

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 443 230**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.02.2008 E 08709178 (1)**

97 Fecha y número de publicación de la concesión europea: **23.10.2013 EP 2129796**

54 Título: **Nuevos marcadores para el cáncer**

30 Prioridad:

**21.02.2007 DK 200700273**  
**24.04.2007 DK 200700601**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**18.02.2014**

73 Titular/es:

**OSLO UNIVERSITETSSYKEHUS HF (100.0%)**  
**Kirkeveien 166**  
**0450 Oslo, NO**

72 Inventor/es:

**SKOTHEIM, ROLF I.;**  
**LOTHE, RAGNHILD A.;**  
**LIND, GURO E. y**  
**AHLQUIST, TERJE C.**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 443 230 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

Nuevos marcadores para el cáncer

**Campo técnico de la invención**

5 La presente descripción se refiere a marcadores novedosos para la hipermetilación de promotores génicos en el cáncer. En concreto, la presente descripción se refiere a un método para determinar si se está desarrollando un tumor en el sistema aéreo-digestivo, o si un individuo está recayendo tras el tratamiento de un tumor de este tipo. Los métodos de la presente descripción comprenden la determinación del estado de metilación de los sitios CpG en la secuencia/región promotora de uno o más genes concretos. La descripción además se refiere a la utilización de genes metilados de este tipo y a kits de diagnóstico para la detección del cáncer.

**10 Antecedentes de la invención**

15 La regulación epigenética alterada es tan común como las mutaciones génicas en el cáncer humano. Estos mecanismos conducen a cambios cuantitativos y cualitativos en la expresión génica que dan lugar a una ventaja en forma de crecimiento selectivo, la cual puede dar como resultado una transformación cancerosa. Las islas de CpG hipermetiladas de manera aberrante en el promotor génico asociado con la inactivación transcripcional se encuentran entre los cambios epigenéticos más frecuentes en el cáncer. Debido a que la detección precoz de la enfermedad puede dar como resultado un mejor desenlace clínico en la mayoría de los tipos de cáncer, la identificación de la metilación génica aberrante asociada al cáncer representa un biomarcador novedoso y prometedor. Para los diferentes tipos de cáncer del sistema aéreo-digestivo, incluido el cáncer colorrectal, los estudios iniciales han identificado la presencia de ADN metilado de manera aberrante en las heces y sangre del paciente. Los genes hipermetilados de manera aberrante muy frecuentemente en los tumores benignos y solo de manera ocasional en la mucosa normal serían buenos candidatos de biomarcadores diagnósticos debido al beneficio clínico potencial de la detección temprana de adenomas de riesgo alto así como los estadios de riesgo bajo de los carcinomas.

25 En general, sin embargo, la sensibilidad y especificidad de los marcadores precoces existentes para el cáncer en el sistema aéreo-digestivo siguen siendo poco satisfactorias y, hasta la fecha, tan solo unos pocos de los genes que, de hecho, se han cribado para analizar la metilación han mostrado una sensibilidad y especificidad razonablemente altas. Se ha detectado hipermetilación específica en *VIM* (vimentina) y *SFRP2* tal como han publicado Muller *et al.* y Chen *et al.* y, recientemente, en un informe de Lind *et al.* se ha sugerido que *ADAMTS1* y *CRABP1* presentan una frecuencia elevada de hipermetilación asociada al cáncer en los tumores colorrectales mientras que la frecuencia de hipermetilación de *NR3C1* era considerablemente menor. Lind *et al.* además identificaron 18 genes como posibles marcadores del cáncer colorrectal. En un estudio de Mori *et al.* se concluyó que la proteína MAL de diferenciación de los linfocitos T, uno de los 18 genes candidatos examinados por Lind *et al.*, no sería un biomarcador diagnóstico apropiado para el cáncer del tracto aéreo-digestivo ya que la frecuencia de metilación de *MAL* es baja y se corresponde solamente con un 6% de metilación (2/34 muestras). Los análisis adicionales de varios de los 18 genes candidatos llevados a cabo por los inventores no proporcionaron resultados alentadores: *NDRG1* no se metiló en ninguna de las muestras analizadas, *NR3C1* se metiló pero únicamente con una frecuencia muy baja y en el análisis subsiguiente de la secuencia de *SDHA* fue imposible confirmar la identidad de este gen, lo que lo convierte en un marcador inadecuado del desarrollo del cáncer.

40 Como conclusión, no existe ninguna indicación de que ninguno de los genes examinados por Lind *et al.* proporcionaría mejora alguna en la tecnología actual para la detección del cáncer. En consecuencia, se necesita un conjunto de genes en el cual cada gen se hipermetile con una especificidad y frecuencia elevadas en el cáncer. En concreto, se necesita un conjunto de genes que sea útil en las técnicas no invasivas, tales como las técnicas que implican el uso de muestras fecales, o en técnicas que se pueden utilizar con material de muestra que se obtiene fácilmente, tal como la sangre o la mucosidad. Este tipo de grupo de genes podría mejorar notablemente la posibilidad de una detección precoz de estos tipos de cáncer. El objetivo final sería desarrollar ensayos diagnósticos para determinar la hipermetilación únicamente en unos pocos, tales como 2 o 3, marcadores génicos de frecuencia elevada.

