

PROYECTO GENOMA HUMANO

Dr. José MORDOH

En los 52 años transcurridos desde 1953, en que Watson y Crick describieron la estructura del DNA, se ampliaron en gran medida las perspectivas de estudio del genoma humano (GH) (se denomina **genoma** todo el DNA de un organismo, incluyendo los genes). El objetivo de describir los 30,000 – 35,000 genes presentes en el GH y lograr su secuencia completa comenzó en 1988, cuando el Congreso de los EEUU asignó fondos al DOE (Department of Energy de los EEUU), creando el DOE Human Genome Program (dirigido desde entonces por Ari Patrinos), y al National Institutes of Health (NIH) que creó una Sección especial, el NIH Human Genome Research Institute (dirigido desde entonces por Francis S. Collins), para contribuir al mismo objetivo. Ambos Programas se fusionaron el 1^o de octubre de 1990 para constituir el US Human Genome Program.

Los objetivos del Proyecto Genoma Humano (PGH) son :

- ◆ Identificar los aproximadamente 30.000 genes en el GH.
- ◆ Determinar la secuencia de las 3.000 millones de bases que conforman el DNA haploide humano.
- ◆ Almacenar esta información en bases de datos.
- ◆ Mejorar las herramientas para el análisis de los datos.
- ◆ Transferencia de tecnología al sector privado.
- ◆ Encarar los problemas legales, éticos y sociales (ELSI) que surjan del PGH.

Se estableció entonces un plazo de 15 años para llevar adelante el proyecto a un costo aproximado de 3.000.000.000 U\$S (1 dólar por base). Otros países llevan adelante sus propios Programas sobre este tema, y por lo menos 18 países han establecido programas de investigación sobre GH. Existe también la Human Genome Organization (HUGO), organización que afilia 50 países y donde más de 1,000 científicos de todo el mundo colaboran para monitorear la marcha del programa.

El propósito del Proyecto del Genoma Humano (PGH) es construir un mapa lo más completo posible del GH y lograr su secuenciación completa.

Afortunadamente, la tarea está avanzando a un ritmo más rápido que lo previsto. En marzo de 1999 ya estaba secuenciado un 15 % del GH, y en junio de 2000 se dispuso de un “borrador” de la secuencia completa, aunque dicho borrador no sea obviamente tan útil como la secuencia definitiva completa. En febrero de 2001, números especiales de las revistas Nature y Science publicaron el borrador de trabajo de la secuencia y su análisis. El producto final debe tener cuatro características principales:

- 1) **Precisión:** debería existir una precisión en la secuencia > 99.99 % (menos de una base errónea cada 10.000 bases).
- 2) **Ensamblado:** los trozos más pequeños secuenciados deben ser ensamblados correctamente en piezas genómicas que reflejen fielmente el DNA original.

- 3) **Costo razonable:** el desarrollo de tecnologías deberá abaratar el costo de la secuenciación .
- 4) **Accesibilidad:** la secuencia completa deberá ser accesible dentro de las 24 hr de obtenida a través de bases de datos públicas, como el GenBank en los EEUU.

Para su mejor análisis, este Capitulo se dividira en cuatro Secciones Principales:

- I) Detalles técnicos del desarrollo del PGH
- II) Conclusiones principales obtenidas del borrador de secuencia del PGH
- III) Proyecto Proteoma Humano
- IV) Puntos futuros a investigar

I) DETALLES TECNICOS DEL PGH

Como se ha visto en Secciones anteriores, los **genes** (elementos codificantes de proteínas) están compuestos por **exones**, que contienen las secuencias codificantes propiamente dichas y que a su vez están separados entre sí por **intrones**, que contienen secuencias no codificantes. Dos de las herramientas principales para poder llevar a la práctica el PGH fueron el descubrimiento de las **enzimas de restricción (ER)** y de la **reacción de polimerasa en cadena (PCR)**. Las ER, de las que existen centenares en este momento, fueron descubiertas analizando el fenómeno de la restricción: ciertas bacterias podían degradar DNA ajeno pero no el propio. Las ER reconocen secuencias específicas de bases y producen entonces cortes en el DNA. Cuanto menor sea el número de bases de la secuencia de reconocimiento, mayor será la probabilidad de cortes producidos por las ER, y por lo tanto menor será el tamaño promedio de los fragmentos de DNA obtenidos. Si el número de bases de la región de reconocimiento es 4, el tamaño promedio del DNA cortado es de 246 bases; si es 6 bases el DNA promedio será de 4,000 bases (4 Kb), y si el número de bases de la región de reconocimiento es 8, el tamaño promedio será de 64,000 bases (64 Kb). Por ejemplo una ER típica es Eco RI, que reconoce la secuencia G↓AATTC y corta

