

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 630 804**

51 Int. Cl.:

**C12Q 1/68**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.11.2012 PCT/EP2012/072258**

87 Fecha y número de publicación internacional: **16.05.2013 WO13068528**

96 Fecha de presentación y número de la solicitud europea: **09.11.2012 E 12784585 (7)**

97 Fecha y número de publicación de la concesión europea: **22.02.2017 EP 2776574**

54 Título: **Métodos para determinar repeticiones de secuencias de nucleótidos**

30 Prioridad:

**10.11.2011 GB 201119390**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**24.08.2017**

73 Titular/es:

**DNAME-IT (100.0%)  
Gaston Geenslaan 1  
3001 Leuven, BE**

72 Inventor/es:

**CUPPENS, HARRY**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 630 804 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Métodos para determinar repeticiones de secuencias de nucleótidos

### Campo de la invención

5 La presente invención se refiere a métodos para determinar el número de nucleótidos en tramos de homonucleótidos.

La presente invención se refiere al campo de los denominados métodos de secuenciación de próxima generación.

### Antecedentes de la invención

10 El ADN (ácido desoxirribonucleico) es el portador universal de la información genética. El ADN es una hélice entrelazada de dos cadenas poliméricas, cada cadena constituida por unidades de nucleótido unidas a un esqueleto de azúcares de desoxirribosa y grupos fosfato unidos por enlaces éster. Estas dos cadenas discurren en direcciones opuestas, anti-paralelas. Cada cadena de ADN está constituida por 4 nucleótidos A, C, G y T, en un orden específico para esa molécula de ADN. Es la secuencia de estas cuatro bases a lo largo del esqueleto la que codifica la información genética.

15 Las dos cadenas de ADN tienen una naturaleza complementaria. Un nucleótido A forma un par de bases con un nucleótido T en la cadena opuesta, y viceversa; un nucleótido G forma un par de bases con un nucleótido C en la cadena opuesta, y viceversa.

20 En células eucariotas, el ADN se transcribe a ARN (ácido ribonucleico). Las moléculas de ARN son bastante similares a las moléculas de ADN, las cadenas individuales de los nucleótidos están unidas a un esqueleto de azúcares de ribosa y fosfato. Existen diferentes tipos de moléculas de ARN dependiendo de su función. Las moléculas de ARNm se usan por organismos celulares para transportar la información genética codificada por el ADN para dirigir la síntesis de proteínas. En algunos virus, el ARN se usa incluso como el código genético en lugar del ADN.

25 El ADN puede replicarse por las ADN polimerasas. Una ADN polimerasa solo puede extender una cadena de ADN existente emparejada con una cadena molde. No puede iniciar la síntesis de una nueva cadena como tal. Para comenzar la síntesis, se debe crear un fragmento corto de ADN, un oligonucleótido o molécula de ARN, llamado cebador, y emparejarse con la cadena de ADN molde.

30 La ADN polimerasa sintetiza una nueva cadena de ADN extendiendo el extremo 3' de una cadena de nucleótidos existente, adicionando nuevos nucleótidos a la cadena molde de uno en uno a través de la creación de enlaces fosfodiéster. Los bloques de construcción entrantes son los nucleótidos trifosfato (dNTPs: dATP, dCTP, dGTP, dTTP). El oxígeno del extremo 3'-hidroxilo de la cadena de ADN en crecimiento realiza un ataque nucleófilo sobre el fosfato alfa (el más cercano al azúcar) del dNTP. El resultado es que el dNMP (desoxirribonucleósido monofosfato, o nucleótido) queda unido covalentemente al carbono 3' del azúcar en el extremo de la cadena de ADN, alargando así la cadena por un nucleótido. Por otra parte, se libera pirofosfato y un protón (Figura 1). A continuación, el proceso se repite.

35 Existe un enorme interés en la determinación de la información genética del ADN, como el nucleótido que se encuentra en una posición dada, el orden de secuencia que se encuentra en un determinado locus/loci en el genoma, o incluso el genoma completo. Incluso se puede secuenciar el ARN cuando se convierte primero en cDNA. La información genética se determina mediante tecnologías de secuenciación, como secuenciación de Maxim-Gilbert, secuenciación de Sanger y derivados de las mismas, pirosecuenciación en paralelo (Roche 454 Life Sciences), secuenciación reversible basada en terminador por síntesis (Illumina), Secuenciación por Ligación y Detección de Oligonucleótidos (SOLiD) (Life Technologies), Secuenciación de Ión Semiconductor (Ion Torrent, Life Technologies), secuenciación de Única Molécula a Tiempo Real (SMRT) basada en propiedades de guías de ondas en modo cero (Pacific Biosciences), sensores de nanoporos (Oxford Nanopore Technologies), etc. Dependiendo de la sensibilidad de muchas tecnologías de secuenciación, se secuencian grupos (clones) de moléculas de ADN idénticas en paralelo. La secuenciación de las cadenas de ADN individuales en paralelo solo es posible mediante secuenciación de Única Molécula a Tiempo Real y sensores de nanoporos. En la mayoría de las tecnologías de secuenciación, se desnaturaliza una molécula de ADN de doble cadena y se secuencian entonces una de estas moléculas de ADN de cadena sencilla. En esencia, esta cadena de ADN solo se usa como molde en una reacción de secuenciación para la síntesis de una segunda cadena de ADN complementaria, en base a la naturaleza complementaria del ADN. Se puede sintetizar una nueva cadena de ADN cuando un oligonucleótido se une al ADN molde. Este oligonucleótido es un cebador para la extensión adicional de una nueva cadena de ADN en crecimiento por incorporación de nucleótidos según el principio de complementariedad. Tales oligonucleótidos son normalmente aproximadamente 10-25 nucleótidos de longitud y se pueden sintetizar fácilmente. Mediante monitorización de la síntesis de esta nueva cadena de ADN, es decir, se puede determinar el orden en el que se incorporan los nucleótidos en la nueva cadena de ADN, la secuencia de ADN de esa cadena. Dada la naturaleza complementaria de las dos cadenas de ADN, entonces se puede conocer también la secuencia de la otra cadena de ADN original.

A pesar del progreso en las técnicas de secuenciación, no se puede determinar con precisión la secuencia del número de nucleótidos en tramos de homonucleótidos con ciertas tecnologías de secuenciación de nueva generación. Por ejemplo, la pirosecuenciación se inventó en la década de los noventa, la secuenciación altamente en paralelo se introdujo en 2005, mientras que la secuenciación de Ión Semiconductor se introdujo solo en 2009. La dificultad encontrada de denominación inexacta de tramos de homonucleótidos ya se conoce desde hace casi dos décadas.

### Resumen de la invención

En esta invención, la longitud de los tramos más largos de homonucleótidos se reduce a una serie de partes de nucleótidos más cortos; o bien ambas partes son tramos de nucleótidos más cortos, o una parte es un tramo de nucleótidos más corto y la segunda parte es sólo de 1 nucleótido de largo. Por ejemplo, cuando un tramo de 8 residuos T se modifica en la séptima posición, el tramo de nucleótidos más corto anteriormente indicado contiene 6 nucleótidos T y la T en la posición octava representa la mencionada "segunda parte de solamente 1 nucleótido de largo" anterior. Ambas partes de nucleótidos más cortas se pueden determinar con más precisión, y el análisis combinado preciso de la parte más pequeña permite la determinación más precisa de la longitud del tramo de homonucleótidos original más largo.

Un primer aspecto de la invención se refiere a métodos de generar una o más copias de una molécula de polinucleótido objetivo, o parte del mismo, como se define en la reivindicación 1, que contiene una repetición de nucleótidos idénticos. Estos métodos comprenden las etapas de:

- alterar un nucleótido en esta repetición de nucleótidos idénticos, o
- alterar diferentes nucleótidos individuales separados a intervalos en dicha repetición de nucleótidos idénticos en otro nucleótido,

con el fin de dividir dicha repetición de nucleótidos idénticos en dos o más partes alteradas más pequeñas de nucleótidos idénticos en las moléculas copia, en donde se usa un oligonucleótido para este fin que se extiende en la repetición no alterada de nucleótidos idénticos, ya sea hasta el final de la repetición no alterada de los nucleótidos idénticos o no, y en donde dicho oligonucleótido cebador no es 100% complementario con la secuencia repetida en la secuencia objetivo.

En otras palabras, como se ilustra en la figura 4, el cebador que se usa hibrida con una parte del tramo de homonucleótidos en la molécula objetivo, pero no hibrida completamente con el tramo de homonucleótidos en la molécula objetivo. Dentro de la porción del oligonucleótido cebador que hibrida con el tramo de homonucleótidos en la molécula objetivo, se incorpora al menos una discordancia, en comparación con el tramo de homonucleótidos en la molécula objetivo.

El polinucleótido es ADN. En estos métodos, las partes alteradas más pequeñas de nucleótidos idénticos se generan a través del uso de un cebador en una reacción enzimática, como síntesis, ligación y/o amplificación de ADN.

