

El genoma humano no es como se creía

Un estudio rompe dogmas: los genes no son independientes y el ADN 'basura' tiene funciones

MÓNICA L. FERRADO, Barcelona

El genoma humano todavía alberga muchos incógnitas. Su secuenciación en 2003 fue un gran logro, pero tan sólo un primer paso. El mapa está listo, pero todavía queda mucho por andar para tener el manual de instrucciones completo, y por el camino pueden deshacerse algunas de las ideas más arraigadas hasta el momento sobre su estructura y funcionamiento. Así lo acaban de demostrar los resultados del proyecto internacional Encode, de la gran *Enciclopedia de los elementos del ADN*, un consorcio internacional de investigadores encabezados por el Nacional Human Genome Research Institute (NHGRI), el mismo instituto estadounidense que dirigió el Proyecto Genoma Humano.

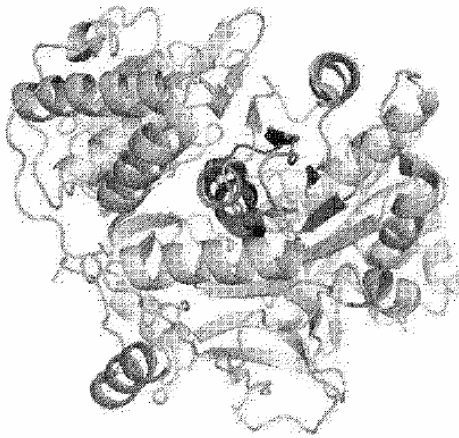
Hasta ahora se creía que los genes eran unidades independientes, separadas por huecos en los que sólo había ADN *basura* (se estima que ocupa el 95% del genoma), insertible y sin ninguna función. También se pensaba que los únicos que podían codificar eran los genes y, en concreto, que cada uno codificaba una sola proteína. Pues bien, los resultados del proyecto Encode han demostrado todo este engranaje biológico.

Los genes no son unidades independientes, sino que muchos se superponen y actúan en red compartiendo información. Además, es muy frecuente que codifiquen más de una proteína y ese 95% de ADN *basura* resulta que importa, y mucho. Por si fuera poco, todos estos elementos interactúan de maneras complejas y sobrepuestas. Así pues, la estructura y el funcionamiento de un solo gen presenta todavía muchas incógnitas.

Con toda esta información sobre la mesa, "la comunidad científica va a tener que replantearse algunas ideas que han mantenido durante mucho tiempo sobre qué son los genes y qué hacen, además de cómo han evolucionado los elementos funcionales: esto podría tener grandes implicaciones en muchas enfermedades humanas", afirma Francis S. Collins, director de NHGRI.

Estas conclusiones son el resultado de cuatro años de investigaciones para desvelar una lista de todos los elementos biológicamente funcionales en el 1% del genoma. En esta pequeña porción del libro de la vida han aplicado todo su conocimiento y tecnología más de 300 científicos de 80 organizaciones de todo el mundo. Es la iniciativa más ambiciosa desde que finalizara el mapa del genoma y ha servido como proyecto piloto, ya que, a partir del próximo mes de octubre, se empezará a hacer lo mismo con el 99% restante, con la intención de acabarlo en dos años.

En el proyecto piloto Encode han participado 14 investigadores españoles del Centro de Regulación Genómica de Barcelona (CRG), de la Universidad de Barcelona y del Centro Nacional de Investigaciones Oncológicas (CNIO). Roderic Guigó, bioinformático e investigador del CRG, ha dirigido uno de los cinco grandes grupos de trabajo del proyecto, el de transcripción y genes, y confiesa que los resultados rompen el dogma central de la biología, según el cual "hasta ahora se pensa-



Modelo de la estructura tridimensional de una proteína junto al esquema de una de las regiones genéticas estudiadas en el proyecto Encode. Fuente: CNIO

El proyecto Encode se centra en 44 regiones, que cubren apenas un 1% de todo el genoma

Los nuevos datos obligan a la comunidad científica a replantear qué son los genes y qué hacen

ba que el ADN generaba ARN mensajeros, que producían proteínas con una función concreta, pero hemos visto que el ARN, aunque no codifique proteínas, también puede tener una función por sí mismo". También se ha comprobado que existe una interacción entre los diferentes genes: "Hemos visto que se tocan, que se sobrepone y que incluso comparten lecturas. Hasta ahora, esta interacción sólo se había probado en casos muy específicos, pero no se sabía que fuese tan común".

En los últimos años, los investigadores han realizado grandes progresos para identificar genes que codifican las proteínas. Sin embargo, los genes tan sólo representan una pequeña fracción del genoma humano (entre el 1,5% y el 2%). Guigó añade que los resultados, que amplían el catálogo de elementos funcionales del genoma, tienen un impacto directo sobre las investigaciones que se es-

tán realizando actualmente en genética: "A partir de ahora tendremos que abrir los ojos y tener un panorama más amplio".

