

CAPÍTULO I

El genoma de los mamíferos

La intención de este capítulo es ofrecer una revisión muy resumida, aunque actualizada, de las características del material hereditario en los mamíferos, y de esta forma, facilitar la comprensión de los términos y conceptos vertidos en el resto del libro. Toda la información detallada sobre los elementos básicos de la genética, así como de las técnicas del análisis del ADN, pueden encontrarse en los libros de texto (la mayoría con versiones en español) listados en la bibliografía general de este capítulo.

I.1. El genoma: estructura y organización

I.1.1. Los ácidos nucleicos

Hasta el año 1944 las bases materiales de la herencia permanecieron desconocidas. Recién en esa fecha, Oswald Avery, Colin MacLeod y Maclyn McCarty, basándose en trabajos anteriores (1928) con extractos de bacteria, pudieron determinar que una sustancia presente en el núcleo y capaz de ser purificada (más tarde identificada como el **ácido desoxirribonucleico** o **ADN**) era el "material genético". Ellos pudieron demostrar que el ADN era el material por medio del cual un rasgo genético podía ser transferido de una cepa de bacteria a otra. En este caso, se trató del pasaje de cepas de *Streptococcus pneumoniae* que forman colonias pequeñas con superficie rugosa (cepa R) hacia cepas con colonias grandes y de superficie lisa (cepa S). La cepa R—que no causa neumonía— puede ser transformada en cepa S—que es patógena— al ser cultivada con extractos de bacteria provenientes de la cepa S. El hecho de que los ácidos nucleicos fueran algo más que un mero componente estructural del núcleo celular fue algo sorpresivo para el ambiente científico de la época¹.

I.1.1.1 Estructura del ADN

La estructura propia del ADN no fue develada hasta el año 1953, año en el cual James Watson (Estados Unidos) y Francis Crick (Inglaterra) reportaron sus sorprendentes trabajos sobre la estructura de la doble hélice. Por dichos trabajos, fueron galardonados con el Premio Nobel de Medicina y Fisiología en el año 1962, junto a Maurice Wilkins². El ADN es una macromolécula compuesta por una alineación de **desoxirribonucleótidos**, cada uno de los cuales está compuesto por un fosfato, un azúcar (**desoxirribosa**) y una base nitrogenada (el azúcar más la base forman un **nucleósido**). El grupo fosfato consiste en un átomo de fósforo (P) uni-

¹ Para una revisión de los acontecimientos más importantes de la historia de la genética, referirse al libro *A History of Genetics (AH Sturtevant)*, recientemente reeditado por *Cold Spring Harbor Laboratory Press, New York, 2001* (<http://www.cshlpress.com/>).

² En enero del 2003 fue publicado en la revista *Nature* un número conmemorativo de los 50 años de estos trabajos (*Nature*, Volumen 421 (6921), <http://www.nature.com/nature/DNA50/>).

do a tres átomos de oxígeno (O). La molécula de desoxirribosa está compuesta por cuatro átomos de carbono (C) y uno de oxígeno, formando un anillo plano; fuera de este anillo sale un quinto átomo de carbono que es el que se une al grupo fosfato.

Las bases en el ADN son de dos tipos: **purinas** [adenina (A) y guanina (G)] y **pirimidinas** [citosina (C) y timina (T)]. Estos nucleótidos, en conjunto, constituyen las “cuatro letras” del **código genético** universal. Cada una de las bases se encuentra unida a la base adyacente en la cadena por el azúcar y el fosfato. La cadena de nucleótidos se mantiene unida debido a uniones covalentes entre el grupo fosfato en el carbón 5' de un nucleótido y el grupo OH⁻ en el carbón 3' del nucleótido adyacente. De esta forma, la cadena de ADN tiene siempre un extremo fosfato 5' (llamado extremo 5') y otro OH⁻ 3' (llamado extremo 3'). Las dos cadenas que se entrelazan se mantienen juntas por uniones de hidrógeno muy específicas entre purinas y pirimidinas. En la estructura de doble cadena, las bases que se enfrentan son siempre una purina, en una cadena, con una pirimidina en la cadena opuesta (G con C y A con T); esto hace que podamos encontrar cuatro tipos de pares de nucleótidos: A-T, T-A, G-C y C-G. De esta manera, las cadenas son complementarias entre sí y la secuencia de bases en una cadena puede ser deducida por la secuencia de bases en la otra. De hecho, el ADN tiene cantidades iguales de los nucleótidos A y T, por un lado, y G y C por el otro. En el ADN de los mamíferos, esta cantidad representa aproximadamente 31% A / 31% T y 19% G / 19% C. Estos porcentajes no son universales: por ejemplo, en la mosca *Drosophila melanogaster*, los porcentajes son 27,5% para A y T y 22,5% para G y C.

Otra consecuencia de este apareamiento de bases es que las dos cadenas forman una estructura en forma de hélices antiparalelas conocida como la **doble hélice**. La estructura tridimensional

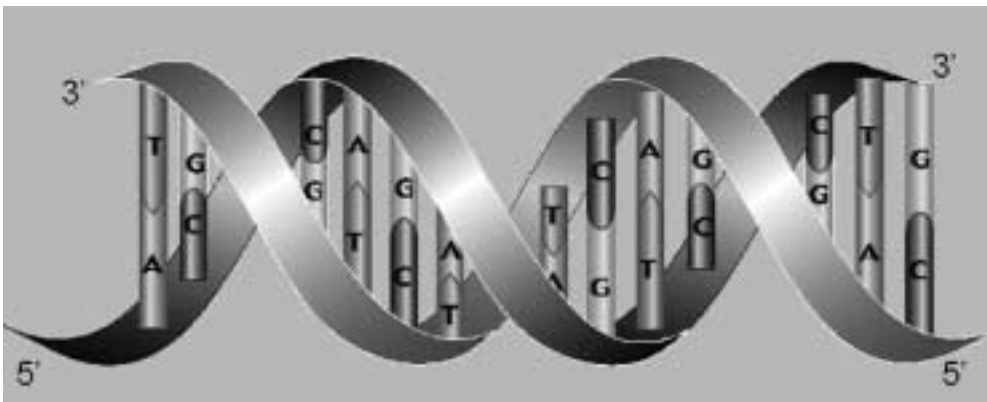


Figura 1.1. La doble hélice del ADN. Figura esquemática de la estructura de la doble hélice del ADN (forma B). Las dos cintas simbolizan las cadenas azúcar-fosfato y las varas transversales los pares de bases complementarios que sostienen las cadenas unidas (por puentes de hidrógeno). Las dos cadenas de ADN se las considera antiparalelas porque una cadena tiene orientación 5'- 3' y la cadena complementaria tiene orientación 3'- 5'. La hélice da una vuelta completa cada 10 pares de bases, lo que equivale a 34 Å (el diámetro de la hélice es de 20 Å). Dibujo por Christine K. Yone, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas, Estados Unidos.

muestra que las dos cadenas se enroscan entre sí, formando una hélice orientada hacia la derecha, conocida como **forma B** o dextrógira, que corresponde con el modelo de Watson y Crick (**Figura 1.1**). Esta no es la única estructura posible, ya que en determinadas circunstancias existen formas alternativas de la hélice orientadas hacia la izquierda, **forma Z** o levógira, e inclusive una **forma A** en la que los pares de bases están desplazados hacia el exterior respecto al eje de la hélice. La longitud de un segmento de ADN se mide en **pares de bases (pb)**, en el caso de segmentos cortos, y kilobases (1 **kb** = 1000 bases) o megabases (1 **Mb** = 1000 kb) para los fragmentos más largos. Las dos cadenas de ADN se pueden separar en forma experimental por la acción del calor o de agentes químicos (por ejemplo, la urea). Este proceso se llama **desnaturalización** y es un fenómeno reversible que es muy utilizado en los protocolos experimentales.

1.1.1.2 La replicación del ADN

La **replicación** del ADN es la capacidad que tiene el ADN de hacer copias (o réplicas) de su molécula. Este proceso es fundamental para la transferencia de la información genética de generación en generación. Las moléculas se replican de un modo semiconservativo; es decir, la doble hélice se separa y cada una de las cadenas sirve de molde para la síntesis de una nueva cadena complementaria. El resultado final son dos moléculas idénticas a la original. Durante la separación de las cadenas se forma una estructura en forma de Y llamada **horqueta de replicación** (en inglés, *replication fork*); la misma se genera gracias a la intervención de varias enzimas, entre las cuales se encuentran la **helicasa** y la **girasa**. El ADN de las células eucariotas se replica por medio de muchas horquetas de replicación y en un estilo de copia bidireccional. El ADN eucariota se sintetiza mucho más lentamente que el de las células procariotas, alrededor se 50 nucleótidos por segundo y por horqueta de replicación.

La principal enzima que interviene en la replicación del ADN es la **polimerasa de ADN**. Esta enzima requiere la presencia de un grupo OH⁻ libre en el extremo 3' de la molécula de ADN, extremo hacia el cual crece la nueva cadena. Debido a esta restricción, la polimerasa de ADN no puede iniciar una cadena nueva sin la ayuda de una molécula accesoria llamada **iniciador** (en inglés, *primer*). Esta molécula iniciadora es en realidad una cadena corta de ARN que sintetiza la enzima **polimerasa de ARN**.

Además de la función de replicación, la polimerasa de ADN tiene otra función muy importante en el control de errores durante el proceso de copia (del inglés, *proof-reading*), donde la enzima comprueba que el nuevo nucleótido agregado a la cadena esté de acuerdo con la complementariedad A-T/G-C.

1.1.1.3 La organización del ADN en cromosomas

En los cromosomas de los eucariotas, el ADN está asociado a proteínas en una sustancia estable denominada **cromatina**. Según sus propiedades ante la coloración artificial en el laboratorio, la cromatina puede clasificarse en **heterocromatina** y **euromatina**. La heterocromatina se tiñe más fuerte y es más condensada, mientras que la euromatina se tiñe débilmente y es menos condensada. La euromatina permanece en forma no condensada durante la interfase y contiene una

proporción mucho mayor de genes que la heterocromatina, la cual es particularmente rica en secuencias repetitivas conocidas como **ADN satélite**. Algunas regiones de la heterocromatina parecen ser estructurales, como la heterocromatina que se encuentra cerca de los **centrómeros** y los **telómeros**; otras, como los **corpúsculos de Barr**, representan regiones inactivadas del cromosoma X.

La estructura de los cromosomas está basada en una serie de enrollados consecutivos (de menor a mayor), lo que genera una compactación muy eficiente del material genético dentro del núcleo. El primer nivel de condensación después de la doble hélice (la cual tiene 20 Å de diámetro), lo constituyen los **nucleosomas**. Estas estructuras en forma de cuentas de un rosario están formadas por la unión del ADN con ciertas proteínas básicas llamadas **histonas**, de las cuales existen cuatro tipos: H2A, H2B, H3 y H4. Cada nucleosoma está formado por una vuelta de 140 pares de nucleótidos unido a dos moléculas de cada tipo de histonas. Los nucleosomas, a su vez,

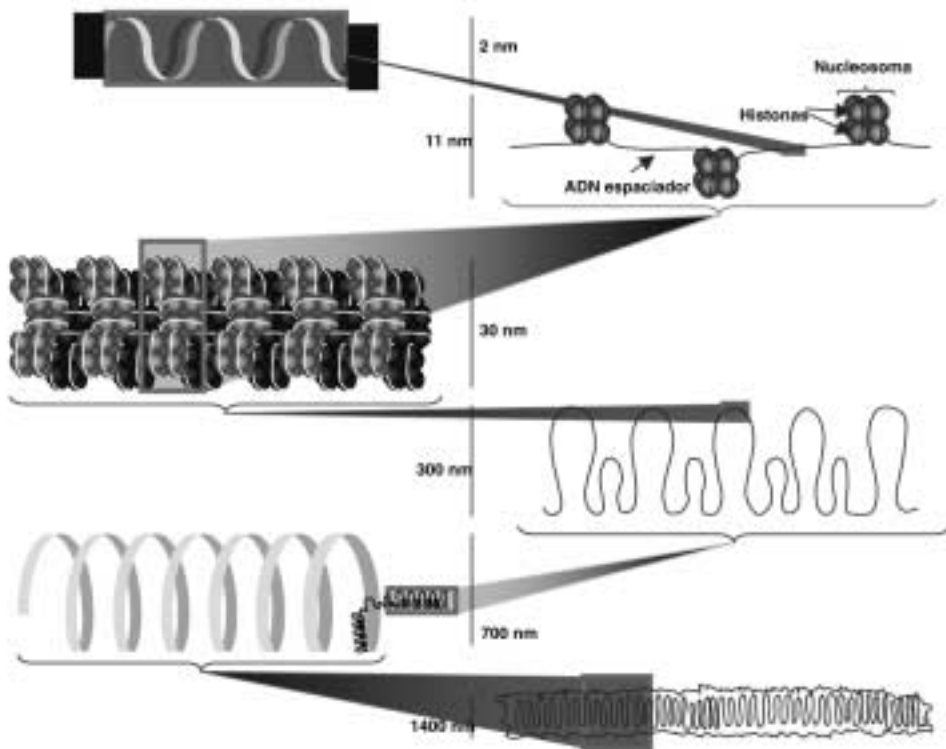


Figura 1.2. Condensación del ADN. La figura muestra el empaquetamiento que sufre la molécula de ADN desde la doble hélice hasta formar la cromátide del cromosoma metafásico. El ADN se encuentra unido a un complejo de histonas formando los nucleosomas (en el microscopio electrónico se ven como “cuentas de rosario”). Los nucleosomas son luego compactados en estructuras llamadas fibras nucleosómicas, solenoides, supersolenoides y finalmente la cromátide (1 nanómetro (nm) = 10 Ångstroms). Dibujo por Christine K. Yone, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas, Estados Unidos.

están unidos por puentes de ADN (en inglés, *spacers*) de entre 15 y 100 pares de nucleótidos. Dentro del núcleo, los nucleosomas están compactados en estructuras mayores llamadas **fibras nucleosómicas**, con un diámetro de aproximadamente 100Å. En el siguiente nivel de enrollado, las fibras nucleosómicas forman una fibra aún mayor (500 Å) llamada **solenoide**; ésta a su vez se enrolla para formar el **supersolenoide**, de 2000 Å de diámetro. Finalmente, el supersolenoide se enrolla, durante la metafase, para formar la **cromátide**, de 6000 Å de diámetro (**Figura 1.2**).

1.1.1.4 El código genético

Como veremos más adelante, el ADN será transcrito, gen tras gen, en una molécula de ARN denominada “mensajero” la cual llevará la información necesaria para la síntesis de proteínas. La información para producir una secuencia específica de aminoácidos está contenida en la secuencia de bases del segmento de ADN. Por convención, al escribir una secuencia de nucleótidos, se coloca el extremo 5' a la izquierda (así como para la secuencia de aminoácidos se coloca el amino terminal también a la izquierda). El orden de las bases constituye la verdadera información (código genético) contenida en el ADN, la cual será transmitida al ARN. Este código está formado por **tripletes de bases** (llamados **codones** en el ARN), de los cuales existen 64 combinaciones posibles (la combinatoria de las cuatro bases es $4^3 = 64$) (**Tabla 1.1**). Al escribir los tripletes de base en el ADN, se usa por convención la cadena que no sirve como molde (cadena

		Segunda Letra				
		U	C	A	G	
Primera Letra	U	UUU UUC	UUU UCC	UAU UAC	UGU UGC	U C
		UUA UUG	UUA UUG	UAA UAG	UGA UGG	A G
		CUU CUC	CCU CCC	CAU CAC	CGU CGC	U C
		CUA CUG	CCA CCG	CAA CAG	CGA CGG	A G
	C	UUU UUC	UUU UCC	UAU UAC	UGU UGC	U C
		UUA UUG	UUA UUG	UAA UAG	UGA UGG	A G
		CUU CUC	CCU CCC	CAU CAC	CGU CGC	U C
		CUA CUG	CCA CCG	CAA CAG	CGA CGG	A G
	A	UUU UUC	UUU UCC	UAU UAC	UGU UGC	U C
		UUA UUG	UUA UUG	UAA UAG	UGA UGG	A G
		CUU CUC	CCU CCC	CAU CAC	CGU CGC	U C
		CUA CUG	CCA CCG	CAA CAG	CGA CGG	A G
G	UUU UUC	UUU UCC	UAU UAC	UGU UGC	U C	
	UUA UUG	UUA UUG	UAA UAG	UGA UGG	A G	
	CUU CUC	CCU CCC	CAU CAC	CGU CGC	U C	
	CUA CUG	CCA CCG	CAA CAG	CGA CGG	A G	

Tabla 1.1. El código genético. El orden de las bases constituye la verdadera información contenida en el ADN, la cual será transmitida al ARN por el proceso de transcripción. Este código está formado por tripletes de bases (codones en el ARN) de los cuales existen 64 combinaciones posibles (la combinatoria de las cuatro bases es $4^3 = 64$). La tabla muestra los codones del ARNm y los aminoácidos para los cuales codifican. Notar las redundancias (muchos codones codifican para el mismo aminoácido) y la presencia de codones que no codifican para ningún aminoácido (codón de terminación o parada). El codón AUG codifica para el aminoácido metionina y actúa a su vez como señal de comienzo de la cadena polipeptídica, por lo que se lo conoce como codón de iniciación.

codificante) que, por lo tanto, será el equivalente en ADN del ARN. Existiendo sólo 20 aminoácidos, queda claro que existe una redundancia del código genético. En algunos casos, las dos primeras bases del codón son suficientes para especificar un aminoácido; por ejemplo, los codones GUU, GUC, GUA y GUG (GTT, GTC, GTA y GTG en el ADN), todos codifican para el aminoácido **valina**. Existen tres codones (UAA, UAG y UGA) que no codifican para aminoácidos y se los conoce como **codón de terminación** (del inglés *stop codon*) debido a que determinan el final de la cadena polipeptídica. El codón AUG (ATG en el ADN) codifica para el aminoácido **metionina** y actúa a su vez como señal de comienzo de la cadena polipeptídica, por lo que se lo conoce como **codón de iniciación** (*start codon*). La secuencia de ADN que se encuentra entre los codones de iniciación y de terminación se llama **marco abierto de lectura** u **ORF** (del inglés *open reading frame*).