En estos métodos, un nucleótido en la repetición de los nucleótidos idénticos complementarios del cebador se sustituye por otro tipo de nucleótido de manera que la repetición no alterada de un número consecutivo de nucleótidos idénticos se divide en dos partes alteradas más cortas de estos nucleótidos idénticos interrumpidas por el tipo de nucleótido sustituido.

En ciertas realizaciones de estos métodos la parte de nucleótidos idénticos en el extremo 3' del cebador es más corta que la parte de nucleótidos idénticos en el extremo 5' del cebador.

En ciertas realizaciones de estos métodos, se sustituyen varios nucleótidos individuales en el cebador en la repetición de nucleótidos idénticos complementarios a intervalos regulares, de modo que todas las partes más cortas de nucleótidos idénticos no superan una longitud dada, por ejemplo, no más de 5, 6 o 7 nucleótidos.

En otras realizaciones de estos métodos la parte más corta obtenida de nucleótidos idénticos no es más de 4-6 nucleótidos.

En realizaciones particulares, el cebador contiene además en el extremo 5' uno o más nucleótidos adaptadores que no son complementarios con la secuencia objetivo. Estos nucleótidos adaptadores se pueden usar para discriminarlos de otro tipo de fragmentos y/o para el procesamiento posterior como amplificación, secuenciación, inclusión de secuencias de códigos de barras.

En ciertas realizaciones de estos métodos se combinan diferentes reacciones de síntesis/ligación/amplificación de ADN del mismo tipo, o de diferentes tipos, en un formato (tipo) multiplex.

Como se define en la reivindicación 13, un segundo aspecto de la invención se refiere a un método para determinar el número de nucleótidos en una repetición de nucleótidos de una molécula de ADN molde en la que se determina la secuencia de una molécula de ADN copiada mediante secuenciación, y que usa el conocimiento previo de los nucleótidos alterados y las posiciones generadas en las moléculas copia como se describe en los métodos citados

anteriormente del primer aspecto, que comprende la etapa de contaje del número de nucleótidos idénticos en las partes más pequeñas generadas y de nucleótidos sustituidos para generar las partes más pequeñas de nucleótidos idénticos.

5 Contar es más importante para la última parte más corta, no necesario para la otra parte donde se puede asumir porque se proporciona por el cebador para el que ya se sabe el recuento exacto, incluso si es más largo de 7 nucleótidos. Estos métodos comprenden además de modo opcional la etapa de realizar un método para determinar la naturaleza cualitativa del tramo de nucleótidos idénticos y sus regiones de ADN corriente abajo y corriente arriba, en donde los fragmentos obtenidos a partir de cada método son discriminados y analizados por separado usando dichas secuencias adaptadoras.

10 Como se define en la reivindicación 14, un aspecto adicional de la invención se refiere a un soporte de datos que comprende instrucciones de programación para analizar y proporcionar los resultados del método descrito anteriormente, cuando se ejecuta en un ordenador.

15 Como se define en la reivindicación 15, un aspecto adicional de la invención se refiere al uso de uno o más oligonucleótidos en una reacción de síntesis, ligación, o amplificación de ADN, para reducir repeticiones más largas de nucleótidos idénticos en dos o más partes más pequeñas de nucleótidos idénticos en un método como se describió anteriormente.

Leyendas de figura

Figura 1: Síntesis de una nueva cadena de ADN. Se libera un pirofosfato del nuevo nucleótido entrante, y se libera un protón de la cadena de ADN que se extiende.

20 Figura 2: Diferencias de señal absolutas y relativas obtenidas en ensayos como una pirosecuenciación y Secuenciación de Ión Semiconductor de diferentes tramos de homonucleótidos.

Figura 3: A. PCR estándar en la que los cebadores, indicados por flechas, flanquean la región de interés, en este caso una región que contiene un tramo de homonucleótidos. [técnica anterior] B. hnr-PCR, de acuerdo con una realización de métodos de la presente invención, en la que uno de los cebadores se extiende en el tramo de homonucleótidos de interés. Este cebador hnr contiene una o más discordancias en el (sitio complementario del) tramo de homonucleótidos.

25

Figura 4.

A. hnr-PCR, de acuerdo con una realización de métodos de la presente invención, en la que uno de los cebadores se extiende en el tramo de homonucleótidos de interés.

30 B. Ejemplo de hnr-PCR, de acuerdo con una realización de métodos de la presente invención, en la que el tramo de homonucleótidos contiene un tramo de 9 residuos T. El cebador de hnr-PCR se extiende 5 nucleótidos en el tramo de homonucleótidos, y contiene una discordancia en la posición 4 en el sitio del tramo de homonucleótidos. Después de la síntesis de una nueva cadena, se obtendrá un tramo de homonucleótidos de 3 residuos T y 5 residuos T, separados por un nucleótido A. Cuando después de secuenciar se encuentran 5 nucleótidos T en el último tramo 3', el tramo original contenía 9 nucleótidos T; si se encuentran 4 nucleótidos T en el último tramo, el tramo original contenía 8 residuos T, etc.

35

Figura 5: Ejemplo, de acuerdo con una realización de métodos de la presente invención como se muestra en la figura 4, en donde se añade un marcador de secuencia (una secuencia de nucleótidos dada) a los cebadores como adaptadores. Cuando se usan ambos estándares y hnr-PCR en un formato de PCR (tipo) multiplex, los cebadores estándar de PCR (uno o ambos) pueden contener otro tipo de adaptador, mientras los cebadores usados en hnr-PCR (uno o ambos) pueden contener otro tipo de marcador. Si se genera más de un amplicón en una PCR (tipo) multiplex, los marcadores de los cebadores usados para la generación de amplicones "estándar" pueden ser idénticos o no. Si se genera más de un amplicón hnr-PCR en un formato de PCR (tipo) multiplex, los marcadores de los cebadores hnr pueden ser idénticos o no. De esta manera todos los amplicones se pueden combinar y separar para el análisis.

40

45

Figura 6: Ejemplo de acuerdo con una realización de métodos de la presente invención como se muestra en la figura 5, en donde una o más secuencias adaptadoras o ligadas a los cebadores para su procesamiento posterior, como un sitio cebador para una segunda síntesis o PCR, secuenciación, código de barras, de ADN, etc., o una combinación de los mismos.

## 50 Descripción detallada de la invención

Las diferentes tecnologías de secuenciación usan diferentes tecnologías para determinar la secuencia de nucleótidos de una cadena de ADN. Una característica común de las actuales tecnologías de secuenciación de alto rendimiento es que muchas cadenas de ADN se secuencian en paralelo, en su mayoría en la superficie de una pequeña placa (placa de picotitulación, célula de flujo), dando hasta más de 600 Gb (giga bases) de información de

secuencia. Normalmente, estos fragmentos de ADN llevan terminaciones de secuencia idénticas, p. ej. a través de la unión de pequeños adaptadores de nucleótidos durante la fase de preparación del ADN para secuenciar, de manera que se puede usar el mismo cebador para la secuenciación de todos los fragmentos de ADN en paralelo.

5 Las actuales tecnologías de secuenciación de alto rendimiento se pueden dividir en líneas generales en dos categorías según qué secuenciación se realiza; secuenciación por síntesis (Illumina, Roche, Ion Torrent, Helicos y Pacific Biosciences) y secuenciación por ligación (SOLiD, life Technologies).

La secuenciación por síntesis puede además dividirse en dos categorías distintas. En la primera subcategoría, cada nucleótido incorporado se detecta como tal, como en el enfoque del terminador reversible en el que cada nucleótido está marcado con un fluorocromo diferente y se analiza al mismo tiempo una elongación de una única base con cualquier base. Solo es posible una incorporación de una única base en cada ciclo/etapa de la reacción de secuenciación. De hecho, los nucleótidos están protegidos, de manera que pueden incorporarse en una cadena de ADN en crecimiento, pero una vez incorporados, no se puede incorporar otro nucleótido en el mismo ciclo/etapa. En un nuevo ciclo, el nucleótido incorporado del ciclo anterior se desprotege primero, de manera que se pueda incorporar un nuevo nucleótido protegido en el nuevo ciclo. En la segunda subcategoría, se monitoriza un producto secundario de la incorporación de nucleótidos, como pirofosfato en la pirosecuenciación o un protón en la Secuenciación de Ión Semiconductor. En las últimas dos estrategias, los nucleótidos nativos individuales se añaden secuencialmente. La Pirosecuenciación o la Secuenciación de Ión Semiconductor se produce después de la progresión de la ADN polimerasa a lo largo de una cadena de ADN permitiendo que sólo esté disponible un único dNTP para su incorporación en un ciclo dado, y luego aproveche la reacción química que se produce cuando el dNTP es incorporado por la polimerasa. Esta reacción se detecta ya sea induciendo una cascada de bioluminiscencia empezando desde el pirofosfato y detectando la luz emitida (pirosecuenciación), o detectando directamente protones liberados durante la incorporación como un cambio en el pH (Secuenciación de Ión Semiconductor). Se inicia entonces un nuevo ciclo para otro dNTP. En consecuencia, los cuatro dNTPs diferentes se añaden por separado en cuatro ciclos diferentes, y esto se repite para un número dado de rondas, en la que cada ronda se construye con el mismo orden de 4 ciclos de administración de los 4 diferentes dNTPs.