Por consiguiente, se desea la identificación de más genes en los cuales las islas de CpG en la región promotora estén hipermetiladas con una frecuencia elevada en los diferentes tipos de cáncer.

50 El documento WO2005/005601 presenta *CNRIP1* asociado con el cáncer pero no como un marcador hipermetilado. Además, el documento refiere que *CNRIP1* se presenta elevado en el cáncer.

El documento WO01/19845 presenta un método para diagnosticar cáncer determinando el estado metilado de *FBN1*.

55 Qunying *et al.* (*Cancer Cell*, vol 9, n.º 5, 2006, páginas 367-378), Und *et al.* (*Molecular Cancer*, vol 4, n.º 1, 2005, página 8) y Chen *et al.* (*JNCI J. Nat. Cancer Inst*, vol 97, n.º 15, 2005, páginas 1124-1132) muestran que los eventos

de metilación y expresión diferencial no siempre están relacionados.

Keshet *et al.* (*Nature Genetics*, vol 38, n.º 2, 2006, páginas 149-153) presenta anticuerpos para detectar islas de CpG metiladas.

**Compendio de la invención**

5 La presente descripción se basa en la comprensión por parte de los inventores de que un subconjunto concreto de genes, los cuales Lind *et al.* identificaron como marcadores potenciales, contienen sitios CpG que se metilan con una frecuencia excepcionalmente elevada en el cáncer aéreo-digestivo.

10 Por lo tanto, es un objeto de la presente descripción proporcionar un conjunto de marcadores diagnósticos del cáncer, p. ej., cáncer en el sistema aéreo-digestivo, en concreto el cáncer de colon y el cáncer colorrectal. Este conjunto de marcadores resuelve los problemas relacionados con la baja especificidad y frecuencia de la metilación en la mayoría de los marcadores del cáncer conocidos.

En consecuencia, un aspecto de la invención se refiere a un método para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras un tratamiento contra el cáncer, que comprende los siguientes pasos:

- 15 a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una muestra de un individuo;
- b) construir un gráfico del percentil del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un gen obtenido a partir de una muestra de una población sana;
- 20 c) construir una curva ROC (curva de eficacia diagnóstica) basándose en el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG determinados en la población sana y en el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG determinados en una población con cáncer;
- d) seleccionar la combinación deseada de sensibilidad y especificidad a partir de la curva ROC;
- 25 e) determinar a partir del gráfico del percentil el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG correspondientes a la especificidad escogida o determinada; y
- 30 f) predecir que el individuo probablemente padecerá cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de dicho gen de la muestra es igual o superior a dicho nivel de metilación, número de sitios CpG metilados o estado de metilación de los sitios CpG correspondientes a la combinación deseada de sensibilidad/especificidad, y predecir si el individuo probablemente padecerá cáncer o no, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de la muestra es menor que dicho nivel de metilación, número de sitios CpG metilados o estado de metilación de los sitios CpG correspondientes a la combinación deseada de sensibilidad/especificidad.

35 La presente descripción además se refiere a un kit diagnóstico para la determinación del cáncer que comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario a una secuencia de ácido nucleico de los genes seleccionados a partir de:

CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927

- 40 a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora, primer exón o intrón, de al menos un gen de una muestra, obtenida de dicho individuo, donde dicho gen es:

45 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927, y donde dicha secuencia comprende una secuencia de ácido nucleico constituida por la SEQ ID NO:7.

El método puede además comprender los siguientes pasos:

- b) comparar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG respecto a una referencia; e
- c) identificar si es probable que dicho individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar

5 cáncer, o esté recayendo tras el tratamiento contra el cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG es más elevado que el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de la referencia e identificar si es poco probable que un individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar cáncer, o esté recayendo tras el tratamiento contra el cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG está por debajo del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de dicha referencia donde el cáncer es un cáncer del sistema aéreo-digestivo.

10 Otro aspecto de la presente descripción se refiere a un método para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer, que comprende el paso de;

SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111

15 FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200

SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e

INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118

20 La invención también trata sobre el uso de CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927

en un ensayo diagnóstico donde se evalúa el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG como un indicador de si el individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer en el sistema aéreo-digestivo, o si un individuo está recayendo tras el tratamiento contra el cáncer en el sistema aéreo-digestivo.

25 Además, la invención trata sobre el uso de una secuencia de ácido nucleico, donde dicho ácido nucleico comprende una secuencia de ácido nucleico seleccionada a partir del grupo constituido por:

- A) Una secuencia de ácido nucleico tal como la define la SEQ ID NO:7,
- B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);
- C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);
- 30 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C),

en un ensayo diagnóstico donde se evalúa el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG como un indicador de si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer.

