

Rafael R. Daga es profesor de genética molecular en la Universidad Pablo de Olavide de Sevilla e investigador del Centro Andaluz de Biología del Desarrollo (CABD), donde dirige su grupo de investigación en arquitectura nuclear y morfogénesis.

Silvia Salas-Pino es profesora de genética en la Universidad Pablo de Olavide de Sevilla.

Paola Gallardo es licenciada en biotecnología y trabaja como investigadora en el CABD.



GENÉTICA

La función reguladora del genoma

El proyecto internacional ENCODE ha revelado que la mayor parte de nuestro ADN, antes denominado «basura», se encarga de organizar y regular los genes que codifican proteínas

Rafael R. Daga, Silvia Salas-Pino y Paola Gallardo

EN EL AÑO 2001 FINALIZÓ EL PROYECTO GENOMA HUMANO Y SE CONOCIÓ LA SECUENCIA de nuestro ADN, esto es, la lista de «letras» que conforman los cromosomas de nuestras células. Pero la gran expectación inicial se desvaneció cuando se descubrió que, a pesar de haber descifrado el libro de instrucciones que sustenta la vida humana, solo se sabía interpretar una pequeña parte de él, alrededor del 1,5 por ciento. Ese porcentaje correspondía, en su mayoría, a genes que codificaban proteínas con una función concreta (estructural, enzimática, hormonal o inmunitaria) en nuestro organismo. El resto del genoma no parecía ejercer ninguna función relevante, por lo que se le denominó ADN «basura».

EN SÍNTESIS

Tras finalizar la secuenciación del genoma humano en 2001 se puso de manifiesto que una gran parte de él no parecía ejercer ninguna función relevante, por lo que se le denominó ADN «basura».

Durante más de diez años, el proyecto internacional ENCODE se ha esforzado en desentrañar el significado de esa enorme proporción del genoma.

Los hallazgos han revelado la enorme complejidad de organización del genoma. En él se han identificado numerosos elementos reguladores responsables de la activación o represión de los genes. Los resultados del proyecto tendrán una importante repercusión en el diagnóstico y tratamiento de múltiples enfermedades humanas.

THINKSTOCK

MÁS ALLÁ DE LA CONFIGURACIÓN en doble hélice del ADN, nuestro genoma presenta una intrincada red de interconexiones entre sus distintas regiones.



Aparte de la incógnita que planteaba la enorme proporción de ADN sin valor aparente, varias preguntas quedaban todavía abiertas. ¿Cómo se utilizaba y controlaba la información contenida en nuestro genoma? ¿Cómo se disponían y regulaban los genes? ¿Cómo la ingente cantidad de códigos daba lugar a un organismo vivo? Asimismo, resultaba de especial interés averiguar la forma de utilizar toda esa información para llegar a diagnosticar y curar ciertas enfermedades humanas. Los esfuerzos para dar significado a la secuencia de los más de 3000 millones de nucleótidos (formados por las bases adenina, citosina, guanina y timina, o A, C, G, T) que componen el genoma humano no habían hecho más que empezar.

Con el objetivo de paliar este vacío en el conocimiento, hace diez años se emprendió un proyecto piloto a escala internacional, financiado y promovido por el Instituto Nacional de Investigación del Genoma Humano de Estados Unidos (NHGRI, por sus siglas en inglés). Bautizado como *Encyclopedia of DNA Elements* (ENCODE), el proyecto trataba de descifrar ese 98,5 por ciento del genoma, cuya función se desconocía. La iniciativa fue desarrollada por un consorcio compuesto por 442 investigadores, pertenecientes a 32 instituciones de países como EE.UU., Reino Unido, Singapur, Japón, Suiza y España, entre otros. En nuestro país, cabe destacar la participación del Centro de Regulación Genómica, en Barcelona, que ha liderado el grupo de análisis de ARN. También han participado algunos investigadores del Centro Nacional de Investigaciones Oncológicas y se ha contado con el apoyo del Instituto Nacional de Bioinformática. El proyecto finalizó en septiembre de 2012, y como resultado se publicaron 30 artículos científicos, seis de ellos en la revista *Nature* y el resto en *Genome Research* y *Genome Biology*.

Mediante el estudio de 147 tipos celulares y el empleo de una combinación de análisis informáticos, anotación manual y validación experimental, el proyecto ENCODE ha revelado que gran parte del genoma humano contiene elementos funcionales. Entre ellos se incluyen las regiones responsables de codificar proteínas (los genes propiamente dichos) y las que regulan la expresión de los genes. De hecho, el proyecto concluyó que una enorme proporción del ADN, el que no contiene genes y que había sido considerado «basura», correspondía a elementos reguladores.

