

R. Palacios Urtasun
P. Villoslada Díaz

Conceptos básicos de genética molecular

Laboratorio de Neuroinmunología
División de Neurociencias
Centro de Investigación Médica Aplicada (CIMA)
Universidad de Navarra
Pamplona

El papel de la genética es clave para comprender las causas de numerosas enfermedades neurológicas, tanto hereditarias como esporádicas. En las últimas décadas se ha producido un espectacular avance en el conocimiento de las enfermedades hereditarias. Los avances recientes en el conocimiento del genoma humano han puesto de manifiesto la complejidad del mismo, pero al mismo tiempo nos está dotando de nuevas herramientas para conocer su papel en las enfermedades neurológicas. A largo plazo esto nos permitirá tener un conocimiento más profundo de las mismas y desarrollar una neurología más personalizada. En esta revisión resumimos los conceptos fundamentales en genética molecular que son básicos para comprender los avances en la neurología.

Palabras clave:
Genoma. Gen. Polimorfismos. ADN mitocondrial. Neurología

Neurol Supl 2005;1(3):4-13

Basic concepts in molecular genetics

The impressive knowledge generated by the genomic research opens a great opportunity to increase our understanding of both hereditary and sporadic neurological diseases. In the last decades we were faced with the spectacular advances in the genetics of neurological hereditary diseases, and the new knowledge and tools developed by the Human Genome Project are going to provide us with a new understanding of the sporadic neurological diseases and with the opportunity to develop a personalized neurological care. In this review we summarize the basic concepts in molecular genetics that a clinical neurologist requires to understand the advances in the genetic susceptibility to neurological diseases.

Key words:
Genome. Gene. Polymorphisms. Mitochondrial DNA. Neurology.

Correspondencia:
Pablo Villoslada Díaz
Laboratorio de Neuroinmunología 2.05
Centro de Investigación Médica Aplicada (CIMA)
Pío XII, 55
31008 Pamplona (Navarra)
Correo electrónico: pvilloslada@unav.es

EL ADN, LA MOLÉCULA DE LA VIDA

El ácido desoxirribonucleico (ADN) es una macromolécula de aspecto filamentosos que contiene la información hereditaria necesaria para el funcionamiento de todos los organismos vivos. Las unidades estructurales básicas del ADN son los desoxirribonucleótidos, los cuales están formados por un azúcar (desoxirribosa) y un grupo fosfato que cumplen una función meramente estructural y una base nitrogenada portadora de la información necesaria para sintetizar las proteínas. Existen cuatro bases nitrogenadas diferentes que dan lugar a cuatro desoxirribonucleótidos: dos bases púricas (con estructura de doble anillo), adenina (A) y guanina (G) y dos pirimidicas (con estructura en anillo simple), timina (T) y citosina (C). Los cuatro desoxirribonucleótidos se encadenan uno tras otro millones de veces mediante la unión del azúcar de uno de ellos con el azúcar del contiguo a través del grupo fosfato formando cadenas de secuencias determinadas de ADN que codifican toda la diversidad de organismos existentes, desde los unicelulares más sencillos hasta el más complejo existente, el hombre.

El ADN presenta una estructura tridimensional en la que dos cadenas o hebras complementarias se retuercen una en torno a otra formando lo que se conoce como la doble hélice¹. Las cadenas complementarias se unen entre sí por las bases nitrogenadas mediante puentes de hidrógeno de tal manera que la estructura de las bases nitrogenadas sólo permite que la G se una con la C (tres puentes de hidrógeno) y la A con la T (dos puentes de hidrógeno). Así, mientras que la secuencia de la primera cadena no responde a ningún orden establecido, el orden de los nucleótidos de la cadena complementaria está condicionado por la secuencia de la primera. La cadena de uniones azúcar-fosfato está construida de manera que el fosfato en el carbono 5' de la desoxirribosa se une al carbono 3' de la siguiente desoxirribosa, diciéndose que tiene una dirección 5' a 3'. Las dos hebras de ADN están dispuestas de manera antiparalela, donde una de ellas va de 5' a 3' y la complementaria de 3' a 5'. La cadena portadora de la información genética, dado que es la que se transcribe a ácido ribonucleico (ARN), se conoce como cadena sentido y su cade-

na complementaria, llamada antisentido, sirve como molde para la reparación del ADN cuando éste se daña, por ejemplo, por el estrés oxidativo o radiaciones. La estructura en doble cadena del ADN supone que en la duplicación de este primero se separen las dos cadenas complementarias y luego, debido al carácter obligatorio de complementariedad, se sintetiza una cadena idéntica a la que se separó, asegurando así que la información contenida en esa molécula de ADN se transmite fielmente de célula a célula, generación tras generación.

LOS CROMOSOMAS, EL ALMACÉN DEL MATERIAL GENÉTICO

El conjunto de todo el ADN de un organismo se conoce con el nombre de genoma y se encuentra organizado en cromosomas (fig. 1). La unidad estructural básica de los cromosomas es el nucleosoma, constituido por un octámero de pequeñas proteínas altamente conservadas llamadas histonas, las cuales interactúan estrechamente con un fragmento de ADN de 146 pares de bases. Los cromosomas de organismos eucariotas están formados por dos cromátidas hermanas unidas por el centrómero. Los extremos de cada cromátida se denominan telómeros y están constituidos por ADN no codificante cuya función principal es la de estabilizar la estructura de los cromosomas. Durante la fase de división celular los cromosomas son visibles al microscopio óptico; sin embargo,

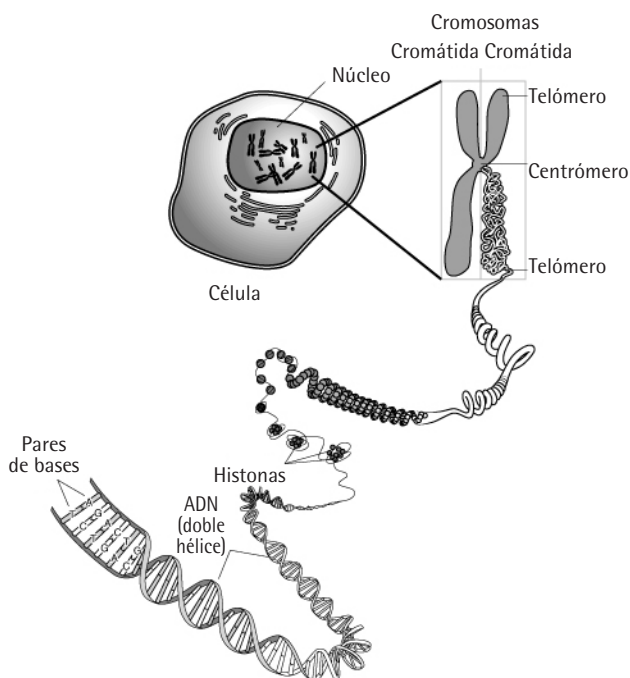


Figura 1 Organización del ADN en cromosomas. El ADN se halla empaquetado en nucleosomas con el soporte de las histonas dando lugar a los cromosomas. Reproducido con permiso del National Human Genome Research Institute (www.genome.gov).

durante la fase de reposo los cromosomas se encuentran des-enrollados en forma de cromatina. La cromatina se presenta de dos formas diferentes, parte como una forma inactiva condensada localizada sobre todo en la periferia del núcleo denominada heterocromatina y el resto como una forma activa diseminada por el resto del núcleo denominada eucromatina en la que se está transcribiendo la información de las moléculas de ADN. Ahora se conoce que para transcribir los genes que participan en una misma función las regiones cromosómicas donde residen pasan de estado de heterocromatina a eucromatina y se colocan en la misma región del núcleo². De esta manera dichos genes comparten mecanismos de regulación a pesar de residir en cromosomas diferentes.