1.1.1.5 Estructura del ARN

El ARN es una molécula lineal (monocatenaria) de cuatro tipos de nucleótidos, donde el azúcar es una **ribosa** y donde la timina (T) es reemplazada por el **uracilo** (U). En el caso del ARN tenemos tres tipos básicos, a saber: **ARN de transferencia (ARNt)**, interviene en la síntesis de proteínas; **ARN mensajero (ARNm)**, también interviene en este proceso y **ARN ribosómico (ARNr)**, el cual forma parte de los **ribosomas**.

El ARNm consiste en una secuencia de nucleótidos que corresponde a la transcripción de una secuencia de ADN. Como veremos, esta transcripción no es un proceso simple ya que el gen contiene secuencias llamadas **exones** e **intrones** y el ARNm sufre una maduración durante la que se cortan los intrones y se empalman los exones. La función del ARNm es la de transportar la información genética del núcleo a los ribosomas (en el citoplasma) donde será traducido a proteínas. El ARNt es una molécula de ARN con estructura cruciforme, cuya función es leer el código del ARNm e ir sintetizando la cadena de proteínas a partir de los aminoácidos asociados a su estructura. Por lo tanto, existen tantos ARNt como aminoácidos puedan ser codificados. Como veremos, el ARNt tiene una parte de su estructura que se une al aminoácido y otra a los codones del ARNm.

El ARNr es un ARN estructural que forma parte de los ribosomas, junto con un grupo específico de proteínas, y le confiere su forma acanalada, posibilitando la unión del ARNm con el ARNt y los aminoácidos en la cadena polipeptídica. Además, tiene una función enzimática que facilita las interacciones en el ribosoma y posibilita la formación de los enlaces peptídicos. Los ribosomas de las células eucariotas tienen cuatro tamaños, 18S, 5S, 5,8S y 28S (S = coeficiente de sedimentación, en honor a Theodor S. Svedberg); y se calcula que el ARNr representa el 50% del ARN total de la célula.

Las células eucariotas poseen además **ARN nucleolar** (o ARN heterogéneo nucleolar), precursor del ARNm maduro, y **snARN** (ARN pequeño nuclear), moléculas de ARN unidas a proteínas que desempeñan un papel importante en el proceso de síntesis de ARNm. Otro grupo importante de ARN lo constituyen los **ARN no codificantes** (del inglés *noncoding RNAs* o *ncRNA*). Entre ellos se encuentran los recientemente descubiertos *micro RNAs* (*miRNAs*), moléculas de

ARN muy pequeñas (entre 20 y 25 nucleótidos de longitud) relacionadas con mecanismos de **interferencia de ARN**.

1.1.2 El ADN genómico es heterogéneo: secuencias codificantes versus no codificantes

Se define como **genoma** al complemento genético de una célula o virus. En los eucariotas, el término hace referencia al grupo haploide de cromosomas, tal como se presenta en las células germinales. No obstante, en la mayoría de los organismos, desde las bacterias hasta los mamíferos, la información genética se encuentra en más de un compartimento celular. La gran masa de ADN, llamado genómico o cromosómico, está contenida en el núcleo de la célula (o su equivalente en el caso de los procariotas), pero también se encuentra ADN en las organelas, los plásmidos (procariotas), los cloroplastos (plantas) y las mitocondrias (eucariotas), aunque esta cantidad es mínima. Por ejemplo, en los mamíferos, el **ADN mitocondrial (ADNmt)** es una molécula circular de alrededor de 17 kb (16,3 kb en el ratón). Este ADNmt se replica de manera similar al nuclear, y posee algunos genes que se transcriben y traduce a proteínas (especialmente los relacionados a la cadena respiratoria), aunque el código genético es diferente al código del ADN nuclear. Los caracteres determinados por el ADNmt (citoplasmático) se heredan exclusivamente a través de la línea materna.

Dentro de la inmensa cantidad de secuencias que forman el genoma podemos hacer una gran división didáctica entre aquellas que codifican para proteínas (genes o secuencias codificantes) y aquellas que no (secuencias no codificantes).

1.1.2.1 Secuencias codificantes: ¿Qué es un gen?

Existen muchas definiciones posibles pero, básicamente, se considera que todo segmento de ADN que se transcribe en ARNm es un **gen** (también llamado gen estructural). Obviamente, más allá de la secuencia de tripletes de bases que será traducida a la cadena polipeptídica, el gen (llamado también "unidad de transcripción") tiene muchas secuencias más, relacionadas a los fenómenos de transcripción y regulación de la expresión (estas regiones son consideradas también parte del gen estructural). Por ejemplo, el ARNm tiene, en sus dos extremos, regiones no traducidas a proteínas, las cuales intervienen en la transcripción y procesamiento del transcrito primario. La región del gen donde comienza la transcripción se llama **sitio de iniciación de la transcripción** y los nucleótidos precedentes (a la izquierda en la convención escrita) se anotan con números negativos. La región inmediata que precede al sitio de iniciación de la transcripción (**aguas arriba** o, en inglés, *upstream*) se conoce como región promotora del gen (o **promotor**) y sirve como sitio de fijación de la enzima polimerasa de ARN. Esta región contiene secuencias muy conservadas en la evolución (secuencia consenso) dentro de las que encontramos las cajas TATA (posición -25), CAAT (posición -75) y GC (posición -90). Estas secuencias funcionan como sitios de pegado (en inglés, *binding*) de proteínas regulatorias llamadas **factores de transcripción**. La terminación es un proceso menos estudiado que la iniciación pero se sabe que la

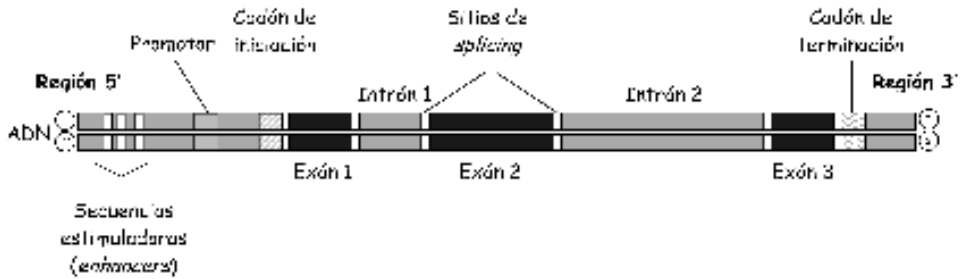


Figura 1.3. El gen eucariota. El esquema muestra la estructura de un gen eucariota con los exones (secuencias propiamente dichas que codifican para las proteínas), los intrones (secuencias que serán removidas en el proceso de maduración del ARNm), los sitios de iniciación y terminación de la transcripción, la región promotora y los codones de inicio y terminación, encuadrando el marco abierto de lectura (ORF). El número de exones en el diagrama (tres) es sólo un ejemplo didáctico, para más detalles sobre la estructura de los genes eucariotas ver el texto.

transcripción se extiende más allá del denominado **sitio de terminación** y que involucra la acción de enzimas específicas. La **Figura 1.3** muestra en forma esquemática la estructura de un gen eucariota estándar.

Hasta el año 1977, se creía que los genes tenían el tamaño necesario para la proteína a codificar, pero nunca mayor. En ese año se confirmó que los genes eucariotas poseen secciones que no son traducidas a proteína. Por lo tanto, los genes son estructuras complejas que contienen varios componentes, entre los que se destacan los exones, los intrones y las secuencias regulatorias. Los exones (el nombre proviene del inglés *expressed regions*) son los segmentos de ADN que contienen las secuencias propiamente dichas que codifican para las proteínas. Según los últimos cálculos, apenas un poco más del 1% del genoma corresponde a las secuencias exónicas. Intercalados entre los exones encontramos los intrones (originalmente fueron llamados *intra-genic regions*) cuya función es aún poco conocida (recordar que los genes procariotas no los contienen). En algunos casos raros, encontramos genes cuyas secuencias no presentan intrones, es el caso de los genes que codifican para el interferón y para las histonas. El tamaño, la secuencia y el número de intrones en un gen varía de una especie a otra, pero el tamaño, la secuencia y el número de exones varía mucho menos entre las especies.

Sorprendentemente, menos del 3% del genoma está formado por secuencias codificantes y sus secuencias regulatorias. El resto del genoma consiste en secuencias no codificantes, en general repetitivas y cuya importancia es desconocida o, al menos, no bien estudiada. Los genes pueden ser tan pequeños como 100 pb (por ejemplo el gen *tRNA^{Tr}*), o gigantescos como el gen de la distrofina humana *DMD* (o su homólogo en el ratón, *Dmd*) —con sus 2,4 Mb— y el gen titina con 178 exones, incluido el exón más largo descrito hasta la actualidad (17 kb). En el caso de la distrofina, las secuencias exónicas representan apenas el 1% del total, siendo el resto intrones y secuencias regulatorias. Más allá de estos extremos, la gran mayoría de los genes tienen un tamaño que oscila entre 10 y 100 kb y se encuentran distribuidos, en promedio, un gen cada 50 kb, aunque existen zonas que están superpobladas de genes y otras más desiertas. (La secuencia inicial —borrador— del genoma del ratón estaría confirmando este hallazgo, ya que se identificaron regiones

completamente desiertas de secuencias codificantes, la mayoría de las cuales corresponden con regiones desiertas en las zonas homólogas del genoma humano.) La mayoría de los genes están representados por un único ejemplar dentro del genoma haploide. No obstante, existen casos de duplicación de genes, como los genes de la globina (varias copias por genoma) y los genes del ARN ribosomal (cientos de copias por genoma). Más aún, existen genes que poseen secuencias muy similares (por tener un origen común) y decimos que pertenecen a **familias multigénicas**. Algunos ejemplos son los genes de las actinas, las inmunoglobulinas, las esterasas y los receptores del olfato (en este caso se han descrito más de mil genes relacionados).

1.1.2.2 Secuencias no codificantes

Las secuencias no codificantes, también llamado **ADN extragénico** (del inglés, *extragenic DNA*), fueron tomadas como **ADN no funcional**, aunque en realidad es poco lo que se conoce de sus posibles funciones. Teniendo en cuenta el número de copias, algunos autores las clasifican en tres tipos de secuencias: de **copia única**, **moderadamente repetitivas** y **altamente repetitivas**. El 60% del ADN extragénico pertenece a secuencias únicas (o presentes en un bajo número de copias), mientras que el 10% son secuencias moderadamente repetitivas y, alrededor del 30%, altamente repetitivas, aunque estos porcentajes varían según las especies analizadas y los autores consultados (**Figura 1.4**). Dentro del grupo de secuencias altamente repetitivas encontramos secuencias repetidas **en tandem** e **intercaladas** (en inglés, *interspersed DNA*).

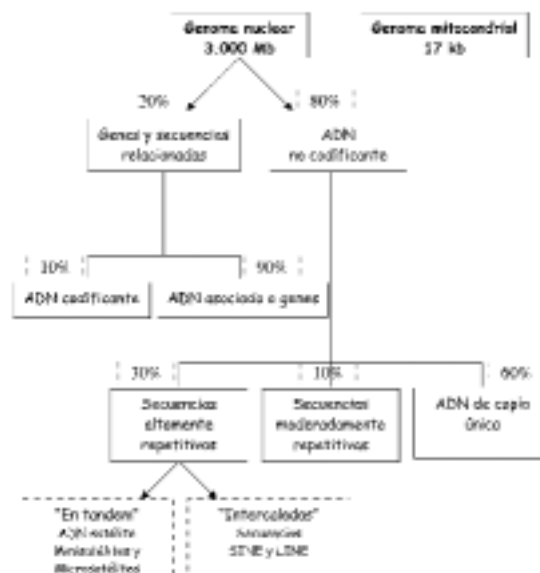


Figura 1.4. Secuencias genómicas. El esquema muestra una clasificación didáctica de las secuencias presentes en el genoma nuclear de los vertebrados. El ADN no codificante o extragénico representa alrededor del 80% del genoma y sólo una fracción mínima (menos del 3%) pertenece a las secuencias codificantes propiamente dichas. Los porcentajes de cada tipo de ADN varían según los genomas e inclusive los autores.

Las secuencias repetidas en tandem están constituidas por el llamado **ADN satélite** (el nombre se debe a la característica banda satélite que se generaba en la centrifugación en gradientes de densidad por Cl_2Cs) y las secuencias **minisatélites** y **microsatélites**. El ADN satélite está formado por unidades de repetición de alrededor de 250 pb y se calcula que existen 700.000 copias esparcidas en los centrómeros y, en menor medida, los telómeros. En el ratón *Mus musculus*, el ADN satélite constituye el 5% del genoma y se conocen dos tipos: el satélite mayor (más abundante) y el satélite menor. Es interesante notar que la especie *Mus spretus* (ver Capítulo III) posee una proporción invertida, ya que porta más copias del satélite menor.

Aunque tienen un origen evolutivo diferente, todas las secuencias en tandem han sido agrupadas bajo el sufijo "satélite", como sucede con los **minisatélites** (unidad de repetición de 10 a 40 pb y tamaño total de hasta 200 kb) y los **microsatélites** (unidad de repetición de 1 a 6 pb y tamaño total de hasta 300 pb). Algunos autores agregan a la lista los **midisatélites** (tamaño de 200 a 500 kb) y los **macro-satélites** (tamaño mayor a 500 kb). Estas secuencias, en particular los **minisatélites** y **microsatélites**, serán vistas en detalle en el Capítulo VI, debido a su utilidad como **marcadores genéticos**.

Dentro de las secuencias intercaladas encontramos dos tipos. El primero lleva el nombre de **SINE** (del inglés *short interspersed nuclear elements*) y se trata de unidades de repetición (o motivos, del inglés *motives*) de un tamaño menor a 400 pb. Las secuencias *Alu* en el humano y las secuencias *B1* y *B2* en el ratón pertenecen a este grupo. Existen entre 90.000 y 150.000 copias de las secuencias *B1* y *B2* en el genoma del ratón, ambas con unidades de repetición de 150 a 200 pb. El segundo grupo se denomina **LINE** (del inglés *long interspersed nuclear elements*) y está representado por secuencias de longitud variable que pueden llegar a medir hasta 6 kb (alrededor de 100.000 copias por genoma haploide). Un ejemplo de este tipo de secuencias lo constituye la secuencia *L1* del ratón.

A este tipo de secuencias intercaladas se las divide también (según su origen) en repeticiones **específicas de linaje** (introducidas en la especie por retro-transposición) o repeticiones **ancestrales** (aquellas presentes en un ancestro común). En el ratón, se calcula que la masa total de repeticiones específicas de linaje sería de alrededor del 32% del genoma (800 Mb), contra sólo 24% en el genoma humano. Por el contrario, las repeticiones ancestrales representan sólo el 5% en el ratón, contra el 22% en el genoma humano. Estos datos explicarían los cálculos que dicen que el 46% del genoma humano está formado por secuencias intercaladas, contra sólo el 37% del genoma del ratón, lo que se debería a la falta de repeticiones ancestrales.

Finalmente, el genoma está sembrado de copias inactivas de genes llamadas **pseudogenes**. Se cree que estas copias surgieron por retro-transposición de moléculas de ARNm dentro del genoma, ya que carecen de algunos o de todos sus intrones. Algunos genes pueden llegar a tener cientos de copias inactivas esparcidas por el genoma, como sucede con el gen *Gapd* (*glyceraldehyde-3-phosphate dehydrogenase*).

1.1.3 La talla del genoma: estimando el número total de genes

El contenido de ADN de los genomas fue siempre una incógnita: la llamada **paradoja del valor C** (del inglés, *C-value paradox*). El valor C es la cantidad de ADN en peso que tiene un organismo en un juego cromosómico (cuando los cromosomas tienen una sola cromátide, antes del período de síntesis). La paradoja del valor C es que no hay relación entre el valor C y la complejidad (evolución) del organismo. Por ejemplo, el genoma de los mamíferos es 200 veces más grande que el de una levadura, pero 200 veces más chico que el genoma del protozoo *Amoeba dubia* (ver <http://www.genomesize.com>). El contenido de ADN del genoma haploide de los mamíferos es de 3 picogramos (pg) (3×10^{-12} g), lo que sería un equivalente a 3×10^9 pb (3.000 Mb).

Considerando a los cuatro nucleótidos A, G, C y T como información en código binario, es fácil calcular que la secuencia completa del genoma puede ser almacenada en el disco rígido de un ordenador personal. Como hemos visto, dentro de esta gran cantidad de información genética, las secuencias que codifican para proteínas y sus respectivas secuencias regulatorias se encuentran diluidas en un gran “océano” de ADN que no tiene función conocida. En las últimas décadas, la cantidad de genes dentro del genoma de los mamíferos fue estimada entre 50.000 y 100.000. A partir de la disponibilidad de las secuencias completas de genomas de organismos superiores –incluidos los genomas humano y del ratón– se reactivó la discusión sobre el número de genes en los mamíferos, el cual, según algunos autores, no sería mayor a 35.000. En la actualidad, se encuentran terminadas las secuencias genómicas completas de virus, eubacterias, arcobacterias y eucariotas. El genoma de la levadura *Saccharomyces cerevisiae* posee 12 Mb y alrededor de 6.000 genes; mientras que la mosca de la fruta (*Drosophila melanogaster*) tiene 180 Mb y alrededor de 13.000 genes. La primera planta secuenciada (*Arabidopsis thaliana*) tiene un genoma de apenas 125 Mb y cerca de 25.000 genes. El genoma humano, el primer mamífero secuenciado, tiene un tamaño de aproximadamente 3.000 Mb y entre 30.000 y 40.000 genes. Para más detalles sobre los genomas secuenciados hasta el momento referirse al Capítulo VI.