### 35 Breve descripción de las figuras

#### Figura 1

40 Muestra resultados representativos de la reacción en cadena de la polimerasa específica de metilación procedentes del análisis de MAL en tres muestras de mucosa normal, tres adenomas y tres carcinomas. Un producto de PCR visible en los carriles U indica la presencia de alelos no metilados mientras que un producto de PCR en los carriles M indica la presencia de alelos metilados. Abreviaturas: A, adenoma; C, carcinoma; N, mucosa normal; POS, control positivo constituido por sangre normal (control para las muestras no metiladas) y ADN metilado *in vitro* (control para las muestras metiladas); NEG, control negativo (contiene agua como plantilla); U, carril del producto de MSP no metilado; M, carril del producto de MSP metilado. La ilustración es una combinación de dos paneles de geles ya que los adenomas se analizaron en un gel diferente.

#### 45 Figuras 2-4

50 Muestran el aumento de la expresión génica tras un tratamiento farmacológico epigenético de líneas celulares metiladas inicialmente. Se observó un aumento de la expresión de ARNm de CNRIP1, INA y SPG20 en líneas celulares de cáncer de colon tras el tratamiento con la 5-aza-2'-desoxicitidina desmetilante, sola o combinada con el inhibidor de desacetilasas tricostatina A. Los gráficos muestran los valores de expresión relativa de CNRIP1, INA y SPG20 respectivamente (escala lineal) en seis líneas celulares de cáncer de colon, HT29, SW48, HCT15, SW480, RKO y LS1034, tratadas con 5-aza-2'-desoxicitidina sola (1  $\mu$ M y 10  $\mu$ M), tricostatina A sola y los dos fármacos combinados (5-aza-2'-desoxicitidina 1  $\mu$ M sola y tricostatina A 0,5  $\mu$ M). Ambas dosis (alta y baja) de 5-aza-2'-desoxicitidina proporcionaron incrementos comparables de los valores de expresión relativos de los tres genes. Esto significa que la desmetilación de las líneas celulares puede conseguirse cultivándolas en

presencia de dosis bajas de 5-aza-2'-desoxicitidina, lo cual es una ventaja considerando la citotoxicidad de este fármaco. En el caso de CNRIP1 e INA, el tratamiento combinado fue más eficaz que el tratamiento individual con 5-aza-2'-desoxicitidina sola y tricostatina A sola. El tratamiento combinado también incrementó la expresión de SPG20, sin embargo, se puede conseguir una reactivación similar o mayor mediante el tratamiento con 5-aza-2'-desoxicitidina sola. Tal como se esperaba, el tratamiento con el inhibidor de desacetilasas tricostatina A solo no aumentó la expresión génica ni de CNRIP1, ni de INA ni de SPG20. Abreviaturas: AZA, 5-aza-2'-desoxicitidina; TSA, tricostatina A.

#### Figura 5

Muestra el estado de metilación del promotor de *MAL* en muestras de mucosa de colon normal y carcinomas colorrectales. Se muestran resultados representativos de la reacción en cadena de la polimerasa específica de metilación. Un producto de PCR visible en los carriles U indica la presencia de alelos no metilados mientras que un producto de PCR en los carriles M indica la presencia de alelos metilados. N, mucosa normal; C, carcinoma; Pos, control positivo (reacción no metilada: ADN procedente de sangre normal, reacción metilada: ADN metilado *in vitro*); Neg, control negativo (que contiene agua como plantilla); U, carril del producto de MSP no metilado; M, carril del producto de MSP metilado.

#### Figura 6

Muestra el sitio de metilación específica en el promotor de *MAL*. La secuenciación con bisulfito del promotor de *MAL* verifica el estado de metilación evaluado mediante la reacción en cadena de la polimerasa específica de metilación. La parte superior de la figura es una presentación esquemática de los sitios CpG amplificados con éxito mediante los dos fragmentos de secuenciación con bisulfito analizados, A (de -68 a +168; hacia la derecha) y B (de -427 a -85; hacia la izquierda). El sitio de inicio de la transcripción se representa mediante +1 y las barras verticales indican la localización de sitios CpG individuales. Las dos flechas indican la localización de los cebadores de MSP. En la parte inferior de la figura, los círculos sólidos representan CpG metilados; los círculos abiertos representan CpG no metilados; y los círculos abiertos con una barra inclinada representan sitios parcialmente metilados (la presencia de aproximadamente un 20-80% de citosina, además de timina). La columna de U, M y U/M en el lateral derecho de esta parte inferior enumera el estado de metilación de las respectivas líneas celulares tal como se determinó mediante análisis de MSP por los inventores. Abreviaturas: MSP, PRC específica de metilación; s, sentido; as, antisentido; U, no metilado; M, metilado; U/M, presencia tanto de la banda metilada como de la no metilada.

#### Figura 7

Muestra la "secuencia con bisulfito" del promotor de *MAL*. Electroferogramas representativos de la secuenciación con bisulfito del promotor de *MAL* en líneas celulares de cáncer de colon. Una subsección del electroferograma de la secuencia con bisulfito, que cubre los sitios CpG entre +11 y +15 respecto al inicio de la transcripción. Una flecha negra indica las citosinas en los sitios CpG, mientras que las citosinas que se han convertido en timinas están subrayadas en rojo. Los electroferogramas de la secuenciación del promotor de *MAL* ilustrados en la presente proceden de la línea celular V9P no metilada y las HCT116 y ALA hipermetiladas.

