

El Genoma Humano

A diez años del lanzamiento del Proyecto Genoma Humano por James Watson, la presentación simultánea de dos versiones del primer borrador de la secuencia casi completa del genoma humano es un acontecimiento sin precedente y de indudable importancia. El 15 de febrero de 2001 la revista *Nature* publicaba el artículo titulado *Initial sequencing and analysis of the human genome*¹ con los resultados de una colaboración entre 20 centros de EE.UU., Inglaterra, Japón, Francia, Alemania y China, involucrando un enorme cantidad de participantes (alrededor de 2500). El grupo se denomina *The International Human Sequencing Consortium* y trata de un emprendimiento de gran envergadura con fondos públicos a partir del Proyecto Genoma Humano originado en los Institutos Nacionales de la Salud de EE.UU. [NIH (*National Institutes of Health*, Bethesda, USA)] y liderado actualmente por Francis Collins. Un día después, aparecía en la revista *Science* con el título de *The sequence of the human genome*² el trabajo de Craig Venter de la empresa *Celera Genomics* en Rockville, MD, USA y sus múltiples colaboradores (274 autores), esta vez totalmente con fondos privados. Se llegaba así a un final feliz de lo que representó una costosa y a veces muy agresiva disputa entre los dos grupos, una verdadera carrera contra el tiempo ya que inicialmente la meta estaba fijada para el año 2005. Hace tres años, la entrada en esta carrera de Craig Venter con fondos privados y una tecnología sofisticada conocida en inglés como *whole genome shotgun sequencing method* obligó a los demás grupos a conseguir todavía más fondos para poder acelerar el paso con el fin de llegar a esta publicación simultánea. Esto es una prueba de la importancia que tiene la competitividad en investigación, como bien lo expresa el Editorial de *Science* " *Thus, we can salute what has become, in the end, not a contest but a marriage (perhaps encouraged by shotgun) between public funding and private entrepreneurship*"³. Queda implícito que todos los datos que componen estos dos trabajos están a disposición mundial y en forma gratuita a través del Internet. En el futuro *Celera Genomics* tiene la intención de comercializar sus nuevos descubrimientos al perfeccionar el mapeo cromosómico.

En investigación, todo emprendimiento es siempre la resultante de muchos experimentos anteriores, es decir, del conocimiento previo adquirido a través de los años. En este caso se puede remontar al principio del siglo XX con el redescubrimiento de las leyes de Mendel, luego el descubrimiento de la estructura del ADN, el advenimiento de la citogenética y la clasificación de los cromosomas, seguido del surgimiento de la biología molecular (de la mano de las enzimas de restricción), las técnicas de amplificación por PCR, y finalmente las técnicas modernas de secuenciación con la indispensable ayuda de sofisticadas computadoras. Sin todos estos conocimientos y todas estas técnicas nunca se hubiera podido emprender el Proyecto Genoma Humano.

Para los que no están familiarizados con la biología molecular, Ridley, en su excelente libro de divulgación titulado *Genoma: autobiografía de una especie en 23 capítulos*³, (ver pág. 251) emplea una ingeniosa comparación. Se imagina el genoma como un libro, con los 23 capítulos que representan los cromosomas; cada capítulo contiene miles de historias llamadas genes; cada historia contiene párrafos llamados exones con anuncios intercalados llamados intrones; cada párrafo está compuesto de pala-

bras llamadas codones y cada palabra está escrita con letras llamadas bases o nucleótidos. Cada palabra esta compuesta de 3 letras ensambladas en las 64 variaciones posibles de las cuatro letras que forman el ADN, es decir los nucleótidos A C G T (adenina, citosina, guanina, timina). La secuencia completa del genoma humano -de ambas fuentes- contiene alrededor de 2.9 mil millones de letras. Lo asombroso es que todo este documento gigantesco (equivalente en extensión a 800 Biblias) cabe en la cromatina del núcleo microscópico de una célula diminuta que a su vez cabe holgadamente en la punta de un alfiler.