Los resultados pusieron también de manifiesto que numerosas alteraciones en la secuencia de ADN asociadas a enfermedades humanas no se producen en los genes sino en las regiones reguladoras, lo que cambia la visión tradicional sobre el origen de la disfunción celular responsable de ciertos trastornos.

Todos los datos generados en el proyecto son de libre acceso y representan una valiosa fuente de información para la comunidad científica, lo que ha contribuido enormemente al conocimiento que se tiene sobre el genoma humano.

UNA NUEVA VISIÓN

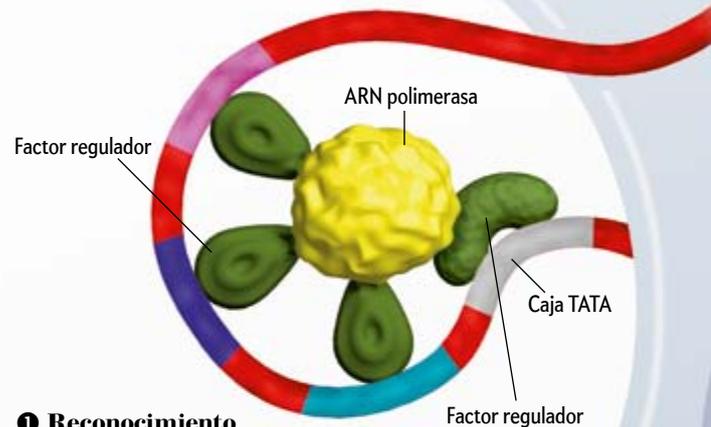
La regulación génica determina cuándo, cuánto y dónde (en qué tejido) se expresa un gen concreto. Se trata de un proceso mucho más complejo de lo que nadie se había planteado; tanto, que cuando finalizó la segunda fase del proyecto se anunció que los datos obtenidos harían modificar varios de los conceptos recogidos en los libros de texto actuales.

Los resultados de ENCODE han demostrado que la expresión de los genes está influenciada no solo por secuencias de ADN reguladoras cercanas a estos, sino por otras regiones muy distantes.

Además, se ha demostrado que una parte de la regulación de los genes está controlada también por moléculas de ARN

¿Cómo se controla la activación de los genes?

Aparte de codificar las proteínas esenciales para el funcionamiento de nuestras células, una gran proporción del genoma humano se encarga de regular si un gen responsable de una proteína se activará o no. El proyecto ENCODE ha profundizado en el estudio de los mecanismos de regulación que se detallan en la ilustración.

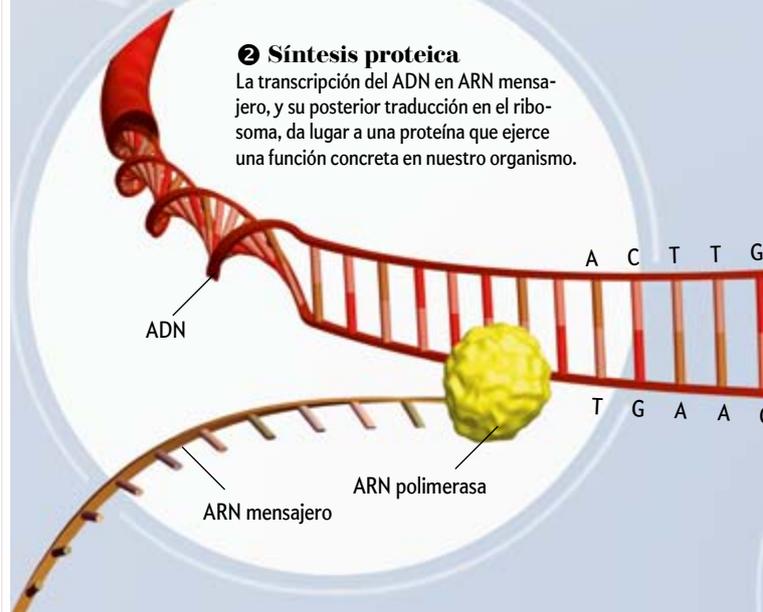


1 Reconocimiento de las secuencias reguladoras

La maquinaria transcripcional reconoce las secuencias reguladoras de ADN (barras de distintos colores sobre la fibra de ADN, en rojo). En el reconocimiento participa la ARN polimerasa, enzima que se asocia a diferentes factores reguladores (verde oscuro). Este complejo se une por un lado al promotor (caja TATA, blanco) y por otro, a otras secuencias (por ejemplo, potenciadores) aguas arriba del promotor.