$$\begin{array}{c} \text{C TTAA}\hat{\uparrow}\text{G} \\ \text{G AATT C} \end{array}$$

selectivamente la doble hélice de DNA en dichos puntos, generando lo que se denomina "**sticky ends**" (**hacer el ejercicio en papel**).

Con respecto a la **PCR**, puede amplificar cientos de millones de veces, en pocas horas, una secuencia de DNA conocida, de origen bacteriano, viral, vegetal ó humano. La PCR es un proceso basado en una DNA polimerasa termoresistente, capaz de sintetizar DNA sobre un templado utilizando los primers adecuados. Mediante ciclos de calentamiento (para separar las hebras sintetizadas) y de enfriamiento (para comenzar un nuevo ciclo de síntesis) la PCR amplifica exponencialmente el DNA sintetizado.

La **clonación** involucra el uso de la tecnología del DNA recombinante para propagar fragmentos de DNA en un huésped extraño. Los fragmentos son generalmente aislados de cromosomas utilizando ER y son unidos a un vector. Introduciendo el vector en células huésped adecuadas, los fragmentos de DNA son reproducidos junto con el DNA de la célula huésped. Los vectores son moléculas de DNA originados en virus, bacterias ó células de levadura. Los vectores pueden acomodar fragmentos de DNA extraño que van desde 12 Kb para vectores bacterianos (plásmidos y cósmidos) hasta 1 Mb para vectores de levaduras (cromosomas artificiales de levadura ó YAC (“yeast artificial chromosomes”). Especialmente útiles para fragmentos de DNA que pueden ser fácilmente clonados y caracterizados ha sido el desarrollo de los BAC (“bacterial artificial chromosomes”).

Antes de 1996, los objetivos de la secuenciación de DNA estaban dirigidos hacia genomas de organismos modelos. El genoma de la levadura *Saccharomyces cerevisiae* fué el primer genoma eucariótico en ser secuenciado. Contiene 12.057.500 bp en el DNA nuclear, con 6.000 genes distribuidos en 16 cromosomas. La secuencia del gusano *Caenorhabditis elegans*, que posee 100.000.000 bp distribuidos en 19.000 genes y 6 cromosomas demostró que 74% de los genes humanos conocidos tienen un homólogo en *C. Elegans*.

El genoma humano diploide está distribuido en 46 cromosomas (23 paternos y 23 maternos), de los cuales 44 son autosomas y 2 son sexuales. Así el complemento diploide para las mujeres es **44 XX** y para los hombres es **44 XY**. Por técnicas de coloración especiales se pueden localizar en los cromosomas “bandas” definidas.

Para visualizar la dimensión del problema de secuenciar en forma completa el GH, tengamos en cuenta que el GH haploide (23 cromosomas) contiene aproximadamente 3.000.000.000 de pares de bases, al igual que la mayoría de los mamíferos. Inmersos en esa enorme cantidad de DNA, y ocupando sólo el 3 % de la misma, se encuentran entre 30.000 – 35.000 genes (Tabla 1). Cuando el tamaño del genoma humano esté escrito, ocupará alrededor de 200 libros del tamaño de una guía telefónica.

Tabla 1. Tamaño comparativo de los genomas de diversas especies

Especie	Genoma Haploide (pares de Bases, bp)	Genes (n)
Homo sapiens	3.000.000.000.	30.000 - 35.000
Mus musculus	3.000.000.000.	30.000 - 35.000
Rattus norvegicus	3.000.000.000.	30.000 - 35.000
Drosophila (mosca de la fruta)	165.000.000.	15.000 - 25.000
C. elegans	100.000.000.	19.000
Levadura y hongos	14.000.000.	8.355 - 8.947
E. Coli	4.670.000.	3.237
H.Influenzae	1.800.000.	
M. Genitalium	580.000.	

¿QUÉ ES UN MAPA GENÉTICO?