Más específicamente para la pirosecuenciación, se hibrida un cebador de secuenciación con un ADN molde de cadena sencilla y se incuba con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa, los sustratos adenosina 5' fosfosulfato (APS) y luciferina. La adición de uno de los cuatro desoxinucleótido trifosfatos (dNTPs) (dATP no se usa, sino más bien dATP $\alpha$ S que no es un sustrato para la luciferasa) inicia la segunda etapa. La ADN polimerasa incorpora solamente el dNTP dado si es complementario con el molde. Esta incorporación libera estequiométricamente pirofosfato (PPi). Luego la ATP sulfurilasa convierte cuantitativamente PPi en ATP en presencia de adenosina 5' fosfosulfato. Este ATP actúa como combustible en la conversión mediada por luciferasa de la luciferina a oxiluciferina que genera luz visible en cantidades que son proporcionales a la cantidad de ATP. La luz producida en la reacción catalizada por la luciferasa se detecta mediante una cámara y se analiza en un programa. Los nucleótidos no incorporados y el ATP se degradan a continuación por la apirasa, y la reacción puede reiniciarse con otro nucleótido. Los cuatro nucleótidos de ADN se añaden secuencialmente en un orden fijo a lo largo de la placa de picotitulación. Durante el flujo de nucleótidos, se secuencian en paralelo miles de copias de ADN unidas a cada una de las perlas. Cuando se añade a un pocillo un nucleótido complementario a la cadena molde, la polimerasa extiende así la cadena de ADN existente añadiendo nucleótido(s). En el caso de que el molde lleve un tramo de nucleótidos idénticos, se puede incorporar un número de nucleótidos igual a la longitud del tramo en el mismo ciclo/etapa. De hecho, se usan dNTPs nativos no protegidos, de manera que se puede incorporar más de un nucleótido en un ciclo dado. La intensidad de señal es proporcional al número de nucleótidos. Sin embargo, la intensidad de tramos de homopolímeros es solamente lineal hasta siete nucleótidos consecutivos. La secuencia de soluciones que producen señales quimioluminiscentes permite al final la determinación de la secuencia del molde.

Más específicamente para la Secuenciación de Ión Semiconductor, se inunda un micropocillo que contiene cadenas de ADN molde clonal en una perla con una sola especie de desoxirribonucleótido (dNTP). Si el dNTP introducido es complementario al molde principal, se incorpora el nucleótido en la cadena complementaria en crecimiento. Esto provoca la liberación de un ión de hidrógeno que dispara al sensor de iones hipersensible, lo que indica que se ha producido una reacción. Si el dNTP introducido no es complementario al nucleótido molde principal, no se incorpora ningún nucleótido y no se liberan iones hidrógeno. El chip se inunda secuencialmente con un nucleótido tras otro. Aquí de nuevo, puesto que se usan dNTPs nativos no protegidos, se incorporarán en un único ciclo múltiples moléculas de dNTP en los tramos de homonucleótidos. Esto conduce a un número correspondiente de hidrógenos liberados y a una señal electrónica proporcionalmente mayor. Aquí de nuevo, la exactitud de la denominación correcta del número de nucleótidos en tramos de homonucleótidos disminuye para tramos de homonucleótidos más largos.

La tecnología de Secuenciación de Ión Semiconductor difiere de otras tecnologías de secuenciación en que no se usa ninguna óptica. La secuenciación de Ión Semiconductor también puede denominarse secuenciación de Ion Torrent, secuenciación mediada por pH, secuenciación de silicio o secuenciación de semiconductores. La Secuenciación de Ión Semiconductor crea una conexión directa entre los mundos químico y digital, permitiendo una secuenciación rápida, simple y masivamente escalable. Los chips de Secuenciación de Ión Semiconductor están diseñados, fabricados y empaquetados como cualquier otro chip de semiconductores. Las placas se cortan en una bola de silicio. Los transistores y circuitos se transfieren entonces por patrones y posteriormente se graban en las

placas usando fotolitografía. Este proceso se repite 20 veces o más, creando un sistema multicapa de circuitos. La Secuenciación de Ión Semiconductor se beneficia de cuatro décadas de mejora exponencial en la tecnología de semiconductores, también conocida como Ley de Moore.

5 La pirosecuenciación y la secuenciación de Ión Semiconductor permite que la secuenciación continúe a una velocidad mucho más rápida que p. ej. la secuenciación de terminador reversible, ya que se requieren menos pasos para detectar una base y continuar la extensión de un molde. Como tal, es común conseguir actualmente lecturas de 100 pares de bases en menos de 3 h con pirosecuenciación o Secuenciación de Ión Semiconductor. Además, se pueden obtener lecturas de secuencias más largas con pirosecuenciación o Secuenciación de Ión Semiconductor, de las que se pueden generar los moldes mediante amplificación clásica por PCR. Esto es altamente deseable en diagnóstico, ya que la PCR es mucho menos intensiva en mano de obra que otras tecnologías de preparación de molde como la construcción de bibliotecas de ADN. Otra ventaja en la preparación del molde de secuenciación por PCR es que la longitud estándar de los amplicones obtenidos, así como las longitudes de lectura obtenidas con estas tecnologías de secuenciación, están en el intervalo de tamaño, o sobre el tamaño, de la longitud media de los exones. Por lo tanto el número de amplicones necesarios para la secuenciación de un gen dado es del mismo orden de los exones en la mayoría de los genes, de modo que la mayoría de los amplicones cubren exones individuales. Cuando se realiza en un formato multiplex, todos los exones de interés pueden enriquecerse en uno o un número limitado de pasos. Esto también se puede conseguir en un formato tipo multiplex, como la amplificación altamente paralela o los circuitos fluidicos integrados (Fluidigm). De hecho, aunque cada reactor de PCR del circuito fluidoico integrado genera solamente un tipo de amplicón, la naturaleza paralela del circuito fluidoico integrado logra y asegura un formato tipo multiplex.

Estos formatos de secuenciación, especialmente la Secuenciación de Ión Semiconductor, tienen el mayor potencial para convertirse en el formato de secuenciación más importante en futuras pruebas genéticas de rutina, especialmente en las pruebas en las que solo se analizan uno o unos pocos genes. Sin embargo, la inexacta denominación de (mayores) tramos de homonucleótidos, es un grave problema. Especialmente los tramos de homonucleótidos de 7 nucleótidos de longitud, no siempre se denominan correctamente. La mayoría de los genes llevan al menos 1 o unos pocos de tramos de homonucleótidos en este rango de tamaño. Además, los tramos de homonucleótidos son propensos a mutaciones, p. ej. debido al deslizamiento de la ADN polimerasa. Por lo tanto, una caracterización correcta de los tramos de homonucleótidos es un requisito previo para una prueba diagnóstica que tiene como objetivo analizar regiones completas de un genoma, como un(os) gen(es). Por ejemplo, la región codificadora y las uniones exón/intrón del gen *CFTR*, en el que las mutaciones causan fibrosis quística, contiene tres tramos de homonucleótidos de al menos 7 nucleótidos. La mayoría de las pruebas comerciales de *CFTR*, las pruebas genéticas más comunes realizadas en la población caucásica y que hasta el momento no se basan en la secuenciación, sólo examinan las 30 mutaciones *CFTR* más comunes, que incluyen tres loci localizados en estos tramos de homonucleótidos. La dificultad de caracterización imprecisa de los tramos de homonucleótidos prohíbe así la implementación de la pirosecuenciación y la Secuenciación de Ión Semiconductor en el diagnóstico genético de rutina. De hecho, tampoco se puede realizar una determinación precisa en estos tramos de homonucleótidos, o se realiza un segundo ensayo como el Sanger clásico a lo largo de estos tramos de homonucleótidos problemáticos. Esto no es deseable en pruebas genéticas de rutina, y se prefiere una prueba "única" o de "centro único".

Para la mayoría de las pruebas genéticas de diagnóstico de rutina, solo existe un interés en el análisis de uno o unos pocos genes. De hecho, sólo se caracterizan los exones y las uniones exón/intrón de estos genes. Por lo tanto, estas regiones de ADN de interés se enriquecen primero a partir del genoma total con el fin de disminuir la "señal de fondo" de las regiones de ADN de no interés, y para aumentar la sensibilidad de la detección.

### Descripción detallada de la invención

Las posiciones vecinas de nucleótidos en una molécula de ADN pueden albergar un(os) nucleótido(s) idéntico(s). En el contexto de la presente invención una repetición de nucleótidos idénticos también se denomina un "tramo de homonucleótidos". Una repetición puede tener tan poco como 2 o 3 nucleótidos, y puede extenderse hasta más de 15 a 25 nucleótidos idénticos. Las repeticiones de 7, 8, 9 o 10 son más comunes, las repeticiones de más de 15 nucleótidos son menos comunes. Debido a las impredecibles conformaciones estructurales, algunas repeticiones de 5 o 6 repeticiones pueden ser incluso difíciles de determinar. Realizaciones particulares incluyen intervalos que tienen como límite inferior y superior cualquiera de los valores citados anteriormente. Por ejemplo, 7 nucleótidos T seguidos se denominaría así un tramo de homonucleótidos de 7 nucleótidos.

