

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 368 767**

51 Int. Cl.:
C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07712741 .3**

96 Fecha de presentación: **19.02.2007**

97 Número de publicación de la solicitud: **1991697**

97 Fecha de publicación de la solicitud: **19.11.2008**

54 Título: **CONFORMACIÓN DE ADN (ESTRUCTURAS DE BUCLE) EN EXPRESIÓN GÉNICA NORMAL Y ANORMAL.**

30 Prioridad:
17.02.2006 GB 0603251

45 Fecha de publicación de la mención BOPI:
22.11.2011

45 Fecha de la publicación del folleto de la patente:
22.11.2011

73 Titular/es:
**ISIS INNOVATION LIMITED
EWERT HOUSE, EWERT PLACE
SUMMERTOWN, OXFORD OX2 7SG, GB**

72 Inventor/es:
**AKOULITCHEV, Alexandre;
RAMADASS, Aroul, Selvam y
NIKITENKO, Leonid, Leonidovich**

74 Agente: **de Elzaburu Márquez, Alberto**

ES 2 368 767 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

Conformación de ADN (estructuras de bucle) en expresión génica normal y anormal

Campo de la invención

La presente invención se refiere a diagnóstico y expresión génica.

5 Antecedentes de la invención

Los procedimientos existentes de diagnóstico de enfermedad son frecuentemente insatisfactorios debido a que no está disponible un marcador adecuado para el diagnóstico fidedigno de la enfermedad o para determinar el estado de la enfermedad. Los presentes enfoques incluyen el uso de detección de proteínas, ARNm o anticuerpos.

10 La literatura científica describe la regulación de la transcripción génica por el cambio dinámico entre bucles de genes cuasi estables formados por yuxtaposición de dos regiones separadas de un gen, tal como marcadores CC (véase Ramadass y col., resumen de póster P037, Transcription UK, Imperial College, Londres, 13-15 abril de 2005).

Compendio de la invención

15 La detección de proteínas, ARNm o anticuerpos no es adecuada en muchos casos de diagnóstico ya que la detección de estas moléculas no representa verdaderamente la expresión de los genes ligados a la enfermedad. La variación estocástica de niveles de expresión de estas moléculas entre células individuales es considerablemente alta, mientras que la semivida varía significativamente y podría ser muy baja, por ejemplo, aproximadamente 15 min para el polipéptido del protooncogén c-myc. Además, la detección de estas moléculas sólo sigue etapas posteriores en el orden de expresión génica - transcripción y traducción.

20 El ajuste conformacional epigenético del gen para posibles rondas reiniciadas de transcripción y expresión proporciona una posibilidad de diagnósticos en una etapa mucho más temprana de expresión génica. Tales estructuras conformacionales también parecen ser estables, es decir, teniendo una alta semivida, que hace que sean más fáciles de detectar.

25 Los inventores han encontrado que el análisis de la conformación cromosómica en ADN genómico puede usarse para diagnóstico de enfermedad. La conformación se forma por la asociación o yuxtaposición de sitios distantes o no adyacentes en el gen. Los sitios pueden ser marcadores CC (que se tratan adicionalmente más adelante). Se ha encontrado que un cambio en la conformación cromosómica de diferentes genes produce un cambio en la expresión de los genes y, por tanto, la detección de la conformación específica puede usarse para detectar expresión anormal de un gen.

30 Por consiguiente, la invención proporciona un procedimiento de detección o diagnóstico de cáncer mediante la detección de expresión génica anormal en un individuo que comprende determinar en una muestra del individuo la presencia o ausencia de una estructura cromosómica anormal en la que dos regiones separadas del gen se hayan puesto en estrecha proximidad por yuxtaposición, para así detectar o diagnosticar si el individuo tiene expresión génica anormal y, por tanto, cáncer.

Descripción de los dibujos

35 La Fig. 1 muestra la identificación de los marcadores para las unidades transcripcionales de ARNPII por el algoritmo de reconocimiento de patrones. (A) Esquema para formar y probar el modelo de marcador. 422 genes humanos anotados se muestrean, se prueban y se retroalimentan durante 10^3 ciclos hasta que se desarrolle un modelo convergente. En los 3' de los genes, la secuencia consenso incluyó una señal previamente desconocida de patrón múltiple, *Checkpoint Charlie*, junto con la señal de poli(A) bien definida y sitios consenso ricos en U. (B) En el extremo 3' del gen beta-globina humano, el marcador CC (marcado con distribución gaussiana) está presente en la dirección 3' del sitio rico en U y se corresponde con el sitio CoTC descrito más adelante. La gráfica muestra una caída en el valor de energía (marcado de gris) en el sitio CC/CoTC en relación con su secuencia vecina. (C) En el cromosoma X en *D. melanogaster*, el marcador CC (distribución gaussiana) coincide con el aislador *gypsy* dentro de la banda 7B2. La densidad de predicciones de CC establece una correlación con sitios de unión a Su(Hw). Estudios anteriores muestran elementos *gypsy* en la banda cromosómica 7B2 y 7B8 yuxtapuestas con formación de bucle alrededor del locus *cut*.

40

45

La Fig. 2 muestra expresión regulada de los genes modelo. (A) La transcripción del gen *hDHFR* está regulada en los promotores principal y secundario en un modo dependiente del ciclo celular. En células quiescentes, un transcrito corto se inicia a partir del promotor secundario en la dirección 5'. La citometría de flujo activada por fluorescencia (FACS) de células U2OS se usa en los experimentos, cultivadas en presencia de 10% de SBF (1) y bajo inhibición de contacto en presencia de 0,5% de SBF. El porcentaje de células G0, G2/M, S y G1 bajo cada condición se muestra en los diagramas. De acuerdo con informes previos, la transferencia Northern confirma la acumulación de ARNm de *DHFR* en células proliferantes (carril 3) con respecto a células quiescentes (carril 4). Análisis por RT-PCR en tiempo real de transcritos iniciados a partir del promotor secundario en células proliferantes (carril 5) y

50

quiescentes (carril 6). Los valores mostrados se calculan a partir de tres experimentos independientes. (B) Los transcritos *hCALCRL* de longitud completa sólo se producen en células endoteliales (HMVEC, carril 1) y no en células no endoteliales (HEK293T, carril 2). Los transcritos no codificantes cortos están presentes en ambos tipos de células como se detecta por RACE en 3' desde el primer exón (carril 3, células endoteliales y carril 4, células no endoteliales). La inmunoquímica confirma que la expresión del receptor está limitada *in vivo* a células endoteliales (flechas negras) y no epiteliales o del estroma (flechas blancas).

La Fig. 3 muestra propiedades de terminación de los marcadores CC. (A) El gen *DHFR* humano contiene tres marcadores CC (triángulos sólidos). Las propiedades de terminación de la transcripción de los marcadores se ensayaron por RT-PCR como se representa en el esquema. Los cebadores inversos (B, C) preceden o siguen la posición del marcador CC probado. La RT-PCR para CC_{DHFR-2} se ensayó en células quiescentes, ya que CC_{DHFR-2} mostró propiedades de terminación reguladas sólo bajo esas condiciones. Los perfiles para energía libre de plegamiento usando el algoritmo de Zuker muestran una caída en el valor (subrayado en gris) para los tres marcadores CC. (B) La estructura del gen *CALCRL* humano incluye tres marcadores CC (triángulos sólidos). Los marcadores CC en *hCALCRL* ($CC_{CALCRL-1}$, $CC_{CALCRL-2}$ y $CC_{CALCRL-3}$) también muestran una posible terminación de la transcripción. Una RACE en 5' desde el primer exón confirma que todos los transcritos se originan en la dirección 3' de $CC_{CALCRL-1}$, con cualquier posible transcrito intergénico satisfactoriamente terminado. La prueba de la transcripción terminada en $CC_{CALCRL-2}$ y $CC_{CALCRL-3}$ se confirmó por RACE en 3'. El número de accesos para los transcritos de RACE se presentan entre paréntesis. En el panel inferior, las gráficas muestran una caída en la energía libre de plegamiento (subrayado en gris) para cada uno de los marcadores CC de *hCALCRL*.

La Fig. 4 muestra las propiedades de conformación cromosómica de marcadores CC. (A) El ensayo 3C integrado se realizó para los sitios CC en el gen *hDHFR* bajo condiciones proliferantes y quiescentes. Los controles indican dependencia completa del ensayo en la reticulación, restricción, ligación, PCR y enriquecimiento de ARNPII por inmunoprecipitación. En células proliferantes se detecta una proximidad espacial entre CC_{DHFR-1} y CC_{DHFR-3} (carril 1+3), pero no entre los sitios CC_{DHFR-1} y CC_{DHFR-2} (carril 1+2) dentro del gen *hDHFR*. En células quiescentes, la proximidad espacial también se detecta entre los sitios CC_{DHFR-1} y CC_{DHFR-2} (carril 1+2). Ilustración esquemática de posibles conformaciones detectadas por el ensayo 3C bajo condiciones probadas. (B) El ensayo 3C integrado se realizó para los sitios CC en el gen *hCALCRL* en líneas de células endoteliales y no endoteliales. Los controles indican dependencia completa del ensayo en la reticulación, restricción, ligación, PCR y enriquecimiento por inmunoprecipitación. En células endoteliales se detectó una interacción entre $CC_{CALCRL-1}$, $CC_{CALCRL-2}$ y $CC_{CALCRL-3}$ que indica una conformación que yuxtapone todos los marcadores (carriles 1+2 y 1+3, véase el esquema). En células no endoteliales sólo pudo detectarse una interacción entre $CC_{CALCRL-1}$ y $CC_{CALCRL-2}$ (carril 1+2, véase el esquema), siendo la interacción entre $CC_{CALCRL-1}$ y $CC_{CALCRL-3}$ única para la transcripción productiva de longitud completa en células endoteliales.

La Fig. 5 muestra las predicciones de Checkpoint Charlie en otros organismos. El modelo estudiado en 422 genes humanos identifica marcadores CC (triángulos rojos) en otras especies. Obsérvese que en el caso de *RGF3* un único marcador CC separa dos genes anotados, que sirve de marcador de 3' para un gen y marcador de 5' para otro. Los exones e intrones se dibujan como rectángulos verdes y grises, respectivamente. La línea continua representa secuencias intergénicas.

La Fig. 6 muestra los principios de la detección de la conformación cromosómica usando el ensayo 3C.

La Fig. 7 muestra el tipado de c-myc para diagnosticar cáncer renal. Los marcadores CC 1 y 2 están posicionados alrededor del promotor P0. La yuxtaposición de CC1-CC2 conduce a la formación de la estructura cerrada que aísla P0 y evita la iniciación de P0, pero no de P1.2. El análisis de la yuxtaposición conformacional CC1-CC2 en muestras de tejido muestra la presencia de producto de PCR específico, confirmando la conformación existente en pacientes con tumor renal (T1-3), pero no en tejidos normales (N1-3). Todas las muestras se probaron independientemente para la presencia de conformación estable en genes sin relacionar, receptor similar al receptor de calcitonina (CRLR). Esta conformación está presente en todos los tejidos y actúa de control interno para el ensayo (marcado como *control*).

La Fig. 8 muestra el perfilado de la conformación cromosómica de cáncer de ovario con MLH1.

La Fig. 9 muestra la desregulación conformacional en líneas celulares de próstata.

Descripción detallada de la invención

La invención proporciona un procedimiento para la detección de expresión anormal de un gen basándose en la determinación de la estructura tridimensional de mayor orden que ha adoptado el gen, y en particular basándose en la posición/patrón de sitios asociados/yuxtapuestos dentro del gen. El procedimiento puede detectar la presencia o ausencia de sitios yuxtapuestos, o una conformación cromosómica producida por tal yuxtaposición, en una o más localizaciones en el gen. La forma normal de expresión de un gen se define normalmente como la expresión de un producto (ARN o polipéptido) en una forma y/o cantidad que permite que el producto realice su función celular/fisiológica.

La expresión anormal puede definirse como un modo de expresión en el que se realiza un producto diferente

(normalmente debido a un cambio en la posición de la terminación de la transcripción) y/o la cantidad de producto se expresa a un nivel alterado (o incluso no se expresa en absoluto). La expresión anormal puede conducir a un estado de enfermedad en el organismo (tal como cualquiera de las enfermedades mencionadas en este documento), y normalmente conducirá a una alteración de la viabilidad y/o funcionamiento de la célula o tejido u órgano en el que se produce la expresión anormal. La expresión anormal se caracteriza normalmente por la expresión de ARN o proteína de longitud aumentada o reducida en comparación con el producto normal y/o la expresión de ARN o proteína a un nivel aumentado o reducido en comparación con niveles de expresión normales.

En una realización preferida, el cambio de expresión normal a anormal se produce debido a un cambio en la estructura cromosómica como se define por marcadores CC. La yuxtaposición estructural de marcadores CC normalmente define la frontera de las unidades de transcripción, y generalmente la expresión anormal impone en exceso fronteras aberrantes (diferentes) a las observadas en expresión normal.

La invención proporciona el diagnóstico de una condición de enfermedad o diagnóstico de la etapa de una enfermedad en un individuo. La enfermedad es normalmente una en la que se produce la expresión anormal de uno o más genes. Tal expresión anormal puede producir o contribuir a la enfermedad. El gen puede ser uno que expresa un polipéptido funcional o ARN que no está traducido (tal como genes de ARN no codificantes y pseudogenes). El gen puede expresar ARN que tiene una función reguladora.

El gen es preferentemente un protooncogén (tal como c-myc) o un gen supresor de tumores (tal como BRCA1). El gen puede ser cualquiera de los genes enumerados en la Tabla 2. El gen puede ser hDHFR, hCALCRL, MLH1, PSA o BORIS (por ejemplo, como se desvela en los n° de acceso de GenBank NM000791, NM005795, NM000249, NM001030047 o NM080618). El gen tiene normalmente 2, 3, 4 o más secuencias de marcadores CC. El gen puede comprender un marcador CC en un intrón proximal al promotor, normalmente en el primer intrón.

La enfermedad puede ser un cáncer tal como un cáncer renal, de ovario, de vejiga, de colon o de próstata. La enfermedad puede ser una enfermedad genética, normalmente producida por expresión de un ARN o producto de polipéptido alterado (como se define anteriormente) y/o producida por la expresión de un nivel diferente de ARN o producto de polipéptido (tal como la ausencia de expresión de un producto tal).

En una realización, el procedimiento se lleva a cabo para determinar la etapa de la enfermedad, particularmente en el caso en el que la enfermedad sea cáncer. El procedimiento puede llevarse a cabo para determinar el riesgo de progresión de cáncer. Por tanto, el procedimiento puede usarse para predecir la tasa o gravedad de tumor o progresión de la enfermedad.

El individuo en el que se realiza el diagnóstico

El individuo que va a diagnosticarse puede tener uno o más síntomas de cualquiera de las condiciones de enfermedad mencionadas en este documento y/o del que se sospecha que tiene una condición de enfermedad tal. El individuo puede estar en riesgo de cualquier condición de enfermedad tal, por ejemplo, debido a que tiene un antecedente familiar de la enfermedad o debido a que vive en un entorno que produce o contribuye al desarrollo de la enfermedad. En el caso de cáncer en un ser humano, el individuo puede tener más de 40 años, tal como más de 50, o más de 60 años de edad. El individuo puede tener antecedentes de tabaquismo.

El individuo puede ser uno que tiene marcadores CC (cuya asociación define la estructura cromosómica) en al menos un gen de su genoma. El individuo es normalmente un eucariota, tal como un eucariota inferior o superior. El individuo puede ser una planta, levadura, insecto, marsupial, ave o mamífero. El individuo es preferentemente un mamífero, tal como un primate, ser humano o roedor.

Diagnóstico

La presente invención proporciona un procedimiento de diagnóstico de expresión génica anormal y, por tanto, un procedimiento de diagnóstico de condiciones de enfermedad particulares. El procedimiento comprende la detección de si hay o no una conformación cromosómica anormal en el ADN del individuo (por ejemplo, tanto directamente por detección de la estructura cromosómica real como indirectamente por detección de los sitios de asociación/yuxtaposición en el gen). Una conformación anormal tal comprenderá generalmente la presencia de una nueva yuxtaposición (o una combinación de yuxtaposiciones) en sitios en un gen (en los que normalmente no se observan, por ejemplo, cuando el gen se está normalmente expresando) o la ausencia de una o más yuxtaposiciones (que se observan normalmente durante la expresión normal). Como se menciona anteriormente, la conformación anormal conducirá a que el gen exprese transcrito de ARN con una diferencia en la secuencia y/o función y/o cantidad, y la diferencia en la expresión puede producir o contribuir a una enfermedad en el individuo, tal como cáncer. La conformación cromosómica anormal puede producir la expresión de una variante de corte y empalme diferente.

Puede usarse cualquier medio adecuado para detectar/examinar la conformación cromosómica del ADN que se analiza. Normalmente, la detección determinará la posición de al menos una estructura similar a bucle en el ADN del individuo. En una realización, el procedimiento puede comprender determinar la presencia o ausencia de un par yuxtapuesto dado de marcadores CC, permitiendo así, por ejemplo, la deducción de que la conformación observada

es diferente de la normal.

Normalmente, el procedimiento se lleva a cabo *in vitro* en una muestra del individuo. La muestra comprenderá ADN del individuo en un estado en el que las regiones del genoma que están asociadas en el estado natural permanezcan asociadas en la muestra (es decir, se preserva el estado cromosómico epigenético), por ejemplo, para regiones asociadas que están separadas menos de 5 kb, 3 kb, 1 kb, 500 pares de bases o 200 pares de bases. La muestra comprenderá normalmente células del individuo. La muestra comprenderá generalmente células de un tejido que participa en la enfermedad que va a diagnosticarse. La muestra comprende normalmente un fluido corporal del individuo y puede obtenerse, por ejemplo, usando un hisopo tal como un hisopo bucal. La muestra es preferentemente una muestra de sangre o una muestra congelada. La muestra puede ser una biopsia, tal como de un tumor. El procedimiento puede llevarse a cabo en una única célula del individuo.

La muestra se procesa normalmente antes de llevarse a cabo el procedimiento, por ejemplo, puede llevarse a cabo la extracción de ADN. El ADN en la muestra puede escindirarse tanto físicamente como químicamente (por ejemplo, usando una enzima adecuada). En una realización se usa anticuerpo específico para ARN polimerasa II para separar el ADN de otros componentes de la célula.

