

ENSAYOS REVISIONES Y OPINIÓN
rev.salud.hist.sanid.on-line

SECUENCIAMIENTO DE GENOMAS DE PATÓGENOS HUMANOS

HUMAN PATHOGENS GENOMES SEQUENCING.

Moreno- Granados Jefer Ivan *
Moreno-Granados Pedro Mauricio **

Moreno-Granados JI; Moreno-Granados PM. Secuenciamiento de genomas de patógenos humanos. Rev.salud.hist.sanid.on-line. 2007. 2(1): 14 - 22

* *Medico cirujano. Magíster en genética Humana. Docente Escuela de Medicina UPTC. Grupo de Investigación en Salud Pública. GISP. jeferuco@yahoo.com.*

** *Medico cirujano. Investigador Health Care Colombia. Docente.*

Resumen: En este trabajo se hace una reseña histórica de los hitos relevantes del secuenciamiento de ácidos nucleicos, genomas y secuenciamiento de genomas de patógenos humanos; se revisan las técnicas de secuenciamiento y las aproximaciones técnicas en el secuenciamiento de genomas completos (extremos rotulados y escopetazo). Así mismo se revisan, en forma breve, conceptos relacionados con el impacto de la genómica de patógenos como son la transcriptómica, islas de patogenicidad, inmunoma, epitomica y vacunología

Además se publica la encuesta hecha a la base de datos de GenBank (www.ncbi.nlm.nih.gov/genomes) por genero y especies de patógenos humanos reconocidos, encontrándose para abril de 2005, Que incluye 187 genomas de patógenos humanos virales secuenciados, 69 genomas de patógenos humanos bacterianos secuenciados y 23 genomas de patógenos humanos eucariotes secuenciados.

Palabras clave: Secuenciamiento, patógeno humano, transcriptómica, islas de patogenicidad, inmunomica, epitomica, vacunología

Abstract: In this paper we review the most important events in nucleic acids sequencings and human pathogen genome sequencing projects; We describe sequencig techniques and methological approaches for whole genome sequencing (sequence tagged sites and shotgun). Also We explain some basical concepts about the impact of pathogen genomics in medicine, such as transcriptomics, pathogenicity islands, immunomics epitomics and vaccinology

In other hand, We show a survey from genbank (www.ncbi.nlm.nih.gov/genomes) of human pathogen genome sequencing projects finished until April 2005. This survey includes data from 187 viral, 69 bacterial and 23 eucariotic human pathogens

Key Words: sequencings, human pathogen, transcriptomics, pathogenicity islands, immunomics epitomics, vaccinology

Las técnicas de secuenciamiento inician con Robert Holley en la determinación de la secuencia de RNAs de cadena corta como los tRNAs.

Hacia 1965 Sanger había dejado de secuenciar proteínas para dedicarse a desarrollar métodos sencillos y rápidos para secuenciar largas cadenas de RNA. Empleando dichos métodos, Sherman Weissman, en la Universidad de Yale y Walter Fiers en Gante, habían secuenciado a finales de 1976 más de la mitad del ADN del virus SV40 que tiene 5226 pares de bases.

El avance decisivo lo obtuvo Sanger 1975 al encontrar sistemas para el secuenciamiento de DNAs de cien a quinientas bases de longitud; basados en la elongación de cadenas de DNA (replicación). Esta técnica y sus derivadas se conocen actualmente como método sintético.

Con uno de estos métodos conocido como el método "mas menos" se determinó fácilmente la secuencia de los 5386 pares de bases del genoma del DNA de ϕ X174.

Allan Maxam y Walter Gilbert desarrollaron en 1977 otro método igualmente poderoso basado en la ruptura química en sitios específicos de mono hebras de DNA. Por este método se conocieron las secuencias de las 5226pb de SV 40 y los 4362pb del plásmido recombinante pBR322 determinados en menos de un año por Greg Sutcliffe en el laboratorio de Gilbert. Esta ultima técnica se conoce actualmente como método químico.

Las dos técnicas el método químico y el método sintético tienen en común varios aspectos.