1.2. Expresión de la información genética: ¿Cómo funcionan los genes?

1.2.1 La transcripción y la traducción del ADN

La síntesis de proteínas comienza con la separación de las dos cadenas de ADN. Luego, la secuencia en una de las dos cadenas (sentido 3' a 5' y complementaria con la secuencia del RNA mensajero que se transcribe), llamada **cadena molde** o **antisentido** (del inglés *antisense*), actúa como molde para la generación del ARNm, o transcripto primario. Este fenómeno se llama transcripción y es dirigido por la enzima **polimerasa de ARN**, la cual agrega nucleótidos en el ex-

tremo 3' de la cadena en crecimiento. La hebra de ADN con la misma secuencia que el RNA mensajero (excepto por la presencia de timinas en vez de uracilos) se denomina cadena **codificante** o **con sentido**.

Antes de que el ARNm migre del núcleo hacia el citoplasma, el mismo sufre varios procesos: (i) el agregado de una guanina metilada (en inglés, *cap*) en el extremo 5' que sirve de protección, (ii) el agregado de una secuencia poliadenilada (en inglés, *poly-A tail*), normalmente entre 100 y 200 adeninas, en el extremo 3' de la molécula, hecho que le brinda estabilidad y, (iii) la remoción de los intrones en un proceso de "corte y empalme" denominado en inglés *splicing*, dando lugar al mensajero "maduro" o transcripto secundario (Figura 1.5). La maquinaria encargada de catalizar las reacciones de *splicing* se conoce como **spliceosoma** (del inglés *spliceo-*

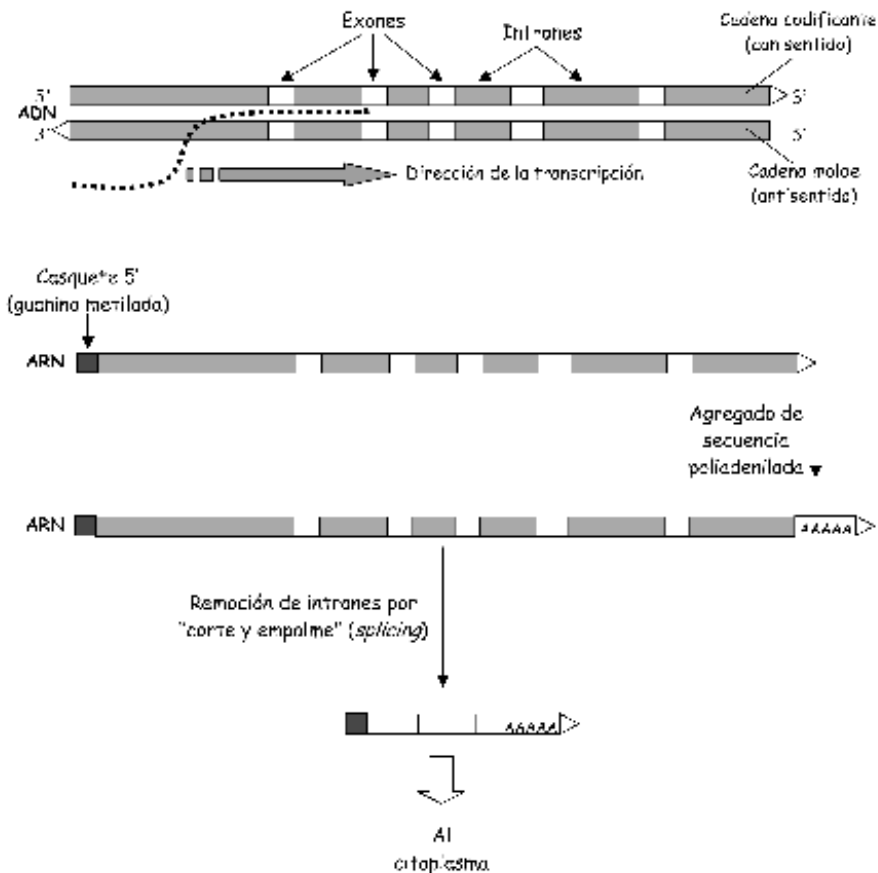


Figura 1.5. La transcripción. El esquema representa la transcripción de un gen eucariota típico con sus tres pasos característicos. Primero, el agregado de una guanina metilada (*cap* o casquete) en el extremo 5', segundo el agregado de una secuencia poliadenilada (cola de poli-A) en el extremo 3' de la molécula y, finalmente, la remoción de los intrones en el proceso de *splicing* ("corte y empalme"). Luego, el transcripto maduro se desplaza al citoplasma donde es traducido a proteína.

some) y está formada por entre 50 y 100 proteínas y snARN. El reconocimiento y la remoción de los intrones se ve favorecido por el hecho de que las dos primeras bases en el extremo 5' (GU) y las dos últimas del extremo 3' (AG) del intrón están muy conservadas en todos los genes. En menor medida, sucede lo mismo con los extremos de los exones en el borde exón-intrón. Estas secuencias forman los llamados **sitios de splicing**: en el extremo 5' se llama **sitio donante** y en el extremo 3', **sitio aceptor**.

El ARN así procesado es usado para dirigir la adición secuencial y precisa de aminoácidos a la cadena polipeptídica, mecanismo conocido como **traducción** y que ocurre en el citoplasma, en unas estructuras llamadas ribosomas. Como veremos más adelante, se pueden generar diferentes ARNm a partir de un mismo gen (por *splicing* alternativo del mismo transcripto primario o por sitios de iniciación alternativos), hecho que permite la existencia de un número mayor de "productos de genes" que de genes propiamente dichos. Los aminoácidos son agregados a la cadena polipeptídica por medio del ARNt, el cual lleva una secuencia de nucleótidos llamada **anticodón**. Los aminoácidos son transportados por el ARNt, específicos para cada uno de ellos, y son llevados hasta el ARNm, donde se aparean el codón de éste y el anticodón del ARNt, por complementaridad de bases, y de esta forma se sitúan en la posición que les corresponde. Una vez finalizada la síntesis de una proteína, el ARNm queda libre y puede ser leído de nuevo. De hecho, es muy frecuente que antes de que finalice una proteína ya está comenzando otra, con lo cual, una misma molécula de ARNm es utilizada por varios ribosomas simultáneamente.

Los procesos descritos se presentan en la mayoría de los genes, pero debemos recordar que existen segmentos de ADN cuya única función es producir (por transcripción) ARNt o ARNr. Estos son llamados genes de ARNr o ARNt y no son traducidos a proteínas. En el caso de los genes de ARNr, y debido al gran requerimiento de este tipo de ARN, existen cientos de copias en el genoma, distribuidas en grupos de genes duplicados. Se conoce como **transcriptoma** (del inglés *transcriptome*) al conjunto de ARN sintetizado por una célula en un momento dado, incluyendo aquellos ARN que no codifican para proteínas.

1.2.2 La regulación de la actividad de los genes

Se entiende por **regulación** a la habilidad de las células para controlar la expresión de los genes en un momento o situación determinados. Como es sabido, no todos los genes se expresan todo el tiempo en todas las células, sino que existen mecanismos de activación y desactivación de genes que orquestan las funciones y la diferenciación de las células. Uno de los ejemplos históricos (aunque no aplicable a los vertebrados), es la capacidad de la bacteria *Escherichia coli* de activar o desactivar los genes que participan en la degradación de la lactosa, según se encuentre presente en el medio o no. Este modelo de la regulación de los genes en los procariontes es conocido como el modelo **operón lactosa**, y fue propuesto en la década de 1960. En el mismo, los genes que codifican para proteínas relacionadas a la degradación de la lactosa se agrupan en unidades conocidas como operón. El operón está formado por se-

cuencias reguladoras (**operador y promotor**) y **genes estructurales**. Los genes estructurales del operón son *LacZ* (enzima **galactosidasa**), *LacY* (enzima **permeasa de lactosa**) y *LacA* (enzima **tiogalactósido transacetilasa**). Aguas arriba encontramos el gen *LacI* (proteína represora) y su promotor. En el estado reprimido, la proteína represora se une al operador y esto impide la transcripción de los genes estructurales. Cuando la lactosa está presente en el medio, actúa como inductora al unirse a la proteína represora y generar en ella un cambio conformacional que le impide unirse al operador, hecho que libera la síntesis de las enzimas (**Figura 1.6**). Por lo

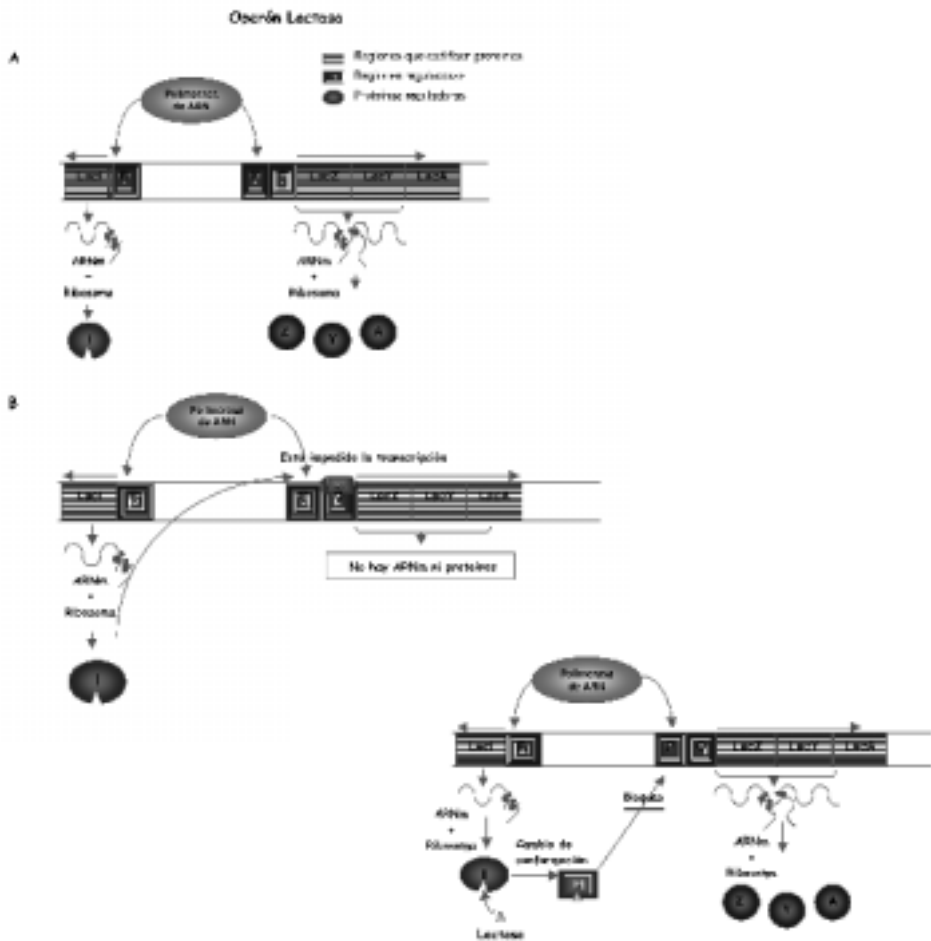


Figura 1.6. Diagrama del operón lactosa. **A**, Los genes estructurales del operón se encuentran a la derecha y son *LacZ* (enzima galactosidasa), *LacY* (enzima permeasa de lactosa) y *LacA* (enzima tiogalactósido transacetilasa). Aguas arriba encontramos el operador (**O**), el promotor de los genes estructurales (**P**), el gen *LacI* (proteína represora **I**) y su promotor (**Pi**). **B**, En el estado reprimido, la proteína represora se une al operador y esto impide la transcripción de los genes estructurales. **C**, Cuando la lactosa está presente en el medio, actúa como un inductor: al unirse a la proteína represora, genera en ella un cambio conformacional que le impide unirse al operador; esto libera la síntesis de las enzimas (**Z**, **Y**, **A**) y la lactosa puede ser metabolizada. Dibujo por *Christine K. Yone, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas, Estados Unidos*.

tanto, de acuerdo al mecanismo de control, los operones pueden ser clasificados en **inducibles** o **reprimibles**. François Jacob, André Lwoff y Jacques Monod del *Institut Pasteur* recibieron el premio Nobel de Medicina y Fisiología en 1965 por estos trabajos.

La regulación de los genes en los organismos multicelulares se complica aún más debido a los procesos de diferenciación que ocurren durante el desarrollo. Cada organismo multicelular comienza como una sola célula que se divide por mitosis; luego las células se diferencian en varios tipos funcionales utilizando algunos genes e ignorando otros. Los **genes homeóticos** (del inglés *homeobox genes*) dirigen la organización general del cuerpo y la posición de los órganos en respuesta a moléculas regulatorias. El momento en el cual se expresa un gen suele seguir una secuencia, tal como veremos en la producción de hemoglobina fetal en los glóbulos rojos de los mamíferos, la cual cambia a hemoglobina adulta poco antes del nacimiento.

Si bien la expresión génica es un proceso muy complejo, se puede organizar su estudio en los llamados **puntos de control**: lugares precisos donde la expresión de los genes puede ser regulada. Las células eucariotas hacen uso de una gran variedad de mecanismos regulatorios; en la **Tabla 1.2** se muestran los puntos de control mejor conocidos en la regulación de los genes eucariotas, entre los que encontramos la transcripción, el procesamiento del ARN y la post-transcripción.

Puntos de control en la regulación de los genes eucariotas	
I.	Regulación por cambios en el ADN
	A. Modificación de bases (metilación)
	B. Amplificación e distribución génica
	C. Transposición (transposones)
	D. Inversiones
	E. Recombinación somática (splicing de ADN)
II.	Regulación de la transcripción
	A. Regulación positiva (factores de transcripción)
	B. Regulación negativa (proteínas represoras)
III.	Regulación del procesamiento del ARNs
	A. Splicing alternativo
	B. Edición del ARNs
IV.	Regulación de la traducción
	A. ARNs no traducidos (mensajeros ocultos)
	B. Estabilidad de los ARNs
	C. Niveles de traducción
V.	Regulación postraducción
	A. Fosforilación
	B. Localización de las proteínas

Tabla 1.2. Puntos de control de la expresión génica. La expresión génica es un proceso muy complejo y su regulación está supeditada a diversos puntos de control, entre los que encontramos la regulación al nivel del ADN, la transcripción, el procesamiento del ARN, la traducción y la postraducción (ver texto).

1.2.2.1 Regulación a nivel del ADN

La regulación en la dosis del producto de un gen es uno de los mecanismos que actúan a nivel del ADN. Un ejemplo de este tipo de regulación lo podemos observar en las histonas, donde una gran cantidad de genes (copias normales del complemento cromosómico) aseguran la gran producción de estas proteínas esenciales para la formación de la cromatina. En otros casos, la dosis génica puede verse afectada en forma temporaria, como respuesta a un estímulo dado, fenómeno conocido como **amplificación génica**. Un ejemplo de esto lo encontramos en el anfibio *Xenopus laevis* donde el número de genes de ARNr (necesarios para la producción de los ribosomas) aumenta hasta 4000 veces específicamente durante el desarrollo del ovocito. Por otro lado, la amplificación génica es muchas veces parte del complejo proceso del desarrollo tumoral. La otra cara de la moneda es la **disminución génica**, proceso estudiado en el nematode *Parascaris equorum* y en el protozoo *Oxytricha sp* pero no descrito en mamíferos.

Uno de los fenómenos más interesantes y únicos de la regulación génica a nivel del ADN es la **recombinación somática** (conocida en inglés como **DNA splicing**), producida en la formación de las inmunoglobulinas y los receptores de linfocitos T (TCR), la cual veremos en forma muy resumida. Por mucho tiempo, el origen de la enorme diversidad que presentan los anticuerpos (con alrededor de 10^9 especificidades) era imposible de explicar con el número de genes presente en el genoma de los mamíferos. En la década de 1970 el científico japonés Susumu Tonegawa descifró que el fenómeno estaba ligado a la presencia de rearrreglos somáticos en el orden de los genes de la línea germinal durante la maduración de los linfocitos. Por estos trabajos, Tonegawa fue galardonado con el premio Nobel de Medicina y Fisiología en el año 1987.

Las inmunoglobulinas se presentan en familias multigénicas con secuencias codificantes que varían según el tipo de inmunoglobulina. Las inmunoglobulinas están formadas por dos cadenas livianas (L) y dos pesada (H); a su vez, estas cadenas tienen porciones constantes (C), con poca variabilidad entre anticuerpos, y otras variables (V), que interviene en el reconocimiento del antígeno. En el caso de las cadenas L, hay cientos de genes que codifican para la zona variable (genes V) y unas decenas de genes para zonas de unión (genes J). Para las cadenas H, existen los mismo genes V y J pero además hay una decena de genes de diversidad (D) entre ellos. Las inmunoglobulinas funcionales se producen durante la maduración de los linfocitos B en un proceso en el que los segmentos V, D y J (en el caso de la cadena H) y V y J (en el caso de la cadena L) se reordenan al azar (por corte y empalme), con pérdida de ADN. Las reordenaciones de segmentos génicos correspondientes a las regiones variables se producen en la médula ósea: primero se organizan los genes de VH y luego los de VL.

Además, durante esta recombinación, se producen pequeña inserciones y deleciones que aumentan las posibilidades de variabilidad en los anticuerpos. La asociación azarosa de una cadena H con otra L y la aparición de mutaciones somáticas en los genes mencionados aumentan aún más la diversidad en la especificidad de los anticuerpos. Finalmente, otro mecanismo particular de la expresión de los genes de inmunoglobulinas es la llamada **exclusión alélica**, don-

de sólo uno de los alelos del par homólogo se expresa, y no ambos como en la mayoría de los genes. Aunque se desconocen los detalles sobre la bioquímica de estas reacciones de recombinación, se han identificado dos genes cuyos productos actúan en la unión de los segmentos V-(D)-J. Se trata de los genes de activación de recombinación *Rag1* y *Rag2*. Son dos genes estrechamente ligados, que desempeñan un papel importante en la maquinaria de recombinación de los segmentos de porciones variables de las inmunoglobulinas; estos genes inducen roturas de cadena doble en el ADN y actúan sobre el ADN de línea germinal de las inmunoglobulinas.