El genoma humano está formado por más de 3.000 millones de pares de nucleótidos repartidos en 46 cromosomas y en el ADN mitocondrial, denominado en ocasiones como el cromosoma número 47. Dos de estos cromosomas, el X y el Y, determinan el sexo y se denominan cromosomas sexuales. Las mujeres tienen dos cromosomas X y los hombres un cromosoma X y otro Y. Los 44 cromosomas restantes se denominan cromosomas autosómicos. El cromosoma 1 es el cromosoma humano más grande y tiene la mayor cantidad de genes, 2.968, mientras que el cromosoma Y es el más pequeño y tiene 231 genes. El número de cromosomas es constante en todas las células de una especie con la excepción de las células sexuales, que poseen la mitad del número normal de cromosomas. El ser humano posee células diploides, es decir, tienen los cromosomas por pares ($2n$), y las células sexuales son haploides (n), lo que permite que los cromosomas sean aportados a partes iguales por el padre y la madre de manera que cada uno aporte la mitad de cada par de cromosomas, esto es, 23 cromosomas autosómicos y uno sexual. De este cromosoma sexual las mujeres siempre aportan un X, mientras que los hombres pueden aportar un X o un Y, por lo que es el hombre quien determina el sexo del hijo. Por el contrario, todos los cromosomas mitocondriales provienen de la madre a partir de las mitocondrias del óvulo ya que en la fecundación el espermatozoide no aporta citoplasma alguno porque solamente el núcleo entra en el óvulo.

Un elemento clave para entender la genética es el fenómeno de recombinación (fig. 2). Durante la primera fase de la meiosis los pares de cromosomas intercambian fragmentos de su ADN. Este mecanismo busca aumentar la diversidad del ADN y por tanto de los rasgos genéticos. El intercambio de ADN entre dos cromosomas está finamente regulado y ocurre en los llamados bloques de ADN, es decir, fragmentos de la secuencia de ADN que son reconocidos por ciertas enzimas ADNasas y ligasas y que son intercambiados. El tamaño de estos bloques varía con cada recombinación, y cuantas más recombinaciones (o generaciones a escala de poblaciones dado que la recombinación ocurre una sola vez en un individuo) hayan ocurrido los bloques son de tamaño más pequeño³. Este concepto es clave para entender el fenómeno de ligamiento genético, por el cual podemos encontrar un marcador asociado con un rasgo o enfermedad, aunque en sí mismo ese marcador sea inocuo, debido a que

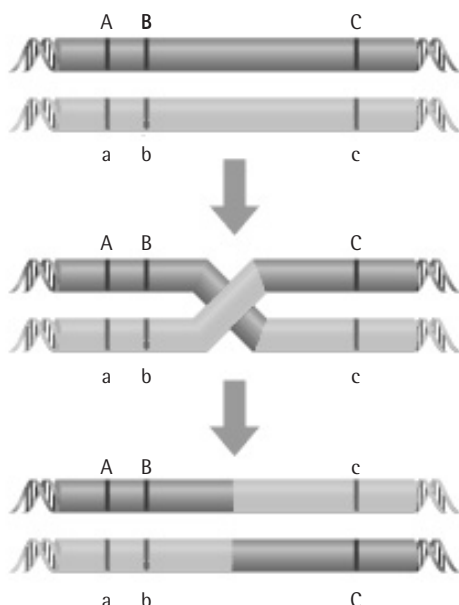


Figura 2 | *Recombinación cromosómica y ligamiento. Representación de un cromosoma paterno (gris oscuro) y uno materno (gris claro) en una célula germinal. Se muestran tres alelos denominados A, B y C. Los alelos paternos se representan en mayúsculas y los maternos en minúsculas. Se produce el fenómeno de recombinación que da lugar al entrecruzamiento de los cromosomas paterno y materno. Una vez terminado el proceso los alelos paterno y materno se han mezclado y constituirán la carga genética de los gametos. Si A es un alelo que predispone a cierta enfermedad y B y C son marcadores genéticos, la recombinación es más probable que ocurra más frecuentemente entre A y C que entre A y B. Esto permite hacer una localización del alelo de susceptibilidad A relativa a los marcadores B y C. Reproducido de la web del Wellcome Trust (www.wellcome.ac.uk) con permiso.*

se halla en el mismo bloque que el gen causante de la enfermedad⁴. El concepto de recombinación y ligamiento es la base de la mayoría de estudios genéticos que han permitido identificar los genes responsables de rasgos o mutaciones. Otro aspecto práctico del mecanismo de recombinación es su empleo como medida del ADN. Dado que los mapas físicos del genoma humano han sido creados en los últimos años, anteriormente para calcular la distancia entre genes y su posición en el genoma se empleaba el concepto de bloque y recombinación, de tal manera que la distancia entre genes o secuencias del ADN se medían en Morgans, una unidad de recombinación genética que equivaldría actualmente a un millón de pares de bases.