Un mapa genético describe el orden de los genes u otros marcadores y el espaciamiento entre ellos en el cromosoma, y puede ser construido a diferentes niveles ó escalas de resolución. Así como existen mapas geográficos topográficos, políticos, y hasta a nivel de carreteras primarias ó secundarias, así también existen diferentes tipos de mapas a nivel genómico.

MAPA GENÉTICO DE UNIÓN (LINKAGE)

En el nivel menos fino se encuentran los “mapas genéticos de unión (linkage)”, basados en un estudio cuidadoso de los patrones de herencia humanas, y que muestran las ubicaciones relativas de marcadores específicos de DNA a lo largo del cromosoma.

¿Qué es un “marcador”? Cualquier característica física ó molecular, heredada, que difiera entre distintos individuos y sea potencialmente detectable, es un marcador genético potencial. Los marcadores deben ser **polimórficos** para ser útiles en el mapeo: esto es, deben existir formas alternativas entre individuos para ser detectables entre distintos miembros en estudios familiares. Los polimorfismos son variaciones en la secuencia de DNA que existen en promedio cada 300-500 bp.

Variaciones en la secuencia de los exones pueden llevar a variaciones observables, como el color de ojos, cabello, grupo sanguíneo. La mayor parte de las variaciones, sin embargo, existen dentro de los intrones y tienen poca ó ninguna influencia sobre el aspecto ó función de los organismos. Sin embargo, dichos cambios pueden ser detectables a nivel del DNA y ser utilizados como marcadores. Ejemplo de ello son los RFLP (fragmentos de restricción polimórficos) que reflejan variaciones en la secuencia de DNA en sitios que pueden ser reconocidos por ER. Otro ejemplo son los números variables de secuencias repetidas en tandem (VNTR: variable number of tandem repeats), cortas secuencias repetidas que varían entre individuos en el número de secuencias repetidas y por lo tanto, en longitud, generando “short tandem repeat polymorphisms (STRPs)”.

El mapa genético de unión se construye observando cuán frecuentemente dos marcadores se heredan conjuntamente. Durante la producción normal de óvulos y espermatozoides, las hebras de DNA ocasionalmente se cortan y reúnen en diferentes partes del mismo cromosoma ó del cromosoma homólogo. Este proceso se denomina “recombinación meiótica”. Dos marcadores ubicados cerca uno del otro en el cromosoma tenderán a ser heredados conjuntamente, dado que es menor la probabilidad de que en la recombinación meiótica se produzca un corte entre ambos. En el mapa genético de unión, la distancia entre marcadores se mide en centiMorgans (cM) (nomenclatura adoptada en honor al genetista Thomas Morgan). Se dice que dos marcadores están separados por 1 cM si son separados por recombinación con una frecuencia del 1%. Una distancia genética de 1 cM equivale a una distancia física de 1 Mb.

En 1994, un consorcio internacional de investigadores de Francia y EEUU publicó un mapa genético de unión conteniendo casi 6.000 marcadores, incluyendo genes implicados en la fibrosis quística, distrofia miotónica, enfermedad de Huntington, enfermedad de Tay-Sachs, varios cánceres, etc. La distancia promedio entre marcadores fue de 0.7 cM.

MAPAS FÍSICOS: son así denominados porque la distancia se mide en unidades físicas “reales”, por ejemplo, en pares de bases (bp), y no en términos genéticos.

Mapas físicos de baja resolución.

Mapa cromosómico. Los distintos tipos de mapas físicos varían en su nivel de resolución. El mapa físico de menor nivel de resolución es el mapa cromosómico ó citogenético, basado en un patrón de bandas distintivas observada por el microscopio común sobre cromosomas teñidos. En él, los genes ú otros fragmentos identificables se asignan a sus cromosomas respectivos, con distancias medidas en bp (pares de bases). Estos marcadores pueden ser asignados a bandas específicas mediante la técnica de hibridización in situ.

Mapa de cDNA. Muestra las posiciones de las regiones de DNA que son expresadas (genes) respecto a regiones cromosómicas particulares ó bandas. El cDNA se sintetiza en el laboratorio a partir del mRNA como templado. Se piensa que los cDNA identifican las regiones del genoma con mayor significación biológica. Estos mapas pueden suministrar información sobre regiones cromosómicas cuya función es aún desconocida.

Mapas físicos de alta resolución.