La conformación cromosómica puede detectarse por determinación de las secuencias que están asociadas, por ejemplo, que forman la base de una estructura similar a bucle. En una realización preferida, el ADN se somete a reticulación antes de una determinación tal. La reticulación comprenderá generalmente un enlace covalentemente ligado a la forma, y generalmente se forma poniéndose en contacto con un agente que produce reticulación. Un agente tal puede ser un aldehído, tal como para-formaldehído, o éster de N-hidroxisuccinimida de ácido D-biotinoil-ε-aminocaproico o éster de N-hidroxisuccinimida de ácido digoxigenin-3-O-metilcarbonyl-ε-aminocaproico. El para-formaldehído puede reticularse con las cadenas de ADN que están separadas 4 Angstroms.

En el procedimiento, el sitio de la yuxtaposición puede establecerse por determinación de las secuencias que se ponen en estrecha proximidad por la formación del bucle. Una determinación tal puede llevarse a cabo por cualquier medio adecuado, y en una realización preferida se realiza usando PCR.

En una realización, el ensayo de captura de la conformación cromosómica se usa, por ejemplo, como se describen en Dekker y col. (2002) Science 295, 1306. En este ensayo, el ADN se reticula (por ejemplo, como se ha descrito anteriormente). Entonces, el ADN reticulado se corta, normalmente por digestión con restricción, y la estructura cortada/digerida se somete a ligación. La ligación producirá los extremos de las cadenas de ADN que se formaron por corte/digestión para ligarse juntos. Por tanto, la ligación producirá generalmente ADN con una nueva secuencia (que no estaba presente en el gen original) que incluye ambas secuencias de los sitios yuxtapuestos. La detección de la nueva secuencia puede usarse como la base de la detección de la conformación (es decir, para detectar la presencia de yuxtaposición en una posición particular).

La secuencia generada por ligación puede detectarse por cualquier medio adecuado. Normalmente se detecta basándose en su secuencia, por ejemplo, usando PCR. En una realización se usa una reacción de detección por PCR en la que cebadores de PCR que se usan se unen en cualquier lado del punto de ligación y producen una reacción de PCR satisfactoria en presencia del producto ligado, pero que no producen una reacción de PCR satisfactoria cuando se lleva a cabo en presencia de un gen que no tiene la estructura relevante (normalmente debido a que los cebadores están unidos demasiado alejados entre sí en la secuencia del gen y la orientación de los cebadores excluye la elección de otros productos (los cebadores se eligen en la misma orientación con el fin de prevenir productos aberrantes)). En esta realización, un producto de PCR no sólo se detectará en presencia del producto ligado (véase la Figura 1). Normalmente, los cebadores de PCR se unirán dentro de 500 pares de bases entre sí cuando se unan al producto ligado.

La secuencia ligada puede detectarse/analizarse por PCR específica de secuencia o por secuenciación directa. La detección puede realizarse usando un sistema basado en gel en el que la secuencia ligada se ejecuta en un gel, y luego el gel se tiñe con un compuesto detectable que se une a polinucleótidos. La secuencia ligada puede detectarse usando una sonda, tal como una sonda de polinucleótidos que se une específicamente a la secuencia ligada.

Los productos de PCR que se forman en las reacciones de PCR mencionadas anteriormente pueden detectarse por cualquier medio adecuado, por ejemplo, por cualquier procedimiento adecuado de entre los procedimientos mencionados anteriormente para la detección del producto ligado.

En una realización, el procedimiento también comprende detectar la estructura cromosómica de otro gen, que es un gen específico de tejido. La detección de la estructura del gen adicional (por ejemplo, por cualquiera de los medios descritos en este documento) permitirá la determinación de si el gen adicional está siendo expresado o no y, por tanto, permitirá la determinación de la especificidad de tejido de la expresión. Esto puede ayudar en el diagnóstico de la enfermedad.

En una realización de la invención, 2, 3 o más genes se analizan con el fin de ayudar en el diagnóstico. En particular, en el caso de diagnóstico de cáncer, el análisis de más de un gen que participa en provocar cánceres puede ayudar en la determinación del cáncer específico.

En otra realización, el análisis de la estructura cromosómica que se lleva a cabo según el procedimiento de la invención se compara con el mismo análisis llevado a cabo en una biopsia de control de tejido con enfermedad (tal como un cáncer/tumor) con el fin de ayudar en el diagnóstico.

5 En una realización, el procedimiento de la invención se lleva a cabo de un modo cuantitativo con el fin de determinar la proporción de células del individuo (por ejemplo, en una localización *in vivo* particular o en un tejido particular) que tienen una expresión génica anormal. Esto puede ayudar en la determinación de la etapa de una enfermedad.

Secuencias en el gen que se asocian para formar la estructura cromosómica

10 Como se ha mencionado en este documento, el procedimiento de la invención comprende detectar la presencia de una conformación cromosómica que se forma por asociación de regiones particulares de un gen. Tales regiones están en el mismo cromosoma, y normalmente están separadas menos de 50.000, tal como menos de 20.000, 10.000, 5000, 1000 o menos de 500 bases. La asociación de las secuencias puede hacer que se forme una estructura de bucle/similar a bucle/topológicamente cerrada. El experto reconocerá lo que se indica por referencia a regiones de un gen que están asociadas. Tales regiones están suficientemente próximas para reticularse juntas, tal como por cualquiera de los agentes de reticulación mencionados en este documento. Por tanto, normalmente 15 estarán separadas una distancia que es del orden de Angstroms, tal como por ejemplo menos 50 Angstroms o menos de 10 Angstroms.

Una o ambas de las secuencias que se asocian pueden:

- causar, regular o contribuir a la terminación de la transcripción, y/o
- ser marcadores CC.

20 El marcador CC normalmente tiene una longitud de 1 a 30 bases de nucleótidos, por ejemplo, 5 a 20 ó 10 a 15 bases.

Los marcadores CC pueden detectarse en cualquier secuencia de genes dada usando la información en la Tabla 1. Una de las secciones posteriores ilustra más adelante en detalle cómo se identifican secuencias del marcador CC. Sigue una breve descripción de cómo se usa la información en la Tabla 1: la tabla muestra 4 conjuntos de pesos. 25 Para cada conjunto de pesos se indica una posición, y los valores de posición para cada tipo de nucleótido se facilitan con referencia a la posición inicial (en la Tabla 1 esto se define como la posición de columna que está en referencia con la posición inicial). Como puede apreciarse, para el primer conjunto de peso, los valores para guanina, citosina, adenina y timina se facilitan para las posiciones 0 a 18. Usando los valores en la Tabla 1 se determina una puntuación para cada base de una secuencia dada en la cadena directa e inversa. Este análisis se hace escaneando la secuencia de izquierda a derecha y luego repitiéndolo en su hebra complementaria. Mientras se escanea, una base se considera como punto de referencia y la puntuación para esa base se determina usando los valores de posición de los 4 conjuntos de pesos y la distancia relativa entre los pesos (es decir, para cada base se determina una puntuación basándose en las secuencias alrededor de esa base cuyas posiciones se definen usando los números de posición en la Tabla 1). Si esta puntuación es mayor que X (valor de entrada dado por el usuario), entonces el par de bases en cuestión está dentro de un marcador CC. Este procedimiento se repite para todas las bases. 35

La puntuación se convierte normalmente en una puntuación (logarítmica inversa) de valor exponencial. En una realización se seleccionan marcadores CC que tienen una puntuación logarítmica inversa superior a 0,9, tal como superior a 0,95 o superior a 0,99 (el cálculo de la puntuación logarítmica se describe en más detalle en una sección posterior). 40

Los inventores han usado la información en la Tabla 1 para detectar marcadores CC en ser humano, levadura y mosca de la fruta (secuencias de *D. melanogaster*).

Kit para llevar a cabo el procedimiento

45 La invención también proporciona un kit para llevar a cabo el procedimiento. El kit comprenderá normalmente un medio para la detección de secuencias yuxtapuestas específicas en un gen. Normalmente, el kit comprenderá un primer par o sonda que puede usarse para detectar una secuencia yuxtapuesta (por ejemplo, detectando un producto ligado como se describe en este documento). Normalmente, uno o ambos cebadores y/o la sonda comprenderán la secuencia que es un fragmento de la secuencia del gen o de la secuencia que es homóloga a la secuencia del gen (se entiende que referencias a la secuencia del gen también incluyen la secuencia complementaria, ya que por supuesto un cebador se unirá a la secuencia del gen y el otro cebador se unirá a la secuencia complementaria). Tal secuencia del gen puede estar en 5' con respecto a la secuencia codificante (por ejemplo, secuencia promotora), secuencia codificante, secuencia de intrón o secuencia 3' con respecto a la secuencia codificante. 50

Los cebadores o sonda tienen normalmente al menos 10, 15, 20, 30 o más bases de longitud, y generalmente comprenden ADN, normalmente en forma monocatenaria. Los cebadores o sondas pueden estar presentes en forma 55

aislada. Los cebadores o sonda pueden llevar una marca reveladora/detectable. Marcas adecuadas incluyen radioisótopos tales como ^{32}P o ^{35}S , marcas fluorescentes, marcas enzimáticas u otras marcas de proteína tales como biotina.

5 El kit puede comprender instrucciones para realizar el procedimiento de la invención. El kit puede comprender un agente de reticulación que puede reticular el ADN, tal como cualquiera de los agentes de reticulación mencionados en este documento.

En una realización, el kit es para llevar a cabo realizaciones de la invención en las que se analiza la estructura cromosómica de más de un gen, tal como 2, 3, 4 o más genes. En tales casos, el kit también puede comprender cebadores o sondas para analizar 2, 3, 4 o más genes diferentes.

10 El kit puede comprender adicionalmente uno o varios reactivos o instrumentos que permiten que se lleve a cabo cualquiera de las realizaciones del procedimiento mencionadas anteriormente. Tales reactivos o instrumentos incluyen uno o más de los siguientes: una marca detectable (tal como una marca fluorescente), una enzima que puede actuar sobre un polinucleótido (normalmente una polimerasa, enzima de restricción, ligasa, RNAsa H o una enzima que pueda unir una marca a un polinucleótido), tampón (tampones) adecuado(s) (disoluciones acuosas) para reactivos enzimáticos, un control positivo y/o negativo, un aparato de electroforesis en gel, un medio para aislar ADN de la muestra, un medio para obtener una muestra del individuo (tal como un hisopo o un instrumento que comprende una aguja) o un soporte que comprende pocillos sobre los que pueden hacerse las reacciones de detección.

Procedimiento de cribado

20 La invención proporciona un procedimiento de identificación de un compuesto para tratar expresión anormal de un gen que comprende determinar si una sustancia candidata puede hacer que la estructura cromosómica del gen cambie de la estructura anormal que se adopta durante la expresión anormal a la estructura normal para así determinar si la sustancia candidata puede ser capaz de tratar expresión anormal. El cambio en la estructura cromosómica puede detectarse usando cualquier procedimiento adecuado descrito en este documento. El procedimiento también puede llevarse a cabo para identificar compuestos que pueden causar un cambio en la expresión de un gen (por ejemplo, un cambio de un modo de expresión a otro modo de expresión), determinando de nuevo si un compuesto candidato puede producir o no un cambio en la estructura del gen.

25 El procedimiento puede llevarse a cabo *in vitro* (dentro o fuera de una célula) o *in vivo* (con un organismo no humano). En una realización, el procedimiento se lleva a cabo en una célula, cultivo celular, extracto celular, tejido, órgano u organismo que comprende el gen. La célula es normalmente una en la que se observa expresión anormal del gen.

El procedimiento se lleva a cabo normalmente poniendo en contacto (o administrando) la sustancia candidata con el gen, célula, cultivo celular, extracto celular, tejido, órgano u organismo y determinar si se produce o no un cambio a la estructura cromosómica normal.

35 Sustancias candidatas adecuadas que se probaron en los procedimientos de cribado anteriores incluyen agentes de anticuerpo (por ejemplo, anticuerpos monoclonales y policlonales, anticuerpos monocatenarios, anticuerpos quiméricos y anticuerpos injertados en CDR). Además, también pueden probarse bibliotecas combinatorias, entidades químicas definidas, péptido y miméticos de péptidos, oligonucleótidos y bibliotecas de agentes naturales tales como bibliotecas de expresión (por ejemplo, bibliotecas de expresión en fago). Las sustancias candidatas pueden ser compuestos químicos que se derivan normalmente de la síntesis alrededor de moléculas pequeñas que pueden tener cualquiera de las propiedades del agente mencionado en este documento (tal como los compuestos orgánicos mencionados en este documento). Los lotes de las sustancias candidatas pueden usarse en un cribado inicial de, por ejemplo, diez sustancias por reacción, y las sustancias de lotes que muestran modulación se prueban individualmente.

45 *Genes y organismos manipulados por ingeniería*

La invención proporciona un procedimiento de cambio del perfil de expresión de un gen que comprende

(i) introducir un marcador CC en el gen, y/o

(ii) eliminar un marcador CC del gen, opcionalmente introduciendo 1, 2, 3 o más mutaciones en el marcador CC, en el que cada mutación es una adición, sustitución o delección de una base de nucleótidos,

50 en el que al menos el 50% de la secuencia codificante del gen permanece invariable en el procedimiento.

En una realización, el número total de secuencias de marcador CC (es decir, secuencias de marcador CC funcionales) permanece invariable en el procedimiento.

Por "eliminar un marcador CC" se entiende que la secuencia del marcador CC entera puede no necesitar eliminarse, pero en su lugar pueden introducirse mutaciones en la secuencia del marcador CC para inactivarla, de manera que

en una realización la secuencia del marcador CC alterada ya no puede producir asociación de regiones del gen.

El ARN o producto de polipéptido del gen retiene la actividad funcional o puede tener una actividad diferente o puede no tener actividad (en comparación con el producto del gen no manipulado por ingeniería). El gen manipulado por ingeniería puede ser cualquiera de los genes mencionados en este documento. El gen manipulado por ingeniería puede replicarse y/o expresarse y/o introducirse en una célula.

La invención proporciona el uso de un polinucleótido que comprende un marcador CC para cambiar la expresión de un gen. Un polinucleótido tal puede usarse para introducir o eliminar un marcador CC de un gen, como en el caso de cualquiera de los genes manipulados por ingeniería descritos en este documento. El polinucleótido es normalmente una molécula de ADN. El polinucleótido puede estar en forma de un vector, tal como un vector vírico. El polinucleótido puede estar en forma de un transposón.

La invención también proporciona un organismo eucariota manipulado por ingeniería no humano que comprende al menos un gen en su genoma cuyo perfil de expresión se ha cambiado por la introducción y/o eliminación de una secuencia del marcador CC, en el que al menos el 50% de la secuencia codificante del gen permanece invariable. Por tanto, el organismo puede comprender el gen manipulado por ingeniería de la invención que se ha descrito anteriormente. El organismo (transgénico) puede ser cualquiera de los organismos mencionados en este documento. La invención también proporciona una parte del organismo que comprende el gen manipulado por ingeniería, tal como una célula u órgano del organismo.

La invención proporciona un procedimiento de preparar el organismo manipulado por ingeniería de la invención que comprende introducir o eliminar un marcador CC en un gen en la célula del organismo, y en el caso de un organismo multicelular, permitir que la célula crezca en el organismo. La introducción o eliminación del marcador CC puede llevarse a cabo en una célula germinal o citoblasto de embrión.

Homólogos

En este documento se citan homólogos de secuencias de polinucleótidos. Tales homólogos tienen normalmente al menos el 70% de homología, preferentemente al menos el 80, 90%, 95%, 97% o el 99% de homología, por ejemplo, con respecto a una región de al menos 15, 20, 30, 100 o más nucleótidos contiguos. La homología puede calcularse basándose en la identidad de nucleótidos (denominada algunas veces en lo sucesivo "homología dura").

Por ejemplo, el paquete UWGCG proporciona el programa BESTFIT que puede usarse para calcular homología (por ejemplo, usado en sus parámetros por defecto) (Devereux y col. (1984) *Nucleic Acids Research* **12**, pág. 387-395). Los algoritmos PILEUP y BLAST pueden usarse para calcular homología o alinear secuencias (tal como identificar secuencias equivalentes o correspondientes (normalmente en sus parámetros por defecto), por ejemplo, como se describe en Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S, F y col. (1990) *J Mol Biol* 215:403-10.

El software para realizar los análisis por BLAST está públicamente disponible del Centro nacional para información biotecnológica (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar primero el par de secuencias de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia de búsqueda que tanto coinciden como satisfacen alguna puntuación T umbral de valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se denomina en lo sucesivo al umbral de puntuación de palabras vecinas (Altschul y col., antes). Estas palabras comunes vecinas iniciales actúan de semillas para iniciar búsquedas para encontrar HSP que las contienen. Las palabras comunes se extienden en ambas direcciones a lo largo de cada secuencia para poder aumentar en la medida de lo posible la puntuación de alineamiento acumulado. Las extensiones para las palabras comunes en cada dirección se detienen cuando: las puntuaciones de alineamiento acumuladas se encuentran fuera de la cantidad X de su máximo valor alcanzado; la puntuación acumulada tiende a cero o por debajo, debido a la acumulación de uno o más alineamientos de residuos de puntuación acumulada; o se alcanza el extremo de cualquier secuencia. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad del alineamiento. El programa BLAST usa por defecto una longitud de palabra (W) de 11, los alineamientos de la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) (B) de 50, esperanza (E) de 10, M=5, N=4, y una comparación de ambas cadenas.

El algoritmo BLAST realiza un análisis estadístico de la similitud entre dos secuencias; véase, por ejemplo, Karlin y Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma mínima (P(N)) que proporciona una indicación de la probabilidad de que se produzca por casualidad una coincidencia entre dos secuencias de polinucleótidos. Por ejemplo, una secuencia se considera similar a otra secuencia si la probabilidad de suma mínima en la comparación de la primera secuencia con la segunda secuencia es inferior a aproximadamente 1, preferentemente inferior a aproximadamente 0,1, más preferentemente inferior a aproximadamente 0,01, y lo más preferentemente inferior a aproximadamente 0,001.

La secuencia homóloga normalmente se diferencia menos de 2, 3, 5 u 8 bases (que pueden ser sustituciones, deleciones o inserciones de nucleótidos). Estos cambios pueden medirse a través de cualquiera de las regiones mencionadas anteriormente en relación con el cálculo de homología.