#### Figura 8

Muestra la expresión de *MAL* en líneas celulares cancerosas y en carcinomas colorrectales. La hipermetilación del promotor de *MAL* se asocia con la pérdida o reducción de la expresión génica en modelos *in vitro*. El nivel cuantitativo de expresión génica de *MAL* se muestra como una relación entre el promedio de los dos ensayos *MAL* (que detectan diferentes variantes de empalme) y el promedio de los dos controles endógenos, *GUSB* y *ACTB*. El valor se ha multiplicado por un factor de 1000. Debajo de cada muestra se presenta el estado de metilación respectivo, tal como se determinó mediante la reacción en cadena de la polimerasa específica de metilación. Los círculos sólidos representan la hipermetilación del promotor de *MAL*, los círculos abiertos representan el *MAL* no metilado y los círculos abiertos con una barra inclinada representan la presencia tanto de alelos metilados como no metilados. Los carcinomas colorrectales se dividen en un grupo no metilado (n = 3) y un grupo hipermetilado (n = 13) y aquí se muestra la expresión de la mediana. En la tabla 1 se puede apreciar el tejido de origen de las líneas celulares individuales.

#### Figura 9

Muestra el aumento de la expresión de *MAL* tras el tratamiento farmacológico. Se observó un descenso de la metilación del promotor de *MAL*, seguido del aumento de la expresión del ARNm en las líneas celulares de cáncer de colon tras el tratamiento con la 5-aza-2'-desoxicitidina desmetilante, sola y combinada con el inhibidor de desacetilasas tricostatina A. El gráfico superior muestra los valores de expresión relativa de *MAL* (escala lineal) en dos líneas celulares de cáncer de colon, HT29 y HCT15, tratadas con 5-aza-2'-desoxicitidina sola, tricostatina A sola y los dos fármacos combinados. La parte inferior muestra los resultados de la MSP de *MAL* para las mismas muestras. Un producto de PCR visible en los carriles U indica la presencia de alelos no

metilados mientras que un producto de PCR en los carriles M indica la presencia de alelos metilados. Abreviaturas: AZA: 5-aza-2'-desoxicitidina; TSA, tricostatina A; Pos, control positivo (reacción no metilada: ADN procedente de sangre normal, reacción metilada: ADN metilado *in vitro*); Neg, control negativo (que contiene agua como plantilla); U, carril del producto de MSP no metilado; M, carril del producto de MSP metilado.

5 Figura 10

Muestra la expresión de MAL en carcinomas colorrectales. Se observó tinción citoplásmica positiva de MAL en los túbulos renales (A), y no se observó tinción en el músculo cardíaco (B), en consonancia con informes anteriores (Marazuela M, *et al.*, *J. Histochem. Cytochem.* 2003, 51: 665-674). Las células epiteliales de los carcinomas colorrectales fueron MAL negativas (C, D), mientras que en el tejido de colon normal se observó expresión citoplásmica de MAL tanto en tejido conectivo como en el epitelial (E, F). Se tomaron todas las imágenes utilizando lentes 40x (amplificación 400x).

10

A continuación se describirá la presente invención más detalladamente.

### Descripción detallada de la invención

#### Definiciones

15 A menos que se definan de otra manera, todos los términos científicos y técnicos utilizados en la presente deberán interpretarse con un significado que sea coherente con su significado en el contexto de la técnica relevante a la que esta invención pertenece. Aunque en la práctica o evaluación de la presente invención se pueden utilizar cualesquiera materiales y métodos similares o equivalentes a los descritos en la presente, se describen los materiales y métodos preferidos. A efectos de la presente invención, se definen los siguientes términos:

20 *Epigenética*

La metilación es un cambio epigenético el cual se define como alteraciones no basadas en la secuencia que se heredan a través de la división celular.

#### *Metilación*

25 La "hipermetilación" en este contexto se refiere simplemente a la metilación en mayor grado que la metilación de referencia. La metilación de referencia es la metilación del gen en una muestra procedente de un individuo sano o procedente de tejido normal. Por lo tanto, un gen metilado el cual no está metilado en tejidos normales se clasificará como hipermetilado.

30 El "estado de metilación" es una medida de la presencia o ausencia de una modificación metílica en uno o más sitios CpG en al menos una secuencia de ácido nucleico. Se sobreentenderá que el estado de metilación de uno o más sitios CpG se determina preferentemente en copias múltiples de un gen particular de interés.

35 El "nivel de metilación" constituye una expresión del grado de metilación en una o más copias de un gen o secuencia de ácido nucleico de interés. El nivel de metilación puede calcularse como una medida absoluta de la metilación en el gen o la secuencia de ácido nucleico de interés. Además, puede determinarse el "nivel de metilación relativa" como la cantidad de ADN metilado, respecto a la cantidad total de ADN presente o como el número de copias metiladas de un gen o secuencia de ácido nucleico de interés, respecto al número total de copias del gen o secuencia de ácido nucleico. Adicionalmente, puede determinarse el "nivel de metilación" como el porcentaje de sitios CpG metilados en el ADN estirado de interés.