Los dos enfoques actuales para generar estos mapas se denominan “arriba hacia abajo” (generando un mapa de macro-restricción” ó de “abajo hacia arriba” generando un mapa de contiguidad. Con ambos enfoques, los mapas representan conjuntos ordenados de fragmentos de DNA generados cortando el DNA genómico con ER y luego amplificados por clonación ó PCR. La fuente de DNA genómico puede ser el genoma entero ó bien cromosomas aislados mediante citometría de flujo.

Mapa de macro-restricción. Se obtienen cortando cromosomas aislados con enzimas de restricción que producen cortes raros. Los fragmentos mayores son luego ordenados. Este mapa tiene mayor continuidad y menores intervalos vacíos que el mapa por contiguidad, pero la resolución es menor y puede no ser útil para hallar genes particulares.

Mapas por contiguidad (abajo hacia arriba): este enfoque consiste en el corte del cromosoma en pequeños fragmentos, cada uno de los cuales es clonado y ordenado. Dichos fragmentos ordenados forman bloques de DNA contiguos ("contigs"). Los clones resultantes de las "bibliotecas" de DNA varían en tamaño de 10.000 bp hasta 1 Mb . Los mapas de este tipo consisten en una biblioteca de clones parcialmente superpuestos que representan un fragmento cromosómico pequeño. Son útiles para ubicar genes en una pequeña área, pero no se extienden por zonas extensas del genoma.

Los adelantos tecnológicos han hecho posible clonar grandes trozos de DNA en fragmentos que van desde 4 Kb en plásmidos comunes, 40 Kb en cósmidos hasta grandes fragmentos de 1 Mb en YAC. Si bien los YAC representan una herramienta clásica para clonar grandes fragmentos de DNA humano, no son perfectos. Algunas regiones del genoma no pueden ser clonados en YAC, y otras regiones están predispuestas a rearreglos. Se han desarrollado nuevos vectores como BAC, que aceptan insertos de 150 Kb en promedio, fagos P1, y sistemas de clonado artificiales derivados del fago P1.

Caminar sobre el cromosoma. Es una de las estrategias más utilizadas para rellenar los huecos. El mapa genético actual tiene alrededor de 41.000 marcadores, ó 1 marcador cada 100.000 bases.

Secuenciación: Finalmente, los mapas físicos y los clones son escalas intermedias al objeto más visible del PGH: la secuenciación completa de las cadenas de 3.000.000.000 de bp que definen nuestra especie. Incluidos, por supuesto, las secuencias de todos los genes así como las secuencias de DNA que codifican funciones mal comprendidas pero que pueden estar involucradas, por ejemplo, en la coordinación de la expresión génica en distintas partes de nuestro cuerpo. El mapa físico provee sets clonados y ordenados de DNA contiguo que representan regiones de un cromosoma, o incluso un cromosoma entero. Una vez que los marcadores genéticos definen la región que contiene el gen buscado, segmentos clonados del mapa físico proveen la oportunidad a partir de la cual se puede aislar el gen en cuestión. Existe una replica copiada de 98 % del genoma humano. Este mapa contiene más de 41.000 marcadores de DNA, llamados "sequence tagged sites" (STS), que alinean adecuadamente las bases. *Si bien se han secuenciado varios cientos de Mbp, la mayoría son de pequeños STS sobre fragmentos clonados. Solo 30 Mbp han sido secuenciados en segmentos más largos, siendo el más grande hasta el momento de 685,000 bp.*

TRANSCRIPTOMA HUMANO.

El plan del PGH que cubrió los años 1993-1998 expandió los objetivos del mapeo e incluyó específicamente nuevos objetivos para mapear los genes en sí mismos. Se comenzó entonces el Programa I.M.A.G.E (Integrated Molecular Analysis of

Genomes and their Expression) para posicionar DNA de regiones genómicas expresadas en el mapa físico. Este mapa genético representa el esfuerzo más intenso para ubicar e identificar los ~ 30,000 genes del GH. A comienzos de 1996, IMAGE había distribuido más de 250.000 clones de cDNA parciales y completos, la mayoría con uno ó ambos extremos secuenciados (**EST: expressed sequenced tags**) para proveer identificadores únicos. La longitud promedio de los EST es 300-500 bp cada uno. Más de 38,000 EST han sido ubicados en el mapa, otorgando a los investigadores que desean ubicar los genes de distintas enfermedades una lista de genes candidatos en la vecindad cromosómica en que se sabe está ubicado el gen responsable de una patología determinada. Como parte de este esfuerzo se han mapeado recientemente 2500 genes.

