

GENÉTICA CUATRO REVISTAS CIENTÍFICAS PUBLICAN EL PROYECTO 'ENCODE', LA ENCICLOPEDIA DE ELEMENTOS DE ADN

El 80% del genoma tiene al menos una actividad bioquímica asignada

→ Hace once años se obtuvo la secuenciación completa del genoma humano y se constató que sólo el 2 por ciento contenía genes. Al resto se le denominó *ADN basura*. Para descifrar su

función, se puso en marcha el proyecto *Encode*, una enciclopedia de elementos de ADN que ha demostrado que el 80 por ciento sí tiene una función específica.

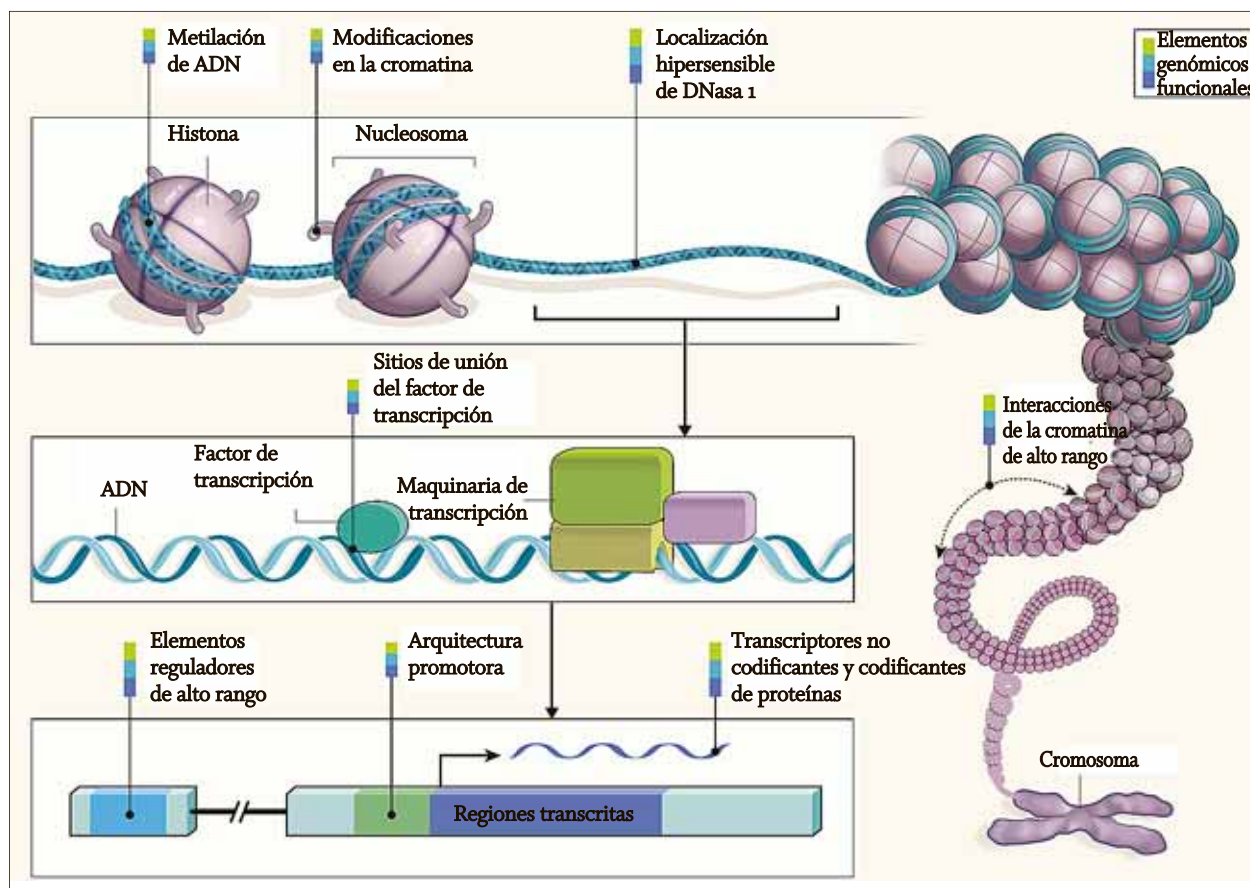
■ Redacción

El proyecto Genoma Humano mostró que sólo el 2 por ciento del genoma contiene genes. Ahora, el *Encode*, una enciclopedia de elementos de ADN, dice que el 80 por ciento del genoma tiene al menos una actividad bioquímica asignada. Los datos de dicho proyecto se publican hoy en *Nature*, en *Genome Research*, en *Genome Biology* y en *Science*.

En su elaboración han colaborado 442 científicos de 32 laboratorios repartidos por Reino Unido, Estados Unidos, España, Singapur, Japón y Suiza, y se ha visto que lo que se llamaba, por ignorancia, *ADN basura* no es tal, puesto que se trata de un conjunto de lo que se podría denominar interruptores que regulan la actividad de los genes. Sin estos elementos los genes no funcionarían y las mutaciones en estas regiones descubiertas podrían ser responsables de diferentes enfermedades.

La información detallada en la enciclopedia muestra un mapa de la función del genoma que identifica 4 millones de interruptores de genes. Esta referencia ayudará a localizar áreas muy específicas de enfermedades humanas.