Los siguientes ejemplos ilustran la invención:

Uso de análisis de reconocimiento de patrones para investigar la organización estructural de genes

Un paradigma emergente de la biología eucariota es que los aspectos estructurales de la organización nuclear desempeñan una función directa en la regulación transcripcional de los genes. A partir de los territorios cromosómicos para bucles de genes - diversos niveles estructurales emergen como componentes importantes de respuestas transcripcionales específicas (1-3). Aquí, los inventores han combinado dos enfoques con el fin de identificar algunas de las propiedades implicadas en la organización estructural de genes transcritos *in vivo*. A partir de las matemáticas aplicadas, los inventores han empleado análisis de reconocimiento de patrones basándose en el modelo lineal generalizado y el teorema de Bayes, y lo han usado para identificar los límites de las unidades transcripcionales de ARN polimerasa II (ARNPII). A partir de biología molecular, los inventores han usado ensayos *in vivo* para analizar y describir el espectro de actividad transcripcional y la organización del dominio subcromosómico estructural en esos sitios.

El análisis de reconocimiento de patrones se ha aplicado ampliamente a diversos campos de estudio tales como medicina, ingeniería y lingüística en los que el análisis de imágenes y la descodificación de datos permiten la identificación de marcadores característicos subyacentes dentro de sistemas complejos. Los inventores han usado metodología de reconocimiento de patrones para analizar datos de genoma humano en relación con las unidades transcripcionales, procesadas por ARNPII. Se usó un conjunto de secuencias de 422 genes manualmente anotados en el cromosoma humano 22 (4) para la identificación computacional de señales reguladoras. Para el estudio dado, de todos los procedimientos disponibles para el reconocimiento de patrones, los inventores encontraron la máquina de vectores de relevancia (RVM) (5-6) como la más satisfactoria. El entrenador de RVM aplica un principio bayesiano escaso que contiene la variación de distancia observada entre las señales reguladoras (7). Del conjunto dado de secuencias, el entrenador escanea marcadores que las definen y construye un modelo lineal generalizado probabilístico. Este modelo "estudiado" puede usarse posteriormente para clasificar secuencias de elección para la presencia de los marcadores definidos. La derivación de este modelo se basa en la probabilidad condicional del teorema de Bayes facilitado a continuación:

$$P(\text{modelo} \mid \text{datos}) = \frac{P(\text{modelo} \mid \text{datos})P(\text{modelo})}{P(\text{datos})}$$

en la que *datos* representa el conjunto de secuencias de ADN. $P(\text{modelo} \mid \text{datos})$ es la probabilidad posterior que da la probabilidad de una secuencia derivada del *modelo*. Depende de la probabilidad de los *datos* dado el *modelo* y las probabilidades del *modelo* y los *datos*.

Cada marcador que define la característica de la secuencia, *x*, se facilita como una matriz de pesos de ADN con respecto al sitio de escisión. Matemáticamente se representa como:

$$f(x) = \sum_{i=-\infty}^{\infty} P(i)W(x,i)$$

en la que *P* es una probabilidad posicional y $W(x,i)$ es una probabilidad de matriz de pesos de ADN para una compensación de *i* con respecto al sitio de escisión. Entonces se usa una combinación de estos marcadores para construir un modelo lineal generalizado:

$$\text{Modelo} = \sum_{m=1}^M \beta_m f_m(x) + k$$

en la que *M* es el conjunto de marcadores que definen el gen y β es los pesos (o importancia) dada de cada marcador.

El modelo estudiado en 422 genes humanos anotados del cromosoma 22 identificó tres tipos de marcadores generales en los extremos 3' (Fig. 1A). Señales de terminación de la transcripción previamente conocidas: señal de poli(A) y sitio rico en U próximo a los extremos 3' de genes transcritos con ARNPII son dos de los tres tipos de marcadores identificados. Este resultado validó en enfoque de los inventores ya que confirmó inequívocamente secuencias ya descritas funcionalmente implicadas en la terminación y el procesamiento del extremo 3' de ARNm (8-10). Interesantemente, el tercer tipo de marcador identificado por el entrenador de RVM fue previamente desconocido. Se posicionó adicionalmente en la dirección 3' del sitio rico en U y comprendió múltiples matrices de pesos de ADN. La variación de distancia observada en cada tipo de los marcadores fue capturada como una distribución gaussiana. Interesantemente, cuando se prueba el modelo en 20 secuencias del cromosoma humano, el marcador no se confinó a los extremos 3', sino que también estuvo presente en los extremos 5' de genes anotados. Debido a la asociación del marcador recientemente definido con las fronteras de las unidades transcripcionales, los inventores lo han llamado según el famoso muro de Berlín de los tiempos de la guerra fría – un marcador *Checkpoint Charlie* (CC).

Interesantemente, a diferencia del sitio poli(A), los inventores fueron incapaces de identificar alguna secuencia consenso primaria para los marcadores CC. Esto sugiere que mediante el análisis de reconocimiento de patrones los inventores han identificado los sitios que podrían compartir propiedades comunes mediante la información codificada en las estructuras secundarias y terciarias de las secuencias correspondientes. De hecho, el análisis de secuencias de marcadores CC usando el algoritmo Zuker de (11) revela energías libres bajas de plegamiento, características de estructuras secundarias y terciarias de alto orden para los transcritos correspondientes.

Para determinar la relevancia funcional de los marcadores CC para la regulación transcripcional, los inventores buscaron cualquier ejemplo de marcadores CC entre elementos reguladores ya definidos. Es importante mencionar que el algoritmo estudiado en genes humanos pudo identificar marcadores CC en eucariotas a través de muchas especies (Fig. 5). Los inventores atribuyeron esto a la función evolutivamente conservada mediada por las estructuras de alto orden del marcador.

Aquí, los inventores presentan dos ejemplos de los marcadores CC funcionalmente asociados a regulación transcripcional. El primer ejemplo del marcador CC se encontró dentro del gen beta-globina humano, estudiado ampliamente por sus propiedades por varios laboratorios. Informes recientes demostraron que la terminación de la transcripción en el gen beta-globina depende no sólo del reconocimiento del sitio poli(A), sino también del sitio de escisión (CoTC) co-transcripcional adicional en la dirección 3' (7, 12-14). Interesantemente, el sitio CoTC coincide con el marcador CC identificado y muestra bajas energías de plegamiento, como se menciona antes (Fig. 1B). Esta observación no sólo confirma la posible relevancia del marcador CC para la frontera del gen transcrito, sino que también sugiere su implicación funcional en el mecanismo de terminación transcripcional regulada.

El segundo ejemplo del marcador CC se encontró en el cromosoma X de *Drosophila melanogaster*, en el que coincidió con el aislador *gypsy* dentro de la banda cromosómica 7B2 (Fig. 1C). *Gypsy* es un elemento aislador de 350 pb bien caracterizado con múltiples sitios de unión Su(Hw), que dirigen estructuras similares a bucle de cromatina de orden superior (15). Un experimento hecho en el locus *cut* en *Drosophila* mostró que los dos sitios del aislador en las bandas cromosómicas 7B2 y 7B8 se ponen juntos en la periferia nuclear haciendo un bucle entre los loci (16) (Fig. 1C). También se ha mostrado una organización similar de fibras de cromatina mediadas por interferencia entre aisladores para las secuencias del aislador *scs* y *scs'* (17). En conjunto, estas observaciones están de acuerdo con el hecho de que, funcionalmente, los marcadores CC también pueden desempeñar un papel en la organización de estructuras de orden superior, que incluyen conformaciones de dominio subcromosómico que podrían detectarse por el ensayo de captura de conformación cromosómica (3C) previamente informado (18).

Con el fin de validar las observaciones anteriores, los inventores realizaron análisis sistemáticos de marcadores CC en dos genes humanos regulados (Fig. 2). Ambos genes modelo - el gen dihidrofolato reductasa (*DHFR*) regulado por el ciclo celular (19) y el gen del receptor similar al receptor de calcitonina (*CALCRL*) específico del tipo de célula (20-22) - muestran modos alternativos de actividad transcripcional regulada. En el análisis, los inventores estuvieron particularmente interesados en saber si los marcadores CC (i) podrían limitar el intervalo de transcripción por ARNPII y (ii) establecer una correlación con cualquier conformación cromosómica específica.

El *DHFR* humano (*hDHFR*) es un gen relacionado con el ciclo celular controlado a partir de los promotores secundarios en la dirección 5' y primarios en la dirección 3'. El gen abarca 28,5 kb en el cromosoma 5 y contiene 6 exones (Fig. 2A). Estudios independientes han mostrado que la expresión de *hDHFR* se induce tras la entrada en la fase S del ciclo celular y se desconecta en células quiescentes (G0) (Fig. 2A) (23). Mientras que en las fases G1/S la transcripción productiva del gen *hDHFR* es accionada por el promotor principal, en células quiescentes, la actividad transcripcional no es derogada, pero cambia a un modo alternativo - a partir del promotor secundario en la dirección 5' y que termina activamente en el segundo intrón. El transcrito del promotor secundario es inestable, pero podría detectarse en abundancia en células quiescentes por RT-PCR.

El gen *hDHFR* contiene tres marcadores CC: (i) en la dirección 5' a partir de ambos promotores (CC_{DHFR-1}); (ii) en el segundo intrón (CC_{DHFR-2}); (iii) en la dirección 3' a partir de la señal de poli(A) funcional (CC_{DHFR-3}) (Fig. 3A). Interesantemente, el análisis paralelo revela más de 40 señales de poli(A) crípticas presentes dentro del mismo gen. Los tres sitios CC descritos mostraron energía libre baja de plegamiento, característica de ácidos nucleicos monocatenarios altamente estructurados (Fig. 3A). Para determinar las propiedades de terminación de cada uno de los sitios CC, los inventores cuantificaron por RT-PCR la abundancia de transcritos *in vivo*, que incluye los inestables y los raros, en la dirección 5' y en la dirección 3' de los sitios CC. En los tres casos, los inventores encontraron pruebas de terminación de transcritos en los sitios CC (Fig. 3A). En el sitio CC_{DHFR-1}, los inventores detectaron terminación de la transcripción de transcritos intergénicos raros. En células quiescentes, el transcrito no codificante corto terminó en el sitio CC_{DHFR-2}. Un sitio AATAAA canónico también está presente próximo al sitio CC_{DHFR-2} y esa parte de la secuencia de ADN coincide con diversas marcas de secuencias expresadas y ADNc de la base de datos pública. En células proliferantes, el sitio CC_{DHFR-3} marcó la terminación de la transcripción productiva mencionada en cualquier sitio. La asociación del sitio CC_{DHFR-3} a la señal de poli(A) funcional es similar a la correlación previamente descrita dentro del gen beta-globina.

El segundo gen modelo de elección fue el gen *CALCRL* humano específico del tipo de célula (*hCALCRL*) (Fig. 2B). Codifica un receptor acoplado a la proteína G de los siete dominios transmembrana (GPCR). Los GPCR de mamífero constituyen una familia amplia y diversa de proteínas cuya función principal es transducir los estímulos

extracelulares en señales intracelulares. La mayoría de los GPCR responden a señales endógenas (endoGPCR) tales como péptidos, lípidos, neurotransmisores o nucleótidos. Los endoGPCR están altamente conservados y sus perfiles de expresión son únicos, dando miles de combinaciones de receptores específicos para tejido y célula para la modulación de procesos fisiológicos. El repertorio de endoGPCRs consiste en 367 receptores en seres humanos.

5 Sin embargo, los mecanismos que regulan su expresión y función específica siguen siendo ampliamente desconocidos. El endoGPCR codificado por el gen *hCALCRL* se considera que es una molécula clave en la regulación de la actividad de miembros de la familia de la calcitonina de péptidos que desempeñan funciones esenciales en el crecimiento, la supervivencia y la navegación celular. El gen *CALCRL* humano (103,15 kb) se localiza en el cromosoma 2 y contiene quince exones y se transcribe en diversos tejidos humanos y tumores. El gen
10 *hCALCRL* se transcribe hasta su longitud completa en células endoteliales y no en no endoteliales como se muestra por transferencia Northern e inmunohistoquímica (Fig. 2B). Sin embargo, en células no endoteliales podría detectarse un transcrito no codificante que termina en el primer intrón (Fig. 2B). Los inventores consideraron el gen *hCALCRL* como un buen modelo de regulación específica del tipo de célula de expresión génica (22).

15 Similar a *hDHFR*, los marcadores CC pudieron detectarse tanto en la dirección 5' del promotor ($CC_{CALCRL-1}$) como en la dirección 3' de la señal de poli(A) funcional ($CC_{CALCRL-3}$) del gen *hCALCRL*. Un tercer marcador CC adicional ($CC_{CALCRL-2}$) está presente en el primer intrón del gen (Fig. 3B). Una RACE en 5' a partir del primer exón confirma que todos los transcritos se inician en la dirección 3' y ninguno a partir de la dirección 5' del marcador $CC_{CALCRL-1}$. Esto sugiere que $CC_{CALCRL-1}$ podría terminar transcritos intergénicos que podrían interferir con la unidad de transcripción *hCALCRL*. Un análisis de RACE en 3' confirma la presencia de transcritos terminados próximos a
20 $CC_{CALCRL-2}$ (en el primer intrón) y sitios $CC_{CALCRL-3}$ (en la región en la dirección 3' del sitio de escisión). Los tres sitios del marcador CC muestran energía libre baja de plegamiento como se muestra anteriormente (Fig. 3B). Por tanto, *in vivo*, tanto en genes *hDHFR* y *hCALCRL*, los marcadores CC muestran propiedades de terminación transcripcionales.

25 Con el fin de validar la segunda propiedad sugerida del marcador CC, los inventores probaron entonces si participaban en cualquier conformación cromosómica específica como se define por el ensayo 3C. Este ensayo se desarrolló para monitorizar conformaciones cromosómicas altamente flexibles *in vivo* detectando la proximidad espacial de sitios distantes que participan en la formación de las estructuras similares a bucle. Los inventores han ajustado las condiciones del ensayo para mejorar el rendimiento y la sensibilidad de la detección en células humanas (véase *Materiales y procedimientos*). Y, lo que es más importante, la etapa inicial del ensayo también
30 implica el enriquecimiento de los loci cromosómicos transcritos con inmunoprecipitación anti-ARNPII (24).

35 Cuando se analizaron para el gen *hDHFR*, se encontró que los sitios de los marcadores CC_{DHFR-1} y CC_{DHFR-3} , posicionados separados más de 29 kb, se yuxtaponían en células proliferantes normales (Fig. 4A). La proximidad espacial de estos dos sitios fue altamente específica (Fig. 4A, compárense 1+2, 1+3 en células proliferantes) y dependiente en presencia de ARNPII, reticulación, restricción, ligación y PCR (Fig. 4A, controles de *hDHFR*). Como se muestra antes (Fig. 3A), estos dos sitios también muestran propiedades de terminación transcripcional en células proliferantes.

40 Los cambios en el modo transcripcional en el gen *hDHFR* bajo condiciones quiescentes se asocian, entre otras cosas, a generación de transcritos cortos que terminan dentro del segundo intrón. Y, lo que es más importante, el gen *hDHFR* contiene un tercer marcador CC posicionado en el mismo sitio. Análisis previos de la transcripción de *hDHFR* en estado quiescente indicaron que el marcador CC_{DHFR-2} se activó como un sitio de terminación para el transcrito no codificante corto (Fig. 3A). Por tanto, los inventores quisieron analizar si en estado quiescente un modo transcripcional diferente establecería una correlación con conformación cromosómica alternativa para el marcador CC_{DHFR-2} . De hecho, como se muestra en la Figura 4A, la conformación *in vivo* que yuxtapone los marcadores
45 CC_{DHFR-1} y CC_{DHFR-2} puede detectarse por el ensayo 3C en células quiescentes. Sólo bajos niveles de esta conformación se detectaron en la población de células proliferantes. Interessantemente, la conformación de $CC_{DHFR-1}:CC_{DHFR-2}$ observada no destruyó la conformación de $CC_{DHFR-1}:CC_{DHFR-3}$ descrita anteriormente para las células proliferantes. Teniendo en cuenta la naturaleza del ensayo 3C, este resultado podría tener varias explicaciones. En primer lugar, la conformación específica de quiescentes podría imponerse en exceso en la conformación de $CC_{DHFR-1}:CC_{DHFR-3}$ retenida. En segundo lugar, el resultado podría representar dos poblaciones de células ya que cambian
50 de una conformación a la otra. Y, lo que es más importante, la conformación de $CC_{DHFR-1}:CC_{DHFR-2}$ fue específica para el modo quiescente de transcripción y de acuerdo con el intervalo de transcritos detectados. Por tanto, los inventores han detectado para el gen *hDHFR* una conformación cromosómica *in vivo* caracterizada por proximidad espacial de marcadores CC. La proximidad de marcadores CC_{DHFR-1} y CC_{DHFR-2} fue específica para el modo transcripcional descrito para el estado quiescente del ciclo celular.

55 Para probar si los marcadores CC participaban o no en cualquier disposición estructural asociada a expresión específica del tipo de célula del gen *hCALCRL*, los inventores estudiaron sus conformaciones en células permisivas (endoteliales, HMVEC) y no permisivas (no endoteliales, HEK293T) para la transcripción. En células HMVEC, el gen *hCALCRL* activo muestra un perfil conformacional en el que los tres marcadores CC_{CALCRL} se yuxtapusieron con estrecha proximidad entre $CC_{CALCRL-1}:CC_{CALCRL-2}$ y $CC_{CALCRL-1}:CC_{CALCRL-3}$ (Fig. 4B; no se muestran los datos para
60 $CC_{CALCRL-2}:CC_{CALCRL-3}$). Y, lo que es más importante, las fronteras de estas dos posibles conformaciones de bucle se correspondieron con las fronteras de los dos transcritos detectados en células HMVEC (Fig. 2B). Para probar si cualquiera de estas conformaciones era única para células HMVEC, los inventores analizaron *hCALCRL* en células

HEK293T transcripcionalmente no permisivas. Aunque los inventores todavía detectaban la yuxtaposición de CC_{CALCRL}-1 y CC_{CALCRL}-2, la interacción entre CC_{CALCRL}-1 y CC_{CALCRL}-3, que engloba la longitud completa del gen *hCALCRL*, ya no estaba presente (Fig. 4B). La conformación de CC_{CALCRL}-1:CC_{CALCRL}-2 coincide con la presencia de transcritos de *hCALCRL* cortos que terminan en el primer intrón en el sitio CC_{CALCRL}-2 en células HEK293T (Fig. 2B).
 5 Por tanto, la expresión específica del tipo de células del gen *hCALCRL* está asociada a una conformación cromosómica única, como se detecta entre marcadores CC_{CALCRL}-1 y CC_{CALCRL}-3. Y, lo que es más importante, esta conformación engloba la longitud completa de los transcritos productivos generados en células HMVEC.