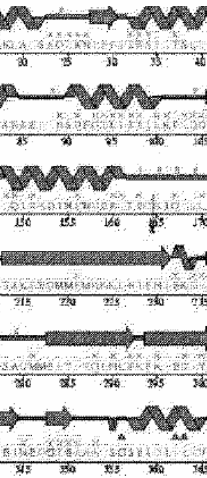
La investigación sobre enfermedades humanas con una base genética va a ser más compleja. Otro investigador español que ha participado en el proyecto es Xavier Estivill, responsable del programa Genes y Enfermedad del CRG. "Durante mucho tiempo se ha estado buscando mutaciones en los genes para enfermedades, pero los resultados del proyecto indican que tenemos que buscar más allá, que es necesaria una aproximación más global a la hora de estudiar enfermedades", concluye Estivill.

El proyecto Encode se centra en 44 regiones que cubren el 1% del genoma, de las que 30 han sido seleccionadas al azar; y 14 se corresponden con genes o grupos de genes ya identificados y relacionados con enfermedades, como el

gen de la fibrosis quística, el de la hemoglobina, o el grupo de genes *hox*, que establecen el patrón anatómico de los vertebrados. "Los principales genes que causan enfermedades como la fibrosis quística o la enfermedad de Alzheimer, ya los tenemos", explica Estivill, "pero quizá haya alteraciones en otras zonas que deberemos estudiar con mayor detalle y afinar en el porcentaje de riesgo real de padecer una enfermedad".

Estivill ha estudiado la presencia de fragmentos de material genético replicado, comparando muestras de 270 individuos europeos, asiáticos y africanos. "Existe una gran variabilidad en el número de fragmentos replicados en función del origen de los individuos", afirma, lo que indica una mayor o menor susceptibilidad a desarrollar ciertas enfermedades.

Alfonso Valencia, director del programa de biología estructural y biocomputación del CNIO y di-



rector del Instituto Nacional de Bioinformática, también ha participado en el proyecto coordinando la participación de Biosapiens, una red europea que se ha centrado en el estudio de las proteínas, y reconoce que la investigación en torno a la proteómica también va a ser más compleja. "Pensábamos que cada uno de los 20.000-25.000 genes del genoma producía una proteína, pero hemos visto que pueden producir hasta cinco, lo que haría un total de hasta 100.000. Además, con Encode hemos visto que son proteínas con variaciones drásticas y que no sabemos qué hacen".

Los datos de Encode también aportan información sobre la evolución de los seres humanos como especie. La mitad de los elementos funcionales del genoma humano no parecen haber estado sometidos a restricciones evolutivas. Esta reserva de elementos funcionales, que no proporciona beneficios específicos en términos de supervivencia, muestra que "todas las funciones esenciales para la célula han estado muy bien atadas, no pueden variar, aunque otras, como el sistema inmune, son muy variables entre individuos", explica Estivill.

España da un paso adelante, y dos atrás

España ha participado en este proyecto con 14 investigadores. Uno de ellos, Roderic Guigó, ha tenido un papel relevante, como uno de los nueve coordinadores principales de este estudio, impulsado desde el NHG, el mismo instituto donde se descifró el genoma humano. "Hasta ahora éramos comparas, ahora ya hemos sido protagonistas", afirma

Roderic Guigó, que participó aportando parte del *software* para la secuenciación del genoma humano en el año 2003. Sin embargo, la presencia española en la coordinación de este gran proyecto no tendrá continuidad en la segunda etapa de Encode, en la que se descifrá el 99% restante del mapa del genoma. "Como no tenemos infraestructuras, a partir de ahora no

podremos liderar el grupo, el liderazgo pasará al Sanger Institute, de Cambridge, aunque continuaremos teniendo un papel importante". Para continuar, sería necesaria tecnología de secuenciación a gran escala. Guigó ha presentado al NHG otro nuevo proyecto para participar en Encode, y para el que no son necesarias grandes infraestructuras. Xavier Estivill opina

que "si España quiere desarrollarse en el campo biomédico, necesita infraestructuras potentes en genómica, y ahora mismo nos pasan por delante países como China o Corea". Según Estivill, "no podemos estar solo analizando datos, sino que tenemos que estar también produciéndolos". Y añade: "La industria farmacéutica nunca vendrá aquí si no tenemos esas

infraestructuras". Alfonso Valencia corrobora que los centros españoles no pueden adquirir compromisos en redes internacionales como centros para la producción de datos, y que tan sólo el Reino Unido, Estados Unidos y Japón cuentan con la tecnología para hacerlo. "Es en la parte computacional donde estamos en una posición competitiva", afirma el investigador, que destaca la colaboración en el proyecto Encode del ordenador español Mare Nostrum.