Ewan Birney, investigador del Instituto de Bioinformática Europeo y coordinador de análisis del proyecto, ha explicado que han encontrado que una gran parte del genoma está implicada en controlar cuándo y dónde se producen las proteínas más



allá de su fabricación.

Se han generado y analizado cerca de 15 terabytes (15 billones de bytes) de datos en bruto, y están disponibles de forma pública. De hecho, *Nature* ha desarrollado una plataforma *on-line* para su publicación.

El estudio ha utilizado alrededor de 300 años en tiempo de ordenador para estudiar 147 tipos de tejido y determinar qué enciende o apaga a genes específicos y cómo ese interruptor difiere entre tejidos o tipos celulares.

Los datos del *Encode* ofre-

cen el conocimiento necesario para analizar más allá de la estructura lineal del genoma y ver cómo toda la red está conectada. Tan importante es saber dónde están ubicados ciertos genes como qué secuencias controlan.

■ Información escrita

A consecuencia de la compleja estructura tridimensional del genoma, en ocasiones dichos controles están lejos del gen que regulan si se lee la secuencia linealmente, pero si se hace de forma tridimensional se observa que se encuentran en-

vuellos alrededor para contactar con ellos. Si no fuera por la enciclopedia de elementos de ADN, probablemente nunca se hubieran analizado estas regiones.

Según Birney, se trata de un gran paso hacia la comprensión del complejo diagrama de cableado del ser humano. "Nos ayudará a estudiar en lo más hondo del circuito de regulación que cuenta cómo todas las partes se unen para configurar un ser complejo". Con el *Encode* se ha visto que el 24,4 por ciento de las variaciones en las secuencias de ADN se se-

asocia con enfermedades autoinmunitarias. En dichas localizaciones se producen alteraciones en los sitios de unión de los factores de transcripción, lo que supone una información relevante sobre las redes reguladoras de las enfermedades autoinmunes. También se ha localizado el 76 por ciento de las variantes asociadas a enfermedades en regiones no genéticas. Esta información ayudará a descifrar gran número de trastornos.

■ (*Nature* 2012; 489: 57-113/*Science* 2012; 337: 1.190-1.196).

CON EL CRG

Veinte científicos del *Encode* son del Centro de Regulación Genómica (CRG) de Barcelona; entre ellos, Roderic Guigó, coordinador de Bioinformática y Genómica, que ha liderado el grupo de análisis de ARN. Según ha explicado a este diario, la tecnología que utilizan para determinar regiones del genoma que producen ARN son muchísimo más sofisticadas que las que el mismo equipo usó en la secuenciación por primera vez del genoma humano (con Celera). "Ahora catalogamos el ARN que se produce en diferentes tipos de células (cerebro, piel, huesos, músculos...). El ADN de un individuo es el mismo en todas sus células pero lo que las diferencia es el ARN", ha precisado. Con la actual tecnología pueden ver cuál es la actividad en cada región del genoma en condiciones normales y en caso de enfermedad, algo que aún tardará décadas en llegar a la clínica. "Aportamos conocimiento fundamental, fundamentos sobre los que construir el edificio. Si entendemos cómo funciona el genoma, sabremos cómo arreglarlo".



Roderic Guigó.

UN 'BANQUETE' DE DATOS QUE DEGUSTARÁN LOS MEJORES 'PALADARES' CIENTÍFICOS

Seis científicos exponen en un artículo del mismo número de *Nature* los retos de la investigación genómica tras la difusión de los datos del proyecto *Encode*.

■ **Redefinición de gen.** Joseph Ecker, del Instituto Médico Howard Hughes, en California, considera que la *digestión* de los nuevos hallazgos obligará a replantearse cuestiones como la definición de gen y de la unidad mínima de herencia y la concepción de que la regulación de un gen está dominada por su proximidad a los elementos reguladores más cercanos.

■ **Control de la expresión.** Los investigadores de *Encode* han centrado sus esfuerzos en un conjunto de líneas celulares bien caracterizadas, según Wendy Bickmore,

de la Universidad de Edimburgo. En su opinión, uno de los retos será hacer un seguimiento de los cambios dinámicos que se producen en la regulación de rutas de desarrollo específicas. También aconseja profundizar en el conocimiento de la estructura de la cromatina en los tejidos con poblaciones celulares heterogéneas.

■ **No codificante pero funcional.** Inês Barroso, del Instituto Wellcome Trust Sanger, en Reino Unido, señala que el proyecto deja patente que al interpretar los resultados de los estudios de asociación del genoma completo deben tenerse en cuenta las regiones no codificantes. De hecho, cuando se estudian sólo secuencias que codifican proteínas (el exoma), se corre

el riesgo de dejar sin analizar regiones cruciales del genoma, comprometiendo la capacidad para descubrir auténticas variantes causales de enfermedades.

■ **Evolución humana.** Los nuevos datos acelerarán la identificación de los cambios clave de secuencia en los elementos reguladores que están detrás de las diferencias funcionales entre los humanos y otras especies. Así lo aseguran Jonathan Pritchard y Yoav Gilad, de la Universidad de Chicago.

■ **Papel de la informática.** Eran Segal, del Instituto Weizmann, en Israel, cree que la informática será esencial para lograr el objetivo -aún lejano- de conocer la función del genoma en cada célula.