

Proyecto Genoma Humano:

SITUACIÓN ACTUAL Y PERSPECTIVAS

L.A.Q.B. Daniel Cervantes García ¹
 L.A.Q.B. Claudia Rebeca González Ruiz ²
 Dr. en C. Netzahualcóyotl Mayek Pérez ³

INTRODUCCIÓN

Las células son las unidades fundamentales de cada sistema de vida. Todas las instrucciones que se necesitan para dirigir sus actividades están contenidas dentro del DNA (el ácido desoxirribonucleico). La secuencia de DNA es el arreglo particular de bases a lo largo de la hebra (por ejemplo, ATTCCGGA). Este orden explica detalladamente las instrucciones exactas requeridas para crear un organismo particular con sus rasgos característicos. El genoma es el juego completo de DNA de un organismo. La palabra genoma es un término acuñado por Winkler en 1920 como una conjunción entre GEN y cromosOMA. Los genomas varían extensamente en tamaño: mientras que el genoma más pequeño conocido de un organismo viviente (una bacteria) contiene aproximadamente 600,000 pares de bases de DNA, el genoma del humano y del ratón tienen aproximadamente $3 \times$

10^6 kb. El DNA del genoma humano se encuentra dispuesto en 23 cromosomas distintos. Los genes son las secuencias específicas de las bases que codifican instrucciones de como hacer proteínas o RNA (ribosomal, de transferencia, o ribozimas). La constelación de todas las proteínas producidas en una célula es llamada proteoma (1, 8).

DESCRIPCIÓN DEL PROYECTO GENOMA HUMANO

El Proyecto Genoma Humano (PGH) fue oficialmente iniciado en los Estados Unidos de América el 1 de octubre de 1990. El objetivo PGH del es conocer la secuencia del DNA de los 23 cromosomas humanos (22 autosomas más los cromosomas Y y X, que determinan el sexo). Durante los primeros años del proyecto se prepararon mapas físicos de cada cromosoma humano con el fin de obtener, para 1998, marcadores de secuencias identificables localizados cada 100,000 bases. Además, del genoma humano, el PGH también incluye secuenciar el genoma de varios organismos "modelo", tales como bacterias (*Escherichia coli* 4.6×10^6 bp); levaduras (*Saccharomyces cerevisiae* 12×10^6 bp), nematodos (*Caenorhabditis elegans* 97×10^6 bp), la mosca de la fruta (*Drosophila melanogaster* 180×10^6 bp) y el ratón (*Mus musculus* 3000×10^6 bp). El objetivo del PGH consiste básicamente en compilar una secuencia de letras y en conocer el significado de dicha secuencia.

¹ Unidad de Laboratorios de Ingeniería y Expresión Genética. Departamento de Bioquímica. Facultad de Medicina. Universidad Autónoma de Nuevo León. Av. Francisco I. Madero esq. Eduardo Aguirre Pequeño s/n. CP 64460. Monterrey, México. Tel. (81) 83294173. E-mail: cervantes.daniel@gmail.com

² Departamento de Fisiología. Centro de Ciencias Básicas. Universidad Autónoma de Aguascalientes. Av. Universidad 940, 20100, Aguascalientes, México.

³ Centro de Biotecnología Genómica. Instituto Politécnico Nacional. Blvd. del Maestro s/n esquina Elías Piña, 87100, Reynosa, México.

El PGH tiene como objetivos específicos:

1. Revelar la secuencia de aminoácidos de cada polipéptido codificado por nuestros genes. Mediante el análisis de las secuencias, se conocerá la función probable de dichos polipéptidos.
2. Obtener información detallada acerca de las secuencias que participan en el control de la expresión de los genes.
3. Identificar los genes que causan enfermedades humanas o que contribuyen a ellas, así como determinar los cambios en la secuencia que determinan la enfermedad.
4. Obtener conocimientos acerca de la evolución al nivel molecular. Los productos de los genes característicos del hombre ayudarán a explicar por qué somos biológicamente diferentes a otros organismos.
5. Disponer de la información para obtener cualquier gen particular en muestras de DNA y continuar la investigación en biología molecular humana.
6. Proporcionar la oportunidad para elaborar proteínas humanas antes desconocidas con valor potencial en medicina (1, 5, 12).