2 Síntesis proteica

La transcripción del ADN en ARN mensajero, y su posterior traducción en el ribosoma, da lugar a una proteína que ejerce una función concreta en nuestro organismo.

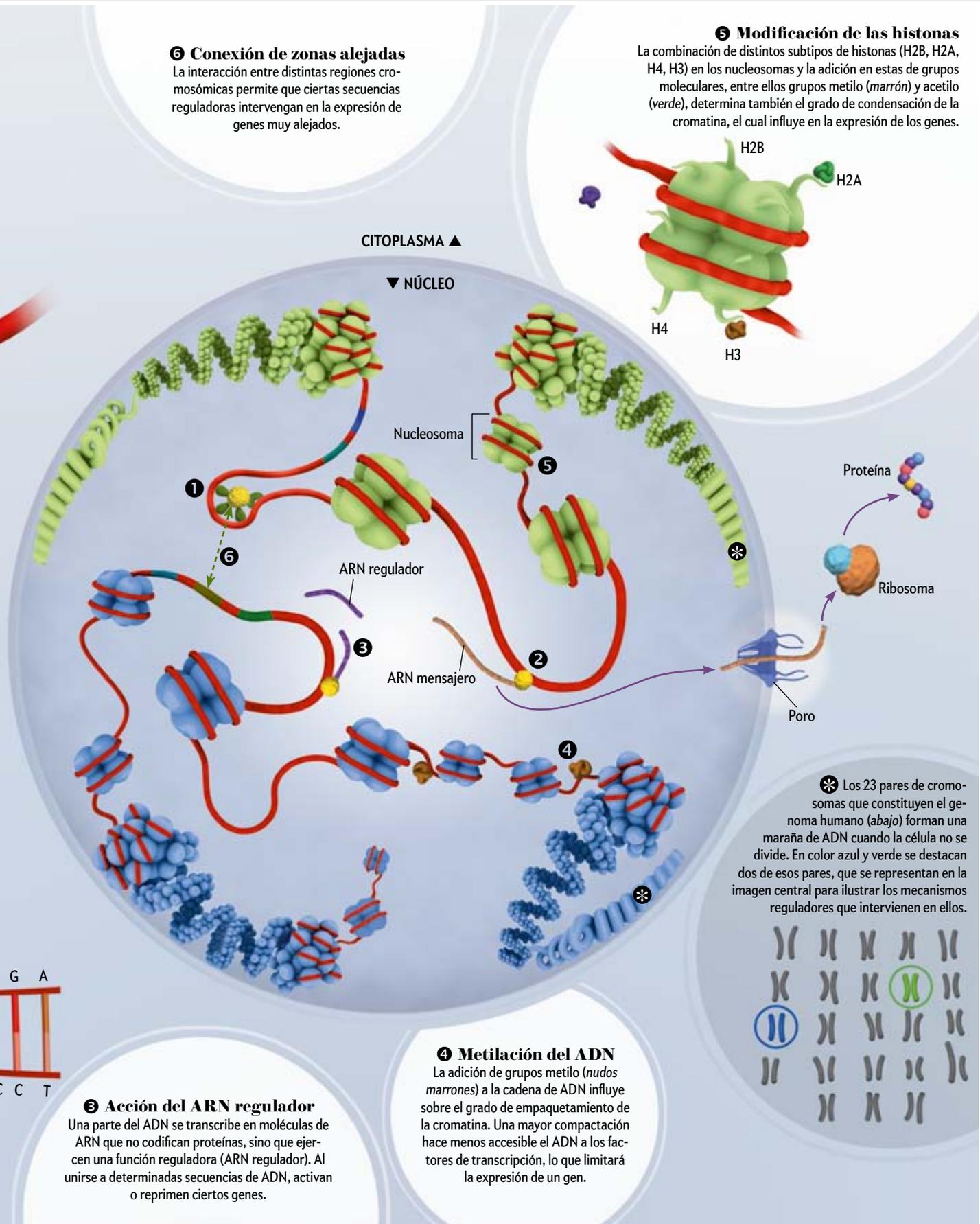


6 Conexión de zonas alejadas

La interacción entre distintas regiones cromosómicas permite que ciertas secuencias reguladoras intervengan en la expresión de genes muy alejados.