40 La expresión "nivel de metilación" también abarca la situación donde uno o más sitios CpG en, p. ej., la región promotora están metilados pero donde el grado de metilación está por debajo del umbral de amplificación. Por lo tanto, el nivel de metilación puede ser un valor estimado del grado de metilación en un gen de interés.

La invención no se limita en modo alguno a los ciertos tipos de ensayos para determinar el estado de metilación o el nivel de metilación de los genes de acuerdo con la invención.

45 En una realización, el nivel de metilación del gen de interés está comprendido entre un 15% y un 100%, tal como entre un 50% y un 100%, más preferentemente un 60-100%, más preferentemente un 70-100%, más preferentemente entre un 80% y un 100%, más preferentemente entre un 90 y un 100%. Por lo tanto, en una realización de la presente invención el nivel de metilación de los genes de acuerdo con la invención es de un 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100%.

#### *CpG*

50 Un "sitio CpG" es una región de ADN donde existe un nucleótido de citosina contiguo a un nucleótido de guanina en

la secuencia de bases lineal a lo largo de su longitud. Las siglas “CpG” se refieren a una citosina y una guanina separadas por un fosfato, el cual une los dos nucleósidos entre sí en el ADN. La notación “CpG” se utiliza para distinguir una citosina seguida por una guanina de una base de tipo citosina emparejada con una guanina.

5 Se puede definir una isla de CpG como un intervalo de ADN contiguo de al menos 200 pares de bases en el cual el contenido G:C es de al menos un 50% y la proporción entre la frecuencia de CpG observada y la frecuencia esperada excede un 0.6. Sin embargo, también puede definirse mediante una definición más rigurosa como un intervalo de 500 pares de bases con un contenido G:C de al menos un 55% y una frecuencia de CpG observada respecto a la esperada de al menos un 0.65.

*Secuencia o región promotora*

10 Una “secuencia o región promotora” comprende una secuencia de ácido nucleico consecutiva que se prolonga hasta 1000 pb en dirección 5’ a partir del sitio de inicio transcripcional de un gen concreto y una secuencia de ácido nucleico consecutiva que se prolonga 300 pares de bases en dirección 3’ a partir del sitio de inicio transcripcional. En el listado de secuencias, la secuencia en dirección 5’ se indica en letras minúsculas mientras que la secuencia en dirección 3’ se indica en letras mayúsculas. En la parte 3’ de las secuencias las letras minúsculas indican una secuencia intrónica.

*Sitio de inicio transcripcional*

20 La expresión “sitio de inicio transcripcional” se utiliza en relación con la presente invención para describir el punto en el cual se inicia la transcripción. La transcripción puede iniciarse en uno o más sitios en el gen, y un único gen puede tener múltiples sitios de inicio transcripcional, algunos de los cuales pueden ser específicos para la transcripción en un tejido o tipo celular particulares.

*Metilación de secuencias de ácido nucleico*

25 Un gen es una región de ADN que es responsable de la producción y regulación de una cadena polipeptídica. Los genes incluyen tanto porciones codificantes como no codificantes, incluidos intrones, exones, promotores, iniciadores, potenciadores, terminadores, microARN y otros elementos reguladores. Tal como se emplea en la presente, se pretende que el término “gen” se refiera al menos a una porción de un gen. Por lo tanto, por ejemplo, a efectos de la presente invención puede considerarse que un promotor es un “gen”. En consecuencia, en una realización de la presente invención, al menos un miembro del grupo de genes comprende una porción no codificante del gen completo. En una realización particular, la porción no codificante del gen es un promotor. En otra realización, todos los miembros del grupo de genes completo comprenden porciones no codificantes de los genes, tales como, sin carácter limitante, intrones. En otra realización particular, las porciones no codificantes de los miembros de los genes son promotores. En otra realización de la presente invención, al menos un miembro del grupo de genes comprende una porción codificante del gen. En otra realización, todos los miembros del grupo de genes completo comprenden porciones codificantes de los genes.

35 La expresión “secuencia de ácido nucleico” se refiere a un polímero de desoxirribonucleótidos tanto en forma mono como bicatenaria.

Una “subsecuencia” es cualquier porción de una secuencia completa. Por lo tanto, una subsecuencia se refiere a una secuencia consecutiva de aminoácidos o ácidos nucleicos la cual es parte de una secuencia más larga de ácidos nucleicos (p. ej., un polinucleótido).