En la pirosecuenciación y secuenciación de Ión Semiconductor, se añaden nucleótidos nativos (desprotegidos) a la reacción de secuenciación en cada ciclo. En un tramo de homonucleótidos dado, se incorporará más de 1 nucleótido en una reacción de secuenciación. El número de nucleótidos complementarios que se incorporarán es igual al número de nucleótidos en el tramo de homonucleótidos, de manera que la intensidad de señal es proporcional al número de nucleótidos. Sin embargo, la intensidad de señal para tramos de homopolímeros es solamente lineal hasta siete nucleótidos consecutivos. De hecho, cuanto más largo es el tramo de homonucleótidos, menor es la diferencia en la intensidad de señal para 1 nucleótido menos o 1 más en el tramo de homonucleótidos (Figura 2). Por ejemplo, un tramo de 3 homonucleótidos da como resultado una señal un 50% más alta que un tramo de 2 homonucleótidos, mientras que un tramo de 8 homonucleótidos solamente da como resultado una señal 14% más alta que un tramo de 7 homonucleótidos. En un momento dado, la diferencia de señal que se necesita detectar con

el fin de determinar la longitud exacta de un tramo de homonucleótidos es del mismo orden que la variabilidad en la señal de fondo, de modo que el sistema alcanza su límite de detección de diferencias de señal. Algunos tramos de homonucleótidos más cortos de 7 nucleótidos podrían incluso denominarse de modo incorrecto ya que las estructuras secundarias específicas del contexto-secuencia pueden dificultar la polimerización del ADN.

5 En esta invención, el tramo de homonucleótidos más largo que no se puede determinar con precisión se reduce a dos, o más (p. ej. 3, 4, 5 o 6), partes más pequeñas, en las que al menos la parte 3' puede determinarse con precisión en ensayos de secuenciación, como pirosecuenciación y secuenciación de Ión Semiconductor. Sólo la última de las repeticiones necesita estar en este intervalo de menos de 5 a 7 nucleótidos idénticos, la primera puede ser aún más de 5 a 7 nt, ya que la primera está completamente localizada en el oligonucleótido y debe tener el número esperado cuando está bien sintetizado y purificado. Sin embargo, para controlar oligos de menor calidad, podría reducirse un tramo muy largo a más de 2 partes de modo que sean todos menores de 5 a 7 nt de largo para que puedan contabilizarse como un control adicional. Su longitud combinada permite entonces una determinación precisa del tramo de homonucleótidos original más largo. Generalmente solo el último tramo de homonucleótidos menor tiene una longitud que se puede determinar con la máxima precisión.

15 La reducción en el tramo de homonucleótidos puede lograrse en la reacción de síntesis de ADN, que se define en esta invención como "síntesis de (hnr) ADN de reducción de tramo de homonucleótidos". Cuando se usa una reacción de PCR para síntesis/amplificación de ADN, la técnica se define en esta invención como "PCR de reducción del tramo de homonucleótidos" (hnr-PCR).

20 De hecho, con el fin de obtener señales de secuenciación suficientemente fuertes y/o secuenciar solamente regiones específicas del genoma total, las moléculas de ADN a secuenciar tienen que ser enriquecidas copiando o incluso amplificando. La hnr-PCR puede ser un medio para copiar/amplificar para que se pueda realizar la reducción del tramo de homonucleótidos y amplificación del ADN objetivo al mismo tiempo y en la misma etapa.

Para la amplificación, los cebadores se diseñan para que se unan a su ADN objetivo complementario mediante hibridación. Los cebadores se diseñan normalmente para ser 100% complementarios a su ADN objetivo. Sin embargo, incluso un cebador que no es complementario al 100% puede unirse a su región de ADN objetivo, especialmente cuando la síntesis de ADN se realiza en condiciones menos astringentes. A continuación se sintetiza una nueva cadena de ADN a través de la extensión del extremo 3' del cebador. Aunque un cebador no necesita ser 100% complementario con el fin de iniciar la síntesis de ADN, el nucleótido 3' necesita ser 100% complementario y estar unido a través de hidrógeno con su ADN objetivo con el fin de iniciar la síntesis de una nueva cadena de ADN. Para la hnr-PCR, se puede diseñar un cebador hnr en un tramo de homonucleótidos que se extiende en el tramo de homonucleótidos (Figuras 3 y 4), ya sea parcialmente en el tramo de homonucleótidos o hasta su extremo completo. Por supuesto, el cebador preferiblemente necesita contener más secuencias complementarias únicas 5' que preceden al tramo de homonucleótidos objetivo con el fin de aumentar la especificidad de unión. En la práctica, se pueden añadir secuencias adaptadoras adicionales 5', para permitir el procesamiento adicional de las nuevas cadenas de ADN sintetizadas como secuencias adaptadoras para cebador necesarias en la secuenciación, segundas etapas de PCR, PCR de emulsión, secuencias de códigos de barras, etc. El extremo 3' de un cebador hnr albergará así un tramo de homonucleótidos, o incluso 1 nucleótido, en el que el tipo de nucleótido es complementario al tipo de nucleótido en el tramo de homonucleótidos objetivo. Dado que un cebador necesita ser necesariamente 100% complementario a su región objetivo, se pueden incorporar nucleótidos no complementarios en ciertas posiciones para obtener un cebador hnr, p. ej. la 2ª posición 3' del cebador hnr, la 3ª última posición 3' del cebador hnr, etc...(en este contexto el nucleótido en el extremo 3' está en la "primera" posición 3'. Cuando un nucleótido se sustituye por un nucleótido no complementario en la región del tramo de homonucleótidos, el tramo de homonucleótidos se rompe y se divide en dos partes más pequeñas de modo que el tramo de homonucleótidos original se reduce a dos partes más pequeñas. Las dos partes podrían ser dos tramos de homonucleótidos más pequeños de la misma longitud, o no. La parte 3' podría incluso llegar a ser un nucleótido largo, y puede, por lo tanto, incluso no ser más un tramo de homonucleótidos. Un cebador hnr que no se extiende hasta el final del tramo de homonucleótidos detectará supresiones e inserciones en el tramo de homonucleótidos. Un cebador hnr que se extiende hasta el final del tramo de homonucleótidos detectará inserciones, pero no supresiones en el tramo de homonucleótidos.

50 Por ejemplo, para un tramo de homonucleótidos de 8 nucleótidos, un cebador hnr que se extiende 6 nucleótidos en el tramo de homonucleótidos y en el que el segundo último nucleótido de la hnr-PCR está sustituido por un nucleótido no complementario, una cadena de ADN recién sintetizada contendrá un tramo de 4 homonucleótidos, seguido por el nucleótido no complementario, y seguido por un tramo de 3 homonucleótidos. Si una reacción de secuenciación detecta en esta última posición un tramo de homonucleótidos de 3 nucleótidos de longitud, el fragmento de ADN bajo investigación alberga entonces 8 nucleótidos en el tramo de homonucleótidos original. Si una reacción de secuenciación detecta en esta posición un tramo de homonucleótidos de 2 nucleótidos de longitud, el fragmento de ADN bajo investigación alberga entonces 7 nucleótidos en el tramo de homonucleótidos. Y si una reacción de secuenciación detecta en esta posición un tramo de homonucleótidos de 4 nucleótidos de longitud, el fragmento de ADN bajo investigación alberga entonces 9 nucleótidos en el tramo de homonucleótidos. De este modo, y en este ejemplo, la determinación imprecisa de 7, 8 o 9 nucleótidos en un tramo de homonucleótidos se transforma en una determinación precisa de 2, 3 o 4 nucleótidos respectivamente.

En otro ejemplo, para un tramo de homonucleótidos de 8 nucleótidos, un cebador hnr que se extiende hasta el final del tramo de homonucleótidos y en el que el segundo último nucleótido de la hnr-PCR está sustituido por un nucleótido no complementario, una cadena de ADN recién sintetizada contendrá un tramo de 6 homonucleótidos, seguido por el nucleótido no complementario, y 1 nucleótido del mismo tipo que el tramo 6 homonucleótidos.

- 5 En otro ejemplo, para un tramo de homonucleótidos de 8 nucleótidos, un cebador hnr que se extiende 7 nucleótidos en el tramo de homonucleótidos y en el que el tercer último nucleótido de la hnr-PCR está sustituido por un nucleótido no complementario, una cadena de ADN recién sintetizada contendrá un tramo de 4 homonucleótidos, seguido por el nucleótido no complementario, y seguido por un tramo 3 homonucleótidos.

- 10 Se puede incorporar más de 1 nucleótido discordante a intervalos regulares en un cebador hnr, de manera que se obtienen diferentes tramos de homonucleótidos más pequeños como control adicional. Para un tramo de homonucleótidos de 15 nucleótidos en el que el 6° y el 12° nucleótido no coinciden en el cebador hnr, el tramo de 15 homonucleótidos se reducirá a dos tramos de 5 homonucleótidos y un tercer tramo de 3 homonucleótidos.