Ambas técnicas trabajan sobre mono hebras de DNA, por lo tanto se inicia el proceso con la clonación del DNA a secuenciar dentro de un vector que tenga parte de su ciclo como monohebra de DNA , por ejemplo, el fago M13

Un extremo de la hebra de secuenciamiento debe estar marcado, actualmente el marcaje no es isotópico pero en aquella época sí.

El producto de la acción de secuenciamiento en ambas técnicas se lee por electrofóresis gracias a la migración diferencial de los fragmentos de diferente tamaño y a la marcación, en aquel tiempo, se usó la autoradiografía actualmente los sistemas moderno usan electroforesis en tubo capilar y lectura por láser de los flourocromos marcadores.

Descripción de las técnicas

Método de secuenciación de Maxam y Gilbert (Método Químico). Se marca el extremo 5 prima de los segmentos mono hebra de DNA con fósforo radioactivo ^{32}P . Este DNA se separa en cuatro alícuotas y se trata cada alicuota con un producto químico diferente que rompe el DNA en una o dos bases específicas. Por ejemplo: el sulfato de dimetilo o dimetilsulfato, un veneno industrial, rompe las cadenas de ADN en los lugares donde existen las bases A o G, y que la hidracina, un compuesto gaseoso, debilita los lugares de las bases C y T. Para discriminar la reacción entre purinas con dimetilsulfato, la reacción con adenina se favorece por acidificación del medio donde se hace la reacción y para diferencia la citosina de la timina se hace la reacción en NaCl 2M.

Controlando las condiciones de la reacción se limita el daño a pocas bases por molécula en cada alicuota.

El tratamiento de las moléculas afectadas con piperidina rompe el azúcar y el enlace fosfato en los sitios donde las bases han sido destruidas. Entonces, se obtienen fragmentos de diferente longitud en cada alicuota de DNA los cuales son resueltos por electroforesis en un gel fino de poliacrilamida urea 8M y se revelan con autoradiografía en una placa fotográfica.

Ordenado los fragmento de cada alicuota de menor a mayor encontramos la secuencia de la monohebra de DNA a estudio; si se registran inconsistencias se puede secuenciar la hebra complementaria.

Método Sintético de Sanger

Para entender el método sintético debemos puntualizar algunos aspectos del proceso de síntesis de DNA.

El DNA se replica en dirección 5' a 3' es decir primero se colocan los nucleótidos del lado del grupo fosfato y se termina con el nucleótido del lado hidroxilo.

Para la replicación, se necesita una hebra molde de DNA, la cual por definición del antiparalelismo debe ir de 3' a 5'. Para poder iniciar la replicación se necesita una estructura complementaria al template y que done un sitio hidroxilo 3' en el extremo; esta estructura es un ácido nucleico anti paralelo y complementario a la sección 3' del template. Esta molécula es conocida como Primer o cebador, en los sistemas naturales se usa un RNA sintetizado por una RNA polimerasa dependiente de DNA llamada Primasa. En los sistemas de síntesis artificial se usa DNA.

Los nucleótidos en la hebra que se está sintetizando se incorporan siguiendo la ley de complementariedad de las bases Adenina frente a Timina y Citosina frente a Guanina, y usando el grupo hidroxilo de nucleótido precedente hacen un enlace fosfodiéster

Como se dice atrás el método de Sanger consiste en una síntesis artificial de DNA que esta marcado en el extremo 5' y que es abortada estratégicamente en una base específica. Para lograr esto, se diseñaron análogos de nucleótidos que fueran capaces de incorporarse al DNA de manera específica; pero que evitaran al ser incorporados la síntesis de más DNA o sea que evitan la elongación de la hebra donde se incorporan.

Es así, como se fabricaron los dideoxinucleótidos trifosfato, unos "nucleótidos" que al contrario de los nucleótidos normales del DNA no tienen 2' deoxiribosa sino 2', 3' dideoxiribosa. Por lo cual, si pueden hacer enlace fosfodiéster con el hidroxilo 3' del nucleótido anterior en la síntesis pero no son donantes de sitios hidroxilo 3' para continuar la síntesis de la hebra.