La aplicación del análisis de reconocimiento de patrones a las fronteras de 422 genes humanos anotados ha identificado y definido varios marcadores, que incluyen un marcador previamente desconocido implicado en la regulación transcripcional. El marcador - *Checkpoint Charlie* - establece coherentemente una correlación con las fronteras de unidades transcripcionales codificantes y no codificantes en diversos espectros de especies (véase también la Fig. 5), muestra estructuras secundarias y terciarias altamente ordenadas para los transcritos correspondientes, se asocia a la terminación de la transcripción regulada por ARNPII *in vivo* y dirige la formación de conformaciones cromosómicas alternativas dependientes de la transcripción. Sorprendentemente, cuando se analiza
 10 en los genes *hDHFR* específicos del ciclo celular y *hCALCRL* específicos del tipo de célula, el marcador se asocia funcionalmente a las distintas conformaciones estructurales de alto orden que son características de uno o de los otros modos de la actividad transcripcional. Los marcadores CC yuxtapuestos no sólo establecen una correlación con estructuras de sub-cromatina cargadas con ARNPII, sino que también perfilan las fronteras de los transcritos sintetizados dentro de aquellas estructuras. Los datos de los inventores están de acuerdo con sugerencias previas de que las estructuras de alto orden se forman en un modo dependiente de la transcripción y podrían ser
 15 importantes para la reiniciación transcripcional.

La regulación transcripcional se realiza a diversos niveles importantes por una multitud de actividades ligadas a reclutamiento específico de secuencias de ADN, modificación de cromatina y remodelación de marcadores CC y organización estructural asociada que están claramente implicadas *in vivo* en el establecimiento de las fronteras
 20 externas para diversas unidades transcripcionales.

Transferencia Northern

Se realizó transferencia Northern para *hDHFR* a partir de ARN total aislado de células U2OS. Las células proliferantes se cultivaron en presencia de 10% de SBF mientras que la célula quiescente se logró bajo inhibición por contacto en presencia de 0,5% de SBF. Como sonda se usaron sondas sintetizadas usando un molde que engloba
 30 secuencias entre el cuarto y el sexto exón de *hDHFR*.

La transferencia Northern para *hCALCRL* se realizó como se ha descrito previamente (25). La CL humana de longitud completa se amplificó por RT-PCR y se clonó en el vector pcDNA3.1. El vector resultante se secuenció usando un analizador Applied Biosystems 377 Genetic y la secuencia se cotejó con la base de datos GenBank. El inserto se escindió y se usó como molde para generar sondas.

35 En cualquier caso, las sondas se marcaron con ³²P-dCTP usando el kit de marcado MegaPrime (Amersham, RU). Después de la hibridación y lavados rigurosos, la transferencia se expuso a Hyperfilm (Amersham, RU) y luego a la pantalla de fósforo. Las señales de hibridación se analizaron usando el software ImageQuant.

Citometría de flujo activada por fluorescencia (FACS)

La citometría FACS de células en crecimiento U2OS y quiescentes se realizó como se ha descrito previamente (26).

Reacción en cadena de la polimerasa por transcripción inversa (RT-PCR)

La PCR por transcripción inversa para determinar la terminación de transcritos en *hDHFR* se realizó en ARN total aislado de células U2OS. Se usaron los siguientes cebadores directos e inversos para los sitios CC_{DHFR}-1, CC_{DHFR}-2 y CC_{DHFR}-3:

CC_{DHFR}-1

45 Cebador directo (A): tggggaactgcacaatatga (SEC ID N°: 1)

Cebador inverso (B): aggggtgcgtcttttaacct (SEC ID N°: 2)

Cebador inverso (C): ccgcacgtagtaggttctgtc (SEC ID N°: 3)

CC_{DHFR}-2

Cebador directo (A): ttccagagaatgaccacaacc (SEC ID N°: 4)

50 Cebador inverso (B): tgtcttttgatcgtggtg (SEC ID N°: 5)

Cebador inverso (C): tggggtatctaaccgatttg (SEC ID N°: 6)

CC_{DHFR}-3

Cebador directo (A): ttggaaaaacccatgaagg (SEC ID N°: 7)

Cebador inverso (B): caacagtctgccagttgtt (SEC ID N°: 8)

Cebador inverso (C): cagggtttgtctgtcacc (SEC ID N°: 9)

5 La RT-PCR se realizó usando el kit de transcripción inversa Omniscript de Qiagen, RU.

Amplificación rápida de extremos de ADNc (RACE)

La RACE se realizó esencialmente como se ha descrito previamente (27). Los cebadores específicos de gen se diseñaron para RACE en 3' (cagagagtgccacctctgcttagg) (SEC ID N°: 10) y 5' (cccacaagcaaggtgggaagagtg) (SEC ID N°: 11) basándose en la secuencia de ADNc de *CALCRL* humano (28). Los transcritos de RACE en 5' y 3' (que terminan en el primer intrón) se secuenciaron y se enviaron a la base de datos GenBank.

Producción y caracterización de anticuerpos

El anticuerpo policlonal de conejo LN-1436 se produjo contra el péptido sintético correspondiente a los residuos 427-461 (HDIENVLLKPENLYN) (SEC ID N°: 12) en el extremo C de la proteína CL humana (hCL) (números de acceso AAC41994 y AAA62158; codificado por el gen *CALCRL*). La especificidad de los anticuerpos se caracterizó por análisis de inmunotransferencia de CL transitoriamente expresada en células HEK293T.

Inmunocitoquímica

Especímenes incorporados en parafina fijados con formalina (n=74) de 20 tejidos humanos normales se seleccionaron de los documentos de archivo del Departamento de patología celular, Hospital John Radcliffe; Universidad de Oxford, Oxford, RU. Se produjeron múltiples micromatrices de tejido (TMA) adquiriendo núcleos cilíndricos (1,0 mm de diámetro) para cada espécimen dispuesto a alta densidad en un bloque de TMA del receptor (29). El procedimiento de recuperación del antígeno se llevó a cabo en secciones sin cera y rehidratadas de 4 µm antes de realizar la inmunohistoquímica usando el anticuerpo anti-hCL LN-1436. La inmunohistoquímica se realizó esencialmente como se ha descrito previamente (30). Se usaron anticuerpos secundarios biotinizados, el kit Vectastain ABC-AP del complejo de estreptavidina-fosfatasa alcalina y el sistema de detección Vector Red (todos de Vector, Burlingame, US). Los controles incluyeron suero de conejo preinmune usado a concentraciones apropiadas.

Captura de conformación cromosómica (3C)

El análisis 3C se realizó como se ha descrito previamente (31) con las siguientes modificaciones. Se reticularon aproximadamente 4×10^6 células completas tratando con 2% de formaldehído a temperatura ambiente durante 10 min. La reticulación se detuvo con cantidad equimolar de glicina y las células se recogieron y se lisaron en tampón hipotónico (Tris-HCl 10 mM [pH 7,2], MgCl₂ 2 mM y 0,5% de Triton X-100). Entonces, los núcleos se resuspendieron y se incubaron durante 20 min sobre hielo en tampón CSK (NaCl 100 mM, sacarosa 300 mM, PIPES 10 mM [pH 6,8], MgCl₂ 3 mM, leupeptina 10 µM, EGTA 1 mM, PMSF 1,2 mM y 0,5% de Triton X-100). La suspensión se centrifugó durante 5000 rpm a 4°C en una centrifuga Hettich Mikro 22R y el sedimento se trató con NaCl 2 M. Después de incubarse durante 10 min sobre hielo se añadió una cantidad suficiente de agua para reducir la concentración de NaCl a 150 mM. Esta muestra se usó para realizar el ensayo de inmunoprecipitación de cromatina con ARNPII como se ha descrito previamente (32). La cromatina inmunoprecipitada con el anticuerpo ARNPII (H-224, Santa Cruz Biotechnology Inc., EE.UU.) se restringió luego con la enzima de restricción *BglIII* (New England Biolabs, RU) y se ligó a ADN ligasa T4 (Roche, RU). Después de digerirse las proteínas con proteinasa K (Roche, RU) y el ARN con ribonucleasa A (Sigma, RU), el ADN se extrajo con etanol. El análisis por PCR en el ADN extraído se hizo usando cebadores específicos de genes con TakaRa LA Taq™ de Takara Bio Inc., Japón.

Diagnóstico de cáncer de ovario y de próstata

Expresión de MLH1 en tejidos normales y con cáncer de ovario (véase la Fig. 8)

Los genes supresores de tumores desempeñan una función vital en la supervivencia y el mantenimiento de células. El silenciamiento de supresores tumorales señala el crecimiento incontrolado que conduce a cáncer. Como mecanismo de seguridad, las células experimentan apoptosis cuando se detectan tales señales para crecimiento incontrolado.

Un homólogo humano del gen mutL de *Escherichia coli*, cáncer de colon sin poliposis, tipo 2 (MLH1), es un gen tal que codifica un gen de reparación del mal apareamiento de ADN. MLH1 señala el mecanismo de reparación iniciado por la lesión del ADN e induce apoptosis de células tumorales. Este gen se localizó en los loci 3p21.3 y acumula diversas mutaciones y modificaciones a medida que envejecen las células. Un cambio tal – aumento de los niveles de metilación en la región promotora de MLH1 – se ha asociado a cáncer hereditario de colon sin poliposis. Por tanto, se ha mostrado que las variantes de corte y empalme de MLH1 alternativas son específicas de tejido y contribuyen a la variabilidad fenotípica en cánceres heredados.

Para ver si las variaciones de corte y empalme inducidas por la mutación de MLH1 están asociadas a cáncer de ovario, los inventores buscaron sitios CC que englobaban la unidad de transcripción. Escaneando la secuencia de MLH1, los inventores encontraron un marcador CC en el 8° intrón y otro en las fronteras formadas en la UTR de 3' de una variante de corte y empalme alternativa. El análisis 3C realizado en estos dos sitios muestra que los sitios CC sólo se yuxtaponen en pacientes normales. Mientras tanto, las muestras de tejido y fluido recogidas de pacientes con cáncer de ovario no revelan yuxtaposición. Por tanto, los sitios CC de MLH1 pueden usarse como marcador para distinguir cáncer de ovario.

Cáncer de próstata

Las pruebas para marcadores de diagnóstico de próstata se realizaron en líneas celulares que representaban tanto estado benigno como tardío de crecimiento tumoral. Los genes de elección fueron PSA y BORIS.

Expresión de BORIS y PSA en tejidos normales y de cáncer de próstata (véase la Fig. 9)

Un novedoso miembro de la familia de los genes de cáncer-testículos, el hermano del regulador de los sitios impresos (BORIS), se expresa sólo en espermatozoides y no en células somáticas normales. Sin embargo, su expresión se ha asociado a varios cánceres humanos que incluyen cáncer de mama y de pulmón. BORIS compete con otro factor de transcripción del dedo de Zn, CTCF, por las perturbaciones epigenéticas en tumores malignos humanos. De ahí que los inventores decidieran probar la asociación de BORIS con carcinoma de próstata humano (LNCaP).

BORIS tiene dos sitios CC que engloban la unidad de transcripción definida en la localización cromosómica 20q13.31. Como el gen se expresa significativamente en tumores malignos, los inventores decidieron probar la yuxtaposición de dos sitios CC en LNCaP. De los resultados mostrados en la figura adjunta, la yuxtaposición de sitios CC sólo se produce en LNCaP y no en líneas celulares de osteosarcoma humano (U2OS). Se estableció otra confirmación por secuenciación del producto de PCR.

Los inventores también miraron en otro marcador de cáncer de próstata bien establecido, el antígeno específico de próstata (PSA). El PSA codificado por el gen calicreína humana 3 (KLK3) se usa para el diagnóstico y el pronóstico de cáncer de próstata detectando los niveles de proteína PSA en sangre. Sin embargo, los inventores usaron aquí la técnica 3C para mirar el gen PSA en células de osteosarcoma humano y líneas celulares de hiperplasia benigna de próstata (BPH1). Como se observa en BORIS, la unidad de transcripción KLK3 se define por dos sitios CC, uno en la UTR de 5' y el otro en la UTR de 3'. Los resultados muestran que estos dos sitios CC sólo interfieren en células BPH1 y no en U2OS.

Por tanto, PSA y BORIS pueden usarse como biomarcadores para identificar células de cáncer de próstata benigno y maligno, respectivamente.

Procedimientos de PCR

MLH1

Enzima de restricción de 3C - *BssSI*

35 Cebadores de MLH1

MF3UTR2 TGGTTTTAGCTGGGATGGAG
MF3UTR1 GAGGCAGGCAGATCACTTGT

MREI2 AGAAGATGCAGGCCAACAAT
MREI1 CTCGTAAAGCCCAAGGAGGT

40 Primera ronda de reacción de PCR

| | |
|--------------------------|--------|
| 2x tampón I | 25 µl |
| dNTP (2,5 mM) | 8 µl |
| ADN | 1 µl |
| <i>Cebadores (25 µM)</i> | |
| Directo (MREI2) | 1 µl |
| Inverso (MF3UTR2) | 1 µl |
| <i>TakaRa LA Taq</i> | 0,5 µl |

Agua 13,5 µl

Total 50 µl

Cebadores

MREI2 - MF3UTR2

Programa de PCR

94°C - 5 min

94°C -1 min

57°C -1 min durante 30 ciclos

72°C - 45 s

72°C - 5 min

Tamaños de producto esperados

5 MREI2 - MF3UTR2 - 527 pb

Segunda ronda de reacción por PCR

2x tampón I 25 µl

dNTP (2,5 mM) 8 µl

ADN 2 µl

Cebadores (25 µM)

Directo (MREI1) 1 µl

Inverso (MF3UTR1) 1 µl

TakaRa LA Taq 0,5 µl

Agua 12,5 µl

Total 50 µl

Cebadores

MREI1-MF3UTR1

Muestras

10 Tomar 48 µl de mezcla y 2 µl de reacción de PCR respectiva de la 1ª ronda

Programa de PCR

94°C - 5 min

94°C - 1 min

59°C - 1 min durante 25 ciclos

72°C - 30 s

72°C - 5 min

Tamaños de producto esperados

MREI1 - MF3UTR1 -325 pb

BORIS

15 Enzima de restricción de 3C - *TaqI*

Cebadores de BORIS

BR5UTR4 GGCTGGAATTGCCCTAAAGT
BR5UTR3 CCTATGAGGGGGCAGTATCA

BR3UTR2 GCTCTTCCTGCTGGGAAAT
BR3UTR1 TACAGGGGTGGAGACAGGTT

5 Primera ronda de reacción de PCR

| | |
|--------------------------|--------------|
| 2x tampón I | 25 µl |
| dNTP (2,5 mM) | 8 µl |
| ADN | 1 µl |
| <i>Cebadores (25 µM)</i> | |
| Directo (BR5UTR4) | 1 µl |
| Inverso (BR3UTR2) | 1 µl |
| <i>TakaRa LA Taq</i> | 0,5 µl |
| Agua | 13,5 µl |
| Total | 50 µl |

Cebadores

BR5UTR4 - BR3UTR2

Programa de PCR

94°C - 5 min

94°C - 45 s

57°C - 30 s durante 30 ciclos

72°C - 25 s

72°C - 5 min

Tamaños de producto esperados

10 BR5UTR4 - BR3UTR2 - 430 ó 784 pb

Nota: Se facilitan dos tamaños de producto porque la enzima de restricción de 3C (Taq I) se escinde en cualquiera de los dos sitios de restricción próximos al marcador CC.

Segunda ronda de reacción de PCR

| | |
|--------------------------|--------------|
| 2x tampón I | 25 µl |
| dNTP (2,5 mM) | 8 µl |
| ADN | 2 µl |
| <i>Cebadores (25 µM)</i> | |
| Directo (BR5UTR3) | µl |
| Inverso (BR3UTR1) | 1 µl |
| <i>TakaRa LA Taq</i> | 0,5 µl |
| Agua | 12,5 µl |
| Total | 50 µl |

Cebadores

15 BR5UTR3 - BR3UTR1

Muestras

Tomar 48 µl de mezcla y 2 µl de reacción de PCR respectiva de la 1ª ronda

Programa de PCR

94°C - 5 min

94°C - 45 s

55°C - 30 s durante 25 ciclos

72°C - 20 s

72°C - 5 min

Tamaños de producto esperados

- 5 BR5UTR3 - BR3UTR1 - 260 ó 564 pb

Nota: Aquí se facilitan dos tamaños de producto porque la enzima de restricción de 3C (Taq I) se escinde en cualquiera de los dos sitios de restricción próximos al marcador CC. La Figura 9 muestra la banda de 564 pb, que se ha verificado por secuenciación.