- 15 Por supuesto, las reacciones de secuenciación de tales amplicones hnr no proporcionan información sobre las posibles mutaciones en el sitio del cebador, ya que la secuencia secuenciada observada en ese sitio se deriva del cebador en lugar del ADN objetivo bajo investigación. En el ejemplo anterior en el que se reduce un tramo de homonucleótidos de 8 nucleótidos de longitud mediante hnr-PCR, en el que el cebador hnr se extiende 6 nucleótidos en el tramo de nucleótidos y en el que el segundo último nucleótido de la hnr-PCR está sustituido por un nucleótido no complementario, los 4 nucleótidos del primer tramo de homonucleótidos más pequeño, el nucleótido discordante, y el primer nucleótido del segundo tramo de 3 homonucleótidos, no proporcionan información de secuencia del fragmento de ADN bajo investigación ya que se deriva de la secuencia del cebador. Sin embargo, el segundo y el  
20 tercer nucleótido del tramo de 3 homonucleótidos proporcionan información de secuencia del ADN bajo investigación. También por esta razón puede ser preferible que los cebadores hnr no se extiendan hasta el final completo del tramo de homonucleótidos, de modo que se obtiene la información de secuencia de la unión del tramo de homonucleótidos.

- 25 En muchos casos, la secuencia aún así se debe determinar en el sitio del cebador hnr o incluso corriente arriba. De hecho, en muchos casos, un tramo de homonucleótidos se localiza en algún lugar en un exón, y se necesita determinar la secuencia completa del exón y de las uniones exón/intrón. Por lo tanto, también se puede realizar una amplificación de PCR clásica usando cebadores que flanquean la región completa de interés, incluyendo el tramo de homonucleótidos (Figura 3). La hnr-PCR puede usarse así como tal, o en combinación con PCR clásica.

- 30 El análisis combinado de un amplicón estándar y un amplicón hnr frente a una región dada finalizará entonces la secuencia real, en la que el amplicón estándar proporciona la información cualitativa de la secuencia, excepto la longitud exacta del tramo de homonucleótidos, mientras el amplicón hnr determinará la longitud exacta del tramo de homonucleótidos.

- 35 Se puede generar y analizar un único tipo de fragmento hnr, o tramo de homonucleótidos, y por lo tanto se usa un solo par de cebadores. Por otra parte, se puede generar y analizar en un formato (tipo) multiplex una combinación de diferentes tipos de fragmentos hnr, dirigidos contra diferentes tramos de homonucleótidos en el que se usan múltiples pares de cebadores. En dicho formato multiplex, podrían incluirse incluso los respectivos amplicones estándar, como amplicones que contienen los tramos de homonucleótidos completos bajo investigación. Cuando se usan formatos de PCR (tipo) multiplex en los que los amplicones estándar y amplicones hnr se amplifican conjuntamente, el cebador no hnr usado para generar el amplicón hnr podría ser el mismo, o no el mismo, que el  
40 usado para la amplificación del amplicón estándar respectivo.

- En tecnologías de secuenciación de nueva generación necesitan analizarse muchas lecturas en cualquier posición y necesitan alinearse con la secuencia de referencia. De hecho, el análisis tiene una naturaleza probabilística y estas tecnologías de secuenciación tienen un error de secuenciación relativamente alto por lectura secuenciada.  
45 Normalmente, se necesita detectar cada nucleótido en una secuencia bajo investigación en 20 a 30 lecturas de secuencia para deducir una secuencia diploide con una alta precisión. Además, todas las secuencias de amplicones múltiples o ADNs objetivo se generan en un único experimento y se obtendrá una mezcla de lecturas de secuencia de todos estos amplicones. Normalmente, se obtiene una mezcla de lecturas derivadas de una mezcla de amplicones derivados de objetivos diferentes, como todos los exones de un gen. Cuando se combinan amplicones estándar y amplicones hnr para el mismo objetivo, la mezcla obtenida será aún más compleja y contendrá ambas lecturas de secuencia para un tramo de homonucleótidos a partir de amplicones de PCR estándar y lecturas de amplicones de hnr-PCR. Aparte de su diferencia en longitud (el amplicón de hnr-PCR sólo comienza o termina alrededor de la secuencia del tramo de homonucleótidos), el amplicón hnr contiene uno o más nucleótidos discordantes cuidadosamente introducidos que se comportan como (una/s) mutación(es) puntual(es). Cuando los  
55 amplicones estándar y los amplicones hnr se alinean juntos, ambos tipos de amplicones se alinearán correctamente con la secuencia de referencia, y el nucleótido discordante se determinará erróneamente como una mutación. Los amplicones estándar y los amplicones hnr necesitan separarse para su análisis y alinearse por separado de modo que su alineamiento y análisis no interfiera. Esto puede realizarse fácilmente cuando se introducen diferentes secuencias adaptadoras 5' en el cebador usado para la amplificación estándar y la amplificación hnr (Figura 5).  
60 Aunque las secuencias adaptadoras pueden diferir para cada amplicón, todos los amplicones estándar contienen

preferiblemente la misma secuencia adaptadora, y si se necesita analizar más de un tramo de homonucleótidos por hnr-PCR, todos los amplicones hnr contienen preferiblemente la misma secuencia adaptadora diferente. El conjunto de lecturas de secuencia obtenidas se puede separar luego en dos grupos basados en la secuencia adaptadora, de manera que se obtiene un grupo de amplicones estándar y un grupo de amplicones hnr. Cada grupo se alinearán por separado con la secuencia de referencia de manera que sus alineamientos y análisis no interfieran. Preferiblemente la secuencia de referencia para el tramo de homonucleótidos puede adaptarse de manera que el nucleótido discordante se recoja en la secuencia de referencia. A todos los cebadores, tanto los cebadores estándar como los cebadores hnr, se les pueden añadir secuencias adaptadoras adicionales para procesamiento adicional, como secuencias para hibridación de oligonucleótidos para otra amplificación, secuenciación, y secuencias de códigos de barras (Figura 6).

En el contexto de la presente solicitud, la PCR, y por lo tanto la hnr-PCR, es el método de síntesis de hnr-ADN. Sin embargo los métodos de la presente invención se pueden realizar con cualquier técnica que sintetice una o más cadenas nuevas de ADN de una cadena de ADN original. Los métodos adecuados conocidos en la técnica son por ejemplo, amplificación isotérmica (amplificación de círculos rodantes), amplificación de desplazamiento de cadena sencilla, amplificación basada en secuencias de ácidos nucleicos (NASBA), PCR en fase sólida (sobre perlas o matrices) (Raindance), sondas candado, sondas selectoras, sondas colectoras, Haloplex PCR (Halogenomics), reacción en cadena de la ligasa y amplificación de los productos oligos unidos por extensión-ligación (TSCA; TruSeq Custom Amplicon, Illumina), o una combinación de los mismos.

#### Listado de Secuencias

20 <110> Harry Cuppens Cuppens, Harry  
 <120> Nuevos métodos para determinar repeticiones de secuencias de nucleótidos  
 <130> C6831-PCT  
 <150> GB 1119390.1  
 <151> 10-11-2011

25 <160> 3  
 <170> PatentIn versión 3.3  
 <210> 1  
 <211> 18  
 <212> ADN  
 30 <213> Artificial  
 <220>  
 <223> Cebador PCR primer  
 <400> 1  
 ggaacgcgtc acgtttat 18

35 <210> 2  
 <211> 52  
 <212> ADN  
 <213> Artificial  
 <220>  
 40 <223> ADN molde teórico  
 <400> 2  
 tggaggtagc ctggaacgc gtcacgtttt ttttgatgg acctagtggc at 52  
 <210> 3  
 <211> 39  
 45 <212> ADN  
 <213> Artificial  
 <220>  
 <223> nueva cadena de AND hipotética modificada  
 <400> 3  
 50 ggaacgcgtc acgtttatt ttgatggacc tagtggcat 39

