillman, C.H., Erickson, K.I. y Kramer, A.F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nat. Rev. Neurosci.* 9, 58-65.
- Horvath, S. y Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371-384.
- Huidobro, C., Fernández, A.F. y Fraga, M.F. (2013). Ageing epigenetics: Causes and consequences. *Molecular Aspects of Medicine* 34, 765-781.
- Instituto Nacional de Estadística (2017). *Encuesta Nacional de Salud 2017*.
- Jones, M.J., Goodman, S.J. y Kobor, M.S. (2015). DNA methylation and healthy human aging. *Aging Cell* 14, 924-932.
- Kanherkar, R.R., Bhatia-Dey, N. y Csoka, A.B. (2014). Epigenetics across the human lifespan. *Front. Cell Develop. Biol.* 2, 1-19.

- Kolodziejczyk, A.A., Zheng, D. y Elinav, E. (2019). Diet-microbiota interactions and personalized nutrition. *Nat. Rev. Microbiol.* 17(12), 742-753. doi: 10.1038/s41579-019-0256-8.
- Ledda, B. *et al.* (2020). Small RNAs in eucaryotes: new clues for amplifying microRNA benefits. *Cell Biosci.* 10, 1. doi: 10.1186/s13578-019-0370-3.
- Lee, K.W.K. y Pausova, Z. (2013). Cigarette smoking and DNA methylation. *Front. Genet.* 4, 132.
- Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y. *et al.* (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10, 573-591.
- Mal, C., Aftabuddin, M. y Kundu, S. (2018). IIKmTA: Inter and intra kingdom miRNA-target analyzer. *Interdiscip. Sci. Comput. Life Sci.* 10, 538-543.
- Marczylo, E.L., Jacobs, M.N. y Gant, T.W. (2016). Environmentally induced epigenetic toxicity: potential public health concerns. *Crit. Rev. Toxicol.* 46, 676-700.
- López Otín, C., Blasco, M.A., Partridge, L., Serrano, M. y Kroemer, G. (2013). The Hallmarks of Aging. *Cell* 153, 1194-1217.
- Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D. *et al.* (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11, 303-327.
- McCartney, D.L., Zhang, F., Hillary, R.F., Zhang, Q., Stevenson, A.J., Walker, R.M., Bermingham, M.L., Boutin, T., Morris, S.W., Campbell, A. *et al.* (2020). An epigenome-wide association study of sex-specific chronological ageing. *Genome Med.* 12, 1.
- Melo, S.F., Barauna, V.G., Junior, M.A., Bozi, L.H., Drummond, L.R., Natali, A.J. y de Oliveira, E.M. (2015). Resistance training regulates cardiac function through modulation of miRNA-214. *Int. J. Mol. Sci.* 16, 6855-6867.
- Mills, S., Lane, J.A., Smith, G.J., Grimaldi, K.A., Ross, R.P. y Stanton, C. (2019). Precision Nutrition and the Microbiome Part II: Potential Opportunities and Pathways to Commercialisation. *Nutrients* 11, 1468.
- Neufer, P.D., Bamman, M.M., Muoio, D.M., Bouchard, C., Cooper, D.M., Goodpaster, B.H., Booth, F.W., Kohrt, W.M., Gerszten, R.E., Mattson, M.P. *et al.* (2015). Understanding the Cellular and Molecular Mechanisms of Physical Activity-Induced Health Benefits. *Cell. Metab.* 22, 4-11.
- Nielsen, S., Scheele, C., Yfanti, C., Akerstrom, T., Nielsen, A.R., Pedersen, B.K. y Laye, M.J. (2010). Muscle specific microRNAs are regulated by endurance exercise in human skeletal muscle. *J. Physiol.* 588, 4029-4037.
- Nieman, D.C. y Wentz, L.M. (2019). The compelling link between physical activity and the body's defense system. *J. Sport Health Sci.* 8, 201-217.
- Nicodemus-Johson, J. y Sinnott, R.A. (2017). Fruit and juice epigenetic signatures are associated with independent immunoregulatory pathways. *Nutrients* 9, E752. doi: 10.3390/nu9070752.
- Pal, S. y Tyler, J.K. (2016). Epigenetics and aging. *Sci. Adv.* 2, e1600584.
- Pandorf, C.E., Haddad, F., Wright, C., Bodell, P.W. y Baldwin, K.M. (2009). Differential epigenetic modifications of histones at the myosin heavy chain genes in fast and slow skeletal muscle fibers and in response to muscle unloading. *Am. J. Physiol. Cell. Physiol.* 297, C6-16.
- Partridge, L., Deelen, J. y Slagboom, P.E. (2018). Facing up to the global challenges of ageing. *Nature* 561, 45-56.
- Pérez, R.F., Santamarina, P., Fernández, A.F. y Fraga, M.F. (2019). Epigenetics and Lifestyle: The Impact of Stress, Diet, and Social Habits on Tissue Homeostasis. En *Epigenetics and Regeneration* (Elsevier), 461-489.
- Raj, K. y Horvath, S. (2020). Current perspectives on the cellular and molecular features of epigenetic ageing. *Exp. Biol. Med. (Maywood)* 245(17), 1532-1542. doi: 10.1177/1535370220918329.
- Rinninella E., Cintoni M., Raoul P. *et al.* (2019). Food Components and Dietary Habits: Keys for a Healthy Gut Microbiota Composition. *Nutrients* 11(10), 2393. doi:10.3390/nu1102393.