40 La expresión “identidad secuencial” indica una medida cuantitativa del grado de homología entre dos secuencias de ácido nucleico con la misma longitud. Si las dos secuencias que se comparan no tienen la misma longitud, deberán alinearse para dar el mejor ajuste posible, lo que permite la inserción de huecos o, de manera alternativa, el truncamiento en los extremos de las secuencias polipeptídicas o secuencias de nucleótidos. La identidad secuencial

$$\frac{(N_{ref} - N_{dif})100}{N_{ref}}$$

45 puede calcularse como  $\frac{(N_{ref} - N_{dif})100}{N_{ref}}$  donde  $N_{dif}$  es el número total de residuos no idénticos en las dos secuencias cuando se alinean y donde  $N_{ref}$  es el número de residuos en una de las secuencias. Por lo tanto, la secuencia de ADN AGTCAGTC tendrá una identidad secuencial de un 75% con la secuencia AATCAATC ( $N_{dif}=2$  y  $N_{ref}=8$ ). Se cuentan los huecos como no identidades del o de los residuos específicos, es decir, la secuencia de ADN AGTGTC tendrá una identidad secuencial de un 75% con la secuencia de ADN AGTCAGTC ( $N_{dif}=2$  y  $N_{ref}=8$ ).

50 Respecto a todas las reivindicaciones de la invención que se refieren a secuencias de nucleótidos, el porcentaje de identidad secuencial entre una o más secuencias también puede basarse en alineamientos utilizando el software clustalW con los parámetros predeterminados. Para los alineamientos de secuencias de nucleótidos estos parámetros son: Alineación = 3Dfull, Apertura del hueco 10,00, Ext. del hueco 0,20, Dist. de separación del hueco 4, Matriz de pesos de ADN: identidad (IUB). De manera alternativa, pueden analizarse las secuencias utilizando el programa DNASIS Max y puede realizarse la comparación de las secuencias con paralign. Este servicio se basa en

los dos algoritmos de comparación denominados Smith-Waterman (SW) y ParAlign. El primer algoritmo, publicado por Smith y Waterman (1981), es un método consolidado que encuentra el alineamiento local óptimo de dos secuencias. El otro algoritmo, ParAlign, es un método heurístico para el alineamiento de secuencias; en Rognes *et al.* se publican detalles del método. Se utilizaron parámetros predeterminados para la puntuación de la matriz y penalizaciones del hueco así como valores E.

En el contexto de la presente invención, el término “complementario” se refiere a la capacidad de dos secuencias de nucleótidos de emparejarse de manera precisa. Por ejemplo, si un nucleótido en una determinada posición de un oligonucleótido es capaz de establecer enlaces de hidrógeno con un nucleótido en la posición correspondiente de una molécula de ADN, entonces se considera que el oligonucleótido y el ADN son complementarios entre sí en esa posición. Se considera que las hebras de ADN son complementarias entre sí cuando un número suficiente de nucleótidos del oligonucleótido puede formar enlaces de hidrógeno con los nucleótidos correspondientes en el ADN diana para permitir la formación de un complejo estable.

En el presente contexto, la expresión “secuencia complementaria” o “complemento” por lo tanto también se refiere a secuencias de nucleótidos las cuales se hibridan con una molécula de ácido nucleico de la invención en condiciones rigurosas.

La expresión “condiciones rigurosas” se refiere a condiciones generales de rigurosidad alta, baja o débil.

El término “rigurosidad” es de uso común en la técnica y se utiliza haciendo referencia a las condiciones (temperatura, fuerza iónica y presencia de otros compuestos tales como disolventes orgánicos) en las cuales se llevan a cabo las hibridaciones del ácido nucleico. Con condiciones “de rigurosidad alta”, el apareamiento de bases del ácido nucleico ocurrirá únicamente entre fragmentos de ácido nucleico que tienen una frecuencia elevada de secuencias de bases complementarias, en comparación con las condiciones de rigurosidad “baja” o “débil”. Las condiciones adecuadas para evaluar la hibridación implican empapar previamente en 5xSSC y prehibridar durante 1 hora a ~40 °C en una solución de formamida al 20%, 5x solución de Denhardt, fosfato sódico 50 mM, pH 6,8 y 50 mg de ADN de timo de ternera desnaturalizado y sonificado, posteriormente se lleva a cabo la hibridación en la misma solución suplementada con ATP 100 mM durante 18 horas a ~40 °C, a continuación lavar tres veces el filtro con 2xSSC, SDS al 0.2% a 40 °C durante 30 minutos (rigurosidad baja), preferentemente a 50 °C (rigurosidad media), más preferentemente a 65 °C (rigurosidad alta) e incluso más preferentemente a ~75 °C (rigurosidad muy alta). Se pueden encontrar más detalles acerca del método de hibridación en Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2.<sup>a</sup> Ed., Cold Spring Harbor, 1989.

### Cáncer

El término “cáncer” designa un grupo de enfermedades en las cuales las células son agresivas (crecen y se dividen sin respetar los límites normales), invasivas (invaden y destruyen tejidos adyacentes) y a veces metastásicas (se extienden a otros lugares del cuerpo). Estas tres propiedades malignas del cáncer lo diferencian de los tumores benignos, los cuales tienen un crecimiento autolimitado y normalmente no invaden ni metastatizan.