DEL ADN A LA PROTEÍNA, EL DOGMA CENTRAL DE LA BIOLOGÍA

La unidad biológica elemental de los cromosomas portadora de la información a la cual se deben los caracteres he-

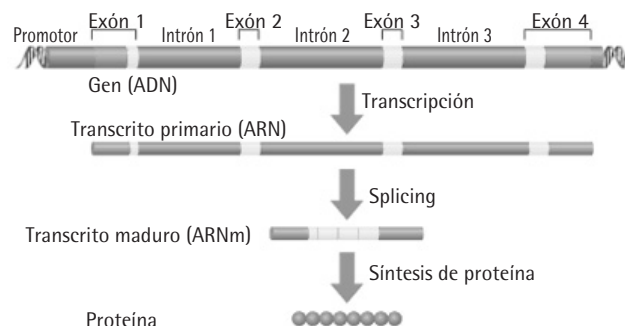


Figura 3 | *Estructura y expresión de un gen. La secuencia de ADN da lugar a un transcrito primario que tras la eliminación de los intrones por el proceso de splicing se convierte en ARNm maduro. Este ARNm se traduce en los correspondientes aminoácidos que formarán la proteína. Reproducido de la web de Exon Hit Inc (www.exonhit.com) con permiso.*

reditarios son los genes (fig. 3). La secuencia de nucleótidos de un gen especifica cómo y cuándo se han de sintetizar las proteínas. Por término medio los genes contienen unos 3.000 nucleótidos, aunque este tamaño varía mucho. El gen más grande conocido en el ser humano es el de la distrofina que tiene 2,4 millones de nucleótidos. La información codificada en un gen es transmitida a una molécula intermedia, denominada ARN mensajero (ARNm) mediante un proceso denominado transcripción. Las unidades estructurales básicas del ARN son los ribonucleótidos, que son similares a los desoxirribonucleótidos con la diferencia de que en el ARN el azúcar es una ribosa y la base pirimidica T es reemplazada por uracilo (U), también pirimidica y que puede formar puentes de hidrógeno con la A. Al contrario que el ADN y su estructura de doble cadena, el ARN casi siempre está formado por una única cadena. El ARNm es capaz de salir del núcleo y entrar en contacto con los ribosomas para iniciar la síntesis proteica en un proceso denominado traducción. Para posibilitar la síntesis de proteínas a partir de ADN, el código genético está diseñado de manera que un triplete de nucleótidos consecutivos, o codón, indica la formación de un aminoácido particular, así como el inicio y el final de la síntesis proteica. Dado que la información del ARN influye de forma profunda en la actividad de la célula, el ARNm es degradado rápidamente una vez terminada la traducción.

Los genes están compuestos de cuatro elementos, una zona promotora en el extremo 5' y una zona reguladora en el extremo 3', que controlan el proceso de transcripción, y exones e intrones. Los exones representan la parte del gen que codifica la proteína durante la traducción y los intrones son fragmentos de ADN que se encuentran entre dos exones adyacentes. Por tanto, un gen está formado por un exón, luego un intrón, después otro exón y así sucesivamente. Los intrones no están presentes en el ARNm ya que durante el proceso de maduración del ARN se produce un fenómeno denominado *splicing* durante el cual los intrones se separan

permitiendo que los exones se empalmen entre sí para dar lugar a la secuencia codificadora de la proteína. Del total de nucleótidos del genoma humano, menos del 2% codifica para proteínas, es decir, son exones. Si bien antes se pensaba que los intrones no contenían información, en la actualidad se comienza a saber que en ellos reside la información que regula el *splicing*⁵.

El proceso de transcripción comienza cuando determinadas proteínas, denominados factores de transcripción, se unen al promotor. Esta unión se produce en secuencias determinadas con una alta conservación evolutiva y se denominan motivos o cajas. Las cajas más frecuentes son la TATA, una secuencia rica en T y A situada unos 25 nucleótidos por delante del sitio de inicio de la transcripción, y la CCAAT, una secuencia fija localizada a unos 70 nucleótidos por delante del sitio de inicio de la transcripción. En la región promotora (unas 1.000 bases por encima del inicio del gen) se halla el conjunto de regiones de unión a los factores de transcripción y cajas de inicio. El número de factores de transcripción es elevado y en cada promotor de cada gen existe una combinación diferente, lo que permite regular la expresión de los genes. Además, la información contenida en los promotores puede indicar diferentes formas de inicio de transcripción del gen que daría lugar a ARN y proteínas diferentes. Cada diferente forma de lectura se denomina marco de lectura abierto u *open reading frame* (ORF). Los ORF sirven como elementos de búsqueda de nuevos genes en los estudios del genoma humano y de otras especies. La regulación de la expresión de un gen depende de su promotor (con su combinación de sitios de unión de factores de transcripción) y de la regulación que se produce en regiones próximas al gen y que suele ser común a un grupo de genes del mismo *locus*, entre los que destacan los potenciadores (*enhancers*) y silenciadores (*silencers*), y en regiones distantes, incluso en otros cromosomas, como las regiones de control de los *locus* (*Locus Control Regions*, LCR)⁶.

Cuando se han sintetizado aproximadamente unos 30 nucleótidos de ARN se añade al extremo 5' un nucleótido modificado constituido por un residuo terminal de 7-metilguanosa unido al primer nucleótido transcrito a través de un puente 5'-5'-trifosfato. Este nucleótido modificado se denomina 5'CAP y su presencia en el ARNm es crítica para el reconocimiento por el ribosoma en la síntesis proteica. Para finalizar la síntesis del transcrito primario se le añade en el extremo 3' una cadena de varios cientos de nucleótidos de A, es la llamada cola poli-A, que interviene en la exportación del ARNm del núcleo al citoplasma y contribuye a la estabilidad del ARNm una vez en él. La cola poli-A se une en el transcrito a través del reconocimiento de una secuencia específica, la señal AAUAAA. Una vez sintetizado el transcrito primario se produce el fenómeno de *splicing* durante el cual se eliminan los intrones. Generalmente, el *splicing* es llevado a cabo por un complejo de ARN y proteínas llamado *spliceosoma*⁷, aunque existen determinadas moléculas de ARN como los ribozimas que son capaces de catalizar su propio *splicing*⁸. El *spliceosoma* reconoce los extremos 5' y 3'

de los intrones del transcrito primario por medio de determinadas secuencias conservadas de manera que los nucleótidos comprendidos entre esos puntos son retirados, permitiendo así la unión de los dos exones.

El ARNm maduro presenta dos secuencias transcritas a partir de exones que no se traducen a proteína. Dichas secuencias se encuentran antes del codón iniciador y después del codón terminador y se denominan regiones 5'UTR (*untranslated regions*) y 3'UTR, respectivamente. Se han atribuido diversas funciones a estas regiones, como la de participar en la estabilidad del ARNm, en su localización celular o en la eficiencia en la traducción, aunque no se ha podido definir una función concreta ya que la habilidad de las UTR para realizar su función depende de su secuencia y ésta difiere entre distintos ARNm^{9,10}.

En los últimos años se han identificado una serie de ARN aparentemente sin potencial codificante denominados ARN no codificante (ARNnc)¹¹. Se ha especulado que estos ARNnc pueden ser de gran importancia en la regulación de las funciones vitales de la célula como es el caso de los ARN de interferencia, transcritos que se unen con una alta especificidad a un ARNm determinado interfiriendo así en la expresión del gen. Sin embargo, hasta la actualidad ha habido pocos avances que demuestren las funciones de estos transcritos.

Por el momento se desconoce el número total de genes del genoma humano y la función del 50% de todos los genes descubiertos hasta el momento. Las últimas previsiones calculan que el genoma humano tiene entre 20.000 y 25.000 genes¹², solamente tres veces más que la mosca, un número sorprendentemente bajo para nuestra especie, lo que hace pensar que la complejidad humana no se encuentra en el número de genes, sino en las interacciones que se dan entre ellos¹³.