De este modo se hace una reacción en cuatro alicuotas con templates mono hebra de DNA a estudio, primers marcados en su extremo 5', 2' deoxinucleótidos trifosfato y un dideoxinucleótido trifosfato específico para cada alicuota o sea dideoxinucleótidos trifosfato de adenina para un tubo, dideoxinucleótidos trifosfato de guanina para otro, dideoxinucleótidos trifosfato de citosina para el siguiente y dideoxinucleótidos trifosfato de timina para el último.

Se tiene entonces cuatro reacciones de amplificación que se van a parar en

Adenina, Guanina, Citosina y Timina respectivamente. Formando DNAs de diferente tamaño con extremo 5' primario marcado. Estos DNAs se pueden resolver por electroforesis en gel de poliacrilamida y revelar por autoradiografía y de la misma forma que en el método químico leyéndose la secuencia.

Estrategias de secuenciamiento de los genomas

En cuanto al secuenciamiento existen dos tipos de aproximaciones:

Secuenciamiento Aleatorio a lo largo del genoma sin importar si la región secuenciada es codificadora o no. Este método es descrito también como el método del escopetazo.

Secuenciamiento de segmentos rotulados: en esta aproximación se parte de secuencias cuyos extremos son conocidos ya sea por que se obtienen por digestión con una enzima de restricción de corte específico o por que se clona respecto a una característica funcional determinada de DNA; o por mapeo inferido por recombinación, de este tipo de secuencias las hay de dos clases los contigs y los DNA complementario.

- El **secuenciamiento por contigs** es el secuenciamiento del genoma que parte de segmentos del mismo que tienen la propiedad de estar traslapados (secuencias que comparten información en los extremos y que al organizarlas en forma concatenada dan lugar a una secuencia más larga) unos con otros y de tener segmentos contiguos y que ya han sido clonados (Figura 1).
- El **secuenciamiento por DNA complementario (cDNAs)** es el

secuenciamiento de DNAs obtenidos por retrotranscripción de secuencias de RNA mensajero que se expresan en un determinado tejido, por ejemplo Cerebro, con lo cual, solo se secuenciarían regiones funcionales del genoma para la síntesis de proteínas de ese tejido (Figura 2).

Principales Hitos de secuenciamiento

- 1964 Robert Holley inicia la historia del secuenciamiento de los ácidos nucleicos: el primer ácido nucleico en ser secuenciado es el tRNA de alanina de *Saccharomices*
- 1977 Fago ϕ x174 5.3kb
- 1978 SV40 5.2kb
- 1979 HVB 3.2kb
- 1981 DNA mitocondrial humano 16.6kb
- 1982 Fago lambda 48kb
- 1984 EBV 172kb
- 1995 Venter Se secuencia el genoma de un ser de vida independiente *Haemofilus influenzae* Rd 1.83Mpb
- 1996 Goffeau anuncia el secuenciamiento del genoma del eucariote *Saccharomyces Cerevisiae* de 12Mpb
- 1997 *E. coli* K12 4.6Mb
- 1998 se secuencian los genomas de las bacterias del tifo y la tuberculosis
- 1998 *C. elegans* 97Mb
- 2000 *D. melanogaster* 120Mb (Eucromatina) Se anuncia el secuenciamiento del genoma de *Drosófila*
- 2000 Cromosoma 22 humano
- 2000 Cromosoma 21 humano
- 2000 Genoma Humano, Se anuncia el secuenciamiento del Genoma Humano Junio 2000 (Bill Clinton y Tony Blair)
- 2001 *A. tumefaciens*

Genomas secuenciados a la fecha

Genomas Virales

Para enero 27 de 2004 en GenBank se registraba el secuenciamiento de 1506 genomas virales. De los cuales unos 187 genomas eran de virus patógenos humanos , podemos encontrar las secuencias de virus tan frecuentes como el de las paperas y el sarampión a genomas de virus exóticos como el de la fiebre de lassa o virus consentidos como SARS coronavirus que en menos de un mes de describirse el cuadro clínico se secuenció el genoma del agente etiológico o de virus tan interesantes como el de la hepatitis C del cual primero se conoció el genoma y luego se buscó el virión lo cual dio origen a la virología sin virus.