PSA

- 10 Enzima de restricción de 3C - *TaqI*

Cebadores de PSA

PR5UTR2 CGTGATCCACCCATCTCAG
 PR5UTR1 CTATTGGGAGACCGAAGCAG

- 15 PF3UTR2 GGGAAAGGGAGAAGATGAGG
 PF3UTR1 TAGGGGAAGGTTGAGGAAGG

Primera ronda de reacción de PCR

2x tampón I 25 µl

dNTP (2,5 mM) 8 µl

ADN 1 µl

Cebadores (25 µM)

Directo (PR5UTR2) 1 µl

Inverso (PF3UTR2) 1 µl

TakaRa LA Taq 0,5 µl

Agua 13,5 µl

Total 50 µl

Cebadores

PR5UTR2-PF3UTR2

Programa de PCR

94°C - 5 min

94°C - 45 s

61°C - 30 s durante 30 ciclos

72°C - 25 s

72°C - 5 min

Tamaños de producto esperados

PR5UTR2 -PF3UTR2 - 481 pb

Segunda ronda de reacción de PCR

| | |
|--------------------------|--------------|
| 2x tampón I | 25 µl |
| dNTP (2,5 mM) | 8 µl |
| ADN | 2 µl |
| <i>Cebadores (25 µM)</i> | |
| Directo (PR5UTR1) | 1 µl |
| Inverso (PF3UTR1) | 1 µl |
| <i>TakaRa LA Taq</i> | 0,5 µl |
| Agua | 12,5 µl |
| Total | 50 µl |

Cebadores

5 PR5UTR1 - PF3UTR1

Muestras

Tomar 48 µl de mezcla y 2 µl de reacción de PCR respectiva de la 1ª ronda

Programa de PCR

94°C - 5 min
 94°C - 45 s
 61°C - 30 s durante 25 ciclos
 72°C - 20 s
 72°C - 5 min

Tamaños de producto esperados

10 PR5UTR1 - PF3UTR1 - 266 pb

Detalles de marcadores CC

MLH1

CC1 - 24367 pb en la dirección 3' de TSS

TAACCCCAT

15 CC2-57357 pb en la dirección 3' de TSS

TAACATAA

(Las letras subrayadas en negrita representan la secuencia del marcador CC)

20 En tejido normal, el gen se expresa con transcritos alternativos. Un transcrito tal empieza en el 8º intrón en el que está presente CC1 y termina en el marcador CC2. En tejido de cáncer de ovario, el gen está regulado por disminución ya que acumula mutaciones, deleciones y metilación que conducen a transcritos defectuosos o a no transcritos. Los inventores encontraron la yuxtaposición de CC1 y 2 en tejidos normales, y no en tejidos de cáncer de ovario. Esto se refiere al cambio en el modo transcripcional del gen en estos tejidos.

BORIS

CC1 - 5282 pb en la dirección 5' de TSS

CTTTGAAAGC

CC2 - 28038 pb en la dirección 3' de TSS

AAAA**T**TGCT

(Las letras subrayadas en negrita representan la secuencia del marcador CC)

- 5 BORIS tiene dos sitios CC, uno en la UTR de 5' y el otro en la UTR de 3'. En células U2OS no se espera la expresión de BORIS y de ahí que no deba observarse yuxtaposición de marcadores CC. Mientras que, en la línea celular de carcinoma de próstata humano (LNCaP), BORIS se expresa. Los inventores encontraron una yuxtaposición de CC1 y CC2 en LNCaP y no en U2OS.

PSA/KLK3

- 10 CC1 - 408 pb en la dirección 5' de TSS

CTGG**TCTC**AGAGT

CC2 - 5843 pb en la dirección 3' de TSS

TACTGTGGTTTA

(Las letras subrayadas en negrita representan la secuencia del marcador CC)

- 15 KLK3 tiene dos sitios CC, uno en la UTR de 5' y el otro en la UTR de 3'. En células U2OS no se espera la expresión de KLK3 y de ahí que no deba observarse yuxtaposición de marcadores CC. Mientras que, en la línea celular de hiperplasia prostática benigna (BPH-1), KLK3 se expresa. De ahí que la yuxtaposición de CC1 y CC2 se observe en pbH-1 y no en U2OS.

Referencias

- 20 1. P. R. Cook, I. A. Brazell, E. Jost, *Journal of Cell Science* **22**, 303 (noviembre, 1976).
 2. T. Cremer, C. Cremer, *Nat Rev Genet* **2**, 292 (Apr, 2001).
 3. D. Carter, L. Chakalova, C. S. Osborne, Y. F. Dai, P. Fraser, *Nature Genetics* **32**, 623 (diciembre, 2002).
 4. J. E. Collins y col., *Genome Research* **13**, 27 (enero, 2003).
- 25 5. T. A. Down, T. J. Hubbard, *Genome Research* **12**, 458 (marzo, 2002).
 6. M. E. Tipping, *Journal of Machine Learning Research* **1**, 211 (1 de junio de 2001).
 7. M. J. Dye, N. J. Proudfoot, *Cell* **105**, 669 (1 de junio de 2001).
 8. N. J. Proudfoot, A. Furger, M. J. Dye, *Cell* **108**, 501 (22 de febrero de 2002).
 9. G. Yeung y col., *Molecular and Cellular Biology* **18**, 276 (enero, 1998).
- 30 10. M. Yonaha, N. J. Proudfoot, *EMBO Journal* **19**, 3770 (17 de Julio de 2000).
 11. M. Zuker, *Nucleic Acids Research* **31**, 3406 (1 de Julio de 2003).
 12. A. Teixeira y col., *Nature* **432**, 526 (25 de noviembre de 2004).
 13. S. West, N. Gromak, N. J. Proudfoot, *Nature* **432**, 522 (25 de noviembre de 2004).
 14. M. Kim y col., *Nature* **432**, 517 (25 de noviembre de 2004).
- 35 15. T.I. Gerasimova, V. G. Corces, *Cell* **92**, 511 (20 de febrero de 1998).
 16. K. Byrd, V. G. Corces, *Journal of Cell Biology* **162**, 565 (18 de agosto de 2003).
 17. J. Blanton, M. Gaszner, P. Schedl, *Genes and Development* **17**, 664 (1 de marzo de 2003).
 18. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* **295**, 1306 (15 de febrero de 2002).
 19. J. E. Slansky, P. J. Farnham, *Bioessays* **18**, 55 (enero, 1996).
- 40 20. B. Fluhmann, M. Lauber, W. Lichtensteiger, J. A. Fischer, W. Born, *Brain Research* **774**, 184 (7

de noviembre de 1997).

21. N. Aiyar y col., *Journal of Biological Chemistry* 271, 11325 (10 de mayo de 1996).
22. L. L. Nikitenko, D. M. Smith, R. Bicknell, M. C. Rees, *FASEB Journal* 17, 1499 (agosto, 2003).
- 5 23. S. L. Hendrickson, J. S. Wu, L. F. Johnson, *Proceedings of the National Academy of Sciences of the United States of America* 77, 5140 (septiembre, 1980).
24. R. Metivier y col., *Cell* 115, 751 (12 de diciembre de 2003).
25. L. L. Nikitenko y col., *Molecular Human Reproduction* 7, 655 (julio, 2001).
26. Z. Darzynkiewicz, *The Cell Cycle. A Practical Approach*. P. Fantes, R. Brooks, Eds. (IRL Press, Oxford, 1993), pág. 45-68.
- 10 27. L. L. Nikitenko, D. M. Smith, R. Bicknell, M. C. Rees, *FASEB Journal* 17, 1499 (agosto, 2003).
28. N. Aiyar y col., *Journal of Biological Chemistry* 271, 11325 (10 de mayo de 1996).
29. J. Kononen y col., *Nature Medicine* 4, 844 (Jul, 1998).
30. L. L. Nikitenko, I. Z. MacKenzie, M. C. Rees, R. Bicknell, *Molecular Human Reproduction* 6, 811 (septiembre, 2000).
- 15 31. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* 295, 1306 (15 de febrero de 2002).
32. R. Metivier y col., *Cell* 115, 751 (12 de diciembre de 2003).

Una descripción de marcadores CC y su detección

El análisis de reconocimiento de patrones se ha aplicado ampliamente a diversos campos de estudio tales como medicina, ingeniería y lingüística en los que el análisis de imágenes y la descodificación de datos permiten la
 20 identificación de marcadores característicos subyacentes dentro de sistemas complejos. Los inventores han usado metodología de reconocimiento de patrones para analizar datos del genoma humano en relación con las unidades transcripcionales, procesadas por ARN polimerasa II. Se usó un conjunto de secuencias de 422 genes manualmente anotados en el cromosoma humano 22 para la identificación computacional de señales reguladoras presentes en las
 25 fronteras de las unidades transcripcionales. Se dio atención particular a la identificación de las señales en el extremo 3' de las unidades de transcripción. Esto demostró ser funcionalmente relevante ya que experimentos posteriores confirmaron que las señales tenían propiedades de terminación *in vivo*.

El patrón encontrado en las fronteras tiene señales múltiplex y se representa en un formato XML que explica 3 aspectos clave.

- a. Los alfabetos del ADN de cada señal identificada
- 30 b. La variación posicional de cada señal como anchura de distribución gaussiana
- c. Distancia entre cada señal en un patrón en par de bases

Como los patrones se observan en las fronteras de unidades transcripcionales, los inventores lo llamaron marcador 'Checkpoint Charlie' (CC).

Los marcadores CC en una secuencia desconocida pueden identificarse usando un conjunto de código identificado
 35 como 'escáner'. El escáner necesita 3 datos de entrada del usuario

- a. La secuencia en estudio
- b. El patrón en formato XML
- c. Un factor de rigurosidad (puntuación logarítmica inversa) para descartar marcadores CC débiles (valor por defecto: 0,99, por ejemplo)

40 El escáner lee el ADN de entrada e intenta ajustar los patrones en la secuencia. Esto se hace pasando por la secuencia de ADN tomando cada base como punto de referencia. El escáner empieza con la primera base como punto de referencia e intenta ajustar el patrón definido en el formato XML. El grado de aptitud se determina por una puntuación. Si esta puntuación es mayor que el factor de rigurosidad suministrado por el usuario, se encontró un marcador CC. La posición del marcador CC identificada se facilita en un formato de GFF patrón y el escáner se
 45 mueve a la segunda base en la secuencia de entrada.

Este procedimiento se repite hasta que el escáner lee todas las bases en el ADN de entrada y su cadena complementaria.

Los resultados finales de este escaneado para el patrón del marcador CC serán un archivo de texto con posibles posiciones del marcador CC en la secuencia de entrada con su puntuación respectiva en el formato GFF.

5 Detección del marcador CC

Para ilustrar la detección del marcador CC en una secuencia dada se considera la siguiente secuencia.

```

ATATTTGTACTATGGCTCTGAATAAATAATAAGGACAGGAAGCCCGGAGAAGGAGAG
TTTTTTTTTTTTTTTTGGTACGAGAACTCTCTGTACTATTTTTTTCAACTTTTCTTTTT
CTTTTCTTTTGAGACGGAGTCTTACTCTTCTTGCCAGGCTGGAGTGCAATGGCGCG
ATCTCGGCTCACTGCAACCTCCACCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCC
TCCAAGTAGCTGGGATTACAGGCATGTGCCACCATGCCTGGCTAATTTTGTATTTT
TAGTAGAGATGGGGTTTCACCATGAGCGCCAGGCTGGTCTTGAACACCTGACCTCG
TGATCCACCTGCCTCGGCCTCCCAAAGTACTGGGACTACAGGTATGAGCCACTGTGC
CCAGCCGACAAAAC
  
```

10 Dada esta secuencia, se hace un escaneado de izquierda a derecha para encontrar el marcador CC. Ahora se considera la 50ª base (subrayada) como punto de referencia de los inventores. Para determinar si esta base es un marcador CC o no, los 4 conjuntos de pesos descritos en la Tabla 1 deberían coincidir con esta secuencia. Por sencillez se muestra un ejemplo en el que los 4 conjuntos de pesos (también subrayados) están presentes.

15 Como se ha descrito antes, los 4 conjuntos de pesos tienen una distancia relativa entre sí con respecto al punto de referencia. Por ejemplo, de la Tabla 1 puede observarse que el primer conjunto de peso empieza en la posición 8 con respecto al punto de referencia. Este primer conjunto de peso tiene 19 valores posicionales para cada tipo de aparición de nucleótidos en esa posición. Por ejemplo, para la primera posición, una guanina conseguirá un valor de 0,19 y una timina puntuará 0,33. Asimismo, para la segunda posición, una guanina puntuará 0,20 y una timina puntuará 0,39. La segunda posición se multiplica por la primera puntuación. Esto se repite hasta que los 19 valores posicionales sean leídos y multiplicados por su valor previo.

20 En el ejemplo de los inventores, los inventores tienen TTTTTTTTTTTTTTTGGT que empieza en una 8ª base en relación con el punto de referencia. De ahí que la puntuación de los inventores para este conjunto de peso sea $(0,33 \times 0,39 \times 0,34 \times 0,35 \times 0,41 \dots)$, etc.

Este procedimiento también se repite para los otros 3 conjuntos de pesos, multiplicando cada vez el valor posicional por la puntuación previa calculada hasta ese momento.

25 La puntuación final de los 4 conjuntos de pesos se convierte en una puntuación (logarítmica inversa) de valor exponencial para su fácil manipulación. La puntuación logarítmica es igual a $1,0/(1+e^{-X})$ en la que X es la puntuación obtenida por el procedimiento anterior usando los pesos en la Tabla 1. Si esta puntuación logarítmica es superior a 0,90 (por ejemplo), entonces esa base se considera marcador CC. En el ejemplo de los inventores, la multiplicación de los valores posicionales de los 4 conjuntos de pesos dio una puntuación logarítmica inversa de 0,99999. Como este valor es superior a 0,99, la 50ª base, A, está dentro de la secuencia del marcador CC. El analizar otras bases
30 en la secuencia permite la identificación de la secuencia desde la base 41ª hasta la base 56ª como marcador CC (con una puntuación final de 0,99968).

Procedimiento usado en detectar la yuxtaposición de marcadores CC *in vivo*

35 El procedimiento descrito más adelante identifica ampliamente las etapas clave en la detección de la yuxtaposición de marcadores CC en muestras de tejido. Esta es la primera metodología desarrollada para analizar muestras congeladas de tejido de pacientes.

- Las muestras de tejido se rebanan en secciones delgadas sobre un portaobjetos de vidrio
- Añadir 1 ml de 1x PBS frío en hielo al portaobjetos y lavar durante 5 min
- Añadir paraformaldehído 0,67 M para reticular proteína y ADN
- 40 • Incubar durante 10 min a temperatura ambiente en una plataforma de balanceo
- Añadir glicina 1 M para extinguir la reacción de reticulación

ES 2 368 767 T3

- Raspar las células y transferir las células a una Eppendorf
- Centrifugar a 13.000 rpm durante 1 min para recoger las células a temperatura ambiente
- Eliminar el sobrenadante y añadir 1 ml de tampón hipotónico helado
- 5 • Pipetear las células una cuantas veces para preparar una fina suspensión de células (si se requiere, centrifugar un poco rápidamente)
- Incubar sobre hielo durante 10 min para hinchar la célula y que los núcleos emerjan
- Centrifugar a 5.000 rpm durante 5 min a 4°C para recoger los núcleos
- 10 • Drenar el sobrenadante de citosol y disolver el sedimento de núcleos en 1 ml de tampón CSK
- Incubar sobre hielo durante 20 min
- Centrifugar a 5.000 rpm durante 5 min a 4°C para recoger los núcleos
- Drenar el sobrenadante en la medida de lo posible y retener el sedimento
- Disolver el sedimento de células en NaCl 2 M (la disolución se vuelve viscosa)
- 15 • Incubar sobre hielo durante 10 min
- Diluir la muestra con agua suficiente para reducir la concentración de NaCl a 150 mM
- Añadir 10 µl de anticuerpo Pol II (H-224) a la Eppendorf
- Incubar a 4°C durante la noche con agitación o rotación
- 20 • Tomar 30 µl de la suspensión de perlas de proteína G-Sepharose para conseguir aproximadamente 20 µl de perlas secas (cortar la punta de la pipeta si se requiere)
- Centrifugar a 2.000 rpm durante 3 min para recoger las perlas
- Lavar dos veces con 1 ml de agua MilliQ y centrifugar a 2.000 rpm durante 3 min para recoger las perlas
- Añadir 1 ml de tampón de lavado de restricción a las perlas
- 25 • Mezclar bien y dispensar a diferentes Eppendorfs (si se requiere), lavar y centrifugar a 2.000 rpm durante 3 min para recoger las perlas
- Transferir el contenido completo a la Eppendorf con perlas y mezclar bien
- Incubar a 4°C durante 1 hora con agitación o rotación
- Centrifugar a 1000 rpm durante 3 min a 4°C y eliminar el sobrenadante. El sobrenadante puede analizarse para fracciones sin unir.
- 30 • Añadir 1 ml de tampón de lavado de restricción, rotar a 4°C durante 5 min, centrifugar a 2000 rpm durante 3 min a 4°C. Eliminar el sobrenadante.
- Añadir 1 ml de tampón de lavado de restricción, rotar a 4°C durante 5 min, centrifugar a 2000 rpm durante 3 min a 4°C. Eliminar el sobrenadante.
- 35 • Añadir 1 ml de tampón de lavado de restricción, rotar a 4°C durante 5 min, centrifugar a 2000 rpm durante 3 min a 4°C. Eliminar el sobrenadante.
- Medir las perlas y la cantidad de tampón de restricción que queda, añadir
 - Tampón de restricción 1x
 - Enzima de restricción 30-60 unidades

ES 2 368 767 T3

- Agua Variable para 100 µl de reacción
- Digerir el ADN incubando a 37°C durante la noche
- Incubar a 65°C durante 10 min para detener la digestión por restricción
- Añadir > 200 µg/ml de RNasa A al tampón
- 5 • Incubar a 37°C durante 30 min
- Añadir 400 µl de agua MilliQ y diluir la reacción de restricción
- Añadir
 - Tampón de ligación 1x
 - ADN ligasa T4 30 unidades
- 10 • Agua Variable para 100 µl de reacción
- Incubar a 16°C durante 4 h
- Incubar a 65°C durante la noche para reticulaciones inversas
- Añadir 450 µg de proteinasa K a cada muestra
- Incubar a 42°C durante 1 hora para digerir las proteínas
- 15 • Añadir 660 µl de fenol, pH 7,9 (volumen igual) a cada muestra y remover con vórtex
- Centrifugar a 13.000 rpm durante 10 min
- Transferir el sobrenadante a la Eppendorf de 1,5 ml
- Añadir 0,3 M de NaCl y 0,5 µg de glucógeno
- Mezclar bien y añadir 1 ml de etanol helado
- 20 • Precipitar el ADN a -80°C durante 1 hora
- Centrifugar a 14.000 rpm durante 20 min a 4°C
- Resuspender el sedimento de ADN en 10 µl de agua libre de RNasa
- Establecer una reacción de PCR TakaRa para cada muestra
 - Tampón de PCR 1x
- 25 • dNTP 200 µM de cada NTP
- ADN 1 µl
- Cebador directo 0,5 µM
- Cebador inverso 0,5 µM
- TakaRa LA Taq 2,5 unidades
- 30 • Agua Variable para 50 µl de reacción
- Ejecutar las muestras en un gel de agarosa al 2%

