

EL PROYECTO GENOMA HUMANO: 1988-2000

«1988.—Mapping and Sequencing the Human Genome. Committee on Mapping and Sequencing the Human Genome. Board on Basic Biology. Commission on Life Sciences. National Research Council. National Academy Press. Washington, D.C.».

«Executive Summary».—«Los humanos llevan tiempo intrigados por las fuerzas que dan forma a ellos y a otros organismos. ¿Qué código dicta el color de los ojos, del pelo o la forma de una flor? Hace más de 100 años Gregor Mendel descubrió que tales rasgos hereditarios están controlados por unidades celulares que, posteriormente, se conocieron como genes. En años recientes, la comprensión de esos genes ha crecido considerablemente tras el conocimiento de la biología molecular del ADN, la molécula gigante de la que están formados los genes. Ahora es posible obtener la descripción última de los genes y del ADN; ello, desde el reciente desarrollo de las técnicas que permiten mapear (localizar) los genes en el ADN de cualquier organismo y luego secuenciar (ordenar) cada una de las unidades de ADN, conocidas como nucleótidos, que constituyen los genes.

Cuanto más de nuestros genes estén mapeados y sus ADNs secuenciados, dispondremos de una fuente cada vez más útil —una base de datos esencial que facilitará la investigación en bioquímica, fisiología, biología celular y medicina—. Esta base de datos tendrá su mayor impacto sobre el cuidado de la salud y la prevención de la enfermedad así como sobre nuestro conocimiento de las células y los organismos. La concepción de organizar un gran proyecto para mapear y secuenciar el ADN en los genes y en las regiones intergénicas que los conectan (el ADN completo o genoma humano) recibe cada vez más atención en el mundo. Varios países han expresado su interés en apoyar tal proyecto. Para evaluar lo que los Estados Unidos de N.A. deberían hacer en esta área, el «Board on Basic Biology» del «National Research Council's Commission on Life Sciences» estableció el «Committee on Mapping and Sequencing the Human Genome», cuyas conclusiones se recogen en este documento.

En este documento el comité explora cómo, cuando y por qué debemos mapear y secuenciar el ADN en el genoma humano. Para llevar a cabo esos objetivos, el comité alcanzó las siguientes conclusiones:

- *Conseguir un mapa, una secuencia y un conocimiento cada vez mayor del genoma humano exige una acción especial que deberá orga-*

nizarse y establecerse para ese propósito. Tal esfuerzo especial en las próximas dos décadas potenciará significativamente el progreso en biología y medicina humanas.

- *Los problemas técnicos asociados con el mapeo y la secuenciación de los genomas humano y de otros organismos son lo suficientemente grandes para que un programa científico como este requiera un esfuerzo diversificado y sostenido para mejorar nuestras posibilidades para analizar las complejas moléculas de ADN. Aunque las capacidades necesarias no existen todavía, la manera en que deben desarrollarse parecen evidentes. Las tecnologías avanzadas requeridas emergerán de un esfuerzo común que recalque proyectos piloto y desarrollo tecnológico. Una vez establecidas, estas tecnologías no solo harán factible completar el análisis del genoma humano y otros, sino que también harán contribuciones importantes a muchas otras áreas de la biología básica y la biotecnología.*
- *Los objetivos iniciales más importantes se esforzarán en adquirir un mapa de alta resolución de ligamiento génico del genoma humano, una colección de clones ordenados de ADN y una serie de mapas físicos complementarios de resolución progresivamente creciente. El objetivo último será obtener la secuencia nucleotídica completa del genoma humano, comenzando a partir del material de la colección de clones ordenados de ADN. La consecución de este objetivo requerirá mayores (pero alcanzables) avances en el manejo y en la tecnología de secuenciación del ADN.*
- *Un enfoque genético comparativo es esencial para interpretar la información en el genoma humano. Más aún, deberán llevarse a cabo en paralelo estudios intensivos de aquellos organismos que proporcionen modelos particularmente útiles para comprender la estructura, función y evolución génica humana.*
- *El esfuerzo para mapear y secuenciar comenzará como una serie de programas encaminados a mejorar el desarrollo tecnológico competitivos y revisados por pares. Los fondos deben incluir tanto ayudas individuales como a grupos de mediano tamaño interdisciplinarios de científicos e ingenieros. Dado que la tecnología requerida para conseguir la mayoría de los objetivos de los proyectos necesita de mejoras importantes, el comité se manifiesta específicamente en contra de establecer, en el momento actual, uno o unos pocos grandes centros de secuenciación.*

- *El proyecto genoma humano deberá diferir de la marcha de la investigación actual en cuanto que los subproyectos deben incrementar en un factor de 5 a 10, sus eficacias de mapeo, secuenciación, análisis o interpretación de la información.*
- *Los progresos hacia los objetivos antes expuestos requerirán el establecimiento de instalaciones centralizadas bien dotadas, incluyendo un banco central para los fragmentos de ADN clonados generados en el esfuerzo de mapeo y secuenciación y un centro de datos para la colección computarizada y la distribución de grandes cantidades de información de las secuencias de ADN. El comité sugiere que los grupos que soporten tales servicios sean seleccionados mediante concurso abierto.*