En 1985 se propuso un proyecto con la finalidad de secuenciar los nucleótidos del genoma humano. En 1990, el PGH se inició oficialmente en los EUA bajo la dirección de los Institutos Nacionales de Salud (NIH) y el Departamento de Energía de EUA (DOE) con un plan de 15 años y 3 billones de dólares para completar la secuencia del genoma (6, 9). El 14 de abril del 2003 el Consorcio Internacional de Secuenciación del Genoma Humano, dirigido en los EUA por el Instituto Nacional de Investigación del Genoma Humano (NHGRI) y el DOE, anunció la terminación exitosa del PGH con dos años de anticipación. La secuencia terminada producida por el PGH cubre más del 99 % de las regiones de genes contenidas en el genoma humano y éstas han sido secuenciadas con una precisión de 99.99 % (9, 13). La secuencia de cada cromosoma humano se encuentran a disposición del público en la siguiente página web: <http://www.ncbi.nlm.nih.gov/genome/guide/human/>

METODOLOGÍA DEL PROYECTO GENOMA HUMANO

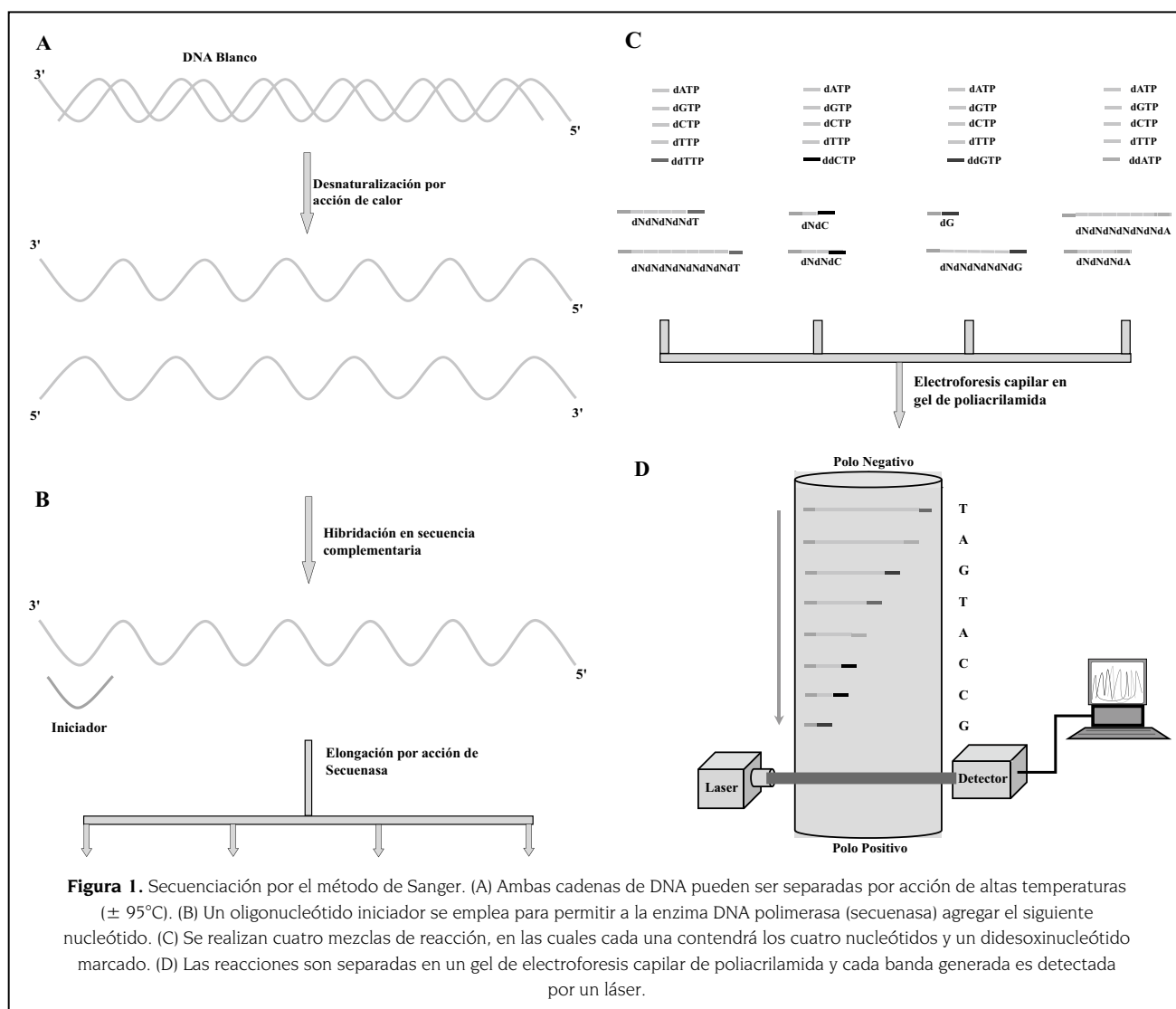
El PGH, al tratarse de un proyecto que pretende identificar la secuencia completa del genoma humano, tanto las secuencias codificantes (exones) como las no codificantes (intrones), necesita de técnicas que permitan identificar el lugar (*locus*) y la distancia en que se encuentran dichos genes. La secuenciación es el proceso por el cual se identifican las secuencias en que están unidas los 3×10^9 pares de bases y, posteriormente, saber qué significan dichas secuencias (1, 9). En un principio se acordó que el PGH se realizaría en dos etapas, una de ligamiento o cartografía genética de todos los cromosomas, etapa que terminó en 1998 y otra que corresponde a la secuenciación que inició en 1998 y concluyó en abril del 2003. Existen dos categorías principales de técnicas de cartografía genética: 1) ligamiento o cartografía genética, que sólo identifica el orden relativo de los genes a lo largo del cromosoma y 2) cartografía física, que son métodos más precisos para determinar las distancias entre genes dentro del cromosoma.