Tabla 1

```

<?xml version="1.0" ?>
- <modelo>
  <peso por unidad="-0,10936629789322752" />
5  <peso por unidad="-9,14545921645492" />
  - <limitación de peso="9,722125061947459" no logarítmico="falso">
    - <posición posicionada="8" máx="verdadero">
      <ancho de la distribución gaussiana="0,7303045966167145"
        compensación="0" />
10  - <inversa de la matriz de pesos="0,0"
      normalizado por máximo="falso" alfabeto="ADN"
      columnas="19">
    - <pos de las columnas="0">
      <símbolo de peso="guanina"
15  <peso="0,19891304347826086" />
      <símbolo de peso="citosina"
        <peso="0,32065217391304346" />
      <símbolo de peso="adenina"
        <peso="0,14673913043478262" />
20  <símbolo de peso="timina"
        <peso="0,33369565217391306" />
      </columna>
    - <pos de las columnas="1">
      <símbolo de peso="guanina"
25  <peso="0,2076086956521739" />
      <símbolo de peso="citosina"
        <peso="0,23804347826086958" />
      <símbolo de peso="adenina"
        <peso="0,1641304347826087" />
30  <símbolo de peso="timina"
        <peso="0,39021739130434785" />
      </columna>
    - <pos de las columnas="2">
      <símbolo de peso="guanina"
35  <peso="0,22934782608695653" />
      <símbolo de peso="citosina"
        <peso="0,2423913043478261" />

```

<símbolo de peso="adenina"
 peso="0,1858695652173913" />
 <símbolo de peso="timina"
 peso="0,3423913043478261" />
 5 </columna>
 - <pos de las columnas="3">
 <símbolo de peso="guanina"
 peso="0,23369565217391305" />
 <símbolo de peso="citosina"
 10 peso="0,29456521739130437" />
 <símbolo de peso="adenina"
 peso="0,12065217391304348" />
 <símbolo de peso="timina"
 peso="0,35108695652173916" />
 15 </columna>
 - <pos de las columnas="4">
 <símbolo de peso="guanina"
 peso="0,22065217391304348" />
 <símbolo de peso="citosina"
 20 peso="0,22065217391304348" />
 <símbolo de peso="adenina"
 peso="0,14673913043478262" />
 <símbolo de peso="timina"
 peso="0,41195652173913044" />
 25 </columna>
 - <pos de las columnas="5">
 <símbolo de peso="guanina"
 peso="0,2554347826086957" />
 <símbolo de peso="citosina"
 30 peso="0,22934782608695653" />
 <símbolo de peso="adenina"
 peso="0,16847826086956522" />
 <símbolo de peso="timina"
 peso="0,3467391304347826" />
 35 </columna>
 - <pos de las columnas="6">
 <símbolo de peso="guanina"

peso="0,22934782608695653" />
 <símbolo de peso="citosina"
 peso="0,21630434782608696" />
 <símbolo de peso="adenina"
 5 peso="0,17282608695652174" />
 <símbolo de peso="timina"
 peso="0,3815217391304348" />
 </columna>
 - <pos de las columnas="7">
 10 <símbolo de peso="guanina"
 peso="0,2597826086956522" />
 <símbolo de peso="citosina"
 peso="0,22934782608695653" />
 <símbolo de peso="adenina"
 15 peso="0,14673913043478262" />
 <símbolo de peso="timina"
 peso="0,3641304347826087" />
 </columna>
 - <pos de las columnas="8">
 20 <símbolo de peso="guanina"
 peso="0,2076086956521739" />
 <símbolo de peso="citosina"
 peso="0,20326086956521738" />
 <símbolo de peso="adenina"
 25 peso="0,1858695652173913" />
 <símbolo de peso="timina"
 peso="0,40326086956
 </columna>
 - <pos de las columnas="9">
 30 <símbolo de peso="guanina"
 peso="0,19021739130434784" />
 <símbolo de peso="citosina"
 peso="0,225" />
 <símbolo de peso="adenina"
 35 peso="0,1815217391304348" />
 <símbolo de peso="timina"
 peso="0,4032609695652174" />

</columna>

- <pos de las columnas="10">

<símbolo de peso="guanina"
 peso="0,29456521739130437" />

5

<símbolo de peso="citosina"
 peso="0,20326086956521738" />

<símbolo de peso="adenina"
 peso="0,15543478260869564" />

10

<símbolo de peso="timina"
 peso="0,34697391304347826" />

</columna>

- <pos de las columnas="11">

<símbolo de peso="guanina"
 peso="0,2510869565217391" />

15

<símbolo de peso="citosina"
 peso="0,19456521739130436" />

<símbolo de peso="adenina"
 peso="0,15108695652173912" />

20

<símbolo de peso="timina"
 peso="0,4032608695652174" />

</columna>

- <pos de las columnas="12">

<símbolo de peso="guanina"
 peso="0,2467391304347826" />

25

<símbolo de peso="citosina"
 peso="0,225" />

<símbolo de peso="adenina"
 peso="0,14673913043478262" />

30

<símbolo de peso="timina"
 peso="0,3815217391304348" />

</columna>

- <pos de las columnas="13">

<símbolo de peso="guanina"
 peso="0,2728260869565217" />

35

<símbolo de peso="citosina"
 peso="0,159782608769565217" />

<símbolo de peso="adenina"

```

        peso="0,225" />
        <símbolo de peso="timina"
        peso="0,3423913043478261" />
</columna>
5    - <pos de las columnas="14">
        <símbolo de peso="guanina"
        peso="0,2858695652173913" />
        <símbolo de peso="citosina"
        peso="0,19456521739130436" />
10   <símbolo de peso="adenina"
        peso="0,1858695652173913" />
        <símbolo de peso="timina"
        peso="0,33369565217391306" />
</columna>
15   - <pos de las columnas="15">
        <símbolo de peso="guanina"
        peso="0,2597826086956522" />
        <símbolo de peso="citosina"
        peso="0,1858695652173913" />
20   <símbolo de peso="adenina"
        peso="0,15978260869565217" />
        <símbolo de peso="timina"
        peso="0,39456521739130435" />
</columna>
25   - <pos de las columnas="16">
        <símbolo de peso="guanina"
        peso="0,3641304347826087" />
        <símbolo de peso="citosina"
        peso="0,19021739130434784" />
30   <símbolo de peso="adenina"
        peso="0,19021739130434784" />
        <símbolo de peso="timina"
        peso="0,2554347826086957" />
</columna>
35   - <pos de las columnas="17">
        <símbolo de peso="guanina"
        peso="0,2684782608695652" />

```

```

        <símbolo de peso="citosina"
            peso="0,27717391304347827" />
        <símbolo de peso="adenina"
            peso="0,20326086956521738" />
5      <símbolo de peso="timina"
            peso="0,2510869565217391" />
    </columna>
    - <pos de las columnas="18">
        <símbolo de peso="guanina"
10      peso="0,2554347826086957" />
        <símbolo de peso="citosina"
            peso="0,22934782608695653" />
        <símbolo de peso="adenina"
            peso="0,15108695652173912" />
15      <símbolo de peso="timina"
            peso="0,3641304347826087" />
    </columna>
</matriz de pesos>
</posicionado>
20 </limitación>
    - <limitación de peso="7,430703153194244" no logarítmico="falso">
    - <posición posicionada="-22" máx="verdadero">
    <ancho de la distribución gaussiana="7,116598573800258"
        compensación="0" />
25 - <inversa de la matriz de pesos="0,0"
    normalizado por máximo="falso" alfabeto="ADN"
    columnas="5">
    - <pos de las columnas="0">
        <símbolo de peso="guanina"
30      peso="0,1261574074074074" />
        <símbolo de peso="citosina"
            peso="0,12152777777777778" />
        <símbolo de peso="adenina"
            peso="0,5520833333333333" />
35      <símbolo de peso="timina"
            peso="0,20023148148148148" />
    </columna>

```

- <pos de las columnas="1">
 <símbolo de peso="guanina"
 peso="0,18171296296296297" />
 <símbolo de peso="citosina"
 peso="0,18634259259259296" />
 <símbolo de peso="adenina"
 peso="0,35300925925925924" />
 <símbolo de peso="timina"
 peso="0,2789351851851852" />

5

10

</columna>

- <pos de las columnas="2">
 <símbolo de peso="guanina"
 peso="0,18171296296296297" />
 <símbolo de peso="citosina"
 peso="0,14004629629629628" />
 <símbolo de peso="adenina"
 peso="0,20949074074074073" />
 <símbolo de peso="timina"
 peso="0,46875" />

15

20

</columna>

- <pos de las columnas="3">
 <símbolo de peso="guanina"
 peso="0,10763888888888888" />
 <símbolo de peso="citosina"
 peso="0,09375" />
 <símbolo de peso="adenina"
 peso="0,5983796296296297" />
 <símbolo de peso="timina"
 peso="0,20023148148148148" />

25

30

</columna>

- <pos de las columnas="4">
 <símbolo de peso="guanina"
 peso="0,16782407407407407" />
 <símbolo de peso="citosina"
 peso="0,1863425925925926" />
 <símbolo de peso="adenina"
 peso="0,5335648148148148" />

35

```

        <símbolo de peso="timina"
            peso="0,11226851851851852" />
    </columna>
</matriz de pesos>
5  </posicionado>
    </limitación>
    - <limitación de peso="20,800433402457763" no logarítmico="falso">
    - <posición posicionada="309" máx="falso">
    <ancho de la distribución gaussiana="123,21919574445802"
10     compensación="0" />
    - <inversa de la matriz de pesos="0,0"
        normalizado por máximo="falso" alfabeto="ADN"
        columnas ="4">
    - <pos de las columnas="0">
15     <símbolo de peso="guanina"
            peso="0,5205858504341704" />
        <símbolo de peso="citosina"
            peso="0,1598047165219432" />
        <símbolo de peso="adenina"
20     peso="0,1598047165219432" />
        <símbolo de peso="timina"
            peso="0,1598047165219432" />
    </columna>
    - <pos de las columnas="1">
25     <símbolo de peso="guanina"
            peso="0,1525271710220418" />
        <símbolo de peso="citosina"
            peso="0,1525271710220418" />
        <símbolo de peso="adenina"
30     peso="0,5424184869338746" />
        <símbolo de peso="timina"
            peso="0,1525271710220418" />
    </columna>
    - <pos de las columnas="2">
35     <símbolo de peso="guanina"
            peso="0,5571014456775343" />
        <símbolo de peso="citosina"

```

peso="0,14763285144082192" />
 <símbolo de peso="adenina"
 peso="0,14763285144082192" />
 <símbolo de peso="timina"
 5 peso="0,14763285144082192" />
 </columna>
 - <pos de las columnas="3">
 <símbolo de peso="guanina"
 peso="0,141013450947342" />
 10 <símbolo de peso="citosina"
 peso="0,5769596947157973" />
 <símbolo de peso="adenina"
 peso="0,1410134350947342" />
 <símbolo de peso="timina"
 15 peso="0,1410134350947342" />
 </columna>
 </matriz de pesos>
 </posicionado>
 </limitación>
 20 - <limitación de peso="13,776783968061828" no logarítmico="falso">
 - <posición posicionada="-29" máx="verdadero">
 <ancho de la distribución gaussiana="17,24379973438928"
 compensación="0" />
 - <inversa de la matriz de pesos="0,0"
 25 normalizado por máximo="falso" alfabeto="ADN"
 columnas ="6">
 - <pos de las columnas="0">
 <símbolo de peso="guanina"
 peso="0,17708333333333331" />
 30 <símbolo de peso="citosina"
 peso="0,21875" />
 <símbolo de peso="adenina"
 peso="0,4270833333333333" />
 <símbolo de peso="timina"
 35 peso="0,17708333333333331" />
 </columna>
 - <pos de las columnas="1">

<símbolo de peso="guanina"
 peso="0,22337962962962962" />
 <símbolo de peso="citosina"
 peso="0,14467592592592593" />
 5 <símbolo de peso="adenina"
 peso="0,3761574074074074" />
 <símbolo de peso="timina"
 peso="0,25578703703703703" />
 </columna>
 10 - <pos de las columnas="2">
 <símbolo de peso="guanina"
 peso="0,19560185185185183" />
 <símbolo de peso="citosina"
 peso="0,13541666666666666" />
 15 <símbolo de peso="adenina"
 peso="0,14004629629629628" />
 <símbolo de peso="timina"
 peso="0,5289351851851851" />
 </columna>
 20 - <pos de las columnas="3">
 <símbolo de peso="guanina"
 peso="0,15393518518518517" />
 <símbolo de peso="citosina"
 peso="0,16782407407407407" />
 25 <símbolo de peso="adenina"
 peso="0,505787037037037" />
 <símbolo de peso="timina"
 peso="0,1724537037037037" />
 </columna>
 30 - <pos de las columnas="4">
 <símbolo de peso="guanina"
 peso="0,2465277777777776" />
 <símbolo de peso="citosina"
 peso="0,16782407407407407" />
 35 <símbolo de peso="adenina"
 peso="0,42245370370370366" />
 <símbolo de peso="timina"

```
                peso="0,16319444444444445" />
</columna>
- <pos de las columnas="5">
    <símbolo de peso="guanina"
5                peso="0,13541666666666666" />
    <símbolo de peso="citosina"
                peso="0,19560185185185183" />
    <símbolo de peso="adenina"
                peso="0,5567129629629629" />
10    <símbolo de peso="timina"
                peso="0,11226851851851852" />
    </columna>
</matriz de pesos>
</posicionado>
15 </limitación>
</modelo>
```

Tabla 2

| Gen ID | Símbolo | Localización | Descripción | Sinónimos | Referencias externas |
|--------|---------|------------------|---|--|---------------------------------|
| 9590 | AKAP12 | 6q24-q25 | Proteína (gravina) 12 de anclaje a la cinasa A (PRKA) | AKAP250 DKFZp686M0430 DKFZp686O0331 | HGNC:370 MIM:604698 HPRD:05263 |
| 208 | AKT2 | 19q13.1-q13.2 | Homólogo 2 del oncogén vírico de timoma murino v-akt | PKBBETA PRKBB RAC-BETA | HGNC:392 MIM:164731 HPRD:01262 |
| 324 | APC | | Poliposis coli adenomatosa | DP2 DP2.5 DP3 FAPI FPC GS | HGNC:583 MIM:175100 HPRD:01439 |
| 578 | BAK1 | 6p21.3 | Antagonista/destructor 1 de BCL2 | BAK BCL2L7 CDN1 MGC117255 | HGNC:949 MIM:600516 HPRD:02744 |
| 581 | BAX | 19q13.3-q13.4 | Proteína X asociada a BCL2 | Bax zeta | HGNC:959 MIM:600040 HPRD:02498 |
| 596 | BCL2 | 18q21.33 18q21.1 | CLL/linfoma 2 de linfocitos B | Bcl-2 | HGNC:990 MIM:151430 HPRD:01045 |
| 10904 | BLCAP | 20q11.2-q12 | Proteína asociada al cáncer de vejiga | BC10 | HGNC:1055 HPRD:16552 |
| 672 | BRCA1 | 17q21 | Cáncer de mama 1, aparición temprana | BRCA BRCC1 IRIS PSCP RNF53 | HGNC:1100 MIM:113705 HPRD:00218 |
| 675 | BRCA2 | 13q12.3 | Cáncer de mama 2, aparición temprana | BRCC2 FACD FAD FAD1 FANCB FANCD FANCD1 | HGNC:1101 MIM:600185 HPRD:02554 |
| 60500 | BRCA3 | 13q21 | Cáncer de mama 3 | BRCAX Cáncer de mama, tipo 3 | HGNC:18617 MIM:605365 |
| 1116 | CHI3L1 | 1q32.1 | Proteína 1 similar a quitinasa 3 (glicoproteína de cartilago 39) | GP39 HC-gp39 HCGP-3 PIYKL40 | HGNC:1932 MIM:601525 BPRD:03314 |
| 1620 | DBC1 | 9q32-q33 | Proteína 1 eliminada en cáncer de vejiga | DBCCR1 FAM5A | HGNC:2687 MIM:602865 HPRD:04181 |
| 1630 | DCC | 18q21.3 | Eliminada en carcinoma colorrectal | CRC18 CRGR1 | HGNC:2701 MIM:120470 HPRD:00391 |
| 8788 | DLK1 | 14q32 | Homólogo 1 similar a delta (Drosophila) | FA1 PREF1 Pref-1 ZOG pG2 | HGNC:2907 MIM:176290 01446 |
| 9170 | EDG4 | 19p12 | Diferenciación endotelial, receptor acoplado a la proteína G lisofosfatídica, 4 | EDG-4 LPA2 LPA2 | HGNC:3168 MIM:605110 |

| Gen ID | Símbolo | Localización | Descripción | Sinónimos | Referencias externas |
|--------|---------|-----------------------------|---|---|---------------------------------------|
| 23566 | EDG7 | 1p22.3-p31.1 | Diferenciación endotelial, receptor acoplado a la proteína G lisofosfatídica, 7 | Edg-7 GPCR HOFNH30 LP-A3 LPA3 LPA3 PAR3 RP4-67813 | HGNC:14298 MIM:605106 HPRD:05486 |
| 2064 | ERBB2 | 17q11.2-17q11.2-q12 17q21.1 | Homólogo 2 del oncogén vírico de leucemia eritroblástica v-erb-b2, homólogo de oncogén derivado de neuro/glioblastoma (avian) | HER-2-HER-2 neu HER2 NEU NGL TKR1 c-erb B2 | HGNC:3430 MIM:164870 HPRD:01281 |
| 2066 | ERBB4 | 2q33.3-q34 | Homólogo 4 del oncogén vírico de leucemia eritroblástica v-erb-a (avian) | HER4 | HGNC:3432 MIM:600543 HPRD:02767 |
| 51013 | EXOSC1 | 10q24 | Componente de 1 exosoma | CGI-108 CSL4 Cs14p SKi4 Sk14p hCs14p p13 | HGNC:17286 MIM:606493 HPRD:16223 |
| 2353 | FOS | 14q24.3 | Homólogo del oncogén vírico de osteosarcoma murino v-fos FBJ | c-fos ASM ASM1 BWSID1 S813E MGC | HGNC:3796 MIM:164810 HPRD:01275 |
| 283120 | H19 | 11p15.5 | H19, ARNm sin traducir expresado maternalmente | 4485 PRO2605 proteína predicha de HQ2605 | HGNC:4713 MIM:103280 |
| 3726 | JUNB | 19p13.2 | Protocogén jun B | - | HGNC:6205 MIM:165161 HPRD:01303 |
| 3814 | KISS 1 | 1q32 | Supresor de la metástasis KISS-1 | KISS-1 MGC39258 | HGNC:6341 MIM:603286 HPRD:04475 |
| 5653 | KLK6 | 19q13.3 | Calicreína 6 (neurosina, cima) | Bsp Kik7 MGC9355 NEUROSINA P RSS18 PRSS9 SP59 ZYME hK6 ERK ERK2 ERT1 MAPK2 P42MA PK PRKM1 PRKM2 p38 p40 p41 p 4-1mapk | HGNC:6367 MIM:602652 HPRD:04037 |
| 5594 | MAPK1 | 22q11.2 22q11.2.1 | Proteína cinasa 1 activada por mitógeno | COCA2 FCC2 HNPC2 HNPC2 MGC5172 hMLH1 ALL- | HGNC:6871 MIM:176948 HPRD:01496 |
| 4292 | MLH1 | 3p21.3 | Homólogo 1 mutL, cáncer de colon, no poliposis de tipo 2 (E. coli) | | HGNC:7127 MIM:120436 HPRD:00390 |
| 4297 | MLL | 11q23 | Leucemia de linaje mielóide/linfóide o mixto (homólogo de Trithorax, Drosophila) | 1 CXXC7 HRX HTRX1 MILL1A IT RX1 | HGNC:7132 MIM:159555 HPRD:01162 50 |