El cáncer se clasifica normalmente de acuerdo con el tejido a partir del cual se originan las células cancerosas, así como el tipo celular normal al cual se parecen más. Un diagnóstico definitivo normalmente requiere un examen histológico de una pieza de biopsia tisular por parte de un patólogo. El pronóstico de los pacientes oncológicos está muy influenciado por el tipo de cáncer, así como por el estadio o la importancia de la enfermedad. Normalmente, se asocia un diagnóstico precoz con un éxito mayor en el tratamiento y una tasa de supervivencia mayor.

Los “genes supresores de tumores” son genes inactivados a menudo en las células cancerosas, lo que da como resultado la pérdida de funciones normales en esas células, tales como una replicación de ADN exacta, control sobre el ciclo celular, orientación y adhesión en los tejidos e interacción con células protectoras del sistema inmunitario. En varios tipos de cáncer, incluido el cáncer colorrectal, se ha observado la inactivación epigenética por parte del promotor de la hipermetilación de las islas de CpG de varios genes supresores de tumores.

Un tumor puede ser cualquier masa, bulto o hinchazón anómalos, sin embargo, según la interpretación que se le da en la presente, el término se refiere a la neoplasia, específicamente la neoplasia sólida. Se define la neoplasia como una proliferación anómala de células alteradas genéticamente. Las neoplasias pueden ser benignas o malignas. La neoplasia maligna o el tumor maligno se considerará en la presente cáncer. La neoplasia benigna o el tumor benigno es un tumor (neoplasia sólida) que normalmente deja de crecer por sí mismo, no invade otros tejidos y no forma metástasis. Sin embargo, los tumores benignos pueden malignizarse.

Los tumores que invaden los tejidos circundantes se considerarán en la presente cáncer. La premalignidad, el precáncer o los tumores no invasivos se considerarán en la presente una neoplasia que no es invasiva pero que tiene el potencial de progresar a un cáncer (y volverse invasivo) si no se trata.

Se pueden emplear métodos de acuerdo con la invención para determinar el grado de gravedad, es decir, los estadios, tales como el sistema Dukes, el sistema Astler-Coller y la determinación TNM del estadio de la AJCC



(siglas en inglés de Comité conjunto americano del cáncer). El sistema Dukes es un sistema de cuatro clases para la determinación del estadio que clasifica el carcinoma colorrectal de A a D basándose en el alcance del tumor: A, penetración en la pared intestinal pero no a través de esta; B, penetración a través de la pared intestinal; C, nódulo linfático implicado independientemente del alcance de la penetración en la pared intestinal; D, cáncer que se ha extendido a órganos lejanos, p. ej., hígado y pulmón. Existen muchas modificaciones de esta clasificación, p. ej., la determinación TNM del estadio.

#### *Biomarcador*

Un biomarcador puede ser cualquier sustancia cuya detección indica un estado de la enfermedad particular. Un biomarcador también puede indicar un cambio en la expresión o estado de una proteína que se correlaciona con el riesgo o la progresión de una enfermedad, o con la sensibilidad de la enfermedad a un tratamiento dado. Un buen biomarcador también puede utilizarse para diagnosticar el riesgo de la enfermedad, presencia de la enfermedad en un individuo, o para individualizar los tratamientos de la enfermedad. En el presente contexto los términos biomarcador y marcador se utilizan indistintamente.

El marcador tumoral, marcador de cáncer y, en este contexto, marcador de la metilación es un marcador para detectar un cáncer y/o tumor. El marcador se puede utilizar para detectar en un individuo la presencia de un cáncer y/o tumor, o un cáncer y/o tumor en desarrollo, o bien si el individuo está predispuesto a un cáncer y/o tumor o si está recayendo en dicho cáncer y/o tumor.

Los genes de acuerdo con la descripción pueden ser un marcador, un biomarcador, un marcador de cáncer o un marcador tumoral respectivamente.

#### *Evolución del tumor*

Además de determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer, los métodos de acuerdo con la descripción también pueden utilizarse para detectar la evolución del cáncer en un sujeto. Esto puede llevarse a cabo determinando el nivel o estado de metilación de uno o más genes en un individuo en diferentes momentos, y a continuación determinar la diferencia en el nivel o estado de metilación de uno o más genes a lo largo del tiempo. Una diferencia en el nivel o estado a lo largo del tiempo puede indicar si el individuo ha desarrollado, está desarrollado o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer.

La presente descripción también proporciona un método para realizar un pronóstico acerca del desarrollo de la enfermedad en un paciente oncológico humano. A efectos de esta invención, el término "pronóstico" se pretende que abarque predicciones y análisis de similitud de la evolución de la enfermedad, en particular la recidiva tumoral, diseminación metastásica y recaída en la enfermedad. Se pretende que los métodos de pronóstico de la invención se utilicen clínicamente en la toma de decisiones acerca de las modalidades del tratamiento, incluidos la intervención terapéutica, criterios diagnósticos tales como la determinación del estadio de la enfermedad, y monitorización de la enfermedad y vigilancia de la metástasis o recidiva de la enfermedad neoplásica. En la presente debe sobreentenderse por tratamiento tanto el tratamiento curativo como el preventivo.