La cartografía mediante ligamiento se desarrolló a principios del siglo XX gracias al trabajo de T.H. Morgan. El mapa de ligamiento de Morgan localiza una serie de secuencias marcadoras, una con respecto a otra y con respecto a un gen defectuoso. De esta manera, se puede determinar la región cromosómica que contiene un gen mutado que se hereda de acuerdo con las Leyes de Mendel. En 1994 el mapa genético del PGH contenía 6,000 marcadores ubicados a menos de un millón de pares de bases uno del otro.

La cartografía física determina la distancia real entre puntos diferenciados de los cromosomas. Las técnicas más precisas combinan robótica, uso de láser e informática para medir la distancia entre marcadores genéticos. Para realizar estos mapas se extrae el DNA de los cromosomas humanos y se rompe aleatoriamente en numerosos fragmentos. A continuación, éstos se duplican varias veces en el laboratorio para analizar las copias idénticas obtenidas, llamadas clones. Los clones que comparten varias marcas proceden por lo general de segmentos superpuestos del cromosoma. Las regiones de superposición de los clones pueden a continuación compararse para determinar el orden global de las marcas

a lo largo del cromosoma y la secuencia exacta que ocupan inicialmente los segmentos de DNA clonados. Los mapas físicos especifican distancias físicas que se pueden medir en pares de bases (pb) o alguno de sus múltiplos. El mapa físico corresponde a la propia secuencia del genoma. Pero antes de llegar a obtenerla, hay que elaborar mapas físicos partiendo de resoluciones bajas y avanzando hacia las resoluciones cada vez mayores. En cierta manera, los mapas físicos de menor resolución son los propios cariotipos: la visualización microscópica de la dotación cromosómica haploide humana teñida con colorante de Giemsa muestra un patrón alternante de bandas claras y oscuras, en el que cada banda tiene una media de unos 7 millones de pares de bases (9, 10).