Existen entonces tres instancias de análisis del GH y sus productos:

GENOMA HUMANO : análisis de la secuencia completa

TRANSCRIPTOMA HUMANO: análisis de las secuencias transcriptas en mRNA (genes)

PROTEOMA HUMANO: análisis de las proteínas presentes en el hombre

El plan 1998-2003 puede ser considerado dedicado al desarrollo de herramientas nuevas y más diversas, que incluyen el catálogo de las variaciones en la secuencia del DNA humano; así como nuevas tecnologías y estrategias para estudiar funciones genómicas a escala del genoma completo.

Los métodos actuales de secuenciación son todavía lentos e imperfectos. Debe tenerse en cuenta que el cromosoma más pequeño (# Y) tiene 50 MB y el mayor (# 1) tiene 250 Mb. Por lo tanto, uno de los objetivos principales del PGH es el desarrollo de tecnologías automatizadas de secuenciación que permitan secuenciar con precisión 100.000 bases por día a un costo inferior a 0.50 \$/base. Los métodos más utilizados para la secuenciación del DNA se basan en el de Sanger, donde la replicación del DNA es interrumpida por la incorporación de un dideoxynucleótido. Recientemente, la introducción de la Electroforesis Capilar ha dado gran velocidad al proceso de secuenciación. Las muestras de DNA son introducidas en un dispositivo de 96 capilares; a medida que los fragmentos separados pasan a través de los capilares, son irradiados al mismo tiempo con laser, y la fluorescencia es determinada con un detector multicanal. Dado que fragmentos de todas las longitudes existen en la muestra, las bases son identificadas en orden de acuerdo al tiempo que requieren para arribar a la región detectora de laser.

II) CONCLUSIONES OBTENIDAS A PARTIR DEL BORRADOR DE TRABAJO DE LA SECUENCIA DEL GH.

Datos cuantitativos:

- ◆ El GH contiene 3164.6 millones de bp
- ◆ El gen promedio consiste en 3.000 bp, pero el tamaño varía considerablemente, siendo el mayor gen conocido el de la distrofina: 2,4 millones de bp.
- ◆ El número total de genes se estima en 30.000-35.000
- ◆ El orden del 99.9 % de las bases de nucleótidos es la misma en todas las personas.
- ◆ Las funciones se desconocen para > 50 % de los genes descubiertos.

Separar la paja del trigo

- ◆ Menos del 2 % del genoma codifica la producción de proteínas.
- ◆ Las secuencias repetidas que no codifican proteínas (“junk DNA”) forman por lo menos 50 % del GH.
- ◆ Se piensa que las secuencias repetitivas no poseen funciones directas, pero brindan información sobre la estructura cromosómica y su dinámica. Con el tiempo, estas estructuras reforman el genoma reordenándolo, creando nuevos genes o modificando genes ya existentes.
- ◆ Durante los últimos 50.000.000 de años, hubo una dramática reducción en la frecuencia de repeticiones en el GH.

¿Como esta dispuesto ?

- ◆ Los “centros urbanos”, densas condensaciones de genes del GH, están compuestas predominantemente por G-C.
- ◆ En contraste, los “desiertos” (pobres en genes) son ricos en A-T. Las regiones GC y AT son frecuentemente vistas en el microscopio como bandas claras y oscuras en los cromosomas.
- ◆ Los genes parecen estar concentrados en áreas al azar a lo largo del genoma, con vastos espacios de DNA entre las mismas.
- ◆ Tiras de hasta 30.000 bases C – G, repetidas una y otra vez, se encuentran adyacentes a zonas ricas en genes, formando una barrera entre los genes y las áreas de “junk DNA”.
- ◆ El cromosoma #1 tiene la mayor cantidad de genes (2968) y el cromosoma #Y la menor cantidad (231).

Como se compara el GH con el de otros organismos

- ◆ Mientras el GH tiene áreas “ricas en genes” dispuestas al azar, en otros organismos los genes están dispuestos más uniformemente.
- ◆ Los seres humanos tienen en promedio tres veces más proteínas que la mosca ó el gusano debido al splicing alternativo de los transcritos de RNA y las modificaciones químicas de las proteínas. Este proceso determina diferentes productos proteicos a partir del mismo gene.