La **metilación** del ADN es otro mecanismo de regulación de la expresión génica. En la mayoría de las células eucariotas, una pequeña proporción de las bases citosina se encuentra modificada por la adición de un grupo metilo (CH_3) en el carbón número 5. Esta modificación es posterior a la replicación del ADN y la realiza la enzima **metilasa de ADN**. Existe una preferencia para la metilación de aquellas citosinas que se encuentran formando el dinucleótido 5'-CG-3'. Una gran mayoría de los genes contiene secuencias ricas en CG, aguas arriba; dichas secuencias son sitios potenciales de metilación. Es sabido que la metilación de estas secuencias puede alterar la transcripción y por lo tanto sirve como instrumento de regulación de la expresión del gen. Los niveles altos de metilación son siempre asociados a genes con bajos niveles de transcripción; un buen ejemplo de esto se observa en el cromosoma X inactivo (recordar que en las hembras, uno de los cromosomas X del par homólogo está inactivo), donde existen elevados niveles de metilación.

La alteración de la secuencia de ADN por **transposición** es otra variante de la regulación a nivel del ADN, ya que puede influenciar la expresión de los genes al implicar secuencias regulatorias. Como fuera descrito por Barbara McClintock (entre 1940 y 1960) en el maíz, en los mamíferos la información genética también puede moverse de un sitio a otro del genoma. Estos fenómenos de transposición son debidos a la presencia de secuencias de ADN capaces de cortarse y pegarse, por sí mismas, llamadas **transposones** (o retrotransposones). Estas secuencias son de origen viral y pueden duplicarse por medio de una molécula de ARN intermediaria que se retro-transcribe a ADN y se inserta en una nueva posición, expandiéndose por el genoma. En forma general, existen dos tipos de retrotransposones: los LTR (del inglés *long terminal repeat*) y los no LTR.

1.2.2.2 Regulación a nivel de la transcripción

Muchos genes que codifican para enzimas o componentes esenciales para la célula son expresados en forma constitutiva, es decir, a niveles constantes y en la mayoría de los tejidos. Estos genes se conocen como **genes constitutivos** (en inglés, *housekeeping genes*). Los **genes de expresión diferencial**, en cambio, modifican su expresión en respuesta a señales procedentes del interior o del exterior de la célula, tales como la variación en la concentración de metabolitos o las condiciones ambientales. Las proteínas que codifican aparecen en la célula sólo bajo determinadas condiciones o momentos del ciclo celular, o si están presentes continuamente, su cantidad varía en gran medida en respuesta a una señal. Estos genes suelen estar regulados particularmente a nivel de la transcripción.

Los **factores de transcripción** son proteínas que interactúan con el ADN y son esenciales para la regulación positiva de la transcripción. Como hemos visto, la región promotora de los genes contiene señales que indican a la polimerasa de ARN dónde comenzar la transcripción. La secuencia TATA en los mamíferos indica que alrededor de 30 pb aguas abajo se inicia la transcripción (la misma funciona mejor si hay un segmento rico en GC y la secuencia CCAAT aguas arriba del sitio de inicio de la transcripción). Los promotores eucariotas, a diferencia de los procariotas, no contienen suficientes señales de reconocimiento para la acción de la polimerasa de ARN. Es por eso que las secuencias reguladoras deben ser reconocidas por factores de transcripción que se unan a ellas y activen la transcripción. Para cumplir con esta función, las proteínas reguladoras deben reconocer secuencias específicas del ADN (secuencias consenso) y unirse a ellas. Esta unión resulta posible por la estructura espacial de las proteínas reguladoras que poseen "motivos" especiales conocidos como **hélice-vuelta-hélice** (en inglés, *helix-loop-helix*), **cremallera de leucinas** (*leucin-zipper*) y **dedos de zinc** (*zinc finger*), todos con dominios de unión al ADN. Algunos genes tienen más de un promotor en su estructura, lo que les permite activarse en distintos tipos celulares o sintetizar transcritos primarios ligeramente diferentes. Esto sucede en el ratón (y otros mamíferos) con el gen *Pgr* (receptor de progesterona), el cual presenta la síntesis de una forma B de 108-120 kilodaltons y otra forma A, truncada, de 80 a 94 kilodaltons, por medio del uso de **promotores alternativos**.

Dentro de las proteínas reguladoras de la transcripción, las más estudiadas en los vertebrados son las hormonas. Entre ellas, las hormonas esteroideas (sintetizadas a partir del colesterol), son un buen ejemplo de activación de la transcripción de genes a través de la unión con receptores específicos. Estas hormonas se unen a sus receptores en el citoplasma y viajan al núcleo de la célula, donde se unen a secuencias de ADN específicas para activar la transcripción. Las células que carecen de los receptores específicos en el citoplasma no se verán afectadas por la presencia de la hormona. Tanto los receptores hormonales como los factores de transcripción se unen a secuencias de ADN conocidas como **secuencias estimuladoras** (en inglés, *enhancers*). Son secuencias cortas (menos de 20 pb) que se encuentran en diferentes sitios en los alrededores del gen (muchas se encuentran aguas arriba, otras dentro de los intrones e inclusive en el extremo 3' del gen). Muchos genes se encuentran bajo el control de más de una secuencia estimuladora permitiéndole responder a una variedad de señales externas e intracelulares.

Aun haciendo uso de un único promotor, existe la posibilidad de producir diferentes proteínas, a partir de un único gen, por cambios en el procesamiento del ARNm, lo que se conoce como **splicing alternativo**. A partir de un patrón de corte y empalme diferencial de exones, donde algunos exones son eliminados, este mecanismo permite generar distintos transcritos a partir de la misma secuencia codificante original. En algunos genes de origen neurológico se describieron miles de formas alternativas generadas por este mecanismo. Uno de los hallazgos más sorprendentes de generación de diversidad de proteínas a partir de un sólo gen se observa en el gen *Dscam* (*Down syndrome cell adhesion molecule*). Este gen de la mosca *Drosophila* posee una estructura singular que combina variantes de un mismo exón con *splicing* alternativo, generando decenas de miles de ARNm alternativos. Estudios recientes sobre el *splicing* alternativo (realizados sobre la secuencia completa del genoma humano) indican que en-

tre el 40% y el 50% de los genes humanos posee formas alternativas debidas a este mecanismo, lo que sugiere que el *splicing* alternativo sería un componente esencial de la complejidad funcional del genoma de los vertebrados.

En algunos casos, el número de exones transcritos no se ve alterado, pero cambia la eficiencia de la transcripción. Un ejemplo de esto último es la transcripción diferencial que presenta la α **amilasa** en el ratón. El nivel de transcripción es mucho mayor en las glándulas salivales que en el hígado, debido a la presencia de secuencias líderes alternativas en la región 5' del gen.

1.2.2.3 Regulación postranscripcional

La **edición del ARN** (del inglés, *RNA editing*) es un procedimiento postranscripcional en el cual ciertos nucleótidos son modificados químicamente. Por ejemplo, la conversión de la adenosina en **inosina** (catalizada por la enzima **adenosina deaminasa**) puede alterar la identidad de un aminoácido en la proteína sintetizada. La modificación de la vida media de una molécula de ARNm es otra forma de regulación, como se observa en el aumento de la vida media del transcrito de la caseína ante la presencia de la hormona prolactina.

En las células procariotas, la mayoría de las moléculas de ARNm son traducidas a proteínas el mismo número de veces, con muy pocas variaciones entre genes. En cambio, en los eucariotas, la traducción es un punto importante de regulación de la expresión génica, involucrando dispositivos muy variados. En algunos casos, la traducción de una molécula de ARNm puede ser activada o desactivada por señales específicas. Además, el control del nivel de síntesis de proteínas puede ser regulado en distintos puntos, como la formación de los complejos ribosomas-ARNm. Este tipo de regulación negativa se observa en los óvulos no fecundados, donde se acumulan transcritos no traducidos, por no ser necesarios en ese estadio.

1.2.2.4 Los genes de globinas: modelo de regulación de genes en mamíferos

La hemoglobina, presente en los glóbulos rojos, transporta oxígeno de los pulmones a los tejidos periféricos. Este pigmento está formado por el **grupo hemo** (conteniendo hierro), al cual se une el oxígeno, y un tetrámero formado, en el adulto, por dos pares de moléculas diferentes: las cadenas α y β de la globina. La regulación de la expresión de las globinas durante el desarrollo embrionario y fetal es un ejemplo interesante de adaptación: cada etapa tiene su hemoglobina característica, la cual está mejor capacitada en el transporte de oxígeno en ese momento del desarrollo. Los genes de las globinas son activados y desactivados en diferentes momentos del desarrollo y en distintos tejidos; y el orden temporal de expresión es el mismo en el cual se encuentran ordenados los genes en el cromosoma.

En los estadios tempranos del embrión, las globinas sintetizadas en el saco vitelino son la cadena ζ (letra griega zeta), molécula similar a la cadena α (en inglés, α -like) y la cadena ϵ (letra griega épsilon), similar a la cadena β (β -like). Por lo tanto, la hemoglobina embrionaria presenta la fórmula $\zeta_2\epsilon_2$. En estadios más tardíos (feto) comienza la transición hacia la expresión

de los genes α , en este caso en el hígado fetal, en conjunto con la síntesis de dos formas alternativas (poseen un polimorfismo glicina/alanina) de la cadena γ ($G\gamma$ y $A\gamma$). Por lo tanto, la hemoglobina fetal (hemoglobina F) tiene fundamentalmente la fórmula $\alpha_2 \gamma_2$. Finalmente, antes del nacimiento, comienza la síntesis de cadenas β en la médula ósea, junto con trazas de la cadena δ (casi idéntica a β). De aquí en adelante se sintetiza sólo un tipo de cadena α , y la estructura de la hemoglobina adulta es $\alpha_2 \beta_2$ (hemoglobina A), con trazas de $\alpha_2 \delta_2$ (hemoglobina A2) (Tabla 1.3). En el caso de los genes de globina, uno de los mecanismos regulatorios parece ser la competencia entre los promotores de los genes fetales y los adultos por una única región *enhancer* (la afinidad del *enhancer* por uno u otro promotor sería la causa de la activación o desactivación de los distintos genes durante el desarrollo). La presencia de secuencias afuera del grupo (del inglés *cluster*) conocidas como **LCR** (del inglés *locus control region*) también son fundamentales en los procesos regulatorios de estos genes (regulación *trans*). Como veremos en el Capítulo IX, la **anemia falciforme** es una enfermedad genética humana producida por una mutación en el gen de la β globina. La **talasemia** es también una anomalía hereditaria de la hemoglobina que se caracteriza por la ausencia de formación de una o más cadenas de globina.

En el ratón, las cadenas α y β de la globina están codificadas por el grupo de genes *Hba* en el cromosoma 11 y el grupo de genes *Hbb* en el cromosoma 7, respectivamente. Dentro del grupo *Hba*, los loci *Hba-a1* (cadena mayor) y *Hba-a2* (cadena menor) codifican para las cadenas α de la hemoglobina adulta. Es interesante notar que, dentro de las líneas consanguíneas de ratones, existen al menos cinco formas alternativas (polimorfismo) de la cadenas α . El locus *Hba-x* (*HBZ* en el hombre) codifica para la cadena ζ de la hemoglobina embrionaria. El grupo *Hbb* cuenta con varios genes: *Hbb-y*, *Hbb-bh1*, *Hbb-b1* y *Hbb-b2*. El gen *Hbb-y* (*HBE1* en el hombre) codifica para la cadena embrionaria ϵ (homóloga a la del hombre); el gen *Hbb-*

Cadenas β -like	Cadenas α -like	Hemoglobina presente	Estadio en el cual es sintetizada
ϵ	ζ	$\zeta_2 \epsilon_2$	Embrión temprano
$G\gamma$ $A\gamma$	α	$\alpha_2 \epsilon_2$ $\zeta_2 \gamma_2$ $\alpha_2 \gamma_2$	Feto
β δ	α	$\alpha_2 \beta_2$ $\alpha_2 \delta_2$	Prenatal en adelante

Tabla 1.3. Las hemoglobinas varían según el estadio del desarrollo. En los estadios tempranos del embrión, las globinas sintetizadas en el saco vitelino son la cadena ζ y la cadena ϵ (β -like), por lo que la fórmula de la hemoglobina embrionaria es $\zeta_2 \epsilon_2$. En el feto, comienzan a transcribirse los genes α (en el hígado fetal) junto con las cadenas $G\gamma$ y $A\gamma$. Por lo tanto, la hemoglobina fetal tiene la fórmula $\alpha_2 \gamma_2$. Finalmente, antes del nacimiento, comienza la síntesis de cadenas β (en la médula ósea), junto con trazas de la cadena δ . Desde este momento se sintetiza sólo un tipo de cadena α , y la estructura de la hemoglobina adulta es $\alpha_2 \beta_2$ (hemoglobina A), con trazas de hemoglobina $\alpha_2 \delta_2$ (hemoglobina A2).

bh1 codifica para la globina fetal δ (β -like), mientras que *Hbb-b1* y *Hbb-b2* lo hacen para las globinas adultas β mayor y menor, respectivamente.

La estructura descrita en el ratón se encuentra muy conservada a lo largo de todas las especies de mamífero, incluido el hombre (donde *HBA1* es el gen que codifica para la cadena α menor y *HBA2* para la cadena β mayor), y son a la vez un modelo para el estudio de la evolución y la regulación de los genes en los mamíferos. Los genes α y β de las globinas surgieron por un evento de duplicación génica que se produjo hace unos 500 millones de años en una secuencia de globina ancestral presente en el linaje de los vertebrados. Por esta razón, la estructura y organización de los genes α y β de las globinas son muy similares; por ejemplo ambos grupos de genes tienen una estructura formada por tres exones y dos intrones.

1.3 El análisis del ADN al nivel molecular

Gracias al perfeccionamiento de las técnicas de análisis de los ácidos nucleicos, el conocimiento del genoma de los mamíferos ha progresado de manera considerable en el curso de los últimos 20 años. En el ratón y en el hombre, por ejemplo, una enorme cantidad de genes han sido clonados e inclusive se conoce la secuencia completa de muchos de ellos. Es más, la secuencia completa del genoma humano y el primer borrador de los genomas del ratón y la rata están actualmente disponibles (ver Capítulo VI). Por lo tanto se comienza a tener una idea más precisa sobre la estructura y la forma en que está organizado el genoma de los mamíferos. Haremos un repaso de las técnicas que han permitido el clonado (o clonaje) y el estudio de esos genes.

1.3.1 Las herramientas utilizadas en biología molecular

1.3.1.1 Enzimas de restricción

Las **enzimas de restricción** (o endonucleasas de restricción) fueron originalmente aisladas de bacterias y se vio que eran capaces de cortar el ADN genómico en fragmentos, sobre la base de una secuencia específica de corte (secuencias de reconocimiento). El tamaño de los fragmentos dependía, a la vez, de la enzima utilizada y del origen de la muestra de ADN. Estas enzimas han sido clasificadas en tres grupos: las de tipo I y III poseen una actividad de restricción dependiente de ATP, las de tipo II (las más utilizadas en biología molecular) poseen actividad endonucleasa y la mayoría de ellas reconoce secuencias específicas (palindrómicas) de 4 a 6 nucleótidos de longitud. Algunas enzimas cortan el ADN sobre el mismo eje, de forma tal que generan fragmentos con extremos romos, mientras que otras cortan cada hebra sobre lados opuestos del eje de simetría, creando fragmentos de ADN con extremos cohesivos. Los nombres de las enzimas de restricción están constituidos por letras y números romanos, donde las letras derivan del nombre de la bacteria donde fue aislada la enzima. Por ejemplo, *EcoRI* fue aislada de *Escherichia coli* y *Hinfi* de *Haemophilus influenzae*.

Cuando se digiere ADN genómico con una enzima de restricción, se obtienen millones de fragmentos imposibles de identificar, aunque pueden ser visualizados como un chorreado fluorescente en un gel de **agarosa** teñido con **bromuro de etidio**. Por ejemplo, cuando el ADN de ratón es digerido por la enzima de restricción *EcoRI* (que reconoce una secuencia de corte de 6 pb) se divide en, aproximadamente, dos millones de fragmentos, con tamaños que varían desde algunas pb a varias kb. Si este corte se realiza con una enzima de corte más frecuente (que tenga un sitio de corte de 4 pb, como por ejemplo *TaqI*) se generarán más fragmentos y éstos serán, en promedio, más chicos. A continuación veremos como individualizar estos fragmentos y que utilidad podemos obtener de las enzimas de restricción.