En el PGH se utilizó primordialmente un método de secuenciación desarrollado por el bioquímico británico y dos veces premio Nobel, Frederick Sanger. Este método consiste en replicar piezas específicas de DNA y modificarlas, de modo que terminen en una forma fluorescente de uno de los cuatro nucleótidos. En los modernos secuenciadores automáticos de DNA, el nucleótido modificado, situado al extremo de una de estas cadenas, se detecta con un haz de láser y se determina el número exacto de nucleótidos de la cadena. A continuación, se combina esta información en un ordenador para reconstruir la secuencia de pares de bases de la molécula original de DNA (Figura 1) (4, 12). Los mapas físicos de mayor resolución se suelen elaborar a partir de genotecas (bibliotecas



de genes) en las que el genoma a estudiar se encuentra fragmentado en multitud de trozos aleatorios y desordenados, cada uno de ellos clonado por separado en un vector adecuado: plásmido, cósmido, cromosomas artificiales de levadura (Yeast Artificial Chromosomes, YACs), cromosomas artificiales de bacteria (Bacterial Artificial Chromosomes, BACs), etc. La elaboración de los mapas físicos es, en cierto modo, similar a la de ensamblar un rompecabezas: consiste en ordenar los fragmentos del genoma mediante la búsqueda de fragmentos que tienen alguna zona en común; es decir, hallar conjuntos de pares de fragmentos parcialmente superpuestos. Ello conduce al concepto contig, que es un conjunto de fragmentos de un genoma que se han clonado por separado, pero que son contiguos y que están parcialmente superpuestos. La cartografía de los contigs se puede realizar buscando la "huella genética" común a distintos clones de una genoteca de DNA humano. Dicha huella puede

consistir en un patrón compartido de secuencias de enzimas de restricción (que se pueden indagar ayudándose de algoritmos y programas de cómputo adecuados). Las estrategias más recientes hacen uso de DNA humano en forma de unos 20,000 fragmentos independientes clonados en los BACs, y buscando la "huella genética común" entre clones con base en la detección de determinadas secuencias repetitivas. El último gran hito en cuanto a elaboración de mapas físicos ha sido el desarrollo de "marcadores físicos universales" fácilmente generables, que permiten que los datos obtenidos en un laboratorio sean rápidamente compartidos y asumidos por toda la comunidad investigadora (Figura 2). Se trata de los llamados "lugares etiquetados por su secuencia" (Sequence Tagged Sites, STS). Los STS consisten en fragmentos cortos de DNA (de entre 100 y 1000 pb) cuya secuencia exacta se conoce y se sabe que es única en todo el genoma. Los STS definen puntos concretos únicos del mapa