| Gen ID | Símbolo | Localización | Descripción | Sinónimos | Referencias externas |
|--------|---------|----------------|---|------------------------------|----------------------------------|
| 94025 | MUC16 | 19q13.2 | Mucina 16 | CA125 FLJ14303 | HGNC:15582 MIM:606154.50 |
| 4609 | MYC | 8q24.12-q24.13 | Homólogo del oncogén vírico de mielocitomatosis v-myc (aviar) | c-Myc | HGNC:7553 MIM:190080 HPRD:01818 |
| 4830 | NME1 | 17q21.3 | Células 1 no metastásicas, proteína (NM23A) expresada en | AWD GAAD NDPKA NM23 NM2.3-H1 | HGNC:7849 NEM:156490 HPRD:01131 |
| 5292 | PIM1 | 6p21.2 | Oncogén pim-1 | PIM | HGNC:8986 MIM:164960 HPRD:01292 |
| 5652 | PRSS8 | 16p11.2 | Proteasa, serina, 8 (prostasina) | CAP1 PROSTASINA | HGNC:949.1 MIM:600823 HPRD:02895 |
| 6667 | SP1 | 12q13.1 | Factor de transcripción Sp1 | - | HGNC:11205 MIM:189906 HPRD:01796 |
| 7124 | TNF | 6p21.3 | Factor de necrosis tumoral (superfamilia de TNF, miembro 2) | DIF TNF-alfa TNFA TNFSF2 | HGNC:11892 MIM:191160 HPRD:01855 |
| 7157 | TP53 | 17p13.1 | Proteína p53 tumoral (síndrome de Li-Fraumeni) | LFS1 TRP53 p53 | HGNC:11998 MIM:191170 HPRD:01859 |
| 54997 | TSC | 12q24.22 | Proteína hipotética FLJ20607 | FLJ20607 | HPRD:11649 |
| 7409 | VAV1 | 19p13.2 | Oncogén vav 1 | VAV | HGNC:12657 MIM:164875 HPRD:01284 |
| 7428 | VHL | 3p26-p25 | Supresor de tumores de von Hippel-Lindau | HRCA1 RCA1 VHL1 | HGNC:12687 MIM:608537 |
| 7490 | WT1 | 11p13 | Tumor 1 de Wilms | GUDI WAGRI WIT-2 WT33 | HGNC:12796 MIM:607102 HPRD:06163 |
| 7535 | ZAP70 | 2q12 | Proteína cinasa asociada a la cadena zeta (TCR) de 70kDa | SRK STDJ TZK ZAP-70 | HGNC:12858 MIM:176947 HPRD:01495 |

REIVINDICACIONES

1. Procedimiento de diagnóstico de un cáncer por detección de expresión génica anormal en un individuo que comprende determinar en una muestra del individuo la presencia de una conformación cromosómica anormal, caracterizándose dicha conformación cromosómica anormal por
- 5 (i) la presencia de una nueva yuxtaposición que no está presente durante la expresión normal, o
- (ii) la ausencia de al menos una yuxtaposición que está presente durante la expresión génica normal,
- en el que dicha yuxtaposición es de dos regiones separadas de un gen, para así diagnosticar si el individuo tiene cáncer.
2. Procedimiento según la reivindicación 1 en el que dicha conformación cromosómica es un bucle o una
- 10 estructura topológicamente cerrada.
3. Procedimiento según una cualquiera de las reivindicaciones precedentes, en el que dos o más genes que participan en cáncer se analizan para permitir el diagnóstico del tipo de cáncer y/o al menos uno de los genes que se analiza es un gen específico de tejido que permite que se identifique el tejido en el que se produce la expresión anormal.
- 15 4. Procedimiento según una cualquiera de las reivindicaciones precedentes que comprende detectar la conformación cromosómica por determinación de si las secuencias en el gen se han puesto en yuxtaposición o no por la asociación de marcadores CC en el gen, y/o en el que la yuxtaposición se detecta por:
- reticulación de AND yuxtapuesto, seguido de
 - detección del ADN reticulado, opcionalmente por medio de un procedimiento de detección basado en
- 20 secuencias,
- en el que dichos marcadores CC tienen 5 a 30 nucleótidos de longitud.
5. Procedimiento según la reivindicación 4, en el que después de reticularse el ADN,
- el ADN reticulado se somete a digestión por restricción,
 - la estructura digerida se somete a ligación, y
- 25 - la estructura ligada se analiza/detecta.
6. Procedimiento según la reivindicación 5, en el que el análisis de la estructura ligada comprende la detección de una secuencia de ADN presente en la estructura ligada que no está presente en el gen.
7. Procedimiento según la reivindicación 6, en el que la secuencia de ADN presente en la estructura ligada se detecta por secuenciación o por PCR, en el que opcionalmente la presencia de la secuencia ligada se detecta
- 30 usando una reacción de PCR en la que los cebadores forman satisfactoriamente un producto de PCR usando la secuencia ligada como molde, pero no amplifican la secuencia del gen bajo las mismas condiciones de PCR.

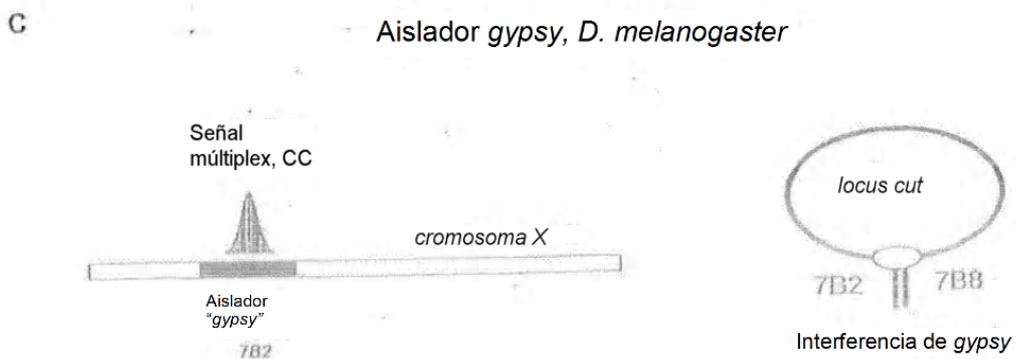
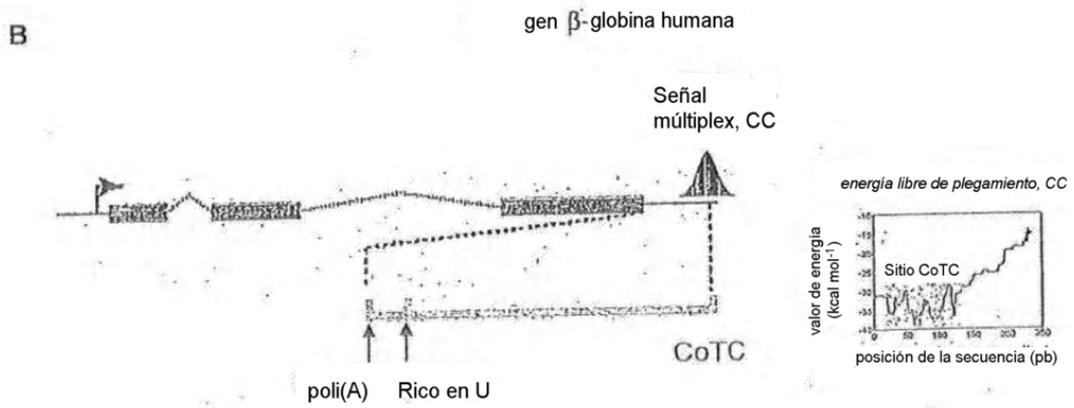
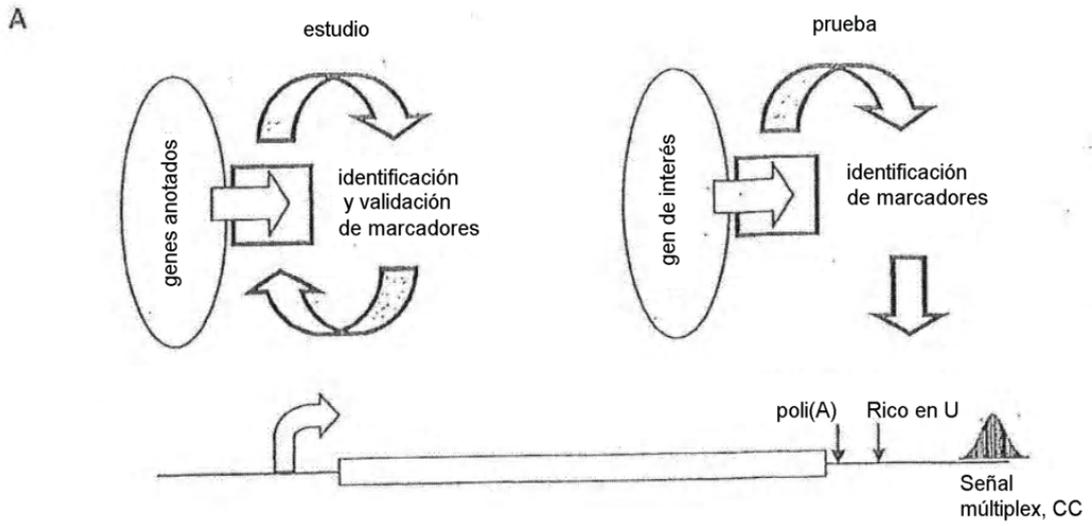


Fig. 1

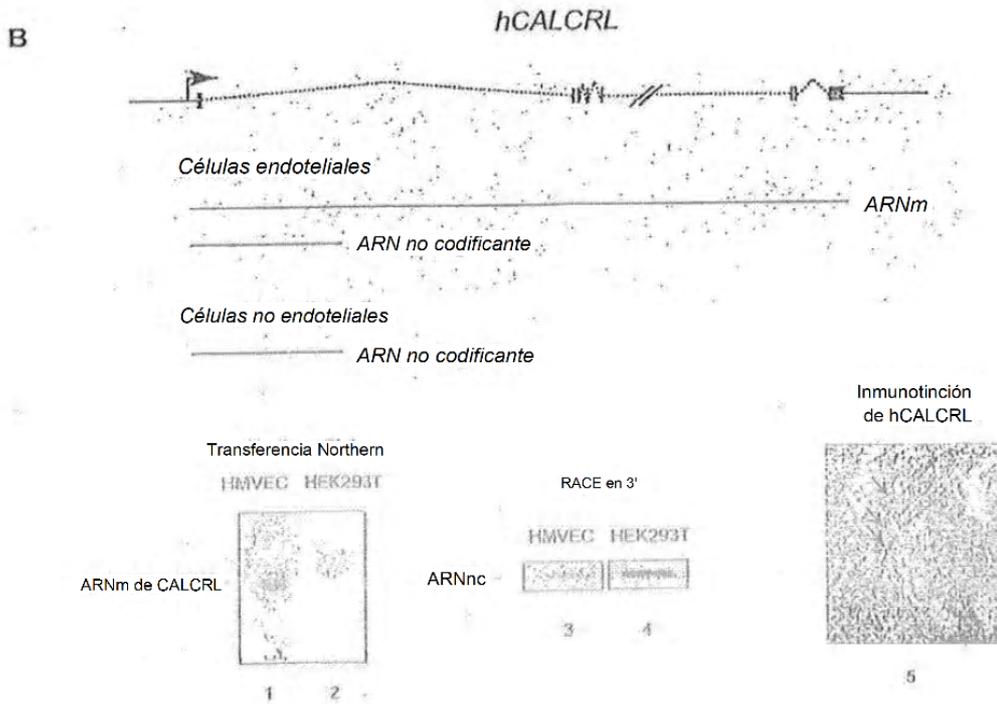
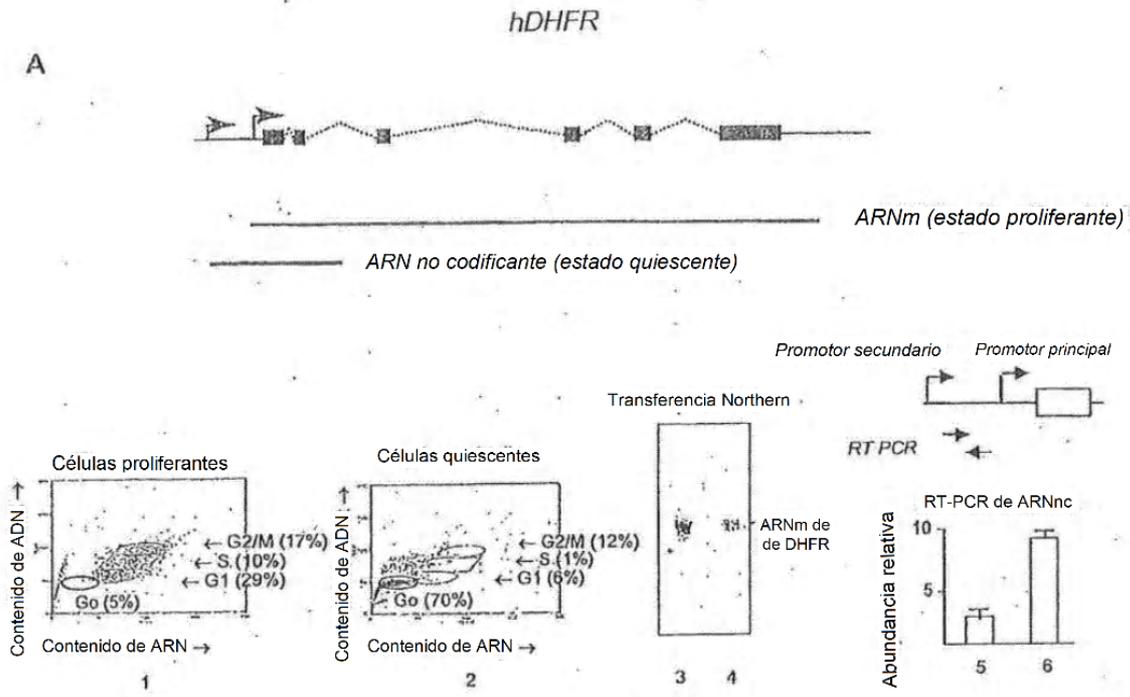