1.3.1.2 Las técnicas de hibridación (Southern blot)

La técnica de **Southern blot** fue descrita a fines de la década de 1970 por E. M. Southern con el fin de detectar un fragmento particular de ADN de entre una gran cantidad de secuencias, gracias al uso de una **sonda molecular** (en inglés, *probe*). Esta sonda es un fragmento de ADN cuya secuencia es complementaria a la que estamos buscando. Los pasos básicos de la técnica son: (i) El ADN blanco (aquel que estamos estudiando) es digerido por una enzima de restricción por lo cual obtenemos una mezcla de cientos de miles de fragmentos de distinto tamaño. (ii) El ADN es migrado en un gel de electroforesis (normalmente de agarosa) para separar los fragmentos en función de su tamaño. (iii) Se transfieren las muestras de ADN del gel de agarosa hacia una membrana (nylon o nitrocelulosa), donde quedarán fijadas tal cual migraron. (iv) Previa desnaturalización del ADN por medio del calor, se coloca la sonda (marcada con una sustancia radioactiva o quimoluminiscente) sobre la membrana. La sonda va a unirse a la secuencia complementaria del ADN que estamos estudiando (fenómeno de hibridación) y, una vez expuesta a un film sensible, dará una marca positiva (una banda/s) en el lugar/es donde se encuentre, según el tamaño (peso molecular) del fragmento (**Figura 1.7**). En los animalarios, esta técnica está muy ligada a la detección de animales transgénicos, en particular a la evaluación de la cantidad de copias del transgén (ver capítulo VIII). Como veremos en el Capítulo VI, la llegada de la técnica de *Southern blot* fue cardinal en el descubrimiento de los primeros polimorfismos en las secuencias de ADN (los **RFLP** o polimorfismos de longitud en fragmentos de restricción), cuya presencia revolucionó el desarrollo de los mapas genéticos. En el caso de querer estudiar la expresión de un gen (usando directamente el ARN), la técnica es básicamente la misma y se denomina **Northern blot**, un guiño al nombre del autor de la técnica de hibridación en ADN.

1.3.1.3 Reacción en cadena de la polimerasa de ADN

Pocas técnicas han tenido un efecto tan profundo sobre la investigación biomédica como lo ha tenido la **reacción en cadena de la polimerasa** o **PCR** (en inglés, *polymerase chain reaction*), siendo actualmente el método experimental más utilizado. En abril de 1983, Kary Mullis dio a conocer la técnica y en el año 1993 recibió el Premio Nobel de Química por el descubrimiento. En forma sencilla podemos decir que se trata de la síntesis enzimática de una gran cantidad de ADN (*in vitro*) de un segmento específico. El ADN blanco (o molde) a copiar está definido por dos oligonucleótidos cortos (generalmente entre 18 y 25 pb), de cadena simple denominados **iniciadores** o **cebadores** (en inglés *primers*). Estos iniciadores se pegan

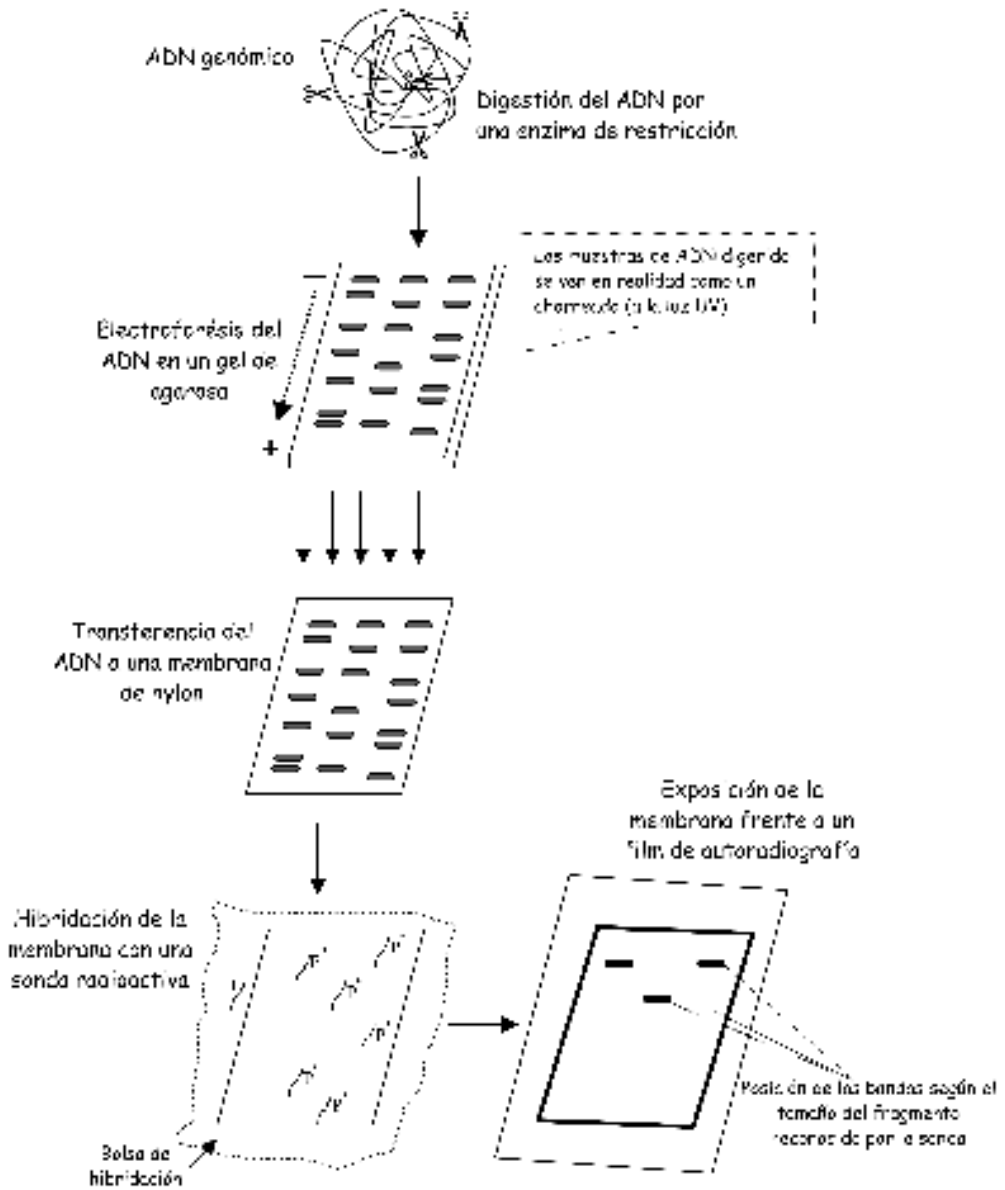


Figura 1.7. La técnica de Southern blot. El ADN genómico se corta con una enzima de restricción y los fragmentos son separados (según su tamaño) en un gel de electroforesis. El ADN es transferido a una membrana de soporte y fijado. Luego se la enfrenta a la sonda específica (marcada con radioactividad o enzimas) donde va a hibridar las secuencias complementarias en el ADN genómico. La membrana se expone a una placa de autoradiografía y se observarán bandas según el tamaño del fragmento que fue reconocido por la sonda.

a secuencias complementarias en las cadenas opuestas del ADN, flanqueando la región a amplificar, e inician la síntesis de nuevas copias, formadas por los oligonucleótidos más la secuencia comprendida entre ellos. La estandarización (y expansión) de la técnica, tal cual la conocemos hoy en día, fue posible gracias a la identificación de una enzima polimerasa resistente a las altas temperaturas, designada **Taq polimerasa** (del inglés *Taq polymerase*). La misma es extraída del microorganismo *Thermus aquaticus* (archaea), el cual vive en fuentes de agua caliente por encima de los 90°C.

El ciclo del PCR consiste en tres etapas bien definidas: (i) La reacción comienza con el calentamiento de la muestra de ADN a amplificar a 90-95°C para lograr que la molécula de doble cadena de ADN se separe en dos moléculas simple cadena (**desnaturalización**). (ii) Luego se permite el pegado de los iniciadores al ADN molde a temperaturas que pueden variar entre 40-60°C (etapa de pegado de iniciadores o **annealing**). (iii) Finalmente, se permite la síntesis enzimática (por medio del uso de la enzima *Taq* polimerasa) de las nuevas cadenas de ADN a 72°C (proceso llamado **extensión**). Este ciclo estándar se repite entre 25 y 40 veces usando termocicladores programados para tal propósito, donde los productos de cada ciclo servirán de molde para los ciclos subsiguientes. Esto resulta en la acumulación exponencial de la secuencia blanco, definida por los iniciadores flanqueantes, llegando a sintetizarse millones de copias (**Figura 1.8**).

El tamaño del fragmento a amplificar es normalmente entre 50 y 4.000 pb (4 kb), aunque con técnicas refinadas y condiciones especiales se puede amplificar segmentos de hasta 30 kb. Es importante recalcar que un aspecto crítico del PCR es el diseño de iniciadores que amplifiquen (en las condiciones apropiadas) en forma única y confiable el fragmento que queremos amplificar, aun cuando éste se encuentre en la enorme mezcla de secuencias que representa el ADN genómico total. Clásicamente, estos fragmentos son analizados posteriormente a la síntesis, por técnicas de electroforesis, aunque ya están en circulación los nuevos equipos denominados de "tiempo real" (del inglés *real time PCR*) en los cuales los productos (obtenidos con *primers* fluorescentes) son analizados directamente durante la reacción de PCR.

Una de las principales variantes de la técnica de PCR es el denominado **RT-PCR** (del inglés *reverse transcription PCR*) y consiste simplemente en la realización de un primer paso de transcripción reversa seguido por los ciclos de PCR. Básicamente, lo que se hace es extraer ARN total y realizar una incubación con la enzima **transcriptasa inversa** (o transcriptasa reversa), la cual transforma el ARN en una cadena complementaria de ADN (se lo nombra como **ADNc**); una vez obtenido el ADNc se procede con un PCR normal (**Figura 1.9**). Existen otras variedades de la técnica de PCR, entre ellas el PCR cuantitativo y semi-cuantitativo que no trataremos en este libro.

1.3.1.4 Técnicas DGGE y SSCP: detección de mutaciones

Las técnicas **DGGE**, electroforesis en geles desnaturalizantes en gradiente (del inglés *denaturing gradient gel electrophoresis*), y **SSCP**, polimorfismos en la conformación de las cadenas sencillas de ADN (del inglés *single strand conformation polymorphisms*), son dos técnicas basa-

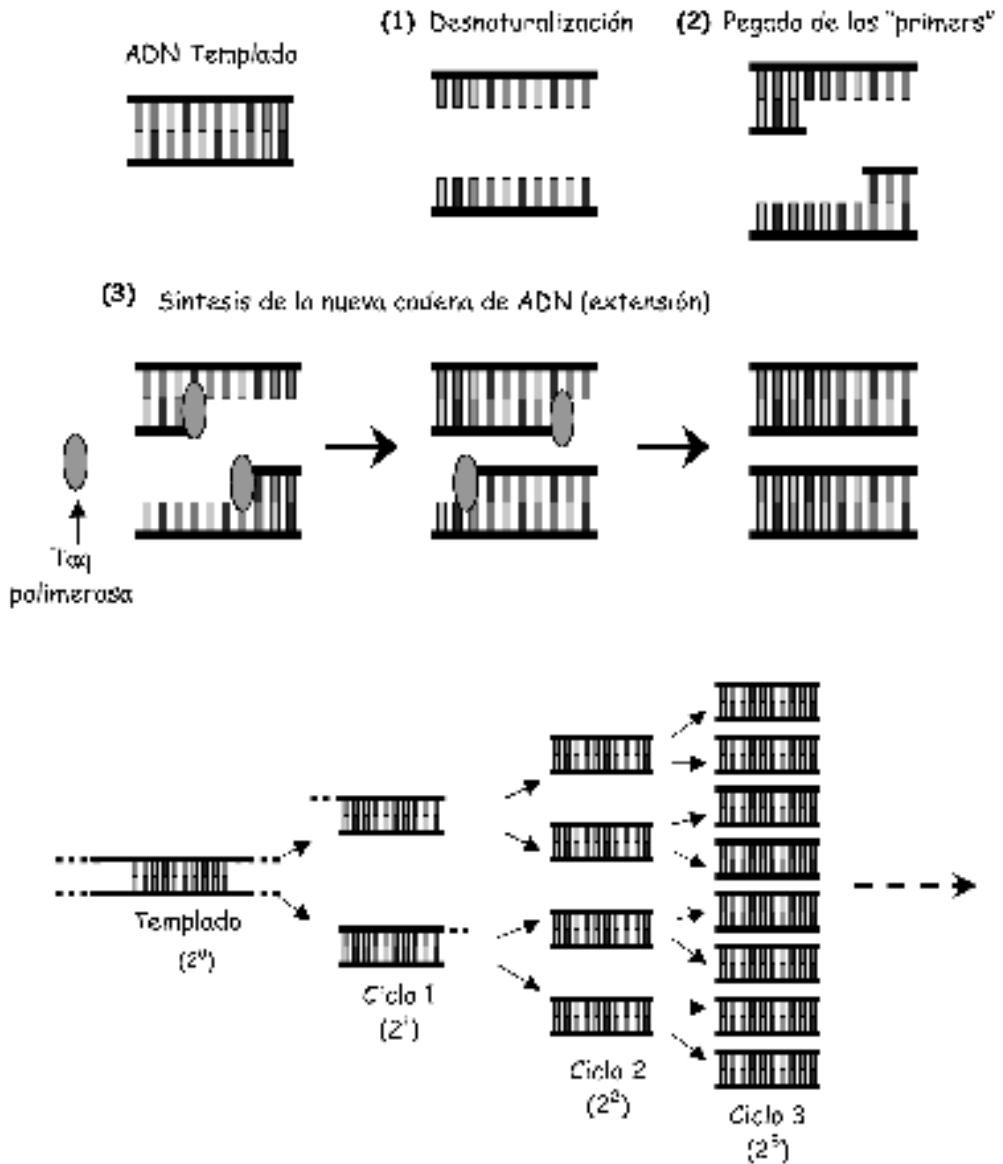


Figura 1.8. Pasos de la reacción en cadena de la polimerasa de ADN (PCR). Primero, la muestra de ADN templado se desnaturaliza por calor (a 94-95°C). La segunda parte del ciclo consiste en el pegado ("annealing") de los primers a la secuencia complementaria en el templado (entre 50 y 70°C). Finalmente, se realiza la extensión (a 72°C), paso en el cual la enzima Taq polimerasa sintetiza una cadena complementaria al molde. A medida que se repiten los ciclos, las copias de ADN aumentan en forma exponencial.

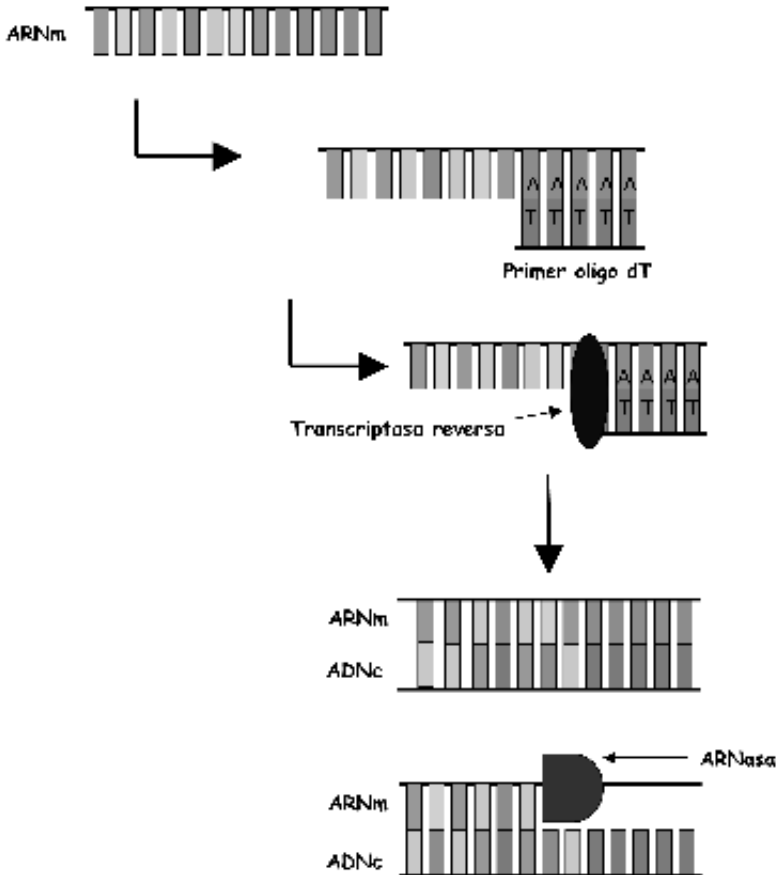


Figura 1.9. RT-PCR. El primer paso consiste en la transcripción del ARNm por medio de la incubación con la enzima transcriptasa reversa y un primer universal (oligo dT o hexámeros). De esta manera se sintetiza una cadena complementaria de ADN (ADNc). Una vez obtenido el ADNc se remueve el ARNm por medio del tratamiento con una enzima y se procede con los ciclos del PCR.

das en la amplificación de ADN por PCR que son muy utilizadas para la detección de mutaciones o polimorfismos (variantes). En el caso de DGGE, la migración del ADN se realiza en un gel de poliacrilamida con concentraciones crecientes de un agente desnaturante (por ejemplo, urea y formamida) con el propósito de imitar un aumento lineal de la temperatura desde la salida de las muestras hasta la base del gel. La disociación transforma el fragmento de ADN en una estructura parcialmente abierta y crea una disminución de su movilidad, lo que hace que el ADN se concentre en un punto del gel. Como la migración de las moléculas depende de la secuencia del segmento amplificado, una sola base cambiada puede alterar el patrón de bandas en la electroforesis. La técnica DGGE permite el análisis de un espectro de mutaciones puntuales y se utiliza habitualmente en estudios de mutaciones espontáneas o inducidas (por ejemplo, mutaciones en genes relacionados al cáncer).

La técnica SSCP esta basada en la migración de las cadenas simples (hebras monocatenarias) de una secuencia de ADN en un gel de poliacrilamida no desnaturalizante. Las muestras se desnaturalizan a altas temperaturas (95°C) y luego el gel migra a bajas temperaturas para evitar que las cadenas de ADN vuelvan a unirse. Según la conformación espacial que las cadenas adopten, y esto depende enteramente de la secuencia nucleotídica del fragmento, las muestras de ADN analizadas presentarán distintos patrones de bandas. El método es tan sensible que es capaz de reconocer una mutación puntual en secuencias de entre 80 y 250 pb de longitud. El SSCP es muy utilizado para detectar mutaciones en genes de importancia médica (por ejemplo, oncogenes) y para diferenciar variantes alélicas en marcadores moleculares.