- Ron, D. y Messing, R.O. (2011). Signaling Pathways Mediating Alcohol Effects. En *Behavioral Neurobiology of Alcohol Addiction*, W.H. Sommer y R. Spanagel, eds. (Berlín, Heidelberg: Springer Berlin Heidelberg), 87-126.
- Ronn, T., Volkov, P., Davegardh, C., Dayeh, T., Hall, E., Olsson, A.H., Nilsson, E., Tornberg, A., Dekker Nitert, M., Eriksson, K.F. *et al.* (2013). A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet* 9, e1003572.
- Schenk, A., Pulverer, W., Koliymitra, C., Bauer, C.J., Ilic, S., Heer, R., Schier, R., Schick, V., Bottiger, B.W., Gerhauser, C. *et al.* (2019). Acute Exercise Increases the Expression of KIR2DS4 by Promoter Demethylation in NK Cells. *Int. J. Sports Med.* 40, 62-70.
- Seaborne, R.A., Strauss, J., Cocks, M., Shepherd, S., O'Brien, T.D., Someren, K.A.V., Bell, P.G., Murgatroyd, C., Morton, J.P., Stewart, C.E. *et al.* (2018 a). Methylome of human skeletal muscle after acute & chronic resistance exercise training, detraining & retraining. *Sci. Data* 5, 180213.
- Seaborne, R.A., Strauss, J., Cocks, M., Shepherd, S., O'Brien, T.D., van Someren, K.A., Bell, P.G., Murgatroyd, C., Morton, J.P., Stewart, C.E. *et al.* (2018b). Human Skeletal Muscle Possesses an Epigenetic Memory of Hypertrophy. *Sci. Rep.* 8, 1898.
- Segabinazi, E., Spindler, C., Meireles, A.L.F., Piazza, F.V., Mega, F., Salvalaggio, G.D.S., Achaval, M. y Marcuzzo, S. (2019). Effects of Maternal Physical Exercise on Global DNA Methylation and Hippocampal Plasticity of Rat Male Offspring. *Neuroscience* 418, 218–230.
- Sharples, A.P., Stewart, C.E. y Seaborne, R.A. (2016). Does skeletal muscle have an 'epi'-memory? The role of epigenetics in nutritional programming, metabolic disease, aging and exercise. *Aging Cell* 15, 603-616.
- Shenker, N.S., Polidoro, S., van Veldhoven, K., Sacerdote, C., Ricceri, F., Birrell, M.A., Belvisi, M.G., Brown, R., Vineis, P. y Flanagan, J.M. (2013). Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum. Mol. Genet.* 22, 843-851.
- Spindler, C., Segabinazi, E., Meireles, A.L.F., Piazza, F.V., Mega, F., Dos Santos Salvalaggio, G., Achaval, M., Elsner, V.R. y Marcuzzo, S. (2019). Paternal physical exercise modulates global DNA methylation status in the hippocampus of male rat offspring. *Neural Regen. Res.* 14, 491-500.
- Sun, Y.V., Smith, A.K., Conneely, K.N., Chang, Q., Li, W., Lazarus, A., Smith, J.A., Almli, L.M., Binder, E.B., Klengel, T. *et al.* (2013). Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum. Genet.* 132, 1027-1037.
- Swann, J.R., Rajilic-Stojanovic, M., Salonen, A., Sakwinska, O., Gill, C., Meynier, A., Faça-Berthon, P., Schelkne, B., Segata, N., Shortt, C., Tuohy, K. y Hasselwander, O. (2020). Considerations for the design and conduct of human gut microbiota intervention studies relating to foods. *Eur. J. Nutr.* 3 de abril. doi: 10.1007/s00394-020-02232-1.
- Talhout, R., Schulz, T., Florek, E., Van Benthem, J., Wester, P. y Opperhuizen, A. (2011). Hazardous Compounds in Tobacco Smoke. *IJERPH* 8, 613-628.
- Tejedor, J.R. y Fraga, M.F. (2017). Interindividual epigenetic variability: Sound or noise? *BioEssays* 39, 1700055.
- Tobi, E. W. *et al.* (2014). DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nature Commun.* 5, 5592. DOI: 10.1038/ncomms6592.
- Xia, J.H., He, X.P., Bai, Z.Y. y Yue, G.H. (2011). Identification and characterization of 63 microRNAs in the Asian Seabass Lates calcarifer. *PLoS One* 6, e17537. doi: 10.1371/journal.pone.0017537.
- Zhang, H. y Gelernter, J. (2017). Revisión: DNA methylation and alcohol use disorders: Progress and challenges: DNA Methylation in Alcohol Addiction. *Am. J. Addict.* 26, 502-515.



En las últimas décadas, se han secuenciado por completo los genomas de cientos de organismos vivos diferentes. La descodificación de esta gran cantidad de información genética promete desvelar los secretos moleculares de la vida en nuestro planeta, así como la base molecular de las enfermedades humanas. Sin embargo, esto no es nada trivial, ya que en muchas especies las secuencias codificadoras de proteínas (es decir, los genes) solo representan una pequeña fracción de su genoma, y el resto de las secuencias no codificantes pueden desempeñar funciones reguladoras. Además, dentro de los organismos multicelulares, el genoma se utiliza de forma exclusiva en cada tipo de célula introduciendo modificaciones epigenéticas o adoptando conformaciones tridimensionales particulares que pueden afectar a la actividad y expresión génicas sin cambiar las secuencias de ADN subyacentes. A pesar de esta complejidad, los recientes avances en diversas tecnologías *ómicas* y de edición del genoma han mejorado notablemente nuestros actuales conocimientos sobre la función del genoma. Por consiguiente, ahora estamos en una disposición única para secuenciar, analizar y modificar los genomas y, de ese modo, mejorar no solo nuestra calidad de vida sino también la de nuestro planeta.