5 Modificación de las histonas

La combinación de distintos subtipos de histonas (H2B, H2A, H4, H3) en los nucleosomas y la adición en estas de grupos moleculares, entre ellos grupos metilo (marrón) y acetilo (verde), determina también el grado de condensación de la cromatina, el cual influye en la expresión de los genes.



especiales, que no codifican proteínas, y por modificaciones epigenéticas. Estas últimas consisten en cambios químicos en la secuencia de ADN o en las proteínas asociadas a él (las histonas), que repercuten en el grado de empaquetamiento del ADN y, por tanto, en la accesibilidad que tienen a él las moléculas que regulan la activación de los genes. Cuando un fragmento del ADN se halla empaquetado, no se pueden leer sus genes; por consiguiente, el grado de compactación del ADN representa un importante mecanismo de regulación de la expresión génica [véase «Evolución de la cromatina», por G. A. Babbitt; INVESTIGACIÓN Y CIENCIA, mayo de 2011].

Estos resultados han supuesto un cambio en la percepción de la regulación génica como un proceso lineal. Se ha demostrado que no solo importa el contexto genómico de un gen, sino también la disposición tridimensional del ADN y la localización de ese gen en una región (o territorio) concreta del núcleo celular. Detallamos a continuación los datos más relevantes que ha obtenido el proyecto ENCODE sobre los elementos funcionales del genoma.

UN ABANICO DE ARN

El 62 por ciento de la secuencia del genoma humano da lugar a transcritos. Estos corresponden a moléculas de ARN de dos tipos principales: los que originan una proteína (transcritos codificantes) y los que no (transcritos no codificantes). En el primer caso, la secuencia de ADN se transcribe en una cadena de ARN mensajero (ARNm), que posteriormente será «traducida» por la maquinaria celular en una proteína, como la insulina o la hemoglobina. El genoma humano contiene 20.687 secuencias codificantes, que representan el 2,9 por ciento del total de la información contenida en el genoma.

En el segundo caso, a partir del ADN se formará ARN como producto final, pero este no dará lugar a una proteína, sino que funcionará como elemento regulador de la expresión de otros genes, en unos casos inhibiéndolos y en otros activándolos, según el tipo de ARN del que se trate. Entre las distintas clases de ARN destacan los largos no codificantes (ARNlnc, *long non-coding*) y los pequeños (ARNs, *small*), que cuentan con 9640 y 8801 tipos, respectivamente. Existen varias subclases de ARN pequeños, como el de transferencia (ARNt), el micro ARN (ARNmi), el ARN pequeño nuclear (ARNsn, *small nuclear*) y el ARN pequeño nucleolar (ARNsno, *small nucleolar*), que se clasifican según su localización celular y su lugar de actuación. Por ejemplo, los micro ARN actúan reduciendo los niveles de expresión de otros genes mediante su unión a moléculas de ARN mensajero parcialmente complementarias.

En los últimos años se ha descubierto la importancia de los ARN no codificantes para el funcionamiento y regulación génica. Estos actúan a lo largo de casi todo el genoma, aunque la mayoría se sitúan cerca de los genes o en intrones (secuencias que forman parte de un gen pero que son eliminadas una vez que este se transcribe en ARN mensajero). La variedad y cantidad de ARN en las células constituyen un indicio de la relevancia de estas moléculas.

Para elaborar un catálogo de todos los tipos de ARN que existen, el proyecto ENCODE se sirvió de distintos tipos celulares, ya que cada subconjunto de ARN reguladores y codificantes confiere identidad morfológica y funcional a cada tipo de célula.

REGIONES REGULADORAS

A pesar de que todas las células de nuestro organismo poseen el mismo libro de instrucciones, esto es, la misma información

genética, nuestro cuerpo posee múltiples tipos celulares que conforman los distintos tejidos y órganos. Mientras que todas las células usan una parte común del genoma, algo necesario para llevar a cabo las funciones básicas y ordinarias (como producir energía, recambiar los componentes esenciales, etcétera), cada tipo celular emplea además una parte específica para llevar a cabo tareas concretas dentro del organismo. De este modo, en las neuronas se activan una serie de genes que permanecen inhibidos en las células musculares y que determinan su morfología y funciones particulares, y viceversa. La parte del genoma que un tipo celular no utiliza se reprime y, por consiguiente, dejan de sintetizarse ciertas proteínas. Lo que distingue a unas células de otras es la expresión diferencial de determinados genes y ARN reguladores.