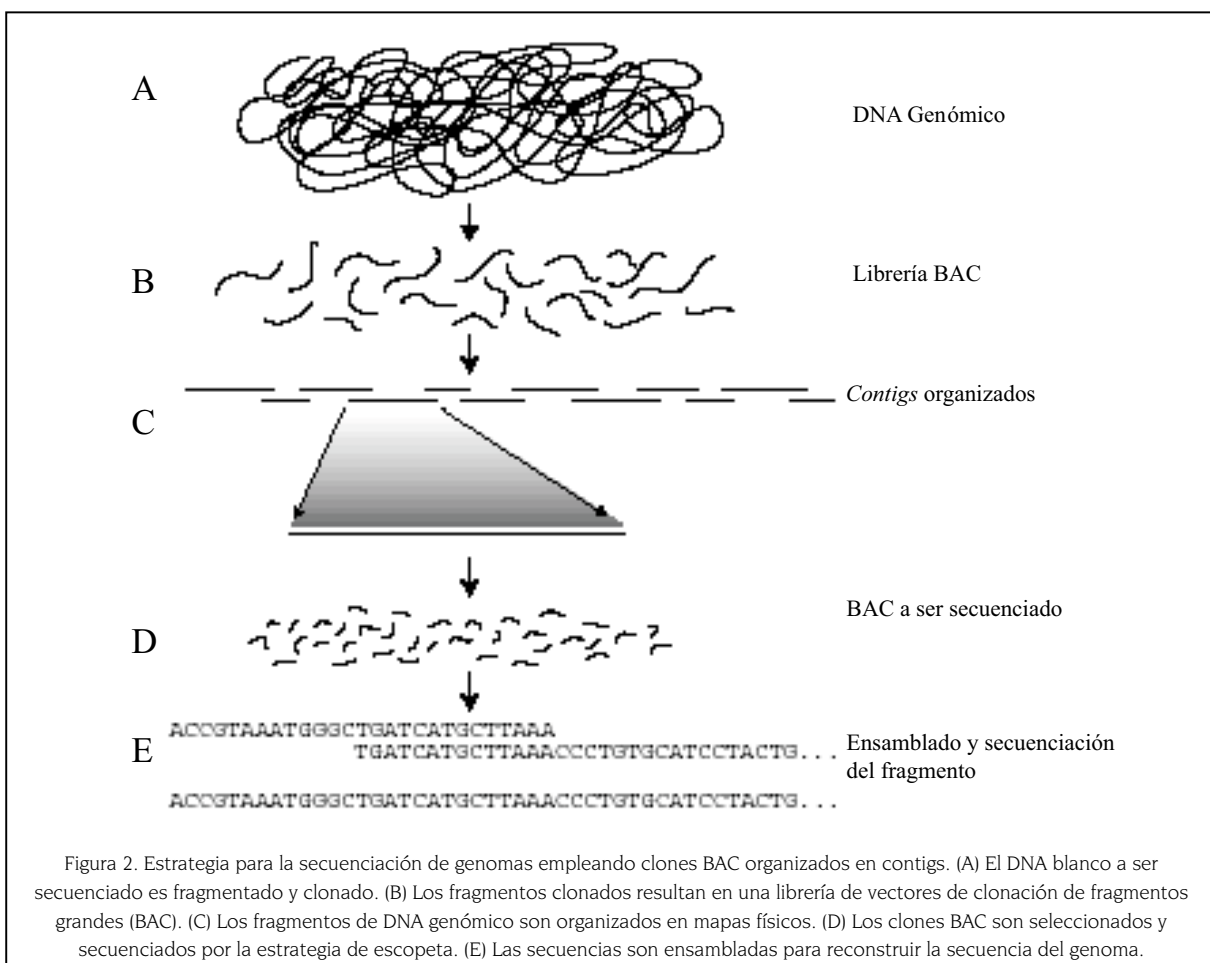


Figura 2. Estrategia para la secuenciación de genomas empleando clones BAC organizados en contigs. (A) El DNA blanco a ser secuenciado es fragmentado y clonado. (B) Los fragmentos clonados resultan en una librería de vectores de clonación de fragmentos grandes (BAC). (C) Los fragmentos de DNA genómico son organizados en mapas físicos. (D) Los clones BAC son seleccionados y secuenciados por la estrategia de escopeta. (E) Las secuencias son ensambladas para reconstruir la secuencia del genoma.

físico y constituyen magníficos "marcadores" fácilmente detectables. El empleo de los STS ayuda a elaborar mapas de contigs según el contenido del STS de los clones superpuestos. Estos mapas de STS permiten la integración de los mapas genéticos y físicos, hacen accesible la fase de secuenciación y facilitan la clonación de genes implicados en enfermedades mediante la llamada estrategia del candidato posicional (6, 12).