- ◆ Los seres humanos comparten la mayoría de las familias proteicas con gusanos, moscas y plantas, pero el número de miembros de cada familia de genes se ha expandido en el hombre, especialmente en proteínas relacionadas con el desarrollo y la inmunidad.
- ◆ El genoma humano tiene una proporción muy superior (50 %) de secuencias repetidas que el gusano (7 %) y la mosca (3 %).
- ◆ Aunque el hombre parece haber detenido la acumulación de DNA repetitivo hace 50 millones de años, dicho proceso no se ha detenido en los roedores. Esto podría explicar algunas de las diferencias fundamentales entre hombres y roedores, aunque el número de genes sea similar en ambas especies.

Variaciones y mutaciones

- ◆ Se han identificado alrededor de 1.4 millones de ubicaciones donde diferencias de una sola base (SNPs) ocurren en humanos. Esta información promete revolucionar el proceso de hallazgo de ubicaciones cromosómicas para secuencias asociadas a enfermedad y el rastreo de la historia humana.
- ◆ La frecuencia de mutaciones en la línea germinal (esperma u ovocito) es 2:1 en hombres vs mujeres. La posible explicación para esta mayor tasa de mutación radicaría en el mayor número de divisiones requerido para la formación de espermatozoides que para los ovocitos.

IV) PROYECTO PROTEOMA HUMANO

El PROTEOMA HUMANO es el análisis del conjunto de proteínas existente en el organismo humano, el próximo desafío a vencer luego del desciframiento del GH. En efecto, la correspondencia entre GENOMA, TRANSCRIPTOMA Y PROTEOMA no es evidente, ya que distintos mensajeros pueden ser transcriptos a partir de un mismo gene, y distintas proteínas pueden ser transcriptas a partir de un mismo mensajero. El interés del análisis del Proteoma respecto al del análisis del Genoma ó del Transcriptoma residen en las siguientes características:

- (i): un mismo genoma puede originar proteomas diferentes; en *M. genitalium* hay presentes 24 % más de proteínas diferentes que de genes; **en el hombre existen el doble al triple de proteínas que de genes.**
- (ii) El nivel de expresión de las proteínas no es totalmente predecible a partir del nivel de expresión del mRNA; la abundancia de una proteína puede ser hasta 50 veces diferente a la de su mensajero.
- (iii) los mensajeros pueden presentar una diversidad importante, tanto por la utilización de promotores diferentes o sitios de splicing alternativos (ver más abajo).
- (iv) Las proteínas pueden ser traducidas a partir de distintos codones de iniciación, y sufren frecuentemente modificaciones post-traduccionales, tales como glicosilación, fosforilación, prenilación, acilación, ubiquitinación, deaminación, etc. Por lo tanto, un mismo gene puede originar distintas proteínas.

¿Cuál es la explicación de que un número limitado de genes (30.000), muy inferior al estimado inicialmente (100.000) produzca tal diversidad de proteínas y una complejidad superior a la observada en especies inferiores con igual número de genes ?

Como hemos visto anteriormente, parte de las respuestas reside en el fenómeno del “**splicing**” del mRNA, mediante el cual un mismo gene puede producir diferentes proteínas.

El procesamiento del mRNA en los eucariotes, transcrito por la RNA polimerasa II, incluye tres procesos principales:

- 1) **5^ˆ-capping**: es el proceso mediante el cual se agrega una 7-metilguanosina a las moléculas nacientes de mRNA. Estas moléculas nacientes (colectivamente denominadas pre-mRNA) no existen como moléculas libres en el núcleo de las células eucarióticas, sino que se asocian a un conjunto heterogéneo de proteínas denominadas “partículas heterogeneas de ribonucleoproteínas” (hnRNPs) que contienen RNA heterogeneo nuclear (hnRNA).
- 2) **Poliadenilacion de los pre-mRNAs**. En las células animales, todos los mRNA, excepto los mRNA de histonas, tienen en el extremo 3^ˆ una cola de poli(A), cuyo tamaño oscila entre 200-250 bases.
- 3) **Splicing**: es el proceso mediante el cual los intrones se pierden en el pre-mRNA, dando origen al mRNA maduro. Dicho proceso es llevado a cabo por el “spliceosoma”.