Algunos genes presentan una expresión diferencial en el tiempo, ya que codifican proteínas solo en determinados momentos de la vida de un individuo. Ello reviste especial importancia durante el desarrollo embrionario: existen genes que se activan en cortos períodos de tiempo, cuando se está formando el organismo, y luego son reprimidos para siempre.

La expresión diferencial que da lugar a cada tipo celular es posible gracias a la existencia en el genoma de numerosas regiones reguladoras y de factores que interaccionan con ellas.

Los factores corresponden a proteínas con función reguladora y pueden ser de varios tipos: de transcripción, modificadores de la cromatina o los que intervienen en la unión de otros factores con la doble hélice de ADN, entre otros. Los más conocidos son los factores de transcripción, que se unen a determinadas secuencias de ADN y regulan la transcripción de este en ARN. Entre los distintos factores de transcripción, unos son necesarios para la síntesis de proteínas, que permiten a las células desempeñar sus funciones vitales; otros controlan la expresión de proteínas o transcritos no codificantes (con función reguladora) específicos de cada tipo celular; y otros se activan como respuesta a determinadas situaciones, como la existencia de un patógeno o la ausencia o presencia de un nutriente o una condición de estrés.

Las regiones reguladoras del genoma, o interruptores génicos, corresponden a secuencias de ADN cortas, que operan a modo de códigos de barras, a las que se unen los factores de regulación específicos. En muchos casos, tales regiones se ubican cerca del gen cuya expresión regulan e incluyen elementos funcionales muy diversos: cabe destacar los promotores, que promueven la transcripción que producirá un ARN; los potenciadores, que aumentan la cantidad de ARN que se produce a partir de un gen; los silenciadores, con un efecto contrario a los potenciadores; y los aislantes, que aíslan los elementos reguladores de ciertos genes o grupos de genes de otros vecinos.

El proyecto ENCODE ha realizado un amplio estudio de los elementos reguladores mediante múltiples aproximaciones experimentales. Uno de sus objetivos ha consistido en identificar las regiones del ADN donde se unen proteínas reguladoras, la mayoría de las cuales corresponde a factores de transcripción, como ya se ha comentado. Para ello se seleccionaron 119 proteínas de unión al ADN (las más representativas, entre ellas 87 factores de transcripción y varios componentes de la ARN polimerasa, la enzima encargada de copiar el ADN en ARN). Este análisis se realizó en 72 tipos celulares y se emplearon diversas técnicas de análisis biológico, como la inmunoprecipitación de cromatina seguida de secuenciación (*ChIP-sequencing*). La técnica consiste en inmovilizar los factores unidos al ADN mediante

EL PROYECTO ENCODE, EN CIFRAS

Con el objetivo de describir todos los elementos funcionales de nuestro genoma, el Instituto Nacional de Investigación del Genoma Humano de Estados Unidos lanzó en 2003 el proyecto ENCODE. Pretendía paliar el vacío en el conocimiento acerca del 98,5 por ciento de nuestro ADN, cuya función no se sabía interpretar después de haber finalizado su secuenciación en 2001.

La magnitud del proyecto, que finalizó en 2012, se refleja en los siguientes datos:



442

investigadores participantes, procedentes de **32** instituciones de EE. UU., Reino Unido, Singapur, Japón, Suiza y España, entre otros países



147

tipos celulares estudiados



288

millones de dólares de financiación



1770

conjuntos de datos generados

Los resultados han generado 30 artículos científicos que arrojan luz sobre la función del genoma no codificante. Vemos ahora que las secuencias que antes se consideraban irrelevantes desempeñan un importante papel regulador. Las conclusiones principales del estudio sobre los elementos funcionales del ADN y su proporción en el genoma se resumen aquí:



80,4 %

Elementos funcionales



2,9 %

Secuencias codificantes de proteínas (20.687 genes)



62 %

Secuencias que se transcriben en ARN



8,1 %

Sitios de unión de factores reguladores



15,2 %

ADN no compactado (con cromatina accesible)



56,1 %

Secuencias enriquecidas con modificaciones o variantes de histonas



98 %

Secuencias promotoras que interactúan con otras regiones del ADN dentro del mismo cromosoma

un tratamiento químico; estos se purifican y, a continuación, se secuencian e identifican los fragmentos de ADN a los que estaban unidos. El análisis permitió establecer que, en promedio, el 8,1 por ciento del genoma corresponde a puntos de unión con proteínas reguladoras. La gran mayoría de los fragmentos de ADN donde se fijan los factores reguladores presentan motivos o secuencias de unión específicas para cada factor, que adquieren una estructura y propiedades que favorecen la interacción con unos reguladores pero no con otros.