SPLICING ALTERNATIVO, CUANDO UN GEN PRODUCE MÁS DE UNA PROTEÍNA

El *splicing* alternativo es el proceso que ocurre cuando el proceso de *splicing* de un transcrito primario da lugar a diferentes ARNm, llamados isoformas, y por tanto a diferentes proteínas¹⁴ (fig. 4). Se estima que el *splicing* alternativo está presente en entre el 35 y el 65% de los genes humanos, lo que explica las predicciones de la existencia de más de 90.000 proteínas distintas. Se ha identificado un gen humano con 3.000 isoformas y otro en mosca con hasta 30.000 isoformas, lo que da idea del poder e importancia que tiene el *splicing* en la regulación génica. Así pues, el *splicing* alternativo contribuye a la identidad celular, dado que una estirpe celular se define por el patrón de expresión de sus genes ya que todas las células de un organismo tienen los mismos genes¹⁵. El *splicing* alternativo puede dar lugar a una nueva definición de la palabra gen ya que se podría considerar no sólo como una secuencia de ADN codificadora

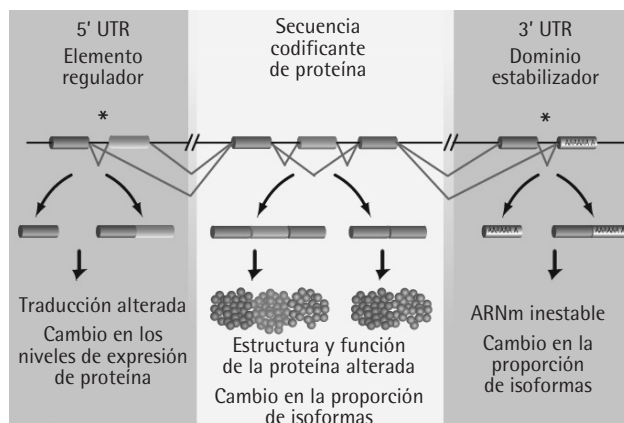


Figura 4 | Splicing alternativo. El splicing alternativo puede ocurrir en cualquier región del transcrito primario, en las regiones 5' o 3' UTR o en la secuencia codificante de proteína. La secuencia 5'UTR contiene regiones reguladoras que controlan la traducción a proteína. Inserciones o deleciones de estas regiones modifican los niveles de expresión de la proteína. La secuencia 3'UTR contiene dominios que participan en la estabilidad del ARNm. La inserción o deleción de dichos dominios modifica la estabilidad del ARNm y por tanto la traducción a proteína. El splicing alternativo dentro de la región codificante tiene como resultado cambios en la estructura y función de la proteína resultante. Reproducido de la web del Wellcome Trust (www.wellcome.ac.uk) con permiso.

de un ARN, sino además como una secuencia adicional de ADN u otro mecanismo regulador que controle el *splicing* alternativo.

LOS POLIMORFISMOS, LAS CLAVES DE LA VARIABILIDAD INTRAESPECIE

Un polimorfismo es una variación en la secuencia del ADN en una posición determinada del genoma entre los individuos de una población que da lugar a la existencia de distintas variantes de un gen en al menos un 1% de la población. Estas variantes se conocen como alelos y generalmente producen variaciones en las características hereditarias tales como el color de los ojos o el grupo sanguíneo. El hecho de tener el material genético duplicado, mitad materno y mitad paterno, hace que cada gen posea dos alelos. Los individuos que poseen los dos alelos iguales se conocen con el nombre de homocigotos, mientras que aquellos que poseen los dos alelos diferentes se denominan heterocigotos. La secuencia del genoma humano es idéntica en un 99,9% para toda la población, por lo que sólo el 0,1% del genoma da lugar a polimorfismos y produce la gran variabilidad existente entre individuos. Los polimorfismos que ocurren con una frecuencia inferior al 1% de la población se denominan mutaciones y generalmente tienen un efecto patológico¹⁶. Estas mutaciones son producidas por fallos en el proceso de replicación del ADN

durante la división celular o por la exposición a radiaciones ionizantes, a determinados agentes químicos o a determinados agentes biológicos como virus o bacterias. Los fallos que pueden ocurrir son la eliminación o inserción de un nucleótido, copiado incorrecto, y eliminación o inserción de una cadena de nucleótidos. La evolución ha empleado los polimorfismos como mecanismo básico para generar nuevos rasgos¹⁷. Los tipos de polimorfismos se clasifican a continuación según su tamaño.

Macropolimorfismos

Variaciones cromosómicas numéricas

Cuando se cuenta con uno o varios cromosomas de más o de menos del número cromosómico normal. Se habla entonces de individuos aneuploides, es decir, su constitución cromosómica no es exactamente completa. Se denomina individuos monosómicos ($2n-1$) a aquellos a los que les falta un cromosoma completo, nulisómicos ($2n-2$) a aquellos a los que les falta una pareja completa de cromosomas y trisómicos ($2n+1$) a aquellos que tienen un cromosoma extra. Por ejemplo, la presencia de tres cromosomas 21 en lugar de dos, la llamada trisomía 21, da lugar al conocido síndrome de Down. En general los cambios en el número de genes son inviables, salvo en el caso de cromosomas con pocos genes como el cromosoma 21.

Variaciones cromosómicas estructurales

Cuando existen variaciones en el número o en la disposición de los genes (fig. 5A). La pérdida de genes se conoce con el nombre de *deleción* y el aumento en el número de genes como duplicación. Cuando un segmento cromosómico concreto cambia de orientación dentro del cromosoma se denomina inversión y cuando dos cromosomas no homólogos intercambian segmentos cromosómicos se denomina translocación. En general este tipo de modificaciones se detectan mediante técnicas de citogenética, pero para ello han de ser de un tamaño grande. Si son pequeñas son difíciles de detectar por citogenética y mucho más mediante biología molecular. Actualmente se piensa que las duplicaciones de los genes han sido la herramienta más importante para la evolución, dado que el gen duplicado puede ser mutado sin que se pierda la función del mismo gracias a la copia existente.