El equipo de la compañía Celera Genomics, que incluye a unos 282 investigadores de instituciones de EUA, Australia y España, en representación de 12 organizaciones académicas, sin fines de lucro y comerciales, obtuvo muestras de DNA de cinco personas, tres mujeres y dos hombres. Este grupo incluyó un afroamericano, un chino, un hispano y dos blancos. Luego de seleccionarlos para librarlos de contaminantes y conjuntarlos en bibliotecas de DNA, se analizaron los trozos del código genético de estos donantes.

Un elemento central de este análisis fue la estrategia de "shotgun" (o escopeta). Este sistema computarizado empieza fragmentando el genoma en un conjunto al azar de piezas de una longitud conocida de 2,000 pb, 10,000 pb y 50,000 pb. Luego de su secuenciación, se usan operaciones matemáticas para ensamblar los fragmentos en bloques continuos y asignarlos al lugar correcto que ocupan en el genoma. En comparación, el método de Sanger duplica grandes trozos del código humano en forma de clones BACs los cuales pueden ser colocados en el mapa del genoma en la región apropiada. Esta estrategia, usada por los investigadores del PGH y financiado con fondos públicos, concentra al comienzo más tiempo y esfuerzo en la generación de clones y mapas, mientras que la estrategia de Celera Genomics usa al final con más intensidad las computadoras.

Celera Genomics empleó el método de escopeta para establecer la secuencia del DNA cubriendo, por lo tanto, todo el genoma cinco veces (11, 12). A continuación, la información del genoma de los BACs, depositada en el banco de datos público, fue dividida en segmentos cortos de 550 bases y cargada en la fórmula de escopeta para cubrir el genoma otras 2.9 veces. El grupo de investigación ensambló luego varias veces la secuencia del genoma humano, usando dos fórmulas u operaciones matemáticas: un

enfoque de Ensamblado del Genoma Íntegro (EGI), que permitió trabajar de inmediato con toda la secuencia, y un enfoque de Ensamblado de Escopeta Dividido en Compartimientos (EEC), diseñado para aclarar segmentos específicos (6, 11). La secuencia producida por Celera cubre más del 99 % del genoma. Según Celera, alrededor del 85% del genoma está en segmentos correctamente ordenados de, por lo menos, 500,000 pb. La secuencia indica la existencia de 26,383 genes codificantes de proteínas y, débilmente, indica la existencia de otros 12,731 genes humanos hipotéticos. Por lo tanto, el total de genes varía de 26,383 a 39,114. Para hacer una doble comprobación de la precisión de su secuencia, se comparó el trabajo descrito con secuencias completadas de los cromosomas 21 y 22 y se encontró una "excelente concordancia". Cuando se comparó con la secuencia del resto de los cromosomas, el análisis insinuó que hay muchos más puntos de ruptura (errores de ensamblado o variaciones genéticas) en el proyecto de ensamblado (financiado con fondos públicos) que en el ensamblado de Celera (9, 11, 12).

En últimas fechas se crearon los chips de hibridación que sirven como un método de secuenciación por hibridación en chips con oligonucleótidos. Se basa en sintetizar distintas sondas de oligonucleótidos y unir las en disposiciones ordenadas (arreglos) a una fina placa de nylon o vidrio. El chip se prueba frente a un DNA marcado fluorescentemente, de modo que el patrón y cantidad de fluorescencia suministra información sobre la secuencia del DNA en cuestión. A modo de estudio piloto sobre sus posibilidades, la empresa Affimetrix ha logrado re-secuenciar por este método los 16 kb de DNA mitocondrial humano, con un dispositivo formado por 135,000 oligonucleótidos (2, 13).

RESULTADOS OBTENIDOS POR EL PROYECTO GENOMA HUMANO

Al momento, el PGH ha arrojado algunas cifras:

1. El genoma humano contiene 3.1647×10^6 kb de nucleótidos.
2. El tamaño del gen promedio consiste en 3000 bases, pero los tamaños varían enormemente. El gen humano más grande

conocido es que codifica para la distrofina con 2.4 millones de bases.