1.3.1.5 Los chips de ADN

Una de las áreas de la investigación genética que más se está desarrollando en estos últimos años es la relacionada a la utilización, para el análisis del ADN, de matrices miniatura de ADN (en inglés, *DNA microarrays*), conocidos normalmente como **chips de ADN** o **micromatrices de ADN**. En esencia, varios métodos estándar de la genética molecular (como el PCR, la hibridación ADN-ADN etc.) son implementados a nivel de microescala, permitiendo realizar un número enorme de análisis paralelos. Por definición, las micromatrices de material genético son una serie ordenada (microscópica) de ácidos nucleicos que permiten el análisis paralelo de muestras complejas.

Uno de los principales tipos de chips de ADN contiene una matriz de alta densidad de oligonucleótidos cortos de cadena simple, los cuales son inmovilizados durante su síntesis (por máquinas robóticas) por un proceso conocido como fotolitografía. De esta manera se pueden fijar hasta 400.000 oligonucleótidos en un área de 1,28 cm². El ADN a ser analizado (con fluorescencia incorporada) se hibrida con los oligonucleótidos inmovilizados y el patrón de hibridación es capturado con la ayuda de un microscopio y analizado con programas de computación adecuados. El otro tipo de chip de ADN consiste en una membrana de nylon que contiene miles de fragmentos cortos de ADN (en este caso doble cadena) productos en general de reacciones de PCR. Estos fragmentos de ADN son pegados por estaciones robóticas. Una de las implementaciones más comunes es el desarrollo de matrices de ADNc (ADN complementario sintetizado a partir del ARNm) a una densidad de más de 1.000 por cm², para el análisis de los niveles de transcripción.

Ambos métodos permiten estudiar la expresión de un número gigantesco de genes en paralelo y hacer un análisis de polimorfismos, o la detección de mutaciones a gran escala. De la misma forma, los chips de ADN permiten la determinación del nivel de expresión de genes en forma masiva: durante el desarrollo, durante los procesos patológicos o ante tratamientos alternativos. Aunque aún están usándose en entornos de la investigación científica, en los próximos años se prevé el uso clínico de algunos de estos sistemas. Entre los usos más destacados tenemos el diagnóstico de enfermedades; el descubrimiento de nuevos medicamentos de acuerdo al organismo y/o paciente tratado (farmacogenómica); y la investigación toxicológica (toxicogenómica) para el estudio de la correlación entre agentes tóxicos y el perfil genético del individuo afectado. Finalmente, las desventajas de los chips de ADN son que no siempre

existe una correlación entre expresión génica y niveles activos de proteínas, y que la optimización de la técnica es laboriosa y requiere de equipos sofisticados de lectura.

Actualmente, las técnicas descritas presentan pocas dificultades para los genetistas moleculares e inclusive la mayoría de las firmas comerciales venden kits "listos para usar". Gracias a éstas y muchas otras técnicas moleculares es que podemos, hoy en día, analizar la estructura del genoma en los niveles más elementales de su estructura. Todas las técnicas descritas en este capítulo pueden encontrarse con más profundidad en los manuales técnicos citados en la bibliografía general.

1.3.2 Clonado de ADN y construcción de bibliotecas genómicas

El punto de partida de la técnica de **ADN recombinante** (moléculas de ADN híbridas, de distinto origen) fue el descubrimiento de las enzimas de restricción, y la primera molécula de ADN recombinante fue presentada por Paul Berg y colaboradores en 1972. Veremos en forma muy esquemática las etapas del clonado (**Figura 1.10**).

La primer etapa del clonado consiste en digerir el ADN genómico con una enzima de restricción. Recordemos que, esencialmente, vamos a clonar todo el ADN (al azar), para finalmente elegir el pedazo de ADN que contiene el gen que estamos buscando. Estas colecciones de ADN genómico total (con secuencias representadas por millones de copias) llevan el nombre de **genotecas** o **bibliotecas de genes**. Se pueden usar enzimas que generan fragmentos de ADN con extremos cohesivos, como *EcoRI*, o sintetizar (por medio de una acción enzimática) este tipo de extremos *in vitro* para agregarlos al ADN. La segunda etapa consiste en incorporar los fragmentos de ADN obtenidos por la digestión enzimática en estructuras que permitan la posterior amplificación de esos fragmentos. Para ello se usa, en general, **plásmidos** bacterianos o virus **bacteriófagos**, denominados genéricamente como **vectores de clonación**.

Los plásmidos son pequeñas moléculas de ADN circular, de algunas kb de largo, que se replican dentro de las bacterias y les confieren la resistencia a ciertos antibióticos o metales pesados. Son elementos de replicación autónoma dentro de la bacteria y se encuentran en número variable, pero bajo circunstancias experimentales pueden ser amplificadas en forma casi ilimitada. Los virus bacteriófagos, o simplemente "fagos", son virus de ADN doble cadena (aunque algunos son ARN simple cadena) que infectan las células bacterianas. El más conocido, y uno de los más utilizados para la clonación, es el bacteriófago λ (letra griega lambda), el cual posee un genoma de ADN de 49 kb.

Ya sean plásmidos o fagos, los vectores son cortados con la misma enzima de restricción que se utilizó para fragmentar el ADN genómico, elegida sobre la base de que realice un sólo corte en el vector. Luego, el ADN genómico fragmentado y el ADN del vector son mezclados *in vitro*. Con esta mezcla de ADN, en las proporciones correctas, más el agregado de una enzima **ligasa** (que favorece la unión de los extremos cohesivos) se forman las moléculas híbridas

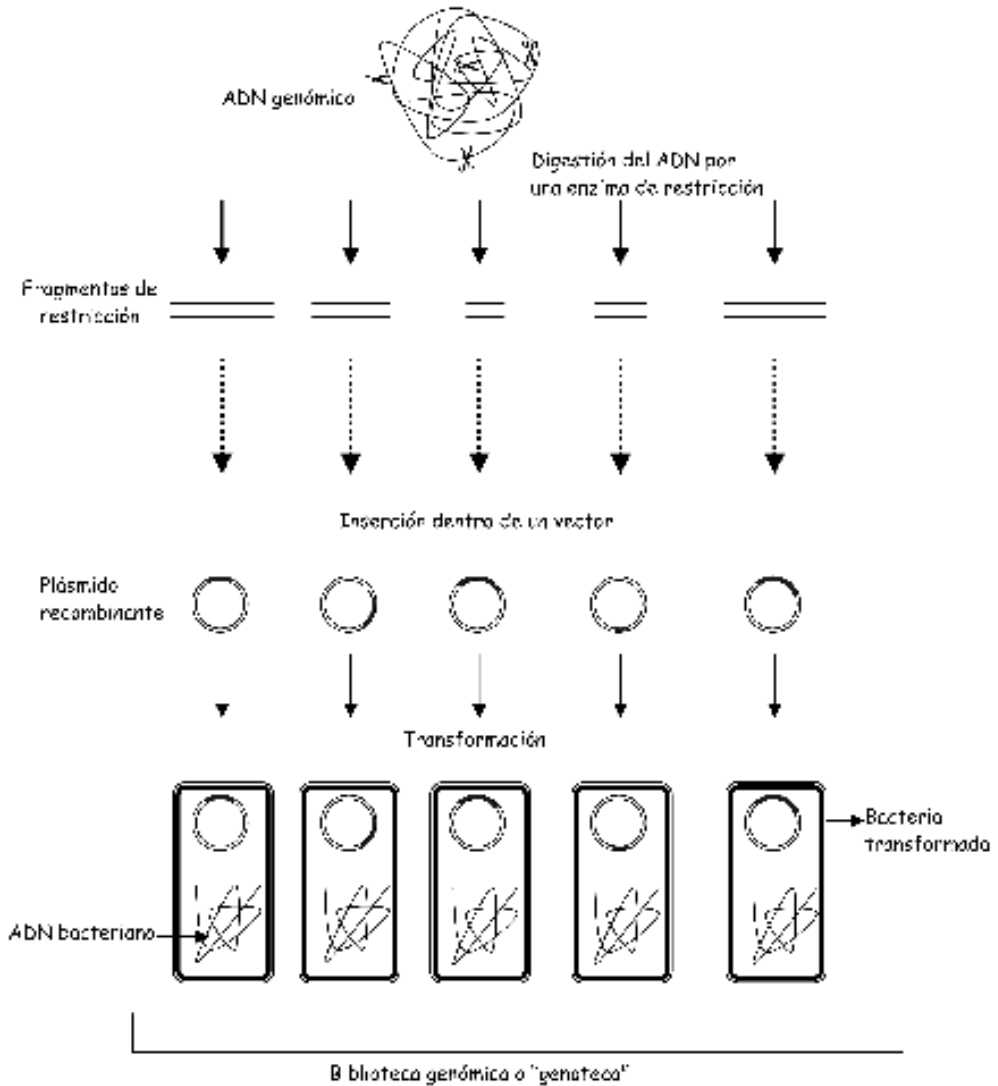


Figura 1.10. Construcción de una biblioteca genómica. El ADN genómico se corta con una enzima de restricción y los fragmentos se insertan en un vector (en la figura un plásmido bacteriano). Estas moléculas "recombinantes" se introducen en una célula huésped, normalmente la bacteria *E. coli*, (transformación). Los clones se cultivan *in vitro* en forma independiente.

de ADN eucariota/ADN plasmídico (o ADN eucariota/ADN del fago). En realidad, la suspensión se convierte en una mezcla de moléculas de distinto tipo: (i) Plásmidos (o fagos) normales que luego de haber sido cortados por la enzima de restricción se han vuelto a unir por sus extremos (la ligasa repara el corte que realizó la enzima). (ii) Plásmidos recombinantes que incorporaron un fragmento de ADN eucariota (de diferentes tamaños) y que se han re-

circularizado formando una molécula única. (iii) Estructuras inestables de ADN en tandem (cabeza-cola), formadas por ADN plasmídico, eucariota, o una mezcla de ambos en proporciones variables.

Las estructuras que interesan son, por supuesto, las segundas, con fragmentos que van desde algunas pares de base hasta 10 kb. Para aislar e identificar estas estructuras se debe realizar primero la denominada **transformación bacteriana**, donde se incorpora, con la ayuda de sales de calcio, los plásmidos tratados *in vitro* en bacterias permisivas (libres de plásmidos). La identificación de las colonias bacterianas infectadas por los plásmidos recombinantes se realiza sobre la base del crecimiento diferencial en medios de cultivos con dos antibióticos (generalmente ampicilina o tetraciclina). Las bacterias que no crecen es porque no incorporaron ningún plásmido y por lo tanto no presentan ningún interés. Las bacterias que crecen en los dos medios (es decir que son resistentes a los dos antibióticos) son aquellas que incorporaron plásmidos no recombinantes. Finalmente, las bacterias que crecen sólo en un medio son las que han incorporado los plásmidos recombinantes. Estos han perdido (por disrupción del gen) la resistencia a un antibiótico, ya que el sitio de corte de la endonucleasa de restricción se encuentra dentro de la secuencia del gen. De esta manera, queda aislado el segmento de ADN de interés, el cual podemos amplificar (por medio del cultivo de esas bacterias) infinitamente y volver a recuperar a voluntad. Con los bacteriófagos la historia es un poco más complicada pero, en esencia, se trata del aprovechamiento del ciclo lítico del virus. En esta fase de la infección el ADN simple cadena del virus penetra en la bacteria, se circulariza y se replica en forma autónoma hasta hacer estallar la célula bacteriana. De la misma manera que con los plásmidos, se puede intercalar un fragmento de ADN eucariótico dentro del genoma del fago y amplificarlo a expensas de su propia replicación en un cultivo de bacterias, previa selección de los fagos recombinantes.

La última etapa del clonado consiste en identificar, entre las innumerables secuencias aisladas dentro de los vectores recombinantes, a aquellas secuencias que nos interesan para nuestro trabajo. Se trata de buscar algunas kilobases entre millones de bases, trabajo que se puede realizar por medio de diversas estrategias. En el caso de que el gen este completo, y si es capaz de expresarse normalmente en las bacterias que incorporaron el plásmido recombinante, se puede pensar en identificar el producto del gen eucariota. Por ejemplo, una bacteria que segrega insulina de ratón no existe en estado natural y sólo es posible en los vectores recombinantes. Si el gen no se expresa pero conocemos su producto, podemos idear y sintetizar una sonda de ADN a partir de la secuencia de proteínas del producto. Con esta sonda molecular podemos identificar aquellas colonias que portan secuencias relacionadas con la de nuestro interés. Desde el momento en que un gen es clonado es posible estudiar su secuencia nucleotídica.

El límite impuesto por el tamaño del fragmento a clonar fue siempre una traba a la hora de generar bibliotecas de genes. En la segunda mitad de la década de 1980 las posibilidades de clonar fragmentos de mayor tamaño (en el orden de Mb) se vio incrementada con el uso de nuevos vectores: los cromosomas artificiales de levadura, **YAC** (del inglés *Yeast Artificial Chromosome*). Además, el avance en el análisis -a gran escala- de los genomas se debe, en buena

medida, a una nueva generación de vectores bacterianos capaces de clonar fragmentos de hasta 300 kb (usando distintas cepas de *E.coli*). Nos referimos a los vectores **BAC** (del inglés *Bacterial Artificial Chromosome*) y **PAC** (del inglés *P1-derived Artificial Chromosome*), los cuales fueron descritos a principios de 1990. Si bien estos vectores tienen mucha más capacidad de clonación que los plásmidos y cósmidos, están aún muy lejos de la capacidad de los YAC. Sin embargo, la mayor facilidad de manipulación y la menor tendencia de los BAC a generar fragmentos quiméricos (secuencias formadas por fragmentos de distinto origen) han hecho de estos vectores la herramienta de elección para generar bibliotecas genómicas y para crear mapas físicos (ver Capítulo VI).

1.3.3 La secuenciación del genoma

1.3.3.1 Las técnicas de secuenciación

En el año 1977 fueron descubiertos, casi simultáneamente, dos métodos de secuenciación de ácidos nucleicos: el primero en *Cambridge*, Inglaterra, por Frederick Sanger, y el segundo en *Harvard*, Estados Unidos por Allan Maxam y Walter Gilbert. Estos métodos, particularmente el de Sanger, han revolucionado la biología molecular del gen de tal manera que, hoy en día, es mucho más fácil secuenciar un gen y deducir la secuencia de aminoácidos de la proteína para la cual codifica que de secuenciar directamente esa proteína. Por estos trabajos, Sanger y Gilbert recibieron el Premio Nobel de Química en 1980.

El método de Sanger consiste en sintetizar, gracias a la enzima ADN polimerasa, una de las cadenas de ADN tomando a la hebra complementaria como molde. Para esto se utiliza una mezcla de desoxirribonucleótidos trifosfato (A, G, C y T) y de **di-desoxirribonucleótidos** trifosfato. Las moléculas "di-desoxi" son análogos de los verdaderos nucleótidos pero tienen la particularidad de parar la síntesis de la nueva cadena, una vez que son incorporados. Esto se debe a que su estructura molecular no permite el encadenamiento 3'...azúcar-fosfato-azúcar-fosfato...5', ya que la enzima no puede enganchar nada en el extremo 3' de una azúcar "di-desoxi". Para llevar a cabo el método de Sanger se necesita realizar, sobre el mismo ADN a secuenciar, cuatro experimentos paralelos en los que un di-desoxinucleótido diferente se mezcla con cuatro nucleótidos "normales". En general, uno (a veces dos) de estos nucleótidos normales están marcados con isótopos radioactivos (P^{32}) para poder ver en un gel el resultado de la síntesis de ADN en las cuatro muestras. El tamaño del fragmento de ADN sintetizado estará limitado por la incorporación de un di-desoxinucleótido ya que la síntesis se detiene después de esta incorporación. Por lo tanto sabemos que, en cada reacción, el tamaño de los fragmentos será muy variable (dependiendo de la incorporación al azar de un "di-desoxi") pero que todos estos fragmentos terminan en el mismo nucleótido. Haciendo migrar estas cuatro muestras independientes en un gel de poliacrilamida de alta resolución (conocidos como geles de secuencia), en cuatro calles consecutivas, se puede deducir la secuencia de la muestra de ADN analizada. El método de Sanger es muy astuto y simple de realizar pero presenta el único inconveniente de necesitar una muestra de ADN simple cadena como molde inicial (**Figuras 1.11 y 1.12**).

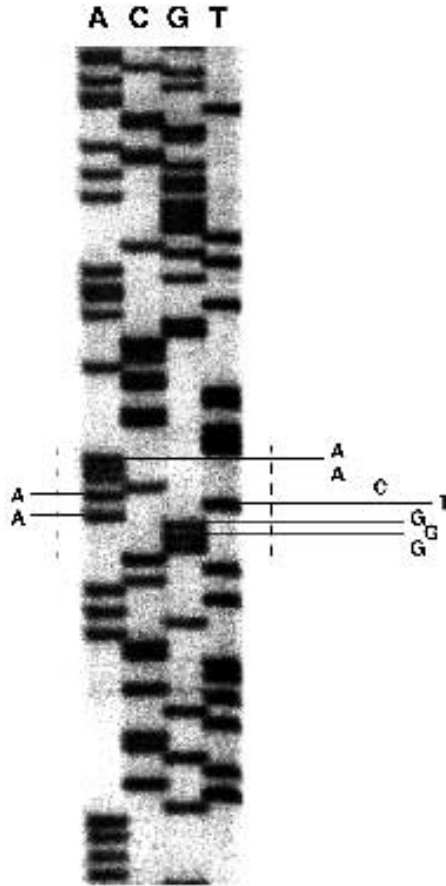


Figura 1.12. Secuenciación del ADN. La fotografía muestra una placa de autoradiografía obtenida a partir de un gel de secuenciación con la técnica de Sanger. En este caso, los di-desoxinucleótido estaban marcados con fósforo radioactivo (P^{32}). La pequeña sección que se señala con las flechas corresponde a la secuencia AACATAGGG.