Polimorfismos medianos

Secuencias repetitivas

Denominados microsatélites, son secuencias de dos, tres o cuatro nucleótidos que se repiten entre 10 y 100 veces. Uno de los microsatélites más comunes, y el primero que se identificó, está formado por repeticiones del dinucleótido

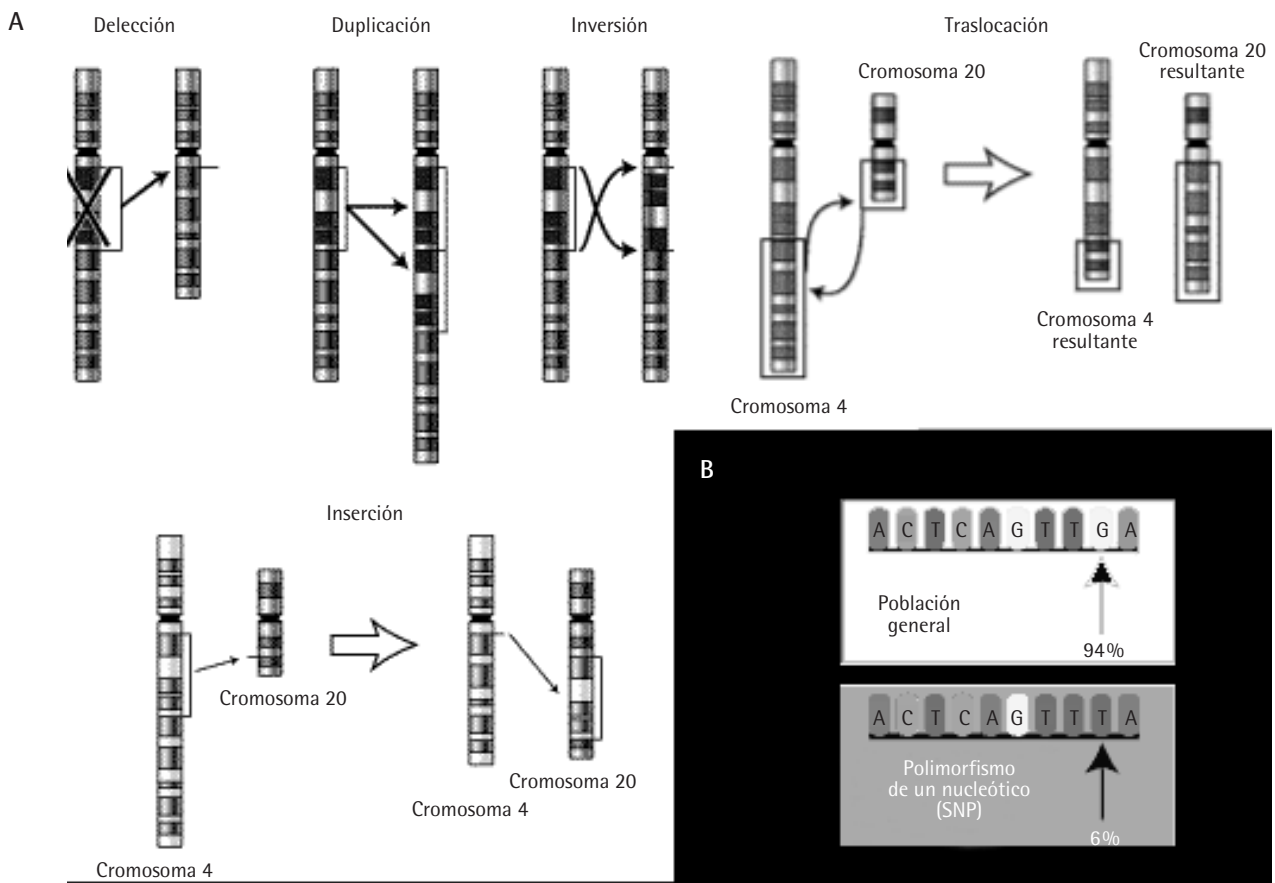


Figura 5 | *Diferentes tipos de polimorfismos. (A) Variaciones cromosómicas estructurales. (B) Polimorfismos de nucleótido único (SNP). Reproducido con permiso del National Human Genome Research Institute (www.genome.gov).*

CA. Los microsatélites son muy frecuentes en el genoma humano y normalmente se localizan en regiones no codificantes. La variabilidad que presentan estos polimorfismos entre individuos es muy alta y se obtiene del diferente número de repeticiones de la secuencia en concreto. Secuencias entre 10 y 15 repeticiones suelen ser polimórficas y generalmente cuanto mayor es el número de repeticiones que presenta un microsatélite, mayor es el número de alelos que presenta. Esta gran variabilidad es fruto de errores durante la replicación del ADN ya que la elevada repetición de las secuencias confunde a la maquinaria encargada de la replicación del ADN produciendo la pérdida o inserción de repeticiones adicionales. Además muchos se crearon por el efecto de los retrovirus que tras insertarse en nuestro genoma y manteniendo la actividad de la transcriptasa inversa han producido ampliaciones de estas pequeñas secuencias. Hoy se sabe que hasta el 10% de nuestro genoma tiene un origen retroviral. Dada la abundancia y el carácter altamente polimórfico de los microsatélites, éstos son usados para realizar pruebas de identificación y de paternidad. Hasta hace poco se pensaba que estas secuencias repetidas no tenían una función directa y que solamente mantienen la estructura y el dinamismo de los cromosomas; sin embargo, en los últimos

años se ha demostrado que los microsatélites tienen gran importancia en la regulación y por tanto en la variabilidad de la expresión génica¹⁸.

Elementos translocados

Estos polimorfismos son causados por los transposones. Los transposones son secuencias de ADN de origen retroviral que se pueden mover por diferentes posiciones del genoma en un proceso denominado transposición¹⁹. Los transposones componen aproximadamente el 45% del genoma humano y pueden haber colaborado de manera importante en la evolución de la diversidad de producción de anticuerpos del sistema inmunitario²⁰.

Micropolimorfismos

Inserciones o deleciones de un nucleótido

Cuando dentro de la secuencia del ADN se introduce o elimina un solo nucleótido.

Polimorfismos de un nucleótido

Llamados comúnmente SNP (*Single Nucleotide Polymorphisms*) ocurren cuando se produce el cambio de un nucleótido por otro (fig. 5B). Se calcula que en el genoma humano hay unos tres millones de SNP, constituyendo aproximadamente el 90% de todos los polimorfismos existentes. Los SNP aparecen tanto en regiones codificantes como no codificantes y dos de cada tres SNP sustituyen una C por una T. Los SNP que se encuentran en regiones codificantes pueden dar lugar a mutaciones sin sentido si dan lugar a un codón terminador, a mutaciones con sentido perdido si forman un codón que codifica para un aminoácido diferente y a mutaciones silenciosas si dan lugar a un codón que codifica para el mismo aminoácido. Los SNP con alelos menos frecuentes generalmente tienen un origen más reciente que los alelos más comunes, por lo que se piensa que el origen de los SNP se encuentra en mutaciones puntuales que conllevan una mejora evolutiva de manera que su frecuencia aumenta significativamente en la población. Muchos SNP no tienen ningún efecto sobre las funciones celulares, pero se ha demostrado que otros pueden predisponer a padecer determinadas enfermedades o modificar la respuesta a determinados fármacos bien porque producen cambios en las proteínas o porque participan en la regulación de la expresión génica. Un buen ejemplo de cómo los SNP afectan a la predisposición de padecer determinadas enfermedades es el gen de la apolipoproteína E (ApoE) y la enfermedad de Alzheimer. El gen de la ApoE contiene dos SNP que dan lugar a tres alelos posibles: E2, E3 y E4. Cada alelo difiere en un nucleótido de ADN y la proteína producida difiere en un aminoácido²¹. La presencia de un alelo E4 aumenta significativamente el riesgo de padecer enfermedad de Alzheimer ya que aparentemente el aminoácido producido por el alelo E4 altera la estructura y función de la proteína²². Por el contrario, la presencia de un alelo E2 reduce significativamente el riesgo de padecer la enfermedad. Otro aspecto importante de los SNP es su efecto en la expresión génica. Los SNP reguladores serían aquellos que modifican la cantidad de ARN sintetizado o condicionan el *splicing* de un gen¹⁸.