ACCESIBILIDAD DE LA CROMATINA

El ADN se encuentra enrollado sobre unas proteínas (histonas) y el conjunto forma los nucleosomas, la estructura básica de la cromatina [véase «El papel clave de las histonas», por R. González Romero et al.; INVESTIGACIÓN Y CIENCIA, diciembre de 2011]. A su vez, la cromatina puede hallarse más o menos compactada, lo que determina la accesibilidad de factores implicados en la regulación de la expresión génica. Una mayor compactación puede limitar el acceso de los factores responsables de transcribir el ADN en ARN y, por consiguiente, dificultar la activación de un gen.

Las zonas donde el ADN se asocia a las histonas pueden identificarse mediante la técnica de hipersensibilidad de la cromatina a la DNasa I (DHS). La DNasa I es una endonucleasa, una enzima capaz de cortar regiones de ADN desnudas. Las regiones del ADN unidas a las histonas resultan inaccesibles a la enzima, por lo que quedan intactas tras la digestión. Mediante esta técnica, el proyecto ENCODE ha determinado que el 15,2 por ciento del genoma consiste en regiones de cromatina accesibles (ADN no compactado) y que el 94,4 por ciento de los factores

reguladores unidos a sus correspondientes regiones reguladoras se hallan dentro de estas regiones no compactadas.

Este análisis ha constituido otra de las herramientas experimentales para identificar las regiones del ADN con función reguladora. Gracias a él, se ha creado una colección de motivos reguladores. En 41 tipos de células humanas se han descrito más de 8,4 millones de secuencias reguladoras (unos 200.000 por tipo celular) y 45 millones de sitios de unión de proteínas reguladoras. La identificación de tales elementos reviste una importancia especial, puesto que el origen de numerosas enfermedades humanas podría hallarse en estas pequeñas secuencias.

EL CÓDIGO EPIGENÉTICO

Las histonas que contribuyen al empaquetamiento del ADN pueden ser modificadas químicamente mediante múltiples mecanismos, entre ellos la metilación y la acetilación (adición de un grupo metilo o acetilo, respectivamente). Asimismo, existen distintas variantes de histonas (subtipos de histonas con una secuencia y estructura proteica similares). Tanto las modificaciones de las histonas como la existencia de diferentes variantes aumentan o disminuyen la afinidad de estas proteínas con el ADN, lo que confiere a la cromatina distintas propiedades de accesibilidad y regulación.

Una de las metas del proyecto ENCODE ha consistido en determinar las modificaciones que presentan las histonas a lo largo de todo el genoma. En este caso se empleó también la técnica de inmunoprecipitación de cromatina seguida de secuenciación para detectar de forma precisa a qué regiones del genoma se hallaban unidas varias formas modificadas de

histonas. En total, se estudiaron 12 variantes de histonas en 46 tipos celulares humanos y se determinó que el 56 por ciento del genoma se halla altamente enriquecido en estas. Es decir, las variantes de histonas y sus modificaciones químicas no se distribuyen de forma homogénea, sino que tienden a acumularse en zonas concretas del genoma. De este modo, las histonas metiladas suelen concentrarse en las regiones promotoras de genes reprimidos, mientras que las acetiladas se sitúan con preferencia en torno a genes activos.

Por otro lado, el análisis reveló que los estados modificados de las histonas, así como la actividad transcripcional de las regiones asociadas a ciertas variantes de histonas, varía enormemente de un tipo celular a otro; es decir, el hecho de que una célula se diferencie en un tipo u otro dependerá en parte de la composición en sus diferentes variantes de histonas y de la modificación química que experimenten estas.

Otra forma mediante la que se regula la expresión de los genes es la modificación química de algunos nucleótidos que componen la secuencia de nuestro genoma. La metilación del ADN, junto con la modificación de las histonas mencionada, forman parte del código epigenético.

La metilación del ADN consiste en la adición de un grupo metilo en las bases citosina, generalmente en las secuencias genómicas donde se repite el dinucleótido CG, llamadas islas CpG, que suelen aparecer dentro o cerca de los promotores de los genes o sitios de inicio de la transcripción. La presencia de estas metilaciones en las regiones reguladoras hace que unos genes se expresen y otros permanezcan inactivos. Determinadas enfermedades se producen no porque un gen esté dañado, sino porque el patrón de metilación de la región reguladora (promotor o interruptor génico) está alterado, lo que provoca que un gen se exprese cuando no debe, o viceversa.