Finalmente, los modernos equipos de **secuenciación automática** por fluorescencia detectan en tiempo real la migración de los fragmentos en el gel de electroforesis por medio de rayos láser. La disponibilidad de iniciadores marcados con fluorescencia (además de una nueva forma de electroforesis llamada "capilar") hacen de esta última variante una herramienta ideal para los trabajos de secuenciación a gran escala, como los proyectos conjuntos de secuenciación de genomas (ver Capítulo VI). Básicamente, en la secuenciación automática de ADN los nucleótidos terminales (aquellos donde se detuvo la cadena) están marcados con distintos colores de fluorescencia (por ejemplo: G, negro; A, verde; T, rojo y C, púrpura). Debido a que cada color distingue los productos que terminan con un mismo nucleótido, los productos de todas las reacciones de síntesis pueden ser sembrados juntos en una misma calle del gel de secuencia (poliacrilamida) y ser separados por electroforesis. Igual que en las técnicas clási-

La otra estrategia es la secuencia completa por escopetazo (*whole-genome shotgun sequencing*). Este **escopetazo generalizado** (o genómico total), se basa en la aplicación directa de la secuenciación sobre el total del ADN genómico (a partir de clones pequeños de 2, 10 y 50 kb), obteniendo una enorme cantidad de lecturas de secuencia (de aproximadamente 500 pb) que necesitan ser ensambladas y ubicadas en el genoma. Este problema de ensamblado y ubicación resulta particularmente difícil considerando el gran tamaño y el alto porcentaje de secuencias repetidas de los eucariotas superiores (más del 50% en el caso de los mamíferos). Sin embargo, para superar este problema específico se puede usar información proveniente de marcadores anclados en los mapas físicos y de las secuencias públicas (*GenBank* y otros). Tanto el genoma de la rata como el del ratón han sido secuenciados por un sistema mixto que combina el sistema jerarquizado y el genómico total. Más detalles de los organismos secuenciados hasta la actualidad serán presentados en el Capítulo VI.

1.4 Elementos de citogenética

Se calcula que el genoma completo de los mamíferos posee una cantidad de ADN que puesto “en línea” mide cerca de dos metros. Pero en la célula, como hemos visto, el ADN lo encontramos dividido en cromosomas, los cuales pueden ser visualizados por medio del microscopio óptico como cuerpos coloreados (al encontrarse empaquetados en asociación con proteínas durante la mitosis). En este momento pueden ser fotografiados para formar un **cariotipo** (número de cromosomas y forma que adquieren cuando se condensan) en orden decreciente de tamaño. Debido a que se trata de células diploides, encontraremos dos copias de cada cromosoma, lo que se denomina un **par cromosómico**.

1.4.1 El cariotipo normal en los roedores de laboratorio

El ratón de laboratorio posee un juego de cromosomas **acrocéntricos**, 19 pares **autosómicos** más el par sexual ($2n=40$), muy difícil de diferenciar sobre la base de su tamaño (**Figura 1.14**). Ordenados de mayor a menor, el cromosoma 1 representa el 7,25% del juego haploide, mientras que en el otro extremo, los cromosomas Y y 19 son los más chicos, representando cada uno 2,6% del total haploide. La rata noruega (*Rattus norvegicus*) posee 21 pares de cromosomas ($2n=42$), pero a diferencia del ratón, no son todos acrocéntricos, sino que muchos de ellos presentan un brazo corto “p” y un brazo largo “q” (**metacéntricos**), y se los agrupa en cuatro grupos designados con las letras A-D (**Figura 1.15**). El hámster Sirio (*Mesocricetus auratus*) y el jerbillo (*Meriones unguiculatus*) tienen 22 pares ($2n=44$), el hámster chino (*Cricetulus griseus*) 11 pares ($2n=22$), y el cobayo (*Cavia porcellus*) 32 pares de cromosomas ($2n=64$). En el Capítulo III veremos más detalles sobre la genética de estas especies.

Desde la década de 1970 es posible colorear los cromosomas de manera precisa generando bandas transversales, rasgo morfológico que se utiliza para su clasificación, además del tamaño

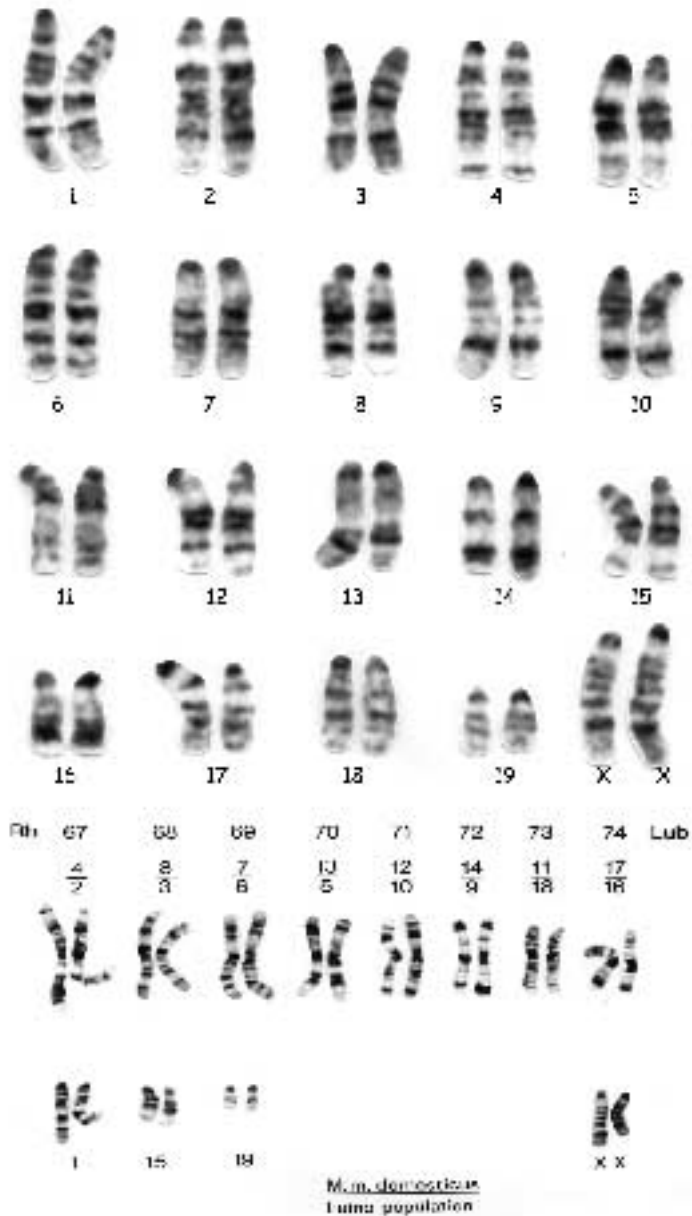


Figura 1.14. El cariotipo del ratón de laboratorio. La foto de arriba muestra el cariotipo normal del ratón (*Mus musculus*). Lo que se observa es un set completo de cromosomas metafásicos de una hembra, teñidos con el colorante de Giemsa (bandeo G). La foto de abajo muestra un ejemplo del tipo de translocaciones Robertsonianas que se encuentran en algunas poblaciones salvajes de ratones (atrapados en Italia y Suiza). Este tipo de rearreglos cromosómicos transforma los cromosomas acrocéntricos en metacéntricos (ver texto). Fotos cortesía del Dr. Heinz Winking, Medizinische Hochschule Lübeck, Alemania.

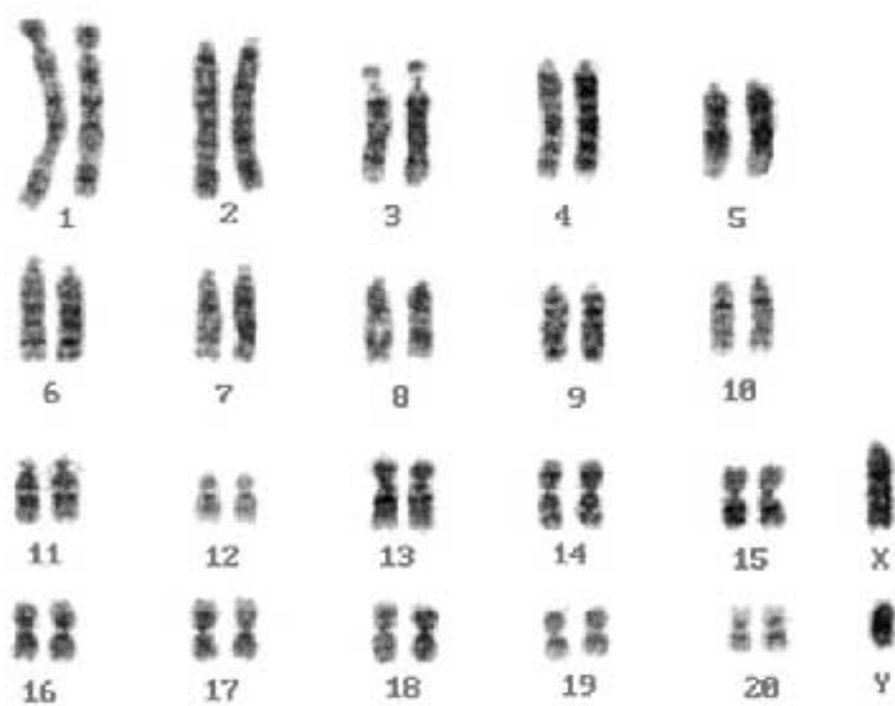


Figura 1.15. El cariotipo de la rata de laboratorio. La foto muestra el cariotipo normal de la rata (*Rattus norvegicus*). Se observa un set completo de cromosomas metafásicos de un macho, teñidos con el colorante de Giemsa (bandeo G). A diferencia del ratón, no son todos acrocéntricos, sino que algunos de ellos son metacéntricos. Foto cortesía del Profesor Göran Levan CMB-Genetics, Göteborg University, Suecia.

y la posición del centrómero. Estas bandas son perfectamente reproducibles por las mismas técnicas de “bandeo” y, a la vez, existen varias coloraciones específicas posibles (bandas G, R, C y Q). Para facilitar el trabajo de presentación y comparación entre investigadores se realiza un esquema o dibujo (en base a las bandas obtenidas) denominado **ideograma**. La **Figura 1.16** muestra un ideograma representando las 312 bandas G (Giemsa) que caracterizan el cariotipo normal del ratón de laboratorio y la **Figura 1.17** muestra un ideograma normal de la rata de laboratorio. Con la utilización de diferentes sondas fluorescentes podemos colorear los cromosomas en forma aún más específica, lo que se denomina en inglés *chromosome painting* (ver Capítulo VI). Finalmente, es posible separar cromosomas del ratón según su tamaño utilizando los equipos de **citometría de flujo** (en inglés, *Fluorescence Activated Cell Sorter-FACS*).

En los ratones en estado salvaje existe una enorme variación de cariotipos, ya sea teniendo en cuenta el número de brazos cromosómicos o el número de centrómeros presentes. En algunos ratones salvajes de Europa y del norte de África se encuentra una constelación de “razas” de ratones, presentando cada una un cariotipo particular, aunque todas pertenecen a la misma especie *Mus musculus domesticus* (ver Capítulo III) y todas esas variaciones son consideradas como normales.

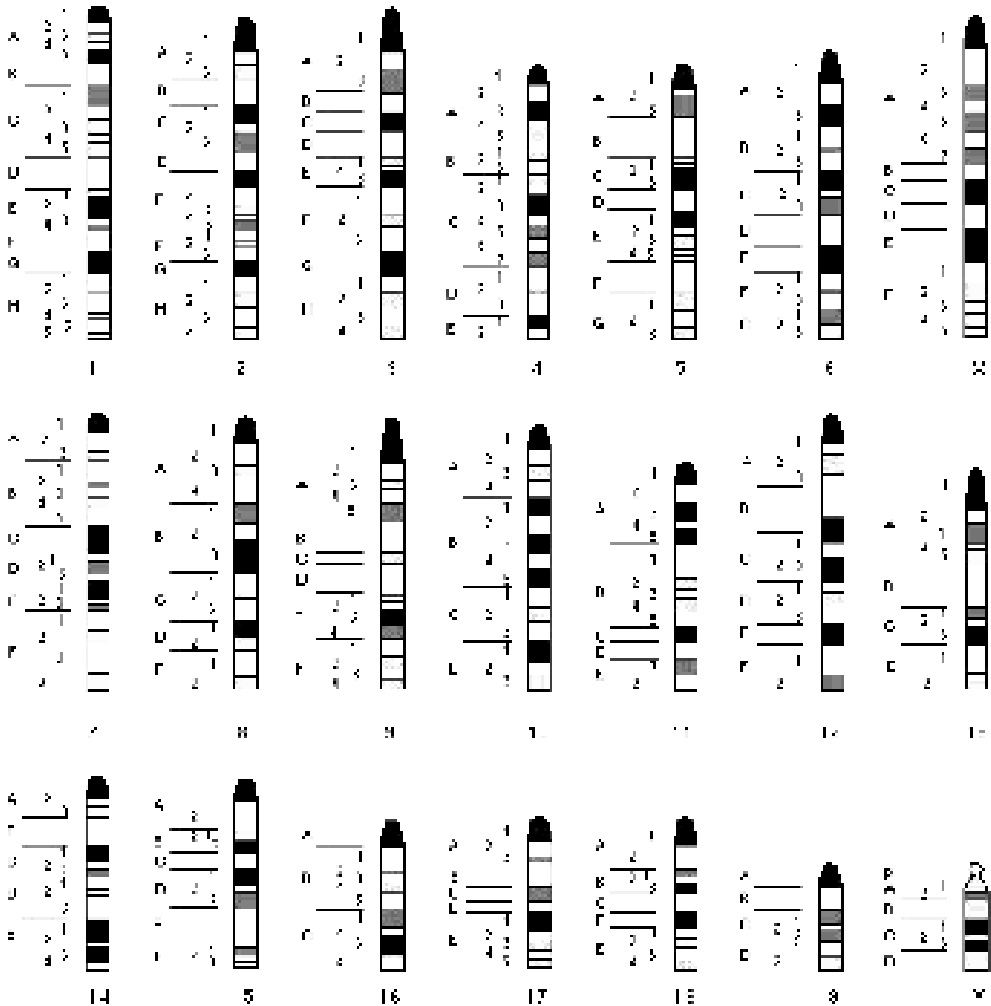
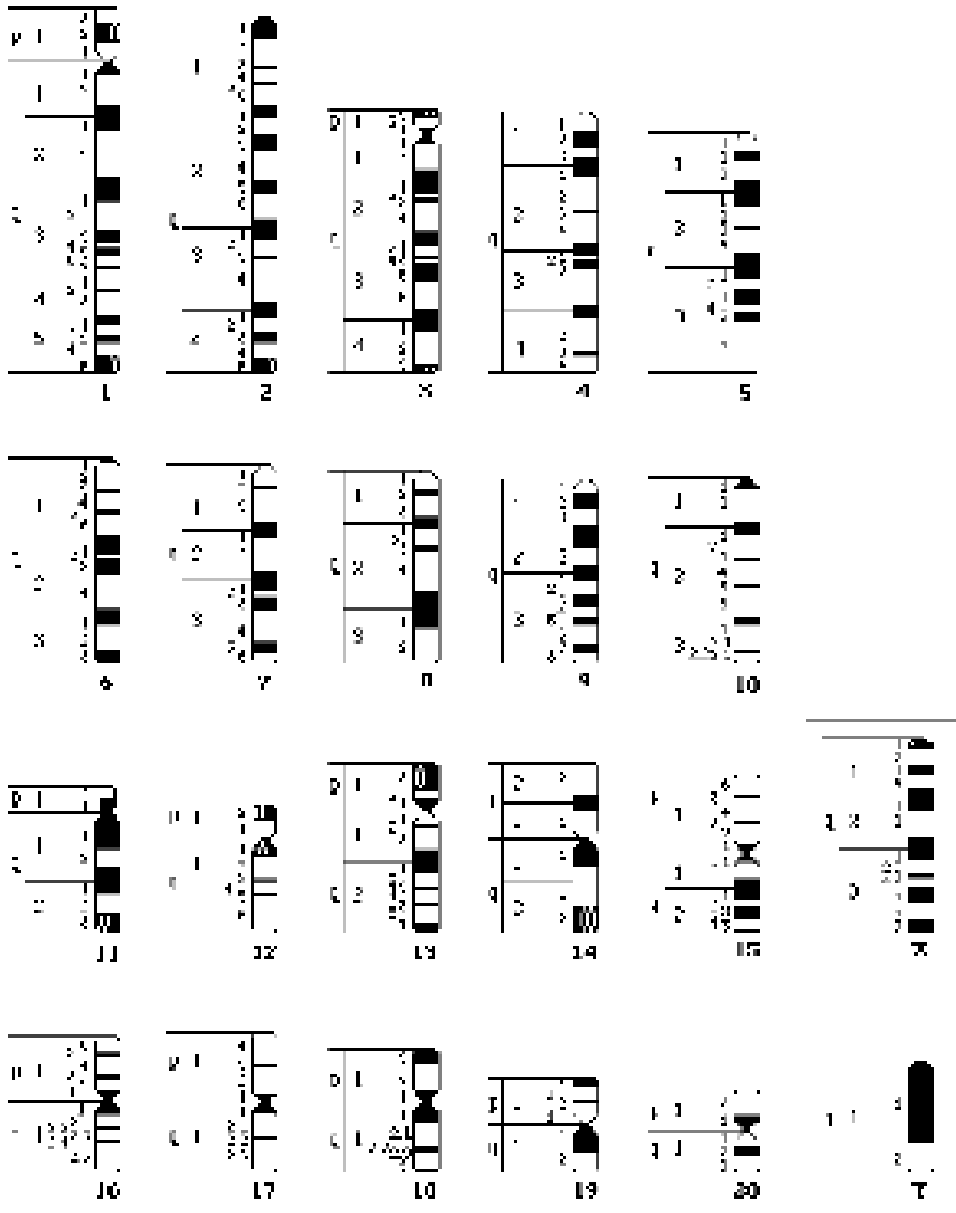


Figura 1.16. El ideograma del ratón de laboratorio. La figura muestra el ideograma del ratón (*Mus musculus*). Lo que se observa son representaciones gráficas de los patrones de bandas G asociados con cada cromosoma en el cariotipo normal. Ilustración cortesía del Dr. Heinz Winking, Medizinische Hochschule Lübeck, Alemania.