Hasta ahora, el genoma más grande que se había secuenciado en forma completa era el de la mosca de la fruta (*Drosophila melanogaster*) con 180 Mb (megabases) y alrededor de 13 000 genes, recientemente secuenciado por *Celera Genomics*. El genoma humano –el primer genoma secuenciado de un vertebrado- tiene un tamaño de 3000 Mb (se espera que todos los mamíferos tengan un genoma de tamaño similar) conteniendo alrededor de 35 000 genes, un número sorpresivamente inferior a los 50.000-100.000 calculados con anterioridad. Llama también la atención el hecho de que el genoma humano es aproximadamente 30 veces más grande que el de la mosca pero no llega a tres veces el número de genes. Tanto en el hombre como en la mosca (o el pequeño nematode terrestre *Caenorhabditis elegans*), cada región codificante de cada gen tiene aproximadamente el mismo tamaño, por lo cual se sabe que en el hombre, contrariamente a lo pensado originalmente, cada gen puede codificar para más de una proteína usando diferentes combinaciones de sus exones. También es sorprendente el hecho que solamente entre el 1.1 al 1.5% del genoma codifica propiamente para las proteínas. Otra curiosidad del genoma humano, difícil de explicar en términos evolutivos, es el hecho que compartimos 223 genes con las bacterias que no existen ni en la mosca, el nematode o la levadura. Además, se encuentran en el genoma humano grandes regiones de secuencias repetidas y “desiertos” de ADN no codificante (en otra época llamado *junk DNA*) que componen del 40 al 46% del genoma. También se ubicaron transposones (genes saltarines) y amplias regiones copiadas y re-ubicadas por ARN retrotransposones (de los retrovirus endógenos) que ocuparían alrededor del 10% del genoma. Existe un 99.9% de homología genómica entre un individuo y otro. Sin embargo, cada uno de nosotros tiene un fenotipo diferente y cada individuo es único como puede comprobarse con el estudio de las secuencias repetidas del genoma conocido como fingerprinting de ADN. Con respecto al análisis de la función de los genes, lo que se denominó en inglés *Functional Genomics*, va a ser muy importante la comparación entre diversos genomas de vertebrados, en particular mamíferos, entre los cuales se espera contar en 2001 o, a lo sumo, en 2002 con el genoma completo del ratón⁵.

¿En que consistió la secuenciación del genoma humano? Se emplearon varios dadores que proporcionaron semen y sangre de la cual se obtuvieron cultivos de linfocitos inmortalizados con EBV (virus Epstein-Barr). De estas dos fuentes se obtuvo ADN que fue tratado con enzimas de restricción para obtener miles de fragmentos que luego fueron incorporados por medio de un vector en bacterias y amplificados para luego ser secuenciados. Así se fueron alineando las letras, ubicándolas en los distintos cromosomas y mapeando sus regiones. Se ha comparado este estudio a un enorme rompecabezas con piezas que faltan y otras que no encajan perfectamente y necesitan ser ajustadas.

El proyecto financiado con fondos públicos adoptó una metodología llamada *shotgun* jerarquizado o clon-por-clon, que incluye el previo clonado del ADN genómico en vectores de alta capacidad (cromosomas artificiales bacterianos o BACs y cromosomas artificiales derivados del virus bacteriófago P1 o PACs), la construcción de contigs (colección de clones superpuesta en sus extremos que forman parches de información) y finalmente la secuenciación por *shotgun* de cada uno de estos grandes

clones que forman los contigs ya «anclados» al genoma (lo que se denomina mapa físico). El método de *shotgun* consiste en la secuenciación de clones pequeños (500-700 pb) que representan al azar un segmento definido de ADN y posterior ensamblado de esas lecturas individuales por análisis de superposición de secuencia. Este abordaje de *shotgun* jerarquizado confina el problema de ensamblado de secuencias a la complejidad presente en segmentos de clonados de 100-200 kb.

Por otra parte, el abordaje utilizado por *Celera Genomics*, llamado *shotgun* generalizado o genómico total, se basó en la aplicación directa de la secuenciación *shotgun* sobre el total del ADN genómico humano (a partir de clones pequeños de 2, 10 y 50 kb), obteniendo una casi infinita cantidad de lecturas de secuencia (aproximadamente 550 pb) que necesitan ser ensambladas y ubicadas en el genoma. Este problema de ensamblado y ubicación resultó particularmente difícil considerando el gran tamaño y el alto porcentaje cubierto por secuencias repetidas del genoma humano. Sin embargo, para superar este problema específico y permitir el armado de contigs y su anclado al genoma, *Celera* se valió de información proveniente de dos fuentes: (a) de los marcadores ubicados como mojones sobre cada una de las bandas cromosómicas del genoma humano y (b) de las secuencias que fueron siendo publicadas como resultados parciales por el proyecto público (en las bases de datos GenBank y otros).

Muchas innovaciones tecnológicas desarrolladas dentro y fuera del Proyecto Genoma Humano posibilitaron no sólo la aceleración de la fase de secuenciación del ADN, sino también la automatización de los procedimientos de clonado, el manejo de la información y ensamblado de secuencias con paquetes de software diseñados especialmente. Entre los adelantos en el proceso de secuenciación, merece la pena citar algunos puntos de crítica importancia: el desarrollo de la secuenciación automática de ADN con nucleótidos terminales marcados con distintos colores de fluorescencia (los mismos son captados por una luz láser que excita el producto fluorescente, es leído por una célula fotoeléctrica y registrado en una computadora), la alta calidad de esos marcadores fluorescentes, el diseño de polimerasas especializadas para reacciones de secuenciación y el desarrollo de sofisticados equipos de electroforesis capilar asociados a complejos programas de computación.