Sobre la base de esas conclusiones, el comité recomienda lo siguiente:

- *En vista de la importancia y magnitud de la tarea, se recomienda unos fondos adicionales de \$200 millones anuales. Tales fondos no deberán retrotraerse del presupuesto federal actual para investigación en ciencias biomédicas.*

La mayoría del comité recomienda:

- *Una agencia federal única deberá liderar el proyecto. Esta agencia deberá recibir y administrar los fondos para el proyecto y deberá ser responsable de las operaciones del banco central y del centro de datos, así como administrar el sistema de revisión por pares utilizado para determinar los beneficiarios de los fondos. Deberá trabajar en estrecha colaboración con un «Scientific Advisory Board» para desarrollar e implementar un alto estándar en la revisión por pares. El «Scientific Advisory Board», compuesto en principio por científicos expertos de renombre en campos relevantes, deberá asesorar no solo en cuanto a la revisión por pares sino sobre el control de calidad, cooperación internacional, coordinación de esfuerzos entre los laboratorios y las operaciones del banco central y los centros de datos ...».*

Y doce años después ...

«2000.—Reseña de la conferencia de prensa que tuvo lugar en la «White House» en Washington D.C. USA, el 26 de junio, con motivo del anuncio de la consecución de un “borrador de trabajo” del genoma humano».

El consorcio público «Proyecto Genoma Humano» da a conocer, hoy, la conclusión de un borrador de trabajo de la secuencia del genoma humano, la huella genética del ser humano.

Este mojón clave implica dos tareas: colocar largos fragmentos de ADN en orden correcto para completar todos los cromosomas, y determinar la secuencia de ADN de esos fragmentos. El ensamblaje que hoy se da a conocer consta de fragmentos solapantes que cubren el 97% del genoma humano, representando un 85% del genoma la secuencia hasta ahora ensamblada.

La secuenciación se disparó durante el último año; el 60% de la secuencia se ha conseguido en los últimos seis meses. Durante este tiempo el consorcio ha proporcionado mil bases de secuencia bruta por segundo —7 días a la semana, 24 horas cada día—. La calidad media del «borrador de trabajo» supera con mucho las expectativas originales del consorcio para este producto intermedio. Los centros del consorcio han proporcionado mayor cantidad de secuencia que la esperada.

Consecuencia de todo ello es que el «borrador de trabajo» está mucho más cerca de la versión final de lo que el consorcio había previsto para este momento. Aproximadamente el 50% de la secuencia del genoma se encuentra en forma casi definitiva, y el 24% lo está definitivamente. La secuencia genómica está organizada en segmentos contiguos de, aproximadamente, 200.000 bases. La fiabilidad media de la secuencia de ADN en este ensamblaje es del 99.9 %.

La información de la secuencia por parte del proyecto público ha sido continua, inmediata y de libre disposición, sin restricciones para su utilización o redistribución. La información es consultada a diario por científicos de la academia y de la industria.

Hasta este momento ya han sido identificadas algunas decenas de miles de genes a partir de la secuencia del genoma. El análisis de la secuencia disponible muestra los 38.000 genes predichos y confirmados por evidencia experimental. Hay muchos miles de predicciones génicas adicionales que han de ser comprobados experimentalmente. Docenas de genes involucrados en enfermedades han sido identificados accediendo al borrador de trabajo.

Objetivos del consorcio. El objetivo del consorcio para la primavera de 2000 fue producir una versión tipo borrador de trabajo de la secuencia humana, un ensamblaje que contiene fragmentos solapantes que cubren aproximadamente el 90% del genoma y que se secuencia en forma de bo-

rrador de trabajo; esto es, con algunas lagunas y ambigüedades. El objetivo último del consorcio es producir una secuencia completamente finalizada, i.e. - sin lagunas y con el 99.9% de fiabilidad. La fecha para este último objetivo había sido el año 2003, pero a la vista de los resultados actuales es completamente seguro que la obtención de la secuencia definitiva se acortará significativamente.

Estrategias complementarias. En un anuncio paralelo, J. Craig Venter, Ph.D., presidente y director científico de Celera Genomics, anunció hoy que ha completado su propio ensamblaje de la secuencia del ADN humano: «La terminación del primer ensamblaje del genoma humano es un logro fascinante para Celera y su personal científico»..

Los proyectos público y privado utilizan automatismos y tecnología de secuenciación similares, pero emplean distintos enfoques para secuenciar el genoma humano. El proyecto público usa una estrategia denominada «abordaje jerárquico» («hierarchical shotgun») por la que largos fragmentos individuales de localización conocida se someten a secuenciación (por ej., troceados en cortos fragmentos que son secuenciados y recompuestos sobre la base de solapamientos de secuencias). El proyecto de Celera utiliza un enfoque denominado «abordaje genómico global» («whole genome shotgun») por el que el genoma completo se trocea en pequeños fragmentos que son secuenciados y recompuestos sobre la base de solapamientos de secuencias.