**REIVINDICACIONES**

1. Un método de generar una o más copias de un ADN objetivo o parte del mismo, que contiene una repetición de nucleótidos idénticos, que comprende las etapas de:
  - 5 - alterar un nucleótido en esta repetición de nucleótidos idénticos, o
    - alterar diferentes nucleótidos individuales separados a intervalos en dicha repetición de nucleótidos idénticos en otro nucleótido, con el fin de reducir dicha repetición de nucleótidos idénticos en dos o más partes alteradas más pequeñas de dichos nucleótidos idénticos en las moléculas copiadas,
  - 10 en donde esta alteración se realiza usando un cebador de oligonucleótidos que se extiende en la repetición no alterada de nucleótidos idénticos del polinucleótido objetivo, en donde dicho cebador de oligonucleótidos comprende en el lado 5' una secuencia que hibrida con la secuencia que precede a la repetición de la secuencia objetivo en la secuencia de ADN objetivo, pero en donde dicho oligonucleótido no es complementario al 100% con la repetición de secuencia en la secuencia objetivo.
2. El método según la reivindicación 1, en donde la parte alterada no es mayor de 4, 5 o 6 nucleótidos.
- 15 3. El método según la reivindicación 1, en donde dicho cebador de oligonucleótidos no se extiende hasta el final de la repetición no alterada de nucleótidos idénticos del polinucleótido objetivo.
4. El método según cualquiera de las reivindicaciones 1 a 3, en donde las partes alteradas más pequeñas de nucleótidos idénticos se generan a través de una reacción enzimática, como síntesis, ligación y/o amplificación de ADN, o una combinación de las mismas.
- 20 5. El método según cualquiera de las reivindicaciones 1 a 4, en el que un nucleótido en la repetición de nucleótidos idénticos complementarios del cebador se sustituye por otro tipo de nucleótido de modo que se reduce la repetición no alterada de número consecutivo de nucleótidos idénticos a dos partes alteradas más cortas de estos nucleótidos idénticos interrumpidas por el tipo de nucleótido sustituido.
- 25 6. El método según cualquiera de las reivindicaciones 1 a 5, en donde la repetición de nucleótidos idénticos en el extremo 3' del cebador es más corta que la repetición de nucleótidos idénticos en el extremo 5' del cebador.
7. El método según cualquiera de las reivindicaciones 1 a 6, en donde el nucleótido en la repetición de nucleótidos idénticos complementarios del cebador que está sustituido por otro tipo de nucleótido está en la segunda, tercera, cuarta o quinta posición contado desde el extremo 3' del cebador de oligonucleótido.
- 30 8. El método según cualquiera de las reivindicaciones 1 a 7, en donde el cebador contiene además en el extremo 5' uno o más nucleótidos adaptadores que no son complementarios con la secuencia objetivo.
9. El método según cualquiera de las reivindicaciones 1 a 8, en donde se combinan diferentes reacciones de síntesis/ligación/amplificación de ADN del mismo tipo, o de diferentes tipos en un formato (tipo) multiplex.
10. El método según cualquiera de las reivindicaciones 1 a 9, que comprende además una amplificación del ADN objetivo, que incluye la repetición de nucleótidos idénticos.
- 35 11. El método según la reivindicación 10, en donde se realizan juntos el método según se describe en la reivindicación 1, que usa el cebador de oligonucleótidos que se extiende en la repetición sin modificar de nucleótidos idénticos del polinucleótido objetivo, y el método como se describe en la reivindicación 10.
12. El método según cualquiera de las reivindicaciones 1 a 11 que comprende además la etapa de realizar un método para determinar la naturaleza cualitativa del tramo de nucleótidos idénticos y sus regiones de ADN corriente abajo y corriente arriba, en donde los fragmentos obtenidos de cada método se discriminan y se analizan por separado usando dichas secuencias adaptadoras.
- 40 13. Un método para determinar el número de nucleótidos en una repetición de nucleótidos de una molécula de ADN molde en la que se determina la secuencia de una molécula de ADN copiada mediante secuenciación, y que usa el conocimiento previo de los nucleótidos alterados y posiciones generadas en las moléculas copiadas obtenidas realizando el método según cualquiera de las reivindicaciones 1 a 12, que comprende la etapa de contar el número de nucleótidos idénticos en las partes más pequeñas generadas y los nucleótidos sustituidos para generar las partes más pequeñas de nucleótidos idénticos.
- 45 14. Un soporte de datos que comprende instrucciones de programa para analizar y proporcionar los resultados de un método según la reivindicación 13, cuando se ejecuta en un ordenador.

15. Usar de uno a más nucleótidos en una reacción de síntesis, ligación, o amplificación de ADN, o combinaciones de las mismas, para reducir repeticiones mayores de nucleótidos idénticos en dos o más partes más pequeñas de nucleótidos idénticos en un método según cualquiera de las reivindicaciones 1 a 13.

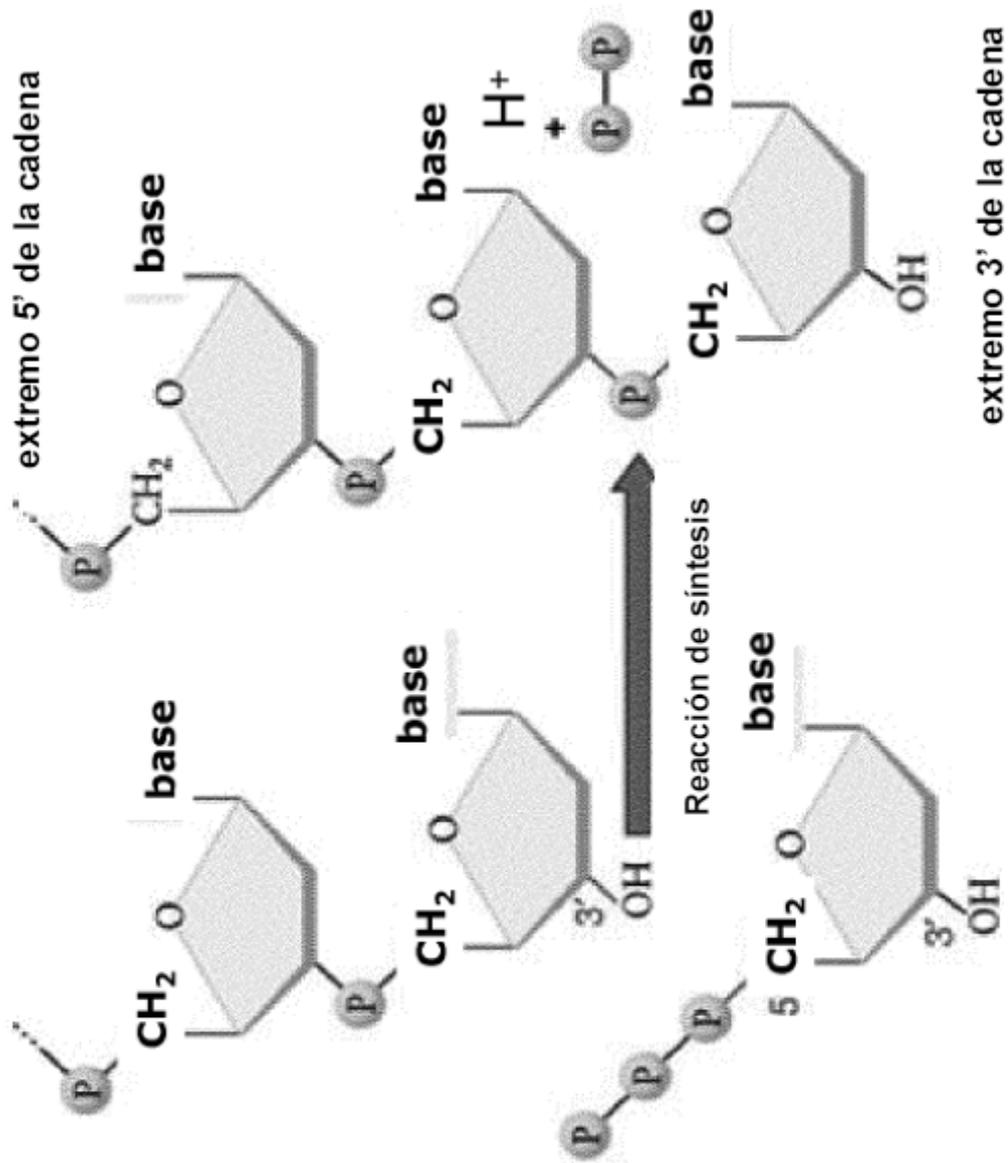
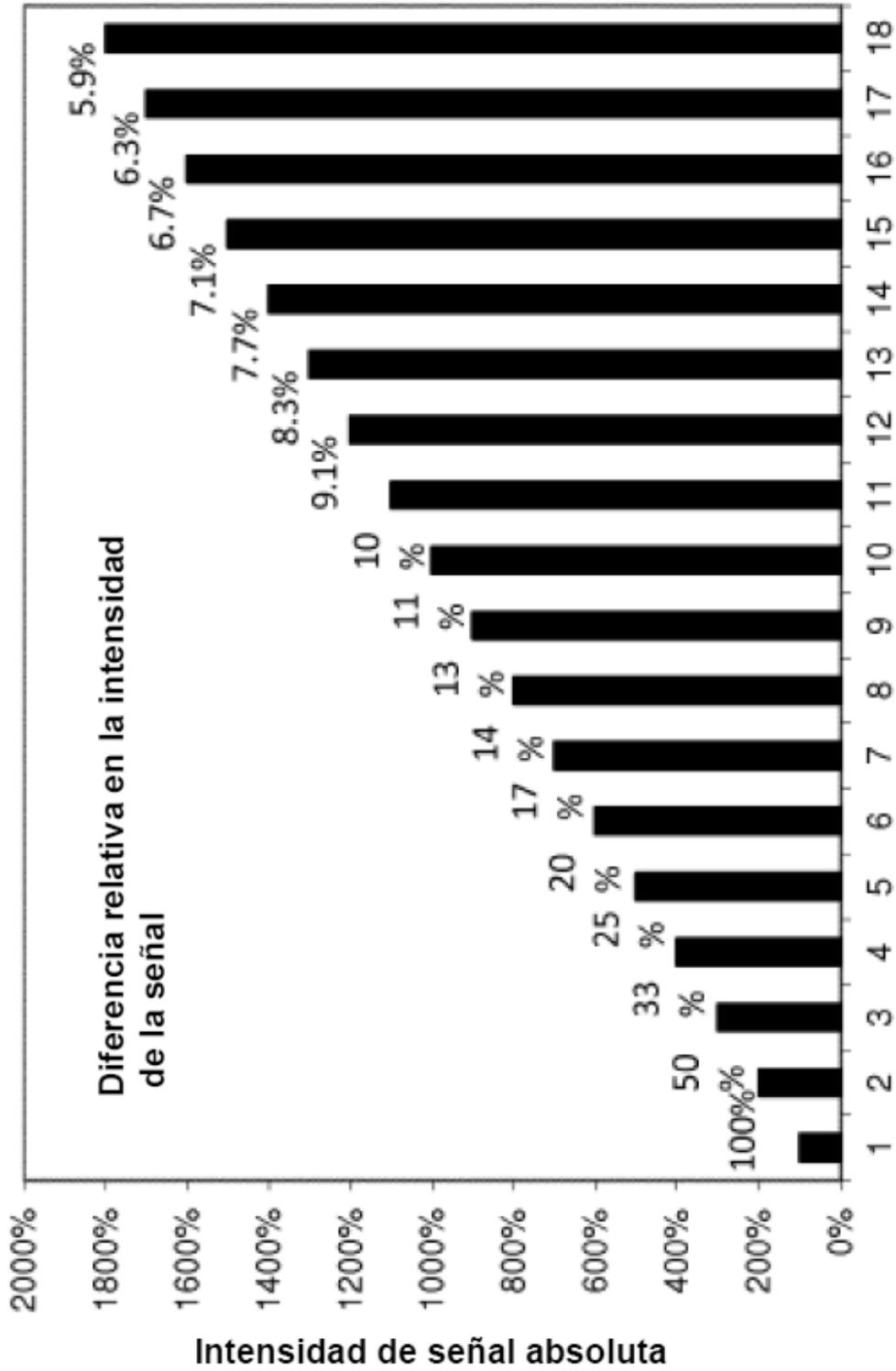