De esta primera mirada sobre el genoma humano, con más de 100 páginas entre las publicaciones originales de las revistas *Science* y *Nature*, sólo se puede concluir que apenas se ha empezado a conocer su contenido en genes e inferir sus funciones; se lo ha comparado a un libro en un idioma extranjero donde lo primero es aprender el idioma.

¿Que aporta el Proyecto Genoma Humano a la medicina? Si bien los genes no tienen como función causar enfermedades, fueron los defectos genéticos asociados a distintas enfermedades los que principalmente sirvieron de marcadores para catalogar los genes correspondientes. Así en los mapas cromosómicos de los artículos de *Science* y *Nature* se aprecian fácilmente, por ejemplo, el gen de la fibrosis quística en el cromosoma 7 y el gen BCRA2 asociado al cáncer de mama en el cromosoma 13. Desde 1980 hasta el año 2000, empleando técnicas cada vez más sofisticadas de mapeo cromosómico, se ubicaron 1430 mutaciones asociadas a enfermedades monogénicas, sin incluir las neoplasias causadas por translocaciones y fusiones de genes⁶. Este catálogo de genes asociados a enfermedades y neoplasias se encuentra por Internet⁷ con una actualización diaria. Sin embargo, ningún gen opera en un vacío; cada gen interactúa directamente o indirectamente a través de las proteínas que codifica con muchos otros genes y sus productos. Esto explicaría las marcadas variaciones en los síntomas de pacientes con la misma enfermedad. Más aún, existen factores epigenéticos y ambientales que participan en el diseño del fenotipo de cada individuo. Curiosamente, estos últimos están tomando cada vez más importancia y explicarían la variabilidad en cuanto a enfermedades y la suscep-

tibilidad a ciertos tipos de cáncer que se observa aún en gemelos monocigotas que son verdaderos clones humanos.

La mayoría de las enfermedades representan la culminación de múltiples interacciones entre el genoma y el ambiente a lo largo de la vida. Predecir tanto la contribución de los genes como la participación de los factores ambientales en enfermedades como la hipertensión, la insuficiencia cardíaca, los trastornos psiquiátricos, es la meta final de muchos de los que emprendieron el estudio del genoma humano pero no será tarea fácil. Lo que sí puede llegar a perfeccionarse rápidamente es el diagnóstico genético de alteraciones monogénicas tanto en la vida uterina como adulta. En cuanto a las neoplasias se tratará de perfeccionar el estudio de los distintos genes oncogénicos, los genes supresores de tumores, los genes de susceptibilidad y sus interrelaciones, dentro del marco del *Cancer Genome Project*⁸. Además, de las comparaciones a gran escala entre el genoma normal y el genoma de células cancerosas surgirán muchas pistas sobre los genes involucrados en el proceso neoplásico.

Los proyectos para el futuro incluyen el estudio de la función de los genes con la determinación de las proteínas que codifican y su mapeo en la forma de lo que se conoce como "proteoma". Se espera que esto permitirá eventualmente profundizar el mecanismo molecular de las enfermedades e individualizar las reacciones de cada paciente frente a diversos tratamientos.

Sin entrar en los problemas éticos, sociales y legales que surgirán en el futuro, se puede concluir que esta presentación duplicada del primer borrador de la secuencia completa de los cromosomas humanos tiene un significado enorme, según algunos comparable en importancia con la llegada del hombre a la luna en 1969. Sin embargo, sólo estamos al comienzo de una fascinante etapa que se completará seguramente más pronto de lo previsto dado el enorme número de investigadores involucrados y el cada vez más abultado aporte de fondos públicos y privados.

Christiane Dosne Pasqualini

Instituto de Investigaciones Hematológicas
Academia Nacional de Medicina, Buenos Aires
e-mail: chdosne@otmail.com

1. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409: 860-914.
2. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001; 291: 1304-48.
3. Jasny BR, Kennedy D. The Human Genome (Editorial) *Science* 2001; 291: 1153.
4. Ridley M. Genome: the autobiography of a species in 23 chapters. New York: Harper-Collins, 1999. (Genoma: autobiografía de una especie en 23 capítulos. Trad. Irene Cifuentes. Buenos Aires: Taurus, 2000). (ver pág. 251).
5. Benavides F, Guénet JL. Modelos murinos de enfermedades humanas. *Medicina (Buenos Aires)* 2001; 61: 215-231.
6. Peltonen L, McKusick VA. Dissecting human disease in the postgenomic era. *Science* 2001; 291: 1224-9.
7. OMIM: www.ncbi.nlm.nih.gov/omim.
8. Interview. Richard Wooster on Cancer and the Human Genome Project. *Lancet Oncol* 2001; 2: 176-8.

Este Editorial no podría haberse concluido sin la ayuda del Dr. Fernando Benavides y de los biólogos Marcelo de Campos Nebel y Carlos De Brasi con quienes fue un placer aprender y discutir.