El abordaje jerárquico tiene la ventaja de que la localización global de cada secuencia individual se conoce con certeza, pero requiere construir un mapa de largos fragmentos que cubran el genoma. El abordaje global no requiere este paso, pero tiene otros requerimientos en la fase de ensamblaje. Ambas estrategias alinean la secuencia a lo largo de los cromosomas humanos utilizando mojones localizados en el mapa físico producido en el Proyecto Genoma Humano.

«Las dos estrategias son bastante complementarias. El proyecto público y Celera plantean discutir los méritos científicos relativos de los métodos utilizados en ambos proyectos. Al final, la mejor estrategia puede ser muy bien una combinación de metodologías para la secuenciación de futuros genomas» - dijo Francis Collins, M.D., Ph.D., director del «National Human Genome Research Institute» de los «National Institutes of Health». En efecto, los planes actuales del proyecto público para secuenciar el genoma del ratón de laboratorio contemplan esta estrategia híbrida.

El Proyecto Genoma Humano se centra, ahora, en convertir el «borrador de trabajo» y las secuencias «casi conseguidas» a una «forma defini-

tiva». Ello se conseguirá rellenando las lagunas existentes en la secuencia del «borrador de trabajo» e incrementado la fiabilidad de la secuencia global hasta el 99.99%. Aunque la versión denominada «borrador de trabajo» es útil para la mayor parte de la investigación médica, una secuencia tan perfecta como sea posible es crítica para obtener toda la información contenida en los datos de la secuencia humana. Ello ya se ha conseguido para los cromosomas 21 y 22, así como para el 24% del total del genoma.

La producción de secuencias más allá de lo esperado ha ido pareja de una sorprendente cosecha de variaciones génicas humanas - los llamados polimorfismos nucleotídicos simples o SNPs. El Proyecto Genoma Humano tenía el objetivo de descubrir 100,000 SNPs antes del año 2003. Hasta ahora, con las secuencias ensambladas y otros datos acumulados por el Consorcio SNP, los científicos ya han identificado más de 300,000 SNPs y seguramente dispondrán hasta 1 millón de SNPs a finales del año. Los SNPs proporcionan una poderosa herramienta para estudiar las enfermedades y la historia humanas.

La secuenciación —determinar el orden exacto de las cuatro bases químicas del ADN denominadas A, T, C y G— se ha conseguido en el Proyecto Genoma Humano merced a los avances tecnológicos en el desciframiento del ADN y la naturaleza colaborativa del esfuerzo, que ha incluido 1,000 científicos de casi todo el mundo que han trabajado juntos con eficacia.

El Proyecto de Secuenciación del Genoma Humano pretende determinar la secuencia de la porción eucromática del genoma humano. La eucromatina excluye ciertas regiones consistentes en largas porciones de ADN repetitivo que codifican muy poca información genética, y que no se recuperarán en el sistema de vectores utilizado en el proyecto genómico. Tales regiones suponen el 10% del genoma y se denominan heterocromáticas (por ej., el centro de los cromosomas, llamado centrómero, consta de ADN heterocromático.).

El Consorcio internacional para la secuenciación del genoma humano incluye científicos de 16 instituciones en Alemania, China, Estados Unidos, Francia, Gran Bretaña y Japón. Los principales cinco centros están localizados en: «Baylor College of Medicine», Houston, Texas; «Joint Genome Institute», Walnut Creek, CA; «Sanger Centre», cerca de Cambridge, Inglaterra; «Washington University School of Medicine», St. Louis, Whitehead Institute, Cambridge, Massachusetts. Juntos, estos cinco centros han generado el 82% de la secuencia.

El proyecto ha sido tan estrechamente coordinado que ninguna región del genoma ha sido desatendida a la vez que se han minimizado las duplicaciones. Los participantes en el consorcio internacional se han adherido a los estándares de calidad del proyecto y a la política de publicidad de los datos. El proyecto se financia mediante ayudas de las agencias gubernamentales y fundaciones privadas en varios países, que incluyen el «US National Human Genome Research Institute», los «US National Institutes of Health», el «Wellcome Trust» en Inglaterra y el «US Department of Energy».

El coste global de la secuenciación del borrador de trabajo ha sido, aproximadamente, 300 millones de dólares USA, de los que, aproximadamente, 150 millones han sido financiados por los «US National Institutes of Health». El coste de la secuenciación del genoma humano se refiere frecuentemente a 3 mil millones de dólares USA. Sin embargo, esta cifra se refiere a la estimación inicial del Proyecto Genoma Humano para un periodo de 15 años (1990-2005) y para una amplia gama de actividades científicas relacionadas con la genómica, que incluye estudios de enfermedades humanas y de organismos experimentales (como bacterias, levaduras, gusanos, moscas y ratón); el desarrollo de nuevas tecnologías para investigación biomédica y de métodos computacionales para el análisis de genomas, y aspectos éticos, legales y sociales relacionados con la genética».

La ciencia española fue espectadora de esta corta pero apasionante historia. Ahora, el reto de la acción «...Comenzamos a vislumbrar las oportunidades que se abren con la investigación biogénica, un hallazgo que afectará de forma muy especial al bienestar de las nuevas generaciones. Ahí debemos estar en el futuro...».

Pedro García Barreno