



EL CODIGO VITAL / AHORA ESTA EN NEGRO SOBRE BLANCO. LAS DOS MEJORES REVISTAS CIENTIFICAS DEL MUNDO PUBLICAN ESTA SEMANA EL PRIMER GRAN BORRADOR DEL MAPA DEL GENOMA DEL HOMBRE. ESTA ACABADO. QUEDAN AUN LAGUNAS QUE RESOLVER, PERO NADIE DUDA DE QUE AYER HA SIDO UN DIA CRUCIAL PARA LA BIOLOGIA

MENOS DE 40.000 GENES

JOSE LUIS DE LA SERNA
David Baltimore es premio Nobel de Medicina y miembro del Instituto de Tecnología de California. Lleva más de 40 años investigando y asegura haber sido testigo en su carrera de avances biomédicos de un gran interés. No obstante, reconoce en *Nature* que «se le sigue poniendo la carne de gallina al recordar el momento en que empezó a leer en esta revista la descripción del primer gran borrador de nuestro genoma». Una gran mayoría de la comunidad científica pensará hoy igual que Baltimore. A pesar de que el mapa del genoma que presenta el Consorcio Internacional, por un lado, y la empresa privada Celera, por otro, no está completamente acabado —y quedan aún

muchos detalles importantes por desvelar—, se puede decir que la primera parte del Proyecto Genoma se ha acabado. Y no hay una segunda parte definida. De ahora en adelante, y con todas las secuencias de ADN del hombre y de muchos animales en Internet, comienza una nueva revolución social. Igual que han existido revoluciones agrícolas, industriales y de las comunicaciones, la que viene —de una transcendencia que muchos aún no vislumbran— será la revolución genómica. Quizá la primera sorpresa que ha dado el borrador del genoma ha sido comprobar que el número de genes es menos de la mitad de lo que se ha estado diciendo tanto tiempo: menos de 40.000. Otra característica que sorprenderá a los no

iniciados es la similitud que tiene el genoma de la especie más superior del mundo, no ya con sus antecesores inmediatos, los primates, sino con seres tan alejados en la escala evolutiva como pueden ser las moscas o los gusanos. De hecho, los investigadores confían en que lo que hará avanzar con pasos de gigante a la Ciencia será la comparación del ADN del hombre con el de otras especies. No obstante, al Consorcio y a Celera les queda mucho trabajo por hacer de a corto plazo. Lo primero es precisar los errores que existen e intentar llenar los vacíos que aún quedan. Luego, hay que comparar. Con el genoma del ratón secuenciado y con el de los monos en estado avanzado, la genómica comparativa probará su

valor. El uso de nuevas herramientas hará posible descifrar el significado del libro de la vida. Convertir toda esta información en una maquinaria que revele el íntimo porqué de todas las patologías, y lograr nuevas curas para ellas, será un objetivo irrenunciable. Otro gran desafío es el de elevar la cultura científica. Hasta que la sociedad no admita que el analfabetismo biomédico será en el siglo XXI tan perjudicial como lo fue en el XX no saber leer, hablar inglés o poder navegar por Internet, no se conseguirán obtener los grandes beneficios que puede aportar el genoma. Probablemente ayer ha comenzado la aventura más importante que ha vivido el hombre: la de conocerse a sí mismo.

En 1997 se describía el genoma del primer organismo, el de la levadura, *Saccharomyces cerevisiae*; dos años después fue el del gusano, *Caenorhabditis elegans*. A mediados del 2000 el mundo científico se veía gratamente sorprendido por la descripción del genoma de la mosca del vinagre, *Drosophila melanogaster*, y a finales del mismo año, el de una planta, *Arabidopsis thaliana*. Ya sólo quedaba el de un vertebrado, el del *Homo sapiens* y... ¡por fin ha llegado! El Consorcio Internacional Genoma Humano, integrado por 20 grupos de diferentes países entre los que no está España, y la empresa privada Celera, acaban de hacer público el mapa provisional del genoma humano (GH), que aporta una extraordinaria y amplia información acerca de las bases genéticas del ser humano, de sus orígenes y evolución, y de las implicaciones que tendrá para la medicina del futuro.

En 1990 nace oficialmente el Proyecto Genoma Humano con el objetivo de llegar a conocer toda nuestra dotación genética. Se calculaba que el genoma tenía alrededor de 3.000 millones de pares de bases a lo largo de una estructura conocida como ADN, que se visualiza en forma de cromosomas. Esos millones de bases configuran los genes. Se calculaba que podíamos tener unos 100.000.

Los genes se encuentran en todas nuestras células y son los responsables de las diferencias físicas y de un gran número de enfermedades de base genética, ya que dan lugar a las proteínas, unas estructuras que tienen una función concreta y específica en el organismo. Si los genes se alteran, esas proteínas pueden estar ausentes o funcionar de forma defectuosa, dando lugar —en el peor de los casos— a una enfermedad genética de mayor o menor severidad.

Actualmente no existe tratamiento para las más de 5.000 enfermedades genéticas conocidas y la esperanza pasa por conocer la proteína defectuosa, la función que tiene en el organismo, y en diseñar

nuevas estrategias terapéuticas basadas en la terapia génica y en la farmacogenética. De ahí el gran interés en llegar a conocer nuestro genoma.

Pero, ¿cómo abordar y diferenciar en genes esa estructura continua de 3.000 millones de bases? Un primer paso consiste en poder manipular el genoma de manera que se pueda analizar un fragmento de varias Kilobases (1.000 bases = 1Kb), y no de 3.000 millones de bases. La fragmentación del genoma en clones se empezó a hacer en los años 80 insertando (clonando) pequeños fragmentos de ADN de 4 o 5 Kb en estructuras que las pudieran albergar (vectores). Con el perfeccionamiento de esta técnica se crearon librerías genómicas.