Genomas procariotes

Para marzo 10 en GenBank se registraba la existencia de 112 genomas procariotes secuenciados a la fecha con 16 arqueas y 96 Eubacterias. Dentro de esos genomas se encuentran 69 genomas de procariotes patógenos para humanos como el genoma del neisseria , micobacteria tuberculosis, micobacteria leprae entre otros

Genomas Eucariotes

Para el 22 de mayo de 2003 se reportaba la existencia de 447 genomas eucarióticos secuenciados entre DNAs Mitocondriales y genómicos . Entre estos genomas se encuentran unos veintitrés genomas de patógenos humanos

Algunas aplicaciones de estas técnicas

El secuenciamiento de genomas ha producido la aparición de las siguientes disciplinas y apreciaciones de la biología:

La genómica que es el secuenciamiento a gran escala de genomas

La genómica Comparativa Busca comprender los mecanismos evolutivos por la comparación de genomas relacionados y poco relacionados, haciendo inferencias filogenéticas .

La comparación entre los genomas de una misma especie de cepas patogénicas contra los genomas de cepas no patogénicas lleva al descubrimiento de las Islas de patogenicidad

Las islas de patogenicidad son segmentos o regiones dispersas de DNA de organismos patogénicos que codifican factores de virulencia y son la diferencia entre cepas patógenas y no patógenas de un mismo organismo; tienen composición de bases diferentes del resto del genoma, están asociadas a elementos genéticos móviles (fagos, plásmidos o transposones) y presentan varias copias en mosaico de varias adquisiciones; los estudios filogenéticos han demostrado, en algunos casos, la transferencia horizontal de genes de estas regiones entre varias especies de organismos patogénicos aparentemente no relacionados como es el caso de *Escherichia coli* y *Shigella* (gen Shiga).

Otro aspecto rentable de la genómica comparativa es la comparación del genoma y, transcriptoma y proteoma del hospedero contra el germen, lo cual ha llevado por lo menos a la concepción de nuevos blancos terapéuticos, dirigidos a los puntos no comunes mejorándose el desarrollo guiado de fármacos.

La anotación de genomas es descubrir el significado de cada uno de los nucleótidos de un genoma así como el significado de la posición de dicho nucleótido en otras

palabras dónde empieza y dónde termina un gen..

La Transcriptómica es el análisis de secuenciamiento de RNAs a gran escala de un organismo en una fase de su vida dado, es decir el análisis del transcriptoma o conjunto de RNAs de un organismo; este proceso abarca técnicas cuantitativas y técnicas automatizadas como los microarrays o chips de DNA

Con esta disciplina se pueden obtener los perfiles de expresión génica de un determinado estadio del Ciclo vital e inferir perfiles de expresión y de interacción hospedero - germen

Estos análisis nos aproximan al conocimiento de los mecanismos moleculares que hacen que la interacción hospedero- germen producen enfermedad es decir el patonoma y su estudio a gran escala: la Patonomica

Los anteriores conocimientos lleva a un mejor entendimiento del proteoma o complemento completo de proteínas de un ser vivo, por medio de la Proteómica o análisis de proteínas a gran escala de un organismo con técnicas como la electroforesis de campo pulsado y resonancia magnética nuclear de péptidos

Un campo provechos y de gran aplicabilidad de la proteómica es el estudio del inmunoma o el complemento completo de antígenos (principalmente proteínas) de un organismo patógeno dado, la Inmunómica o análisis a gran escala de antígenos y epitopes (Epitómica) nos permiten vislumbrar la nueva vacunología.

Como es sabido, para hacer una vacuna eficaz, no solo es necesario que se identifique un antígeno sino que este

desencadene una respuesta celular o humoral especifica y protectora. Es conocido que no en todos los casos el antígeno modales el mas inmunogeno ni por supuesto aquel que desencadena una respuesta protectora o neutralizante.

Con la identificación de diferencias reales entre cepas de un mismo patógeno se hacen mejoras a las técnicas de Epidemiología Molecular de patógenos

Referencias

Brown k. The human genome business today. Scientific American. Julio 2000. Pags 40-47.

Schuler g.d. et al. A gene map of the human genome., Science 1996; 274, 540-545

Hattori I et al. The DNA Sequence of human chromosome 21. Nature 2000; 405, 311-319.