Es muy importante tener en cuenta que el mismo gen puede, mediante splicing alternativo, dar origen a distintas variantes proteicas. Un ejemplo típico de esta situación se da en el caso del gene de la fibronectina, donde hay un splicing tejido-dependiente, siendo por ejemplo la fibronectina de los fibroblastos diferente a la fibronectina hepática. Del mismo modo, múltiples isoformas de numerosas proteínas requeridas para el desarrollo y función neuronal son producidas como resultado de patrones alternativos de splicing de RNA.

Lo que aun no sabemos: un listado para futuras investigaciones.

- ◆ Numero exacto de genes, ubicación y funciones
- ◆ Regulación génica
- ◆ Organización de la secuencia de DNA
- ◆ Estructura y organización de los cromosomas
- ◆ Función del DNA no codificante
- ◆ Coordinación de la expresión génica; síntesis proteica y eventos post-traduccionales
- ◆ Interacción proteica en máquinas moleculares complejas
- ◆ Comparación entre las funciones génicas predichas vs las experimentalmente determinadas
- ◆ Conservación evolutiva entre organismos
- ◆ Conservación proteica
- ◆ Proteomas en los organismos
- ◆ Correlación de los SNPs con salud y enfermedad
- ◆ Predicción de susceptibilidad a enfermedad basado en variación de la secuencia genética
- ◆ Genes involucrados en rasgos complejos y enfermedades multigénicas
- ◆ Biología de sistemas complejos
- ◆ Genética del desarrollo; genómica.

Existen además nuevas áreas de investigación como el Programa ELSI. EL PGH reconoció desde el comienzo mismo su responsabilidad no sólo para desarrollar el hallazgo de genes, sino también para analizar las implicancias sociales más amplias de esta nueva capacidad para descifrar la información genética. 5 % del presupuesto anual del PGH está destinado a estudiar las implicancias éticas, legales y sociales de la investigación genética .

Una de las áreas más activas del programa ELSI ha sido el desarrollo de políticas relacionadas a la privacidad y al uso correcto de la información genética, particularmente en lo que se refiere a los seguros de salud, empleo e investigación médica. Los debates en esta área han enfocado principalmente el potencial de la información genética para predecir las probabilidades aumentadas de desarrollar un fenotipo enfermo en una persona sana.

Finalmente, es importante mencionar que existe el "Human Genome Diversity Project", no relacionado al HGP, y que consituye un esfuerzo multidisciplinario para explicar las bases genéticas de la diversidad entre series humanos.

Bibliografía

Libros

Bernot, A. Analyse de Genomes, Transcriptomes et Proteomes. Dunod, Paris, 3^a ed.(2001)

Baxevanis Andreas D. and Ouellette B.F.F. Bioinformatics. A practical guide to the analysis of genes and proteins. Wiley Interscience, (2001).

Campbell A. M. y Heyer L.J. "Discovering Genomics, proteomics and Bioinformatics"
Ed. Benjamin Cummings, 2003

Artículos

Watson J.D. and Crick F. "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid". Nature 171, 737-734, 1953.

Collins F. "Shatuck Lecture- Medical and Societal consequences of the Human Genome Project". New Engl. J. Med. 341: 28 – 37, 1999.

Science, Feb. 16, 2001

Nature, Feb 15, 2001.

Sitios Internet

Genoma murino

www.broad.mit.edu/media/press/pr_02_mousegenome.html

Human Genome News:

www.ornl.gov/hgmis

GenBank:

www.ncbi.nlm.nih.gov

Human Genome Diversity Project :

www.stanford.edu/group/morrinst/HGDP-FAQ.html

Estructura DNA. Síntesis de RNA. Síntesis de proteínas.

www.johnkyrk.com/DNAreplication.html

Traducción mRNA

www.ncc.gmu.edu/dna/ANIMPROT.htm

Splicing RNA

[//bssv01.lancs.ac.uk/ADS/BIOS336/336L10.html](http://bssv01.lancs.ac.uk/ADS/BIOS336/336L10.html)

PCR – Secuenciación - Polimorfismos

www.dnalc.org/shockwave/pcranwhole.html

Clonado y secuenciación PGH

www.pbs.org/wgbh/nova/genome/sequencer.html#

Clonado y secuenciación PGH

www.genome.gov/Pages/Educationkit/win/Sequence.exe

Filmografía

Video “Gattaca”. Director : Nicholas Andrew.