A partir de ellas se pueden hacer los llamados mapas físicos, es decir, ordenar los clones de una determinada región o cromosoma. Una vez establecido un conjunto continuo de clones ordenados (*contig*) a lo largo del genoma o de un cromosoma, se selecciona un clon y se vuelve a fragmentar en múltiples piezas de ADN de menor tamaño para secuenciarlas o, lo que es igual, leerlas (descifrar su código de bases). Posteriormente se vuelven a unir y se recompone el clon original de 100 o 500 Kb. Esta estrategia, conocida como *shotgun sequencing* [gráfico pág. 6], fue la utilizada por el consorcio internacional para descifrar y secuenciar el genoma.

De los 300.000 clones de partida, fueron válidos 30.000, que secuenciados representan un total de 3.200 Megabases (1Mb= 1000 Kb). Estos resultados, alcanzados en octubre del 2000, representan el 90%

Y... ¡POR FIN EL GENOMA HUMANO!

JAVIER BENITEZ ORTIZ

del genoma y son los que se publican ahora. El 10% restante se pretende completar en el año 2003.

Para obtener estos resultados hubo que realizar una gran innovación y desarrollo tecnológico en dos grandes áreas; en la biología molecular y en la informática. En la primera, los esfuerzos se dirigieron hacia la simplificación tecnológica y la automatización de múltiples usos mediante robots; en la segunda, al desarrollo de *software* para la producción y el análisis de datos y la generación de nuevos ordenadores de mayor capacidad y velocidad.

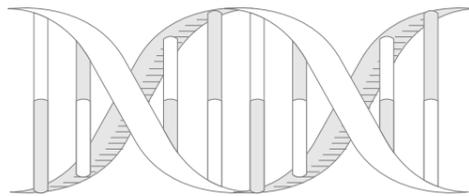
Gracias a todo ello se han conseguido realizar más de 100.000

interés y trascendencia por ser la de nuestra especie y dar una perspectiva global del GH. ¿Qué se puede aprender de este genoma provisional? Son muchos y variados los puntos de interés, pero se pueden destacar algunos de ellos.

a) No todos los genes están distribuidos de forma uniforme a lo largo de los cromosomas. Hay regiones muy ricas en determinadas bases, Citosina y Guanina, que confieren distintas propiedades al genoma. El cromosoma con menor densidad de genes es el Y, el responsable del sexo masculino.

b) A pesar de tener sólo el doble de genes que el gusano o la mosca, el GH es 25 veces mayor que cualquiera de ellos. Esto es porque hay una gran cantidad de material considerado *basura* que está formado por genes que no generan proteínas, secuencias repetidas sin aparente función, o secuencias que saltan de una parte a otra del genoma (trasposones). Se sabe ahora que parte de este material ha tenido y tiene gran importancia en la evolución de nuestro genoma.

c) Los genes constituyen únicamente el 5% del genoma, pero es la parte más importante del mismo. El número de genes puede oscilar en torno a los 31.000, y aunque es una cifra menor de la esperada (100.000), la mitad tiene capacidad para generar no una sola proteína, sino dos o tres diferentes. Aunque parezca paradójico que sólo tengamos el doble de genes que la mosca del vinagre, la realidad es que tenemos capacidad para generar 5 veces más productos proteicos que los otros organismos.



secuencias en 12 horas (una secuencia tiene alrededor de 1.000 bases), cuando los actuales y mejores secuenciadores automáticos tienen una capacidad de 96 secuencias cada tres horas. En el 2000 se trabajaba a un ritmo de 1.000 bases/segundo, 24 horas al día, 7 días a la semana.

La parte secuenciada del genoma (3.200 Mb) es una cifra similar a la esperada. Este conjunto de bases alberga un número de genes, que puede estar en torno a los 31.000. La secuencia obtenida, aunque incompleta, es de enorme

d) Estos genes codifican para proteínas (proteoma) que van a tener una función concreta. El proteoma muestra que no hay grandes diferencias con las proteínas de los otros organismos. Se han definido un conjunto de 1.308 grupos de proteínas comunes con las otras especies, que deben representar el núcleo proteico crítico para las funciones básicas de todas las células. Sólo un 7% de grupos de proteínas son específicas de los vertebrados y son las que marcan las diferencias fisiológicas con el resto. Se puede decir que hay un incremento en la complejidad del proteoma, desde la levadura a los gusanos y a los vertebrados, debida a una innovación proteica a lo largo de la escala evolutiva.

e) La secuenciación del GH muestra que es tremendamente polimórfico en toda su extensión. Los polimorfismos, en este caso moleculares, son variantes que coexisten en la población general sin que esto suponga algo negativo. Son mutaciones no patológicas. Hay un polimorfismo (SNP) cada 1.300 bases, y se han podido anclar 1,42 millones de SNP en el genoma. Esto representa una fuente extraordinaria para buscar genes de susceptibilidad a diversas enfermedades genéticas o para construir los *chips* de SNP que permiten el análisis de miles de fragmentos a la vez.

Al GH todavía le queda un largo camino por recorrer. Para el 2003, la secuencia completa; terminar de completar el catálogo de genes y de sus proteínas correspondientes; pasar de la secuencia a la función concreta; completar el mapa de SNP para abordar los genes de enfermedades comunes (obesidad, hipertensión, etcétera) y terminar la secuencia de los genomas de otros vertebrados (ratón) e iniciar nuevos estudios (primates). A pesar de ese largo camino, hoy podemos decir que ¡por fin tenemos el genoma humano!

Javier Benítez Ortiz es Director del Departamento de Genética Humana del Centro Nacional de Investigaciones Oncológicas Carlos III (CNIO)