Carballo L. Impacto de la revolución genómica en nuestro entendimiento de la enfermedad alérgica, ponencia presentada en el Simposio nacional en genómica, infección, inmunidad y alergia, Medellín, Colombia, Asociación Colombiana de Alergia, Asma e Inmunología 29,30 y 31 de agosto de 2002

Doytchinova I A et al proteomics in vaccinology and immunobiology: An informatics perspective of th immunome . Journal of Biomedicine and Biotechnology 5;267-290, 2003 [http:// jbb.hindawi.com](http://jbb.hindawi.com)

Jwatson J.D. Tooze J. Kurtz D.T. AND Recombinante Introducción a la ingeniería genética , primera edición, editorial labor, Barcelona, España 1986

www.ncbi.nlm.nih.gov

www.genomicglossaries.com/content/omes.asp

www.sciencemag.org science vol 300 ; 11 april 2003

Figura 1. Secuencias conting

Secuencias conting
5'cctgagccacggctctgccaggttatgggccacggcaagaaggtggc3'
5'ccacggcaagaaggtggccgacgcctgaccaacgccgtggcgcacgt3'

Secuencia completa
5'cctgagccacggctctgccaggttatgggccacggcaagaaggtggccgacgcctgaccaacgccgtggcgcacgt3'

Figura 2. DNA Complementario

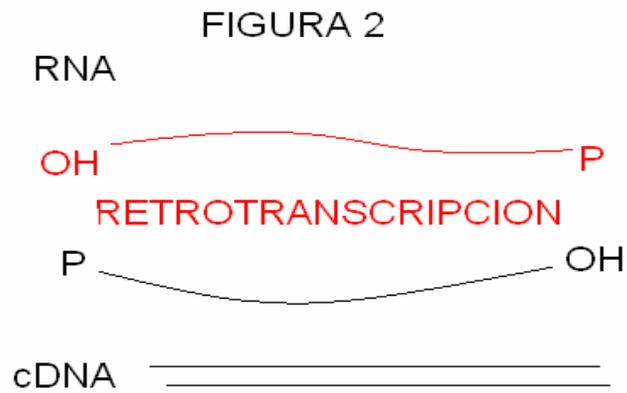


Tabla 1. Cambios de velocidad en el secuenciamiento

Eficiencia persona por año	Año	Hitos relacionados
1	1965	Holley secuencia RNA de Alanina
15	1970	Wu secuencia los extremos cohesivos del fago lambda
150	1977	Método Químico y Método sintético se aplican las técnicas de clonación de ácidos nucleicos
1500	1980	Messing Clonación en Fago M3
15000		
25000	1986	Hood Técnicas parcialmente automatizadas
100000		
1000000	1995	Venter Primer genoma de un ser d vida independiente secuenciado

Tabla 2. Genomas de eucariotes patógenos humanos

Patógeno	Enfermedad
<u>Ancylostoma duodenale</u>	Uncinariasis (anemia tropical)
<u>Ascaris suum</u>	Ascariasis (cerdo)
<u>Brugia malayi</u>	Filariasis
<u>Candida albicans SC5314</u>	Candidiasis
<u>Cryptococcus neoformans var. grubii</u>	Criptococosis (meningitis)
<u>Dermatobia hominis</u>	Miasis
<u>Echinococcus multilocularis</u>	Hidatidosis
<u>Eimeria tenella</u>	Isosporosis
<u>Fasciola hepatica</u>	Fascioliasis
<u>Leishmania major</u>	Leishmaniosis
<u>Necator americanus</u>	Uncinariasis
<u>Onchocerca volvulus</u>	Filariasis oncocercosis
<u>Plasmodium falciparum</u>	Paludismo
<u>Schistosoma japonicum</u>	Esquistosomiasis
<u>Schistosoma mansoni</u>	Esquistosomiasis
<u>Taenia solium</u>	Teniasis
<u>Toxoplasma gondii</u>	Toxoplamosis
<u>Trichinella spiralis</u>	Triquinosis
<u>Trypanosoma brucei</u>	Tripanosomiasis africana (enfermedad del sueño)