ADN MITOCONDRIAL

Las mitocondrias son las organelas encargadas de suministrar energía a las células. Se piensa que su origen es bacteriano por un mecanismo de simbiosis entre la bacteria y la célula eucariota. Debido a dicho origen, las mitocondrias poseen su propio ADN, llamado ADN mitocondrial (ADNmit). Cada mitocondria posee múltiples copias de su ADN. El ADNmit es circular, sin intrones ni regiones reguladoras extras y sin histonas. Todas estas características son propias del ADN bacteriano. El ADNmit codifica un total de 37 genes que participan en la cadena respiratoria. Sin embargo, no todas las proteínas de la cadena respiratoria están codificadas por el ADNmit, si no que se necesita importar algunas otras sintetizadas por genes del núcleo. Las mitocondrias poseen además toda la maquinaria para sintetizar proteínas

a partir de su ADN. Debido a la falta de protección del ADNmit y a la gran cantidad de radicales libres producidos en la mitocondria, el ADNmit posee una elevada tasa de mutaciones que hace que se pierda la eficiencia energética con la edad a pesar de los mecanismos de reparación. Debido a que toda la secuencia del ADNmit es codificante, la mayoría de modificaciones suelen producir mutaciones.

En la fecundación todas las mitocondrias son aportadas por el óvulo y por tanto tienen origen materno (dado que las mitocondrias del espermatozoide que le proporcionan energía durante su carrera no penetran en el óvulo), es lo que se conoce como la herencia materna. Dado que un cigoto recibe un número elevado de mitocondrias del óvulo y dado que algunas de las copias del ADNmit de alguna de sus mitocondrias pueden llevar una mutación, ésta se transmitirá a las estirpes celulares a que dé origen. Este hecho hace que la genética mitocondrial tenga características propias. Durante la formación del embrión en cada mitosis las mitocondrias se distribuyen al azar entre las células hijas. Todas las células de un individuo normal tienen el mismo ADNmit, es lo que se conoce como homoplasmia; por el contrario, si en una célula aparecen dos poblaciones de ADNmit, normal y mutada, se dice que estamos en una situación de heteroplasmia. Si existen mutaciones en el ADNmit de algunas mitocondrias el reparto puede ser equitativo o no (fig. 6). Por tanto cada estirpe celular, que dará lugar posteriormente a un tejido diferente, recibirá una carga de mutación diferente, dando lugar a una situación de mosaico respecto a la carga genética mitocondrial. La proporción de mitocondrias con su ADNmit mutado y las necesidades energéticas del tejido determinarán en gran parte el fenotipo de la enfermedad. Dado que los órganos con mayor demanda energética son el cerebro, músculo, corazón y riñón, dichos órganos son en los que suelen manifestarse las mutaciones del ADNmit, produciendo encefalopatías, miopatías, cardiopatías o síndromes metabólicos.

EPIGENÉTICA, LA REGULACIÓN DE LA EXPRESIÓN AL MARGEN DE LA SECUENCIA DE ADN

La epigenética se define como los cambios heredables en la expresión de los genes que ocurren sin que se modifique la secuencia del ADN. Estos cambios se producen principalmente por modificaciones de la cromatina, que incluye procesos como la metilación del ADN²³ y la acetilación de histonas²⁴, o por cambios transitorios en la accesibilidad a la cromatina. La modificación de histonas parece ser un mecanismo regulatorio universal entre organismos eucariotas desde la levadura al humano, mientras que la metilación del ADN está menos conservada, pero es un común y rápido mecanismo desarrollado entre organismos eucariotas con genomas más complejos. La metilación del ADN se produce en el carbono 5 de los nucleótidos de C y se localiza en regiones ricas en nucleótidos CG, llamadas islas CpG, que suelen encontrarse en las regiones promotoras de los genes. La hipometilación del ADN y la hiperacetilación de las histonas