Figura 1



Número de nucleótidos en el tramo de homonucleótidos

Figura 2

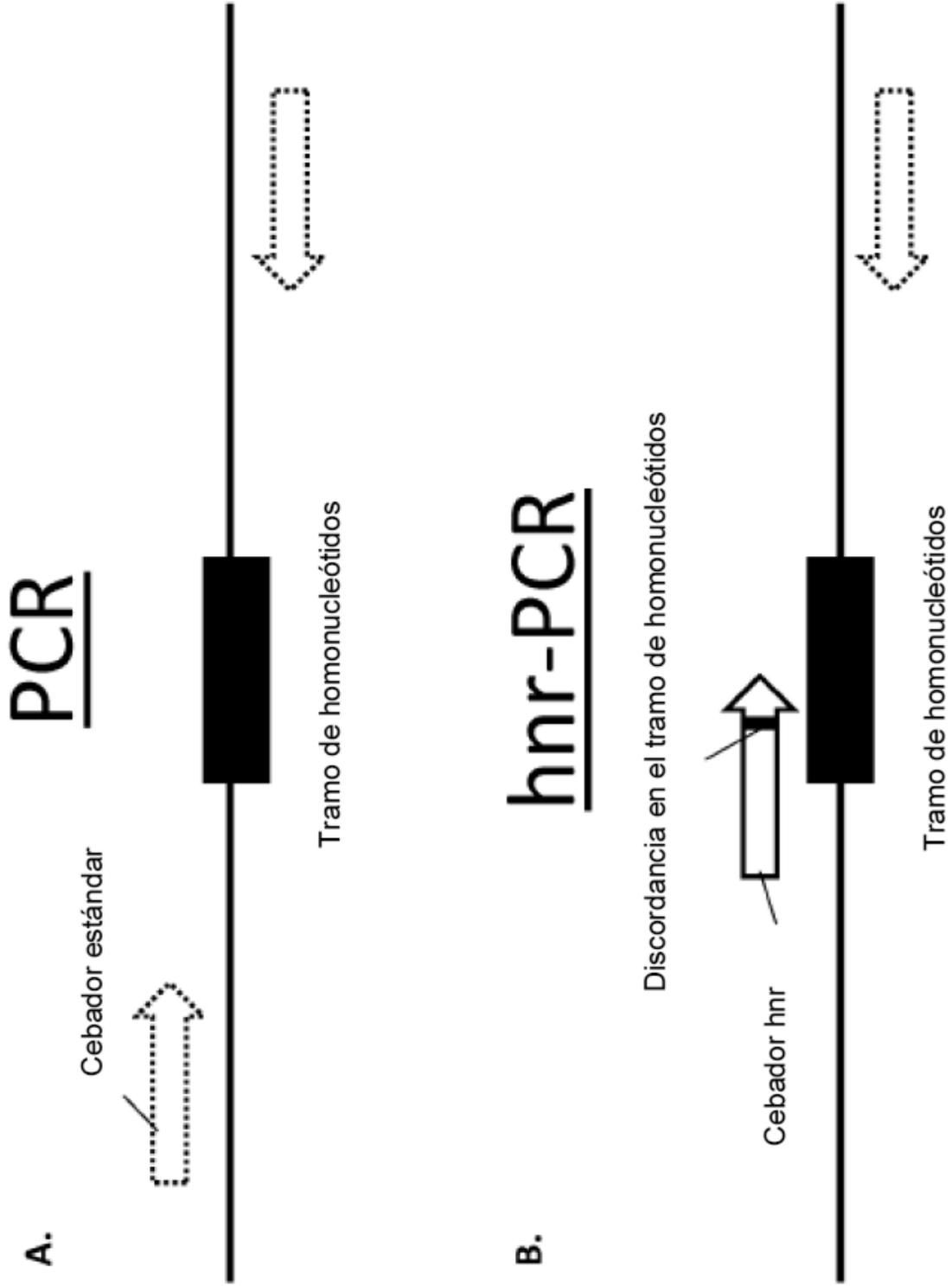


Figura 3