Esti ideograma and G-band nomenclature (Levan 1974)

Figura 1.17. El ideograma de la rata de laboratorio. La figura muestra el ideograma de la rata (*Rattus norvegicus*). Lo que se observa son representaciones gráficas de los patrones de bandas G asociados con cada cromosoma en el cariotipo normal. Ilustración cortesía del Profesor Göran Levan CMB-Genetics, Göteborg University, Suecia.

Otra característica sobresaliente de los cromosomas en el ratón de laboratorio es la presencia de **telómeros** llamativamente largos, en comparación con otros mamíferos. Los telómeros son estructuras que protegen los extremos de los cromosomas y están formados por secuencias cortas de ADN (normalmente TTAGGG) repetidas en tandem. Este alargamiento es debido a la presencia de un número mayor de repeticiones (el ratón tendría 10-20 veces más repeticiones que los humanos); lo que según estudios recientes estaría asociado a la consanguinidad de las líneas de ratones de laboratorio. Si bien reconocemos la importancia de los telómeros y la enzima **telomerasa** en la regulación del envejecimiento celular y el desarrollo del cáncer, el tema excede los objetivos de este capítulo.

1.4.2. Las anomalías en el número de cromosomas (Heteroploidías)

1.4.2.1 Las heteroploidías euploides

Las variaciones en el número de cromosomas (esto suele ocurrir en el momento de la formación del cigoto o a veces en la meiosis), pueden hacer pasar el cariotipo normal ($2n$) a un estado anormal o **heteroploide** ($\neq 2n$). Por ejemplo, si el cigoto recibe más de dos juegos completos de cromosomas será $3n$, $4n$ o $6n$, lo que se califica como triploide, tetraploide o hexaploide. Estas anomalías reciben el nombre de **heteroploidías euploides**. Este tipo de anomalías son bastante raras en el ratón, ya que estos embriones mueren *in utero*, pero pueden ser provocadas en el laboratorio por medio de sustancias químicas que, si bien no interfieren en la duplicación del conjunto cromosómico, impiden la migración de los cromosomas hacia los polos de la célula (por ejemplo, **colchicina**). En el año 1999 se descubrió la primera especie de mamífero salvaje, la rata vizcacha roja (*Tympanoctomys barrerae*) de Argentina, que posee un conjunto extra de pares cromosómicos (cariotipo tetraploide).

1.4.2.2 Las heteroploidías aneuploides

Las **heteroploidías aneuploides** son muy frecuentes y se pueden dividir en dos tipos, según la variación numérica sea en más o en menos. Hablamos de **monosomías** cuando en el cariotipo falta un representante de un par cromosómico determinado y de **nulisomías** cuando faltan ambos cromosomas. En cambio, cuando existen uno o dos cromosomas de más correspondientes al mismo par (o sea supernumerarios), se denominan **trisomías** y **tetrasomías**, respectivamente. Es muy difícil que puedan existir más de cuatro copias de un mismo cromosoma en un cariotipo dado. Normalmente, las nulisomías y las monosomías de los cromosomas autosómicos son letales. Las nulisomías suelen ser incompatibles con la segmentación del cigoto y el embrión muere rápidamente. Las monosomías suelen ser letales en el momento de la implantación o un poco después. En cambio, la monosomía del cromosoma X en el ratón es compatible con la vida. Las hembras XO (que representan el análogo de las mujeres afectadas con el **síndrome de Turner**) se reproducen muy pobremente y transmiten una muy baja proporción de gametas nulas para el cromosoma X (< 50%). Las trisomías resultan de fallas en la disyunción cromosómica durante la gametogénesis (o du-

rante las primeras etapas del desarrollo del cigoto), donde los cromosomas de un mismo par migran los dos juntos hacia la misma célula hija. Como resultado, tenemos una gameta disómica y otra nulisómica: la primera producirá un embrión trisómico ($2n + 1$) y la segunda un embrión monosómico ($2n - 1$). Las tetrasomías ($2n + 2$) y las trisomías dobles ($2n + 1 + 1$) son anomalías cromosómicas muy raras.

1.4.3 Las anomalías en la estructura de los cromosomas

Existe una gran variedad de anomalías estructurales del cariotipo. Algunas son casi normales y se caracterizan, por ejemplo, por una modificación de la compactación de la cromatina en la región vecina al centrómero, dando al cromosoma un aspecto alargado muy particular. Otras, más numerosas y por cierto más graves, resultan de la rotura accidental a nivel de los brazos cromosómicos, lo que genera diversas patologías. Con respecto a estas roturas es necesario hacer dos comentarios: (i) Las roturas de los brazos cromosómicos son eventos que la célula tiende a reparar espontáneamente gracias a una maquinaria enzimática muy elaborada y especializada. Por lo tanto, la mayoría de las roturas son reparadas y no pueden ser detectadas. (ii) Las roturas cromosómicas compatibles con la supervivencia de la célula (y por lo tanto susceptibles de ser transmitidas a la descendencia) son accidentes relativamente poco frecuentes.

1.4.3.1 Rearreglos estructurales resultantes de rotura única

Los accidentes que involucran una rotura única en el cromosoma generan rearrreglos llamados **deleciones** (o deleciones terminales). Estas deleciones resultan de una rotura, no reparada, que separa el cromosoma en dos fragmentos: uno desprovisto de centrómero (fragmento que se pierde) y otro, el "muñón", que representa la parte proximal del cromosoma y que continúa comportándose "normalmente" en las divisiones celulares. Las deleciones terminales provocan una modificación brutal del balance génico de la célula, llevando al estado haploide los genes de la región distal perdida. Normalmente este tipo de deleciones son letales.

1.4.3.2 Rearreglos estructurales resultantes de dos roturas

Estos rearrreglos pueden producirse en un mismo cromosoma o, mucho más frecuentemente, sobre dos cromosomas diferentes:

- (i) Sobre el mismo cromosoma. Pueden ser de dos tipos, según exista, o no, pérdida de substancia. Cuando se producen dos roturas sobre el mismo cromosoma y la reparación se hace luego de la pérdida de un segmento intermedio se trata de una **delección intercalar**. Los efectos deletéreos serán los mismos que aquellos asociados a las deleciones terminales, aunque se conocen algunas deleciones de hasta 30 cM de largo en los cuales los animales sobreviven y hasta son fértiles. En el ratón se conocen muchas deleciones intercalares y además pueden ser inducidas a voluntad en el laboratorio irradiando los testículos (o inyectando el agente alquilante cloranbucilo) de los progenitores algunos días antes

del acoplamiento. Estas deleciones experimentales son muy importantes para los genetistas ya que permiten estudiar en detalle pequeñas regiones del genoma. En el caso de que se produzcan dos roturas sobre el mismo cromosoma pero la reparación se haga sin pérdida de substancia tenemos dos posibilidades: reparación *ad integrum*, donde hay restitución de la estructura normal del cromosoma y por lo tanto ninguna consecuencia, y reparación con giro de 180 grados, formando una nueva estructura denominada **inversión**. Las inversiones son frecuentes en todos los mamíferos e incluso participan en los mecanismos de la evolución. Si la inversión implica un segmento que porta el centrómero se habla de una **inversión pericéntrica** y, en el caso que no comprenda al centrómero, de una **inversión paracéntrica**. Dado que en las inversiones no hay pérdida de material genético son compatibles con la vida de los individuos.

- (ii) Sobre cromosomas diferentes. Los únicos rearrreglos observables (e interesantes para estudiar) serán aquellos que resultan del intercambio de segmentos entre cromosomas diferentes. Estos se denominan **translocaciones** y existen tres tipos en función de los rearrreglos producidos. Las **translocaciones recíprocas** son consecuencia de un intercambio recíproco entre partes terminales de dos pares de cromosomas diferentes y constituyen, por lejos, el rearrreglo más frecuentemente observado. En la mayoría de los casos se ve modificada la morfología del cromosoma y tienen consecuencias profundas sobre la gametogénesis del portador. Por definición, no implica ninguna pérdida de material genético pero en algunos casos de cariotipo no balanceado son letales *in utero* en un estadio temprano del desarrollo. Las **translocaciones en tandem** son translocaciones en donde se forma un cromosoma muy largo y uno complementario muy corto a partir de dos cromosomas de longitud normal. Se trata simplemente de un caso límite de translocación recíproca donde el cromosoma pequeño se pierde en el curso de las divisiones celulares ya que, en general, está constituido sólo por heterocromatina sin función genética vital. Finalmente, las **translocaciones Robertsonianas** (o **fusiones céntricas**) representan un caso de rearrreglo estructural que podemos interpretar como de doble rotura en dos cromosomas (aunque existen otras interpretaciones de este fenómeno). Una fusión céntrica es un rearrreglo que transforma dos cromosomas acrocéntricos en dos cromosomas metacéntricos: uno de brazos grandes y el otro de brazos diminutos. En este caso, como en las inversiones en tandem, el cromosoma pequeño se pierde con las sucesivas divisiones celulares y el cariotipo pierde una unidad (un centrómero). Las translocaciones Robertsonianas son muy frecuentes en la naturaleza, especialmente entre los roedores del género *Mus* (**Figura 1.14**). Entre los ratones de la especie *Mus musculus domesticus* (ver Capítulo III) existen inclusive grupos aislados de animales que presentan un cariotipo con casi todos sus cromosomas en estado metacéntrico, resultado de todas las combinaciones de cromosomas acrocéntricos posibles, a excepción del Y que nunca fue encontrado en fusiones céntricas. Finalmente, debemos resaltar que estos rearrreglos cromosómicos son una gran herramienta para aquellos investigadores que trabajan en genética del ratón.

1.4.3.3 Rearreglos estructurales complejos

El único rearrreglo estructural que implica tres roturas es la **transposición segmentaria** o inserción. Se caracteriza por la transposición de un segmento proveniente de un cromosoma determinado, entre las dos extremidades de una rotura efectuada sobre otro cromosoma. Existen muy pocos rearrreglos de este tipo.

Aquellos lectores interesados en modelos de citogenética pueden consultar el *Cytogenetics Models Resource* del *Jackson Laboratory*, dirigido por Muriel T. Davisson (<http://www.jax.org/resources/documents/cyto/>).

Bibliografía General

- AYALA FJ, KIGER JA JR. *Genética Moderna*. Omega, 1984.
- BALMAIN A. *Cancer genetics: from Boveri and Mendel to microarrays*. *Nature Reviews (Cancer)* 1: 77-82, 2001.
- BAXEVANIS AD. *The Molecular Biology Database Collection: 2003 update*. *Nucleic Acids Research* 31: 1-12, 2003.
- BERGER SL, KIMMEL AK. *Guide to Molecular Cloning*. *Methods in enzymology* 152, 1987.
- BIRREN B, GREEN E, HIETER P, KLAPHOLZ S., MYERS R, RIEBETHMAN H, ROSKAMS J. (Eds.) *Genome Analysis: A Laboratory Manual. Vol 1, 2, 3 y 4*. Cold Spring Harbor Laboratory Press, New York, 1997-1999.
- BRETT D, POSPISIL H, VALCARCEL J, REICH J, BORK P. *Alternative splicing and genome complexity*. *Nature Genetics* 30: 29-30, 2002.
- BROWN SDM. *Whither the mouse genome?* *Mammalian Genome* 13: 481-482, 2002.
- EPSTEIN C. *The Consequences of chromosome imbalance: Principles, mechanisms and models*. Cambridge University Press, Cambridge, 1986.
- GREEN E. *The human Genome Project and its impact on the study of human disease*. En: "Metabolic and Molecular Bases of Inherited Disease", 8th edition. Scriver CR, Sly WS, Beaudet AL, Valle D, Childs B, Vogelstein B (Eds.). McGraw Hill Press, New York, 2000.
- GREENFIELD A. *Applications of DNA microarrays to the transcriptional analysis of mammalian genomes*. *Mammalian Genome* 11: 609-13, 2000.
- GRIFFITHS AJF, GELBART WM, MILLER JH, LEWONTIN, RC. *Genética Moderna*. McGraw Hill Interamericana, 2000.
- GUÉNET J-L. *The Mouse Genome*. En: "Genomes, Molecular Biology and Drug Discovery", pp. 27-51. Academic Press Ltd., New York, 1996.
- HARTL D, JONES E. *Essential Genetics, 2nd edition*. Jones and Bartlett Publishers, Boston, 1999.
- HOWE CJ, WARD ES. *Nucleic Acids Sequencing: A Practical Approach*. IRL Press, Oxford University, 1989.
- INNIS DH, GELFAND JJ, SNIINSKY. *PCR applications. Protocols for functional genomics (Eds.)* MA Academic Press, 1999.
- KAPLAN JC, DELPECH M. *Biologie Moléculaire et Médecine (2 Ed)*. Médecine-Sciences Flammarion, Paris, 1995.
- KLUG, WS, CUMMINGS, MR. *Conceptos de Genética. Quinta Edición*. Prentice Hall, 1999.
- LACADENA, JR. *Genética General*. Conceptos Fundamentales, 1999.
- LEWIN, B. *Genes VII*. Oxford University Press, 2000.
- LYON MF, RASTAN S, BROWN SD (Eds.). *Genetic Variants and Strains of the Laboratory Mouse. 3rd edition*. Oxford University Press, 1996.

- MANNING EL, CROSSLAND J, DEWEY MJ, VAN ZANT G. *Influences of inbreeding and genetics on telomere length in mice*. *Mammalian Genome* 13: 234-238, 2002.
- MASER RS, DEPINHO RA. *Connecting chromosomes, crisis, and cancer*. *Science* 297: 565-569, 2002.
- MCPHERSON MJ, HAMES BD, TAYLOR GR (Eds.). *PCR 2. A Practical Approach*. Oxford University Press, New York, 1995.
- MILLER OJ. *Structure and Organization of Mammalian Chromosomes: Normal and Abnormal*. En: "Medical and Experimental Mammalian Genetics: A Perspective", pp 19-63. ALAN R. Liss, Inc., New York, 1987.
- MODREK B, LEE C. *A genomic view of alternative splicing*. *Nature Genetics* 30: 13-19, 2002.
- MULLIS K. *The polymerase chain reaction*. Birkhauser Press, 1994.
- PERUSKI LF JR., PERUSKI AH. *The Internet and the New Biology: Tools for Genomic and Molecular Research*, ASM, Washington DC, 1997.
- PUERTAS GALLEGUO MJ. *Genética. Fundamentos y Perspectivas*. 2ª Edición. McGraw Hill Interamericana, 1999.
- SAIKI RK, SCHARF SJ, FALOONA F, MULLIS KB, HORN GT, ERLICH HA Y ARNIHEIM N. *Enzymatic amplification of b-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia*. *Science* 230, 1350-1354, 1985.
- SAIKI RK, GELFAND DH, STOFFEL S, SCHARF SJ, HIGUCHI R, HORN GT, MULLIS KB Y ERLICH HA. *Primed-directed enzymatic amplification of DNA with a thermostable DNA polymerase*. *Science* 239, 487-494, 1988.
- SAMBROOK J, FRITSCH EF, MANIATIS T. *Molecular Cloning: A laboratory manual*, 1989.
- SCCHULTZ-SCHAEFFER J. *Cytogenetics*. Springer Verlag, New York, 1980.
- SCHENA M. (Ed). *DNA MICROARRAYS. A Practical Approach*. Oxford University Press, Oxford, 1999.
- SILVER LM. (Ed). *MOUSE GENETICS. Concepts and applications*. Oxford University Press, Oxford, 1995. (Versión online disponible en <http://www.informatics.jax.org/silver/>)
- STORZ G. *An expanding universe of noncoding RNAs*. *Science* 296: 1260-1263, 2002.
- STRACHAN T, READ AP. *Genética Molecular Humana*. Omega, Barcelona (1999) [traducción de: Strachan T, Read, AP. *Human Molecular Genetics*. BIOS, Oxford (1996)]
- STRACHAN T, READ AP. *Human Molecular Genetics 2*. BIOS Scientific Publishers Ltd, Oxford, 1999.
- STRICKBERGER MW. *Genética*. Omega, 1988.
- STURTEVANT AH. *A History of Genetics*. Cold Spring Harbor Laboratory Press, NY, 2001.
- SUZUKI DT, GRIFFITHS AJF, MILLER JH, LEWONTIN RC. *Genética*. Interamericana McGraw Hill, 1992.
- TAMARIN RH. *Principios de Genética*. Editorial Reverté, S.A, 1996.
- WATSON J, HOPKINS N, ROBERTS J, STEITZ J, WEINER A (Eds.). *Molecular Biology of the Gene*, 4th edition. The Benjamin-Cummings Publishing Company, Menlo Park, California, 1987.
- WHEELER DL, CHURCH DM, LASH AE, LEIPE DD, MADDEN TL, PONTIUS JU, SCHULER GD, SCHRIML LM, TATUSOVA TA, WAGNER L, RAPP BA. *Database resources of the National Center for Biotechnology Information: 2002 update*. *Nucleic Acids Research* 30: 13-16, 2002.
- WINTER PC, HICKEY GI, FLETCHER HL. *Instant Notes in Genetics*. 2nd edition. BIOS Scientific Publishers, 2002.