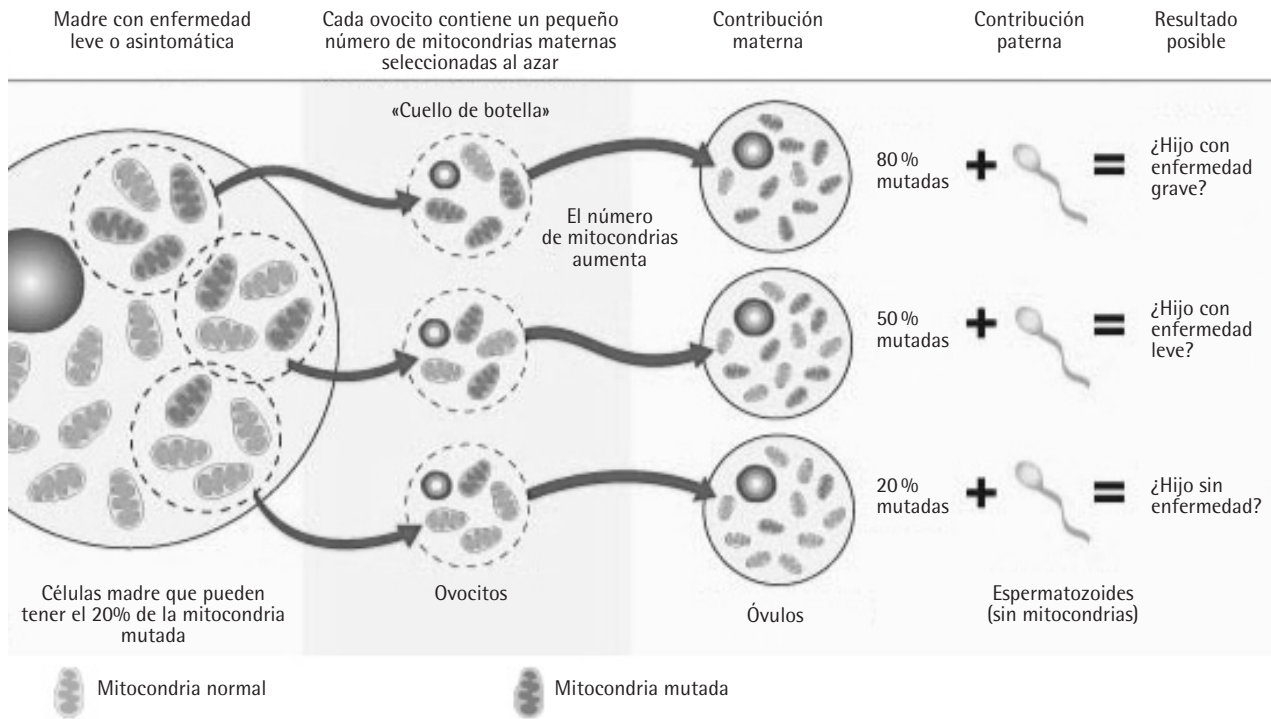


Figura 6 Herencia materna del ADNmit mutado. La mayoría de las personas sanas tienen células homoplásticas, es decir, todas las células presentan el mismo ADNmit. Las personas con ADNmit mutado tienen células heteroplásticas con una mezcla de mitocondrias con ADN normal y mutado. En la fecundación todo el ADNmit es aportado por la madre. Durante la producción de los gametos femeninos en la meiosis las mitocondrias se distribuyen aleatoriamente. Si la madre posee ADNmit mutado esta distribución al azar da lugar a gametos con diferente carga mutada que pueden dar lugar a diferentes manifestaciones fenotípicas. Reproducido de la Mitochondrial Research Society (<http://www.mitoresearch.org/>) con permiso.

dan lugar a la activación de la transcripción, mientras que la hipermetilación del ADN y la hipoacetilación de las histonas la reprimen (fig. 7). Estos y otros mecanismos aún por descubrir regulan la remodelación de la cromatina y la expresión de los genes. Por ejemplo, en las mujeres uno de los dos cromosomas X se inactiva mediante la metilación del ADN para evitar duplicación de la información²⁵. Otro ejemplo sería la existencia de ciertos genes cuya expresión está determinada por la procedencia del padre o madre, denominado en inglés *imprinting*²⁶. De esta manera en dichos genes el alelo que siempre se expresa es el del padre, por ejemplo, y no el de la madre, que está inactivado por metilaciones. Aunque no se conoce el origen de este fenómeno se piensa que ha surgido por un proceso de selección natural de cara a transmitir los rasgos del varón o de la hembra. Otro fenómeno sería la expresión alélica diferencial o haploinsuficiencia²⁷, en la que sólo uno de los dos alelos de un gen se expresa o la expresión de uno de los alelos está muy reducida. En este caso es independiente del origen paterno o materno y se piensa que su función es controlar la dosis de dicho gen (cantidad de proteína que se produce) de cara a regular estrechamente su función. En la actualidad se está procediendo a la caracterización del código epigenético para conocer mejor los mecanismos que regulan el genoma²⁸.

EL PROYECTO GENOMA HUMANO

Con el fin de conocer la secuencia de los 3.000 millones de pares de nucleótidos e identificar todos los genes del genoma humano en octubre de 1990 comenzó una investigación internacional llamada Proyecto Genoma Humano. Dicha investigación se concibió como un recurso para la comunidad científica de manera que además de permitir conocer la secuencia completa permitiese desarrollar herramientas que ayudasen a comprender mejor cómo el genoma humano posibilita y controla la vida. El Proyecto Genoma Humano concluyó parcialmente en 2003 y sus resultados son públicos y se pueden consultar por internet en las web del NCBI Human Genome Sequencing Progress, en la University of California de Santa Cruz, en el Ensemble Genome Browser del Wellcome Trust Sanger Institute y el EMBL-European Bioinformatics Institute, en el DNA Data Bank of Japan y en el EMBL-Bank at the European Molecular Biology Laboratory's Nucleotide Sequence Database. El estudio del genoma humano ayudará en el diagnóstico, prevención y tratamiento de enfermedades, así como en el estudio para entender los mecanismos que confieren la susceptibilidad a padecer determinadas enfermedades. A pesar de que ya se conoce la secuencia completa del genoma humano todavía queda mucho trabajo por hacer como conocer el nú-

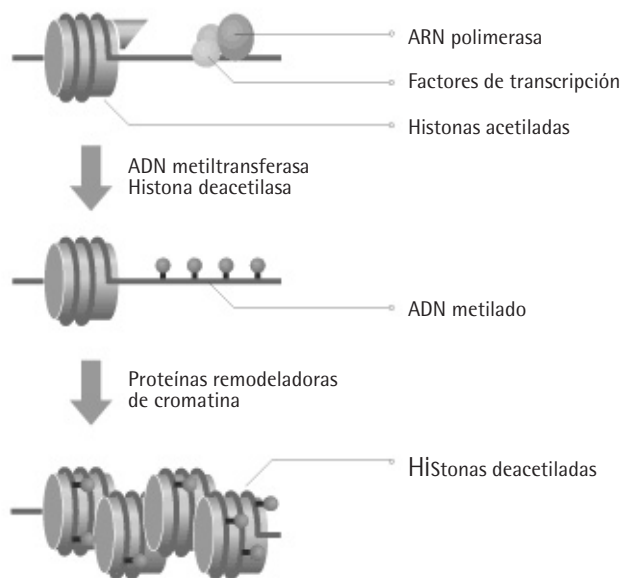


Figura 7 Remodelación de la cromatina. La eucromatina, la forma activa de la cromatina, se caracteriza por la hipometilación del ADN y la hiperacetilación de las histonas. Esto permite el acceso de la ARN polimerasa y de los factores de transcripción al promotor. La metilación del ADN por la ADN metiltransferasa bloquea la unión de los factores de transcripción con el promotor, inhibiendo así la transcripción. La metilación del ADN y la desacetilación de las histonas por la enzima histona desacetilasa da lugar a la compactación de la cromatina, haciendo inaccesible la unión de los factores de transcripción al promotor. Reproducido de la web del Wellcome Trust (www.wellcome.ac.uk) con permiso.

mero exacto de genes e isoformas, su localización y su función, así como la regulación de su expresión y la estructura y organización de los cromosomas. Las exploraciones que se están realizando en el genoma humano están generando datos cuyo volumen y los complejos análisis necesarios para su comprensión no tienen precedentes en la biología. En la actualidad el principal proyecto en marcha como continuación del Proyecto Genoma Humano es el Proyecto HapMap o mapa de haplotipos²⁹. Dado que el ADN se hereda en bloques definidos por la recombinación, el objetivo es conocer las características de dichos bloques de ADN, lo que permitirá desarrollar herramientas más eficientes para analizar el genoma y conocer las diferencias entre poblaciones humanas. Otro proyecto clave es el Proyecto ENCODE³⁰, que busca la caracterización con el máximo detalle de todos los mecanismos reguladores del genoma en ciertos *locus* cromosómicos.