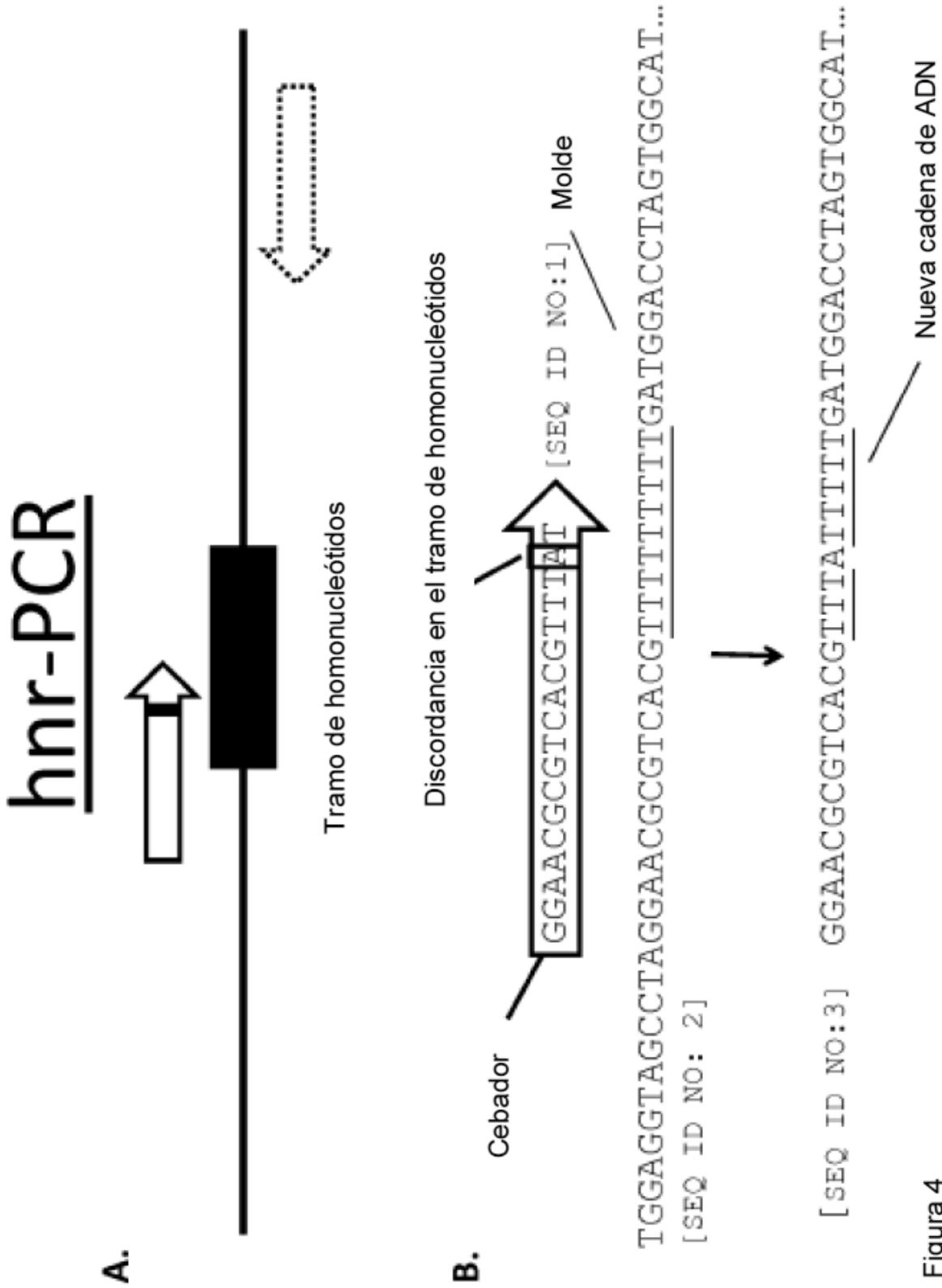


Figura 4

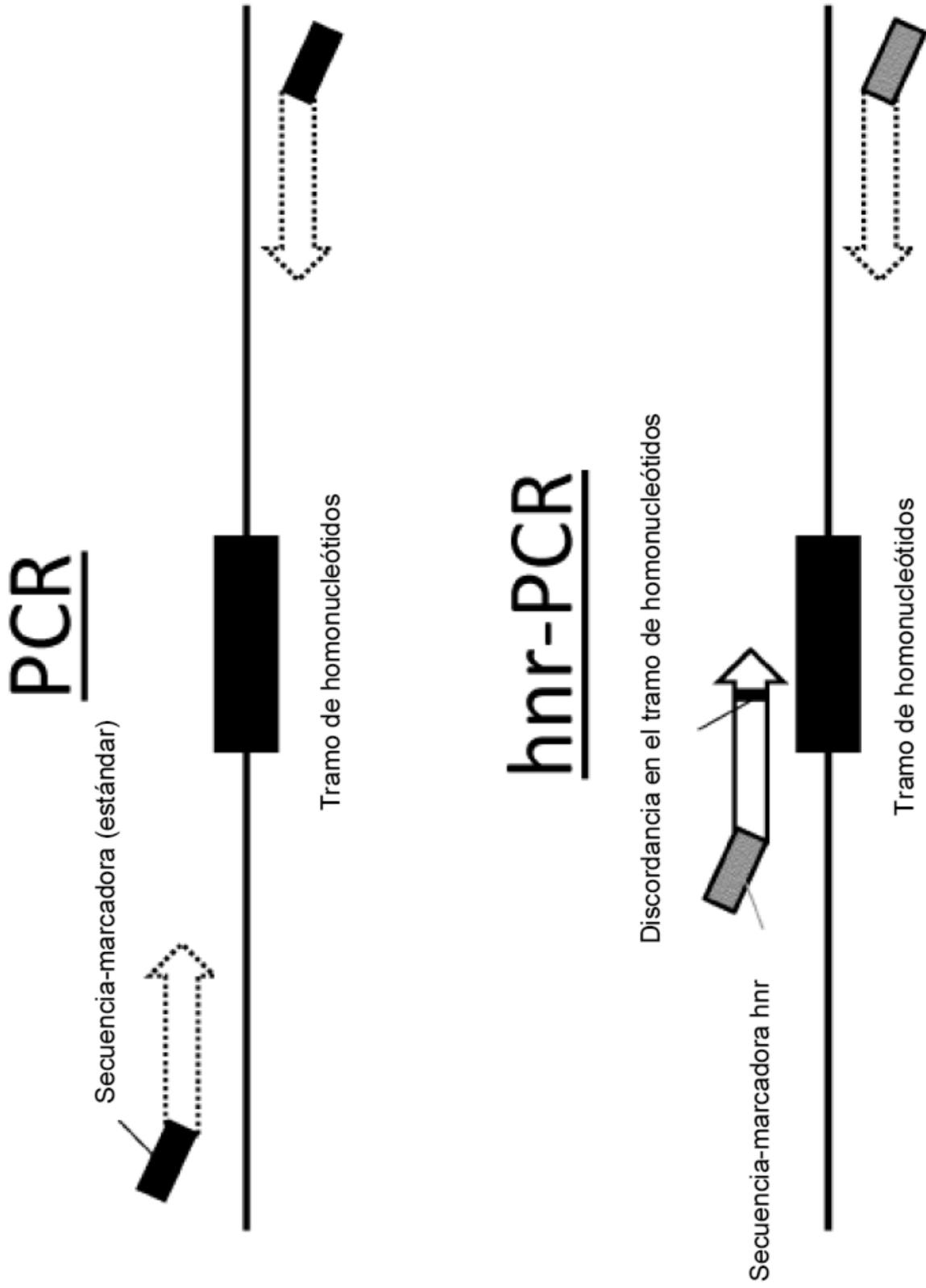


Figura 5

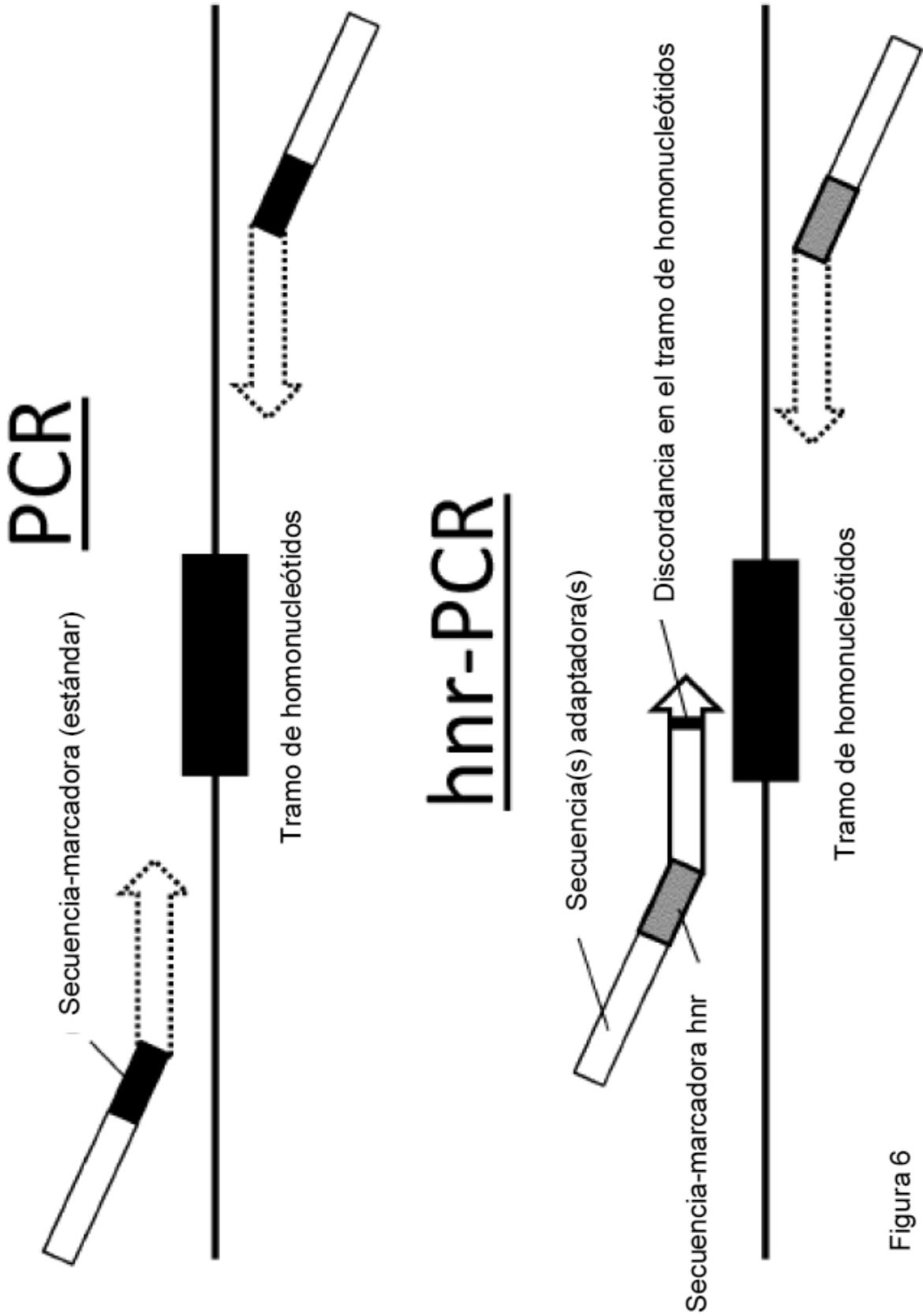


Figura 6