Fig. 2

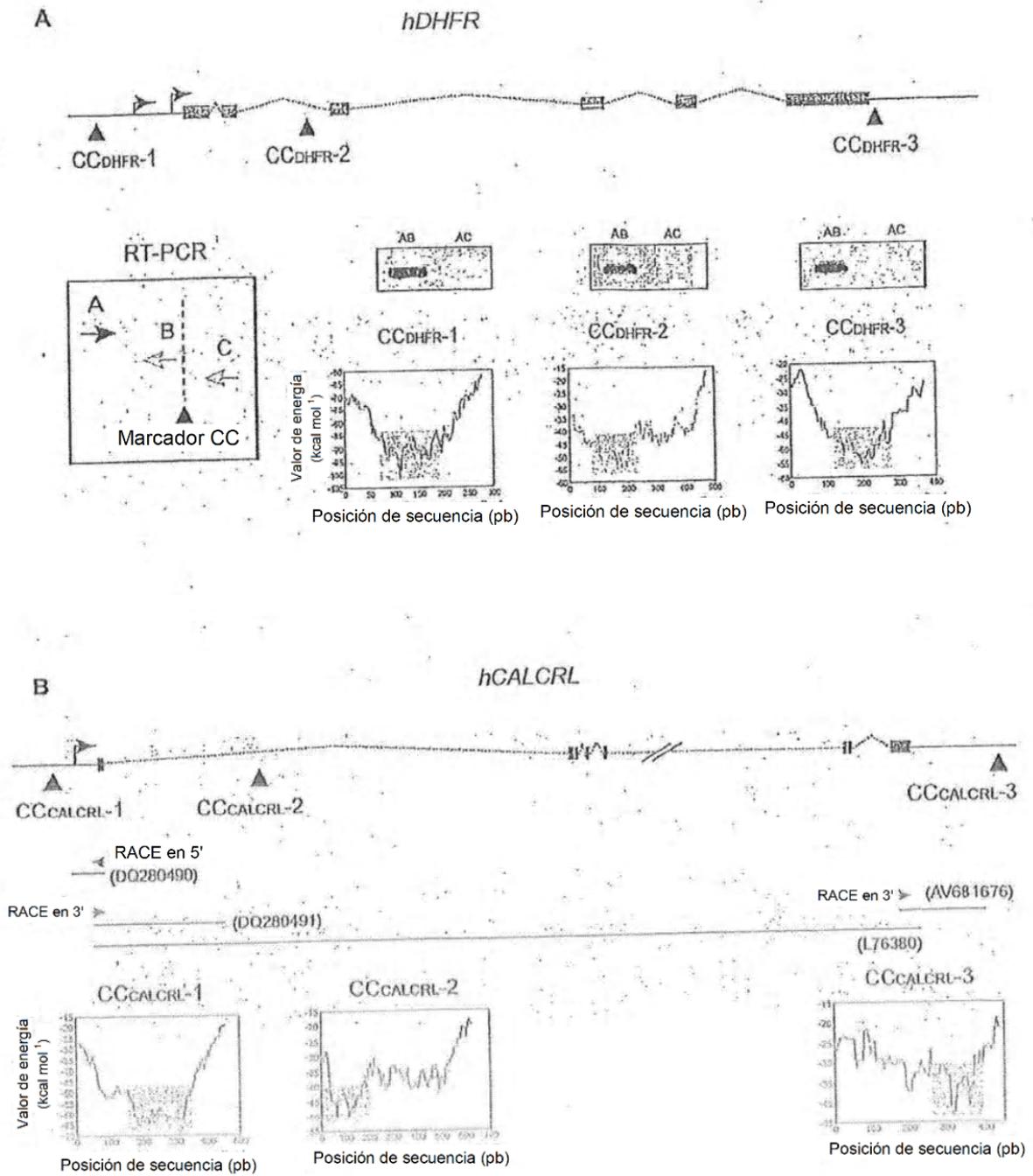


Fig. 3

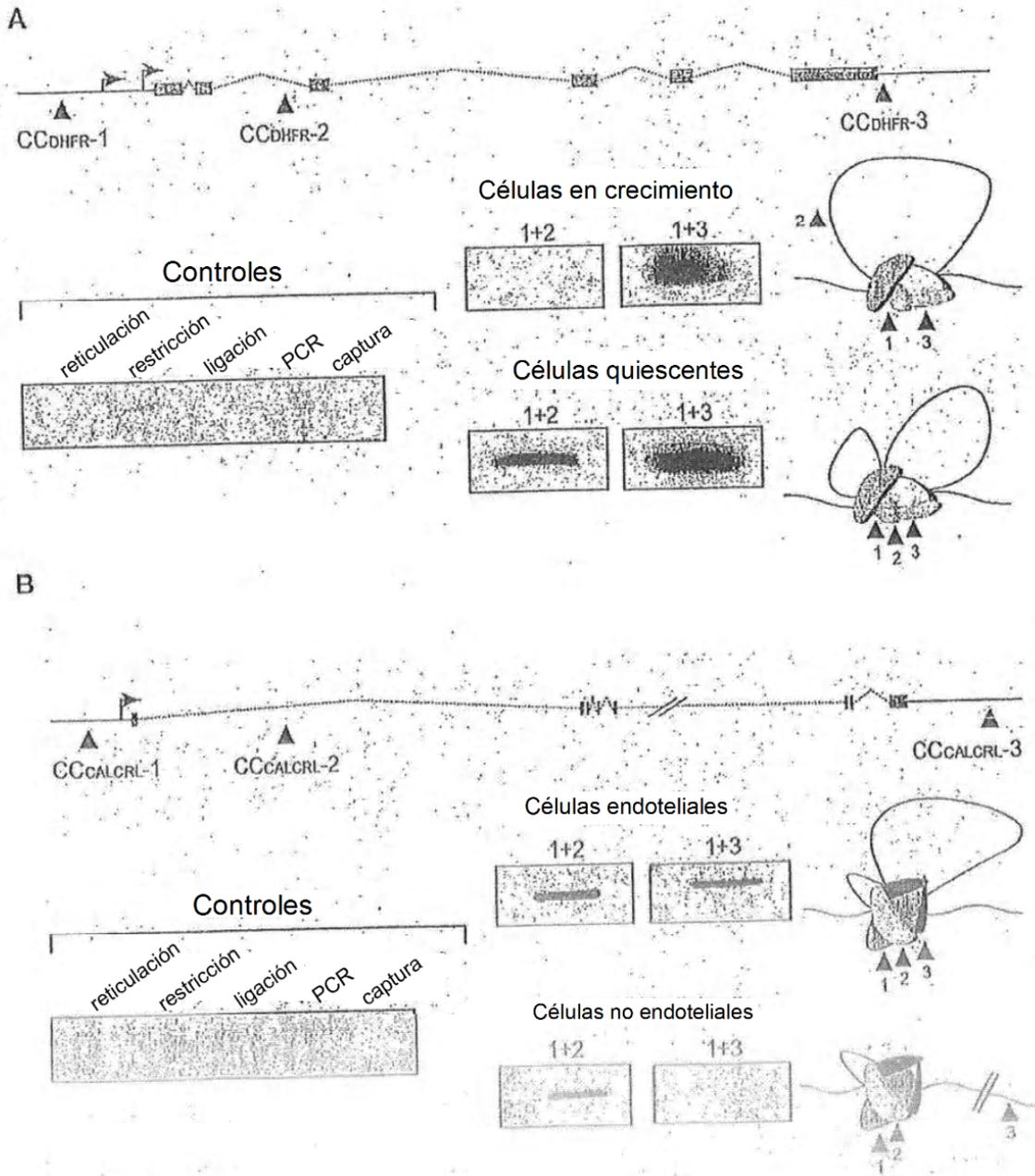


Fig. 4

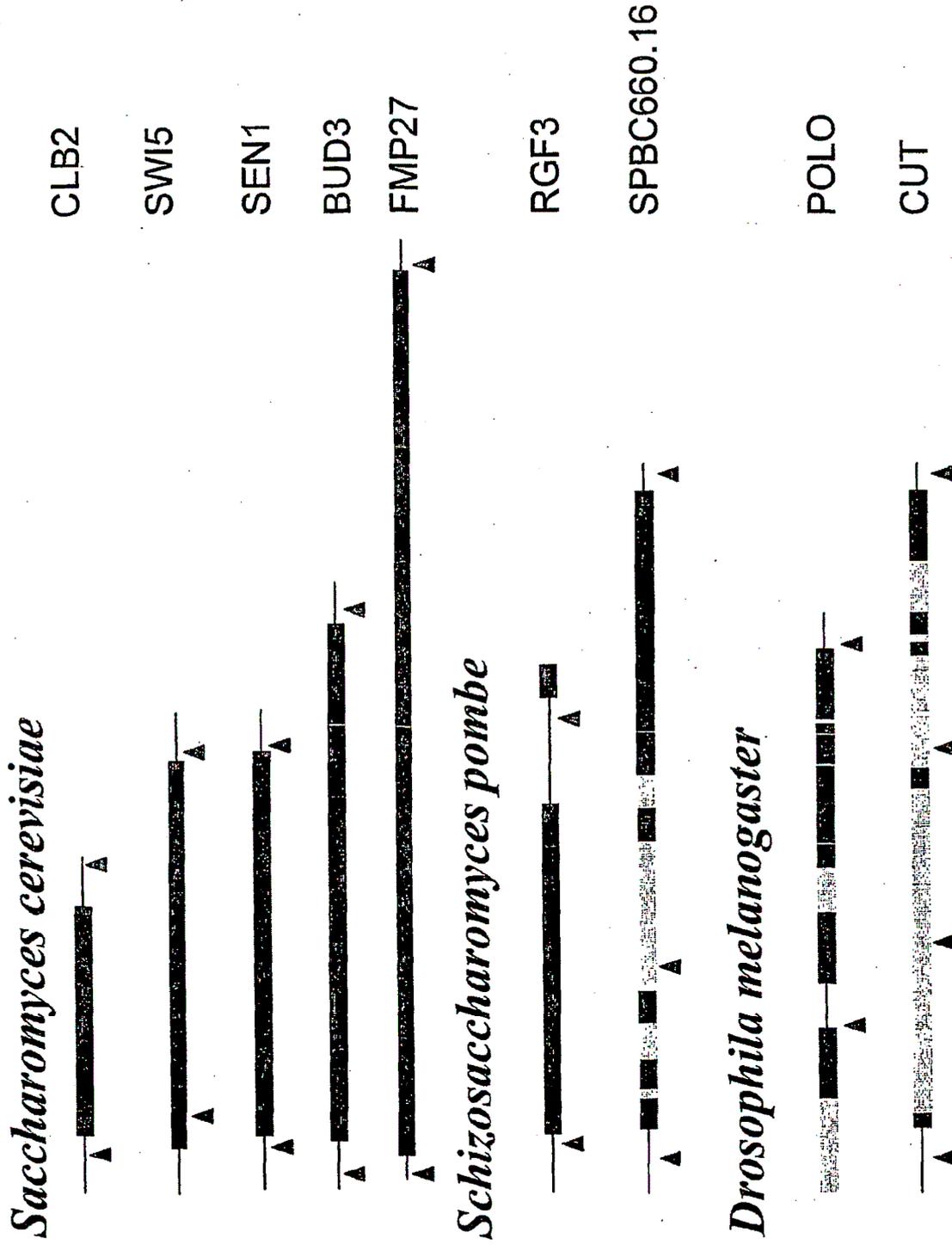


Fig. 5

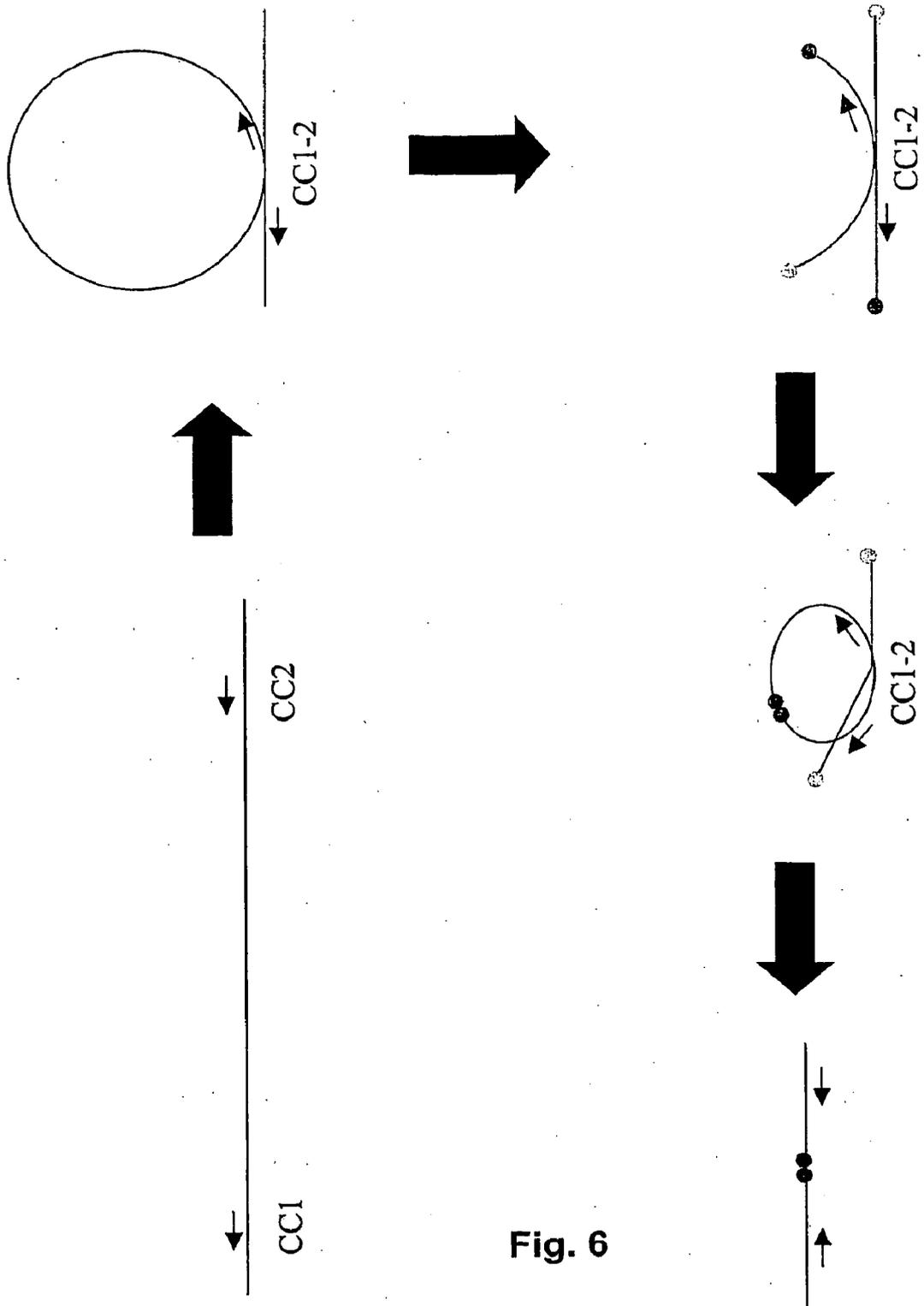
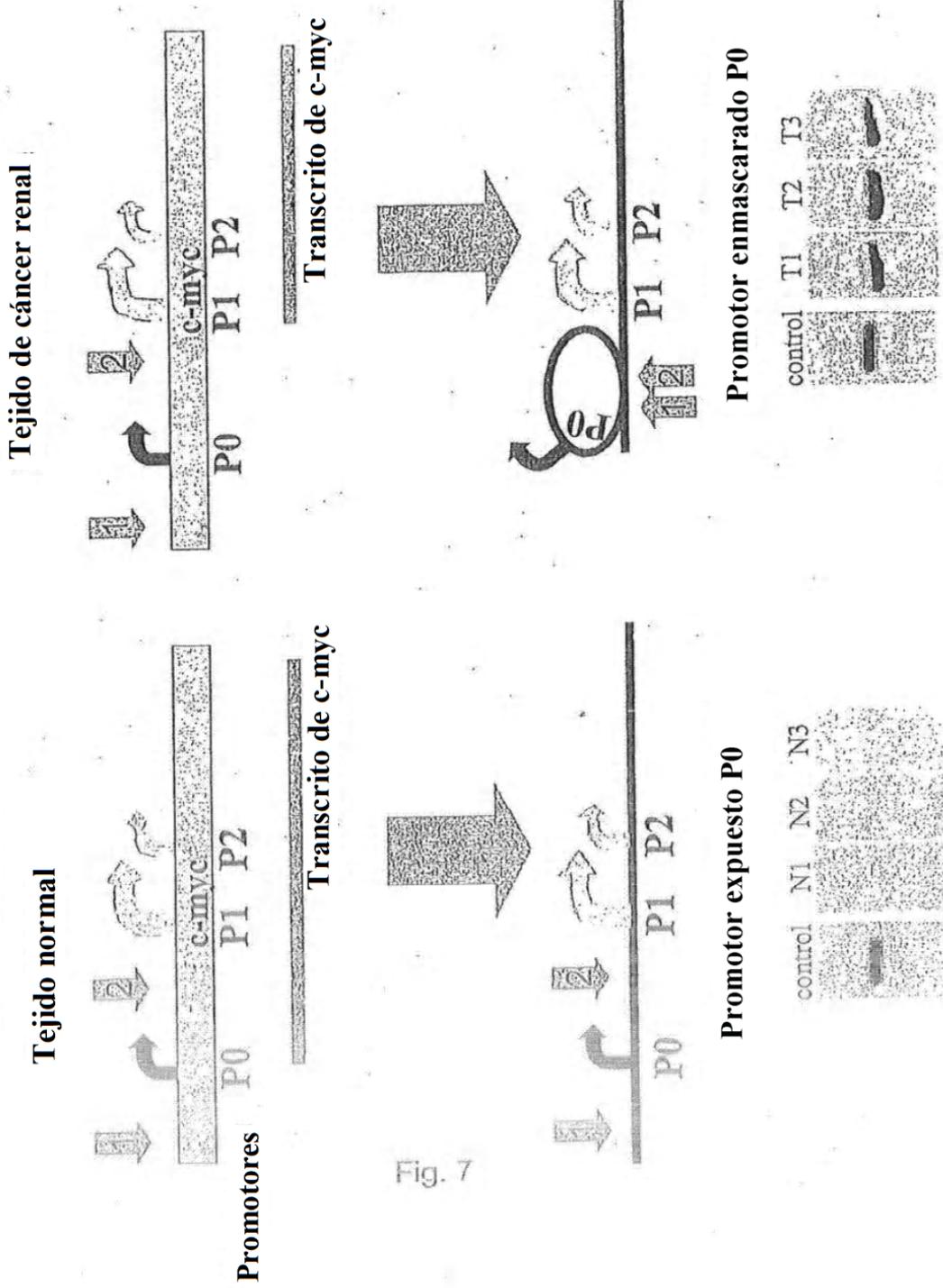
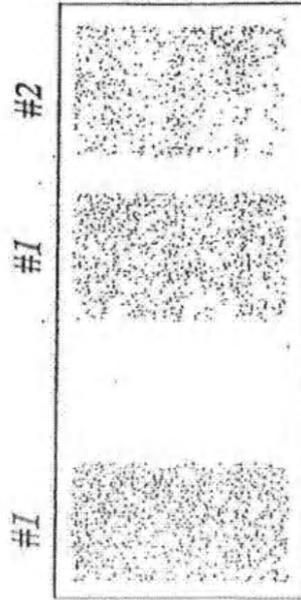


Fig. 6



En tejidos congelados

*Pacientes con cáncer de ovario
Estado temprano Metastásico*



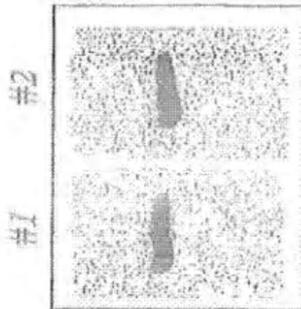
En fluidos

*Pacientes con cáncer de ovario
Estado temprano Metastásico*



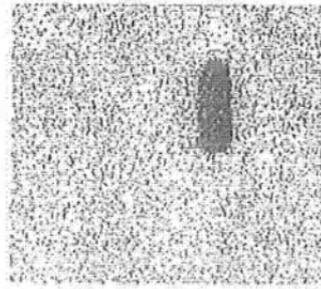
Fig.8

*Pacientes
de control*



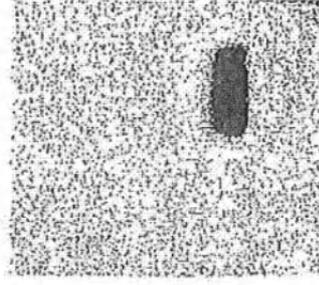
Ejemplos de desregulación conformacional en líneas de células de próstata

U2OS
LNCaP



El marcador BORIS diferencia osteosarcoma humano (U2OS) y carcinoma de próstata humano (LNCaP)

U2OS
BPH1



El marcador PSA/KLK3 diferencia osteosarcoma humano (U2OS) e hiperplasia benigna de próstata (BPH1)

Fig. 9