Mediante un tratamiento químico del ADN (técnica del bisulfito) pueden identificarse las citosinas del genoma que se encuentran metiladas. Con el empleo de este método, el proyecto ENCODE se propuso elaborar un perfil de metilación de todo el ADN del genoma humano. Se analizaron 82 líneas celulares (tipos de células humanas inmortalizadas mantenidas en cultivo), así como distintos tejidos humanos. Los resultados han puesto de manifiesto que las islas CpG presentan una metilación diferencial en los diversos tipos de células; el grado de metilación guarda relación con el empaquetamiento de la cromatina y, por tanto, con la accesibilidad que tienen a ella otras moléculas. Además, se ha observado que ciertas líneas celulares, como las cancerosas, muestran patrones de metilación aberrantes.

INTERACCIÓN CROMOSÓMICA

En los últimos años se ha descubierto un nuevo nivel de regulación de la expresión génica: la interacción física entre distintas regiones cromosómicas, a menudo separadas por cientos de miles de nucleótidos. De acuerdo con esta visión, la expresión de un gen no está regulada por un solo interruptor sino por muchos; una gran parte de ellos son compartidos con otros genes y se encuentran alejados del gen al que regulan. Se ha visto así que la cromatina forma una compleja red tridimensional de interconexión entre distintas partes del genoma. Estas regiones conectadas corresponden a zonas donde se están expresando genes activamente y, por tanto, se hallan menos enrolladas.

Para el análisis de esas interacciones ha sido necesario desarrollar nuevas técnicas moleculares. Cabe destacar la técnica de captura de conformación cromosómica, o 3C, y su variante 5C

(copia en carbón de la técnica 3C). Tales métodos han detectado cientos de interacciones entre distintas partes del genoma. Los pares de locus (genes o secuencias de ADN específicas) que interaccionan muestran una fuerte semejanza en los niveles de expresión génica y en la presencia de elementos funcionales específicos.

De este modo, un sitio de inicio de la transcripción (TSS, por sus siglas en inglés), el lugar donde comienza la transcripción de un gen, puede interaccionar con un promedio de 3,9 regiones de ADN distintas que podrían estar regulando su actividad, entre ellas regiones potenciadoras. Por tanto, estos potenciadores podrían ejercer su función en un gen muy alejado del lugar del genoma que ocupan. Otro dato revelador es que el 98 por ciento de los promotores interaccionan con otras regiones del ADN dentro del mismo cromosoma. Asimismo, el proyecto ENCODE ha determinado que la red de interconexión entre secuencias del genoma posee una arquitectura característica en cada tejido o tipo celular.

INTEGRAR LOS ELEMENTOS REGULADORES

Los distintos mecanismos de regulación mencionados no constituyen elementos aislados, sino que actúan conjuntamente para controlar la expresión de los genes, e incluso dependen unos de otros. El proyecto ENCODE ha elaborado unos complejos mapas de correlación que incluyen todos estos elementos. La herramienta permite conocer cómo influyen unos sobre otros, y cómo ello determina la regulación del genoma.

Por ejemplo, los factores de transcripción no se distribuyen de forma aleatoria, sino que se unen con preferencia a motivos específicos de los promotores que presentan un alto grado de metilación en las secuencias CpG. Los factores de transcripción, a su vez, pueden formar barreras alrededor de las cuales tienen lugar las modificaciones de las histonas y la remodelación de los nucleosomas.

Por otro lado, los ARN pequeños, uno de los tipos de transcritos no codificantes, se concentran con preferencia en las zonas promotoras de los genes, mientras que las metilaciones se sitúan principalmente en las regiones promotoras de genes reprimidos y, en menor medida, en los genes activos.

Como ya se ha comentado, cada tipo celular o tejido humano muestra unas características de regulación intrínsecas, pero también existen numerosas diferencias entre unas personas y otras. Las variaciones individuales se han analizado también en el proyecto ENCODE. Se ha hecho especial hincapié en las variantes alélicas (las diversas formas que puede presentar un gen en distintos individuos), entre ellas los polimorfismos de un solo nucleótido (SNP, de *single nucleotide polymorphism*). Estos consisten en la variación en la secuencia de ADN que afecta a una sola base, lo que en ocasiones puede dar lugar a una pérdida de función. Tales variantes funcionales se ubican no solo en las regiones codificantes, sino también en las no codificantes. Así, se ha comprobado que gran parte de los SNP se asocian a un fenotipo y, lo que resulta más sorprendente, que el 88 por ciento de ellos se sitúan en regiones no codificantes, como intrones y regiones intergénicas. Uno de los puntos importantes del estudio ha sido la identificación de ese tipo de variantes funcionales relacionadas con diversas enfermedades, desde afecciones inmunitarias hasta el cáncer.