PAPEL DEL GENOMA EN ENFERMEDADES NEUROLÓGICAS

La mayoría de enfermedades genéticas afectan al sistema nervioso dado que el cerebro es un órgano muy complejo

que requiere de la integridad del genoma para su función. Por tanto, la mayoría de polimorfismos macro, medios o micro participan en la etiología de enfermedades neurológicas. A nivel de los macropolimorfismos simplemente decir que las dos enfermedades hereditarias más frecuentes como son el síndrome de Down y el cromosoma X frágil producen retraso mental además de otros síntomas. Respecto a los polimorfismos medianos, en los últimos años se ha descubierto toda una familia de enfermedades neurológicas producidas por la expansión de tripletes (o microsatélites) como son la enfermedad de Huntington, ataxia de Friedreich, etc. En el caso de la enfermedad de Huntington, la expansión del triplete conduce a la producción de una proteína con un número muy elevado de glutaminas, lo cual desestructura la proteína huntingtina y la vuelve tóxica. Respecto a los micropolimorfismos, un ejemplo sería la presencia de mutaciones puntuales en diversos genes implicados en la enfermedad de Alzheimer como la proteína precursora del amiloide o la presenilina que da lugar a casos hereditarios de la enfermedad. Otro ejemplo sería la asociación ya comentada del alelo ApoE4 y el Alzheimer esporádico. Finalmente, las mutaciones en el ADNmit dan lugar a las enfermedades mitocondriales, teniendo la gran mayoría de ellas una franca afectación neurológica como el MELAS, MERFF, síndrome de Kearns-Sayre o enfermedad de Leber.

CONCLUSIONES

Los avances en el conocimiento del genoma humano pueden permitirnos en un futuro identificar los elementos genéticos que predisponen a padecer enfermedades neurológicas esporádicas. Con dicha información podríamos identificar personas con riesgo a padecerlas en las que aplicar estrategias preventivas que ayuden a controlarlas. Además, el conocimiento de los mecanismos genéticos que predisponen a padecer enfermedades neurológicas nos ayudaría a mejorar nuestra comprensión de la patogenia de las mismas y a desarrollar nuevos tests diagnósticos o pronósticos o nuevos tratamientos. Finalmente, el desarrollo de la farmacogenética ayudará a identificar los individuos resistentes a una determinada terapia, lo cual permitiría realizar una medicina más personalizada.

AGRADECIMIENTOS

Los autores quieren agradecer al Fondo de Investigaciones Sanitarias del Instituto de Salud Carlos III (FIS 01/0108-01) y a al acuerdo entre FIMA y «UTE proyecto CIMA» por el apoyo recibido para la realización de estudios genéticos en enfermedades neurológicas. Los autores no tienen ningún conflicto de interés.

TEXTOS DE REFERENCIA

Cantor C, Smith C. Genomics: the science and technology behind the human genome project, 1.ª ed. Hoboken: Wiley-InterScience, 1999.

Lewin B. *Genes VIII*. Saddle River: Prentice Hall, 2004.
 Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular biology of the cell*. New York: Garland Publishing (4th ed.), 2002.

WEBS

Human Genome Project: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
 NCBI Human Genome Sequencing Progress: www.ornl.gov/hgmis/project/progress.html
 University of California at Santa Cruz: www.ucsc.edu
 Ensembl Genome Browser: www.ensembl.org
 DNA Data Bank of Japan: www.ddbj.nig.ac.jp
 EMBL-Bank: www.ebi.ac.uk/embl
 HapMap Project: <http://www.hapmap.org/index.html.en>
 ENCODE Project: <http://genome.ucsc.edu/ENCODE>
 Human Epigenome Project : <http://www.epigenome.org>

BIBLIOGRAFÍA

- Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953;171:737-8.
- Sproul D, Gilbert N, Bickmore WA. The role of chromatin structure in regulating the expression of clustered genes. *Nat Rev Genet* 2005;9:9.
- Costas J, Salas A, Phillips C, Carracedo A. Human genome-wide screen of haplotype-like blocks of reduced diversity. *Gene* 2005;349:219-25.
- Amos CI, Zhu DK, Boerwinkle E. Assessing genetic linkage and association with robust components of variance approaches. *Ann Hum Genet* 1996;60:143-60.
- Chamary JV, Hurst LD. Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends Genet* 2005;21:256-9.
- Cook PR. Nongenic transcription, gene regulation and action at a distance. *J Cell Sci* 2003;116:4483-91.
- Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 2003;25:1147-9.
- Doherty EA, Doudna JA. Ribozyme structures and mechanisms. *Ann Rev Biophys Biomol Struct* 2001;30:457-75.
- Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. *Semin Cell Dev Biol* 2005;16:39-47.
- de Moor CH, Meijer H, Lissenden S. Mechanisms of translational control by the 3' UTR in development and differentiation. *Semin Cell Dev Biol* 2005;16:49-58.
- Huttenhofer A, Schattner P, Polacek N. Non-coding RNAs: hope or hype? *Trends Genet* 2005;21:289-97.
- Stein LD. Human genome: end of the beginning. *Nature* 2004;431:915-6.
- Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999;402(Suppl.):C47-52.
- Ast G. How did alternative splicing evolve? *Nat Rev Genet* 2004;5:773-82.
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, et al. Function of alternative splicing. *Gene* 2005;344:1-20.
- Maki H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Ann Rev Genet* 2002;36:279-303.
- Eckhardt RB. Polymorphisms past and present. *Hum Biol* 2003;75:559-75.
- Rockman MV, Wray GA. Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* 2002;19:1991-2004.
- Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol* 2005;1:22.
- Agrawal A, Eastman QM, Schatz DG. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 1998;394:744-51.
- Mahley RW, Rall SC Jr. Apolipoprotein E: far more than a lipid transport protein. *Ann Rev Genomics Hum Genet* 2000;1:507-37.
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261:921-3.
- Volpe P. The language of methylation in genomics of eukaryotes. *Biochemistry (Mosc)* 2005;70:584-95.
- Khan AU, Krishnamurthy S. Histone modifications as key regulators of transcription. *Front Biosci* 2005;10:866-72.
- Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the Mammalian X chromosome. *Ann Rev Genomics Hum Genet* 2005;6:69-92.
- Da Rocha ST, Ferguson-Smith AC. Genomic imprinting. *Curr Biol* 2004;14:R646-9.
- Santarosa M, Ashworth A. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochim Biophys Acta* 2004;1654:105-22.
- Bradbury J. Human epigenome project-up and running. *PLoS Biol* 2003;1:E82. Epub 2003 Dec 22.
- The International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789-96.
- ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004;306:636-40.