3. El número total de genes estimado es de 30,000 a 35,000, mucho menor que las estimaciones hechas anteriormente, que iban de 80,000 a 140,000.
4. Casi todas las bases de nucleótidos (99.9%) son exactamente las mismas en toda la población.
5. Las funciones son desconocidas para más del 50% de los genes descubiertos.
6. Menos del 2% del genoma codifica para proteínas.
7. Las secuencias repetidas que no codifican para proteínas compone al menos el 50% de genoma humano.
8. Las secuencias repetitivas, como se pensaba, no tienen funciones directas, pero apoyan la estructura y la dinámica del cromosoma. Con el tiempo, estas repeticiones reacomoda al genoma, creando nuevos genes y modificando y reorganizando los genes existentes.
9. Durante los pasados 50 millones de años, parece haber ocurrido una disminución dramática en la tasa de acumulación de repeticiones en el genoma humano (6).

La porción de heterocromatina secuenciada del genoma comprende aproximadamente 2.9×10^6 kb de DNA. El tamaño total del genoma es aproximadamente 3.2×10^6 kb, por lo que el restante 10% (0.3×10^6 kb) corresponde a secuencias altamente repetitivas en la heterocromatina. La mayoría de las secuencias altamente repetitivas son elementos transponibles, es decir, que pueden cambiar de lugar, que se mueven a lo largo de todo el genoma como intermediarios de RNA, estos representan el 45% de la secuencia de euromatina. Otro 5% del genoma consiste de segmentos de DNA duplicado, de tal manera que cerca del 60% del genoma humano consiste de secuencias de DNA repetitivas (3).

PERSPECTIVAS DEL PROYECTO GENOMA HUMANO

El PGH influirá en la biomedicina, pues permitirá avanzar en el conocimiento de la base genética de enfermedades (Medicina Genómica) y abrirá perspectivas nuevas en el diagnóstico, pronóstico y tratamiento de enfermedades; aunque esto último requerirá de investigación post-genómica.

Diagnóstico molecular: Detección de mutaciones. El avance en el PGH y en los mapas y secuencias ya disponibles impulsan la necesidad de disponer de técnicas capaces de rastrear DNA en la búsqueda de mutaciones asociadas con enfermedades humanas. En los próximos años se habrán identificado la mayor parte de los genes implicados en enfermedades humanas importantes y en las bases de datos se está introduciendo información sobre mutaciones y sus implicaciones clínicas. Uno de los retos de la medicina será hacer uso de dicha información para mejorar el diagnóstico y pronóstico de las enfermedades dentro de segmentos de DNA con secuencia conocida.

Nuevos y mejores marcadores polimórficos. Los SNPs se encuentran frecuentemente en el genoma y existe un SNP cada 1,000 pb aproximadamente. Si bien la mayoría de los SNPs se encuentran fuera de la región codificante de los genes, algunos contribuyen directamente a la patología por alteración de un gen y como consecuencia, de la proteína correspondiente. Los SNPs son la base de un fuerte desarrollo en el terreno de la disciplina farmacogenómica y para el descubrimiento de nuevos caracteres asociados con enfermedades.

Los "chips" genéticos. Los SNPs proporcionan la base para confeccionar un mapa genético detallado de cada individuo. La posibilidad de evaluar la variación individual en un solo ensayo genético es factible gracias al desarrollo de la microelectrónica aplicada a la genética molecular. Un conjunto de sondas derivadas de SNPs (que representan la variación genómica de un individuo) se hace reaccionar con el chip y permite obtener un patrón de hibridación característico del genoma del individuo estudiado. El perfil genético se interpreta en un lector

de fluorescencia confocal y en su análisis interviene un procesador electrónico. Este tipo de análisis permitirá no sólo continuar el análisis de enfermedades genéticas monofactoriales, sino también comenzar con mejores posibilidades de éxito con el análisis de enfermedades multifactoriales, tales como la hipertensión, la esquizofrenia, las enfermedades maniaco-depresivas o la diabetes tipo II. Otra consecuencia de esta tecnología es la búsqueda de un medicamento adaptado al perfil genético de cada paciente, base de los desarrollos farmacogenómicos (6, 12).