Sin embargo, a pesar de la enorme cantidad de datos obtenidos, aún quedan numerosos elementos funcionales por definir, repartidos entre todos los tipos celulares que componen el cuerpo humano. Por ello, el NHGRI planea impulsar y financiar

durante cuatro años más las investigaciones que profundicen en el análisis de los elementos funcionales catalogados e incluyan otros tipos celulares, además de facilitar el desarrollo de nuevas técnicas que simplifiquen los estudios genómicos.

IMPlicACIONES DE LOS DESCUBRIMIENTOS

Hace tiempo se pensaba que las enfermedades se debían fundamentalmente a defectos en la función de las proteínas, pero cada vez se descubren más enfermedades asociadas a fenómenos de desregulación génica. Dicho de otro modo, los trastornos no tienen solo su origen en las regiones codificantes, sino también en sus regiones colindantes, en zonas del genoma muy alejadas de un determinado gen, en pequeños ARN no codificantes reguladores y en una amplia diversidad de factores externos que repercuten en la epigenética de nuestras células. Los cambios en la regulación de la actividad génica pueden alterar la producción de proteínas y los procesos celulares y desencadenar una enfermedad. Tan perjudicial puede resultar la falta de función de una proteína como su exceso, así como la expresión de una proteína en un momento o en un lugar inadecuado.

Muchos de los elementos funcionales descubiertos a menudo se relacionan con regiones de ADN responsables de varias enfermedades, lo cual sugiere que la regulación de gran parte de estos genes influye en el riesgo o la predisposición a padecer ciertas dolencias. De este modo, se han identificado cinco SNP asociados a la enfermedad de Crohn situados en la región de unión del factor de transcripción GATA2.

La amplia variedad de datos que está proporcionando el proyecto ENCODE está contribuyendo a desentrañar la base

genética de diversas enfermedades. El conocimiento de la relación entre el funcionamiento del genoma y la salud permitirá determinar con antelación los factores de riesgo y ayudará a prevenir sus efectos o a desarrollar terapias que se ajusten a cada paciente. Cuanta más información se disponga de la estructura del genoma y del modo en que este se regula, más fácil resultará identificar los factores de riesgo genéticos relacionados con las enfermedades.

Desde la secuenciación del genoma humano en el año 2001, el conocimiento que teníamos sobre él ha cambiado de modo notable. El ambicioso proyecto ENCODE está teniendo repercusiones importantes, que van desde la redefinición del concepto de gen hasta la obtención de nuevas pistas sobre las causas genéticas de las enfermedades. También ha transformado la visión que teníamos sobre nuestro genoma, dejando atrás la idea de ADN «basura» con la que se definía una gran parte de él y cuya función desconocíamos.

PARA SABER MÁS

ENCODE: More genomic empowerment. George M. Weinstock en *Genome Research*, vol. 17, págs. 667-668, 2007.

A user's guide to the encyclopedia of DNA elements (ENCODE). The ENCODE project consortium en *PLoS Biology*, vol. 9, n.º 4, pág. e1001046, 2011.

An integrated encyclopedia of DNA elements in the human genome. ENCODE project consortium. B. E. Bernstein et al. en *Nature*, vol. 489, n.º 7414, págs. 57-74, 2012.

ENCODE project writes eulogy for junk DNA. E. Pennisi en *Science*, 337, n.º 6099, págs. 1159, 1161, 2012.

Viaje al interior del genoma. Stephen S. Hall en *Investigación y Ciencia*, diciembre de 2012.

Página web del proyecto ENCODE: encodeproject.org

Licencias para instituciones

Acceso permanente a todos nuestros contenidos a través de Internet



INVESTIGACIÓN
Y CIENCIA

MENTE Y CEREBRO

Nuevo servicio para bibliotecas, escuelas, institutos, universidades, centros de investigación o empresas que deseen ofrecer a sus usuarios acceso libre a todos los artículos de *Investigación y Ciencia* y *Mente y cerebro*.

Más información en
www.nature.com/libraries/iyc

nature publishing group 