CONSIDERACIONES ÉTICAS

El desarrollo científico, en lo que respecta al PGH, abre las puertas a diferentes tratamientos que podrían ser benéficos para el hombre; sin embargo, no se debe olvidar que esto implica manipular directamente los mecanismos que transmiten la vida y dirigen la evolución de las especies, incluyendo la humana. Estos hechos exceden los conceptos de ética y humanidad, ya que nunca se había pensado en la posibilidad de que la vida fuera manipulada de este modo. Así surgen preguntas como: ¿se debe prohibir o desaconsejar algún tipo de manipulación genética?, ¿a quién le corresponde la responsabilidad de discriminar entre lo permitido

o no? Así, la UNESCO se compromete a promover y desarrollar la reflexión ética en los avances científicos dentro de las áreas de la biología y la genética (1, 5, 8, 9)

CONCLUSIONES

En este trabajo se presenta un análisis cualitativo y cuantitativo del Proyecto Genoma Humano, con base en la información obtenida de diferentes medios bibliográficos. El PGH es un esfuerzo científico internacional que pretende localizar y secuenciar todos los genes que constituyen el genoma humano y de algunos otros organismos, con el fin de adquirir el conocimiento de la organización, estructura y función de los genes en los cromosomas. El PGH hace uso de la cartografía genética y la cartografía física para caracterizar el genoma. El conocimiento integral del genoma humano tiene aplicaciones potenciales que van más allá de la genética propiamente dicha, pues se podrá valorar la susceptibilidad y la resistencia a enfermedades y diseñar estrategias más poderosas y específicas para su control; así como profundizar en importantes aspectos del orden biológico y de interés económico y social. El conocimiento del genoma humano ofrecerá nuevas formas de prevención, diagnóstico y tratamiento de las enfermedades.

LITERATURA CITADA

1. Anónimo. 2003. *Genomics and Its Impact on Science and Society: The Human Genome Project and Beyond*. http://www.ornl.gov/sci/techresources/Human_Genome/publicat/primer2001/primer11.pdf
2. Butte, A. 2002. "The Use And Analysis Of Microarray Data". *Nature Drug Disc.* 1: 951 - 960.
3. Cooper, G.M. and Hausman, R.E. 2004. *The Cell*. ASM Press. Washington, D.C. 713 p.
4. Goldstein, D. J. 2001. *La era postgenómica. Medicina y biología modernas*. Encrucijadas UBA. Buenos Aires, 1(5): 22 - 33.
5. Iáñez-Pareja, E. 1997. <http://www.ugr.es/~eianez/Biotecnologia/etica.html>
6. Iáñez-Pareja, E. 1997. <http://www.ugr.es/~eianez/Biotecnologia/genoma-1.html>
7. Karp, G. 1998. *Biología Celular y Molecular*. McGraw Hill Interamericana. México. 746 p.
8. Panduro, A. 2002. *Biología Molecular en la Clínica*. McGraw Hill. México. pp. 314.
9. Ruddle, F. 2001. "Hundred-Year Search For The Human Genome". *Annu. Rev. Genomics Hum. Genet.* 2: 1- 8.
10. The International Human Genome Mapping Consortium. 2001. "A physical map of the human genome". *Nature.* 409: 934 - 941.
11. Venter, C.J. *et al.* 2001. "The Sequence of the Human Genome". *Science.* 291: 1304 - 1351.
12. Venter, C.J., Adams, D.M., Granger, S.G., Kerlavage, R.A., Smith, H.O. and Hunkapiller, M. 1998. "Shotgun Sequencing of the Human Genome". *Science.* 280:1540-1542.
13. Weaver, R.F. 2002. *Molecular Biology*. Second edition. McGraw Hill. EUA. pp. 786 - 788.