La presente descripción también proporciona métodos para confirmar los resultados o indicios obtenidos por un método precedente tal como un ensayo o un método de cribado.

Por lo tanto, la frase "ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer" tal como se emplea en la presente abarca la determinación y/o predicción, tal como la estimación o determinación de la probabilidad de la presencia actual del cáncer, de su aparición futura o de una recidiva futura.

#### *Muestra*

Una muestra puede ser, sin carácter limitante, una pieza o corte histológico, tal como una porción de la neoplasia que se está tratando o puede ser una porción del tejido normal circundante. La muestra puede ser preferentemente, sin carácter limitante, sangre, excrementos (heces), orina, líquido pleural, bilis, líquido bronquial, enjuagues bucales, piezas histológicas, ascitis, pus, líquido cefalorraquídeo, muestras obtenidas por aspiración, líquido folicular, tejido o moco. Se puede procesar la muestra antes de someterla a ensayo. Por ejemplo, la muestra puede diluirse, concentrarse o purificarse y/o se puede añadir a la muestra al menos un compuesto, tal como un patrón interno. Los expertos en la técnica conocen los procedimientos para manipular diferentes muestras. Se sobreentenderá que todos los métodos de acuerdo con la invención tratan preferentemente sobre análisis *in vitro* de una muestra.

#### *Frecuencia de metilación de las muestras*

La expresión "frecuencia de metilación de las muestras" se define en la presente como una determinación

cuantitativa de las muestras metiladas, es decir, el número relativo de muestras en las cuales el gen de interés está metilado. A modo de ejemplo, la frecuencia de metilación de las muestras de CNRIP1 es de un 100%, de las 20 muestras de líneas celulares de colon, 20 están metiladas, tal como resulta aparente en la tabla 3. La cantidad relativa de muestras metiladas se compara con un nivel de referencia o valor de corte el cual se estima en función de la sensibilidad y la especificidad de cada gen.

*Referencia*

A fin de determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer, es necesario establecer una referencia o un nivel o valor de referencia. La referencia también posibilita contar en variaciones del método y ensayo, variaciones del kit, variaciones en la manipulación, variaciones relacionadas con la combinación de marcadores entre sí o con otros marcadores conocidos, y otras variaciones no relacionadas directa o indirectamente con la metilación.

En el contexto de la presente invención, el término "referencia" se refiere a un patrón respecto a la cantidad, calidad o tipo con el que se pueden comparar otros valores o características, tal como una curva patrón.

En el presente contexto la referencia o nivel de referencia debe interpretarse como un valor o nivel, el cual se ha determinado midiendo el parámetro (estado de metilación o nivel de metilación) tanto en una población sana de control como en una población con un cáncer conocido y que, de este modo, determina el valor de referencia que identifica la población con cáncer bien con una especificidad predeterminada bien con una sensibilidad predeterminada basándose en un análisis de la relación entres los valores de los parámetros y los datos clínicos conocidos de la población sana de control y la población de pacientes de cáncer.

Como sobreentenderán generalmente aquellos expertos en la técnica, los métodos de cribado del cáncer son procesos de toma de decisión por comparación. Para cualquier proceso de toma de decisión, se necesitan valores de referencia, niveles de referencia o valores de corte obtenidos a partir de individuos que padecen cáncer o una afección de interés y/o individuos que no padecen cáncer, ni una afección de interés.

El nivel de referencia (valor de corte o nivel de corte) se puede establecer teniendo en cuenta varios criterios, incluido el número aceptable de individuos que deberían someterse a ensayos diagnósticos invasivos adicionales, el riesgo promedio de padecer y/o desarrollar, p. ej., cáncer respecto a todos los individuos que se someten a ensayos diagnósticos adicionales, la decisión de que cualquier individuo cuyo riesgo específico de paciente sea mayor que un cierto nivel de riesgo tal como 1 entre 400 o 1:250 (tal como lo define la organización de cribado o el paciente individual) debería someterse a ensayos diagnósticos invasivos adicionales, u otros criterios conocidos por los expertos en la técnica.

Puede ajustarse el nivel de referencia en función de varios criterios tales como ciertos grupos de individuos sometidos a ensayo pero sin restringirse a estos. A modo de ejemplo, se puede fijar un nivel de corte inferior en individuos con inmunodeficiencia y en pacientes con un riesgo elevado de evolucionar hasta una enfermedad activa o se puede fijar un nivel de referencia mayor en grupos de individuos sanos con un riesgo bajo de desarrollar una enfermedad activa.

El nivel de referencia puede ser diferente para los diferentes estadios de la enfermedad (p. ej., tumor benigno o tumor maligno), la procedencia de la mucosa normal (de individuos exentos de cáncer *versus* pacientes oncológicos) o de la procedencia de la sangre o de las heces. Además, el nivel de referencia puede ser diferente para los individuos predispuestos a una enfermedad o los individuos que está recayendo tras el tratamiento de esta.

Los niveles de referencia se pueden adaptar para ajustarse a una especificidad o sensibilidad específicas: si se desea un ensayo con una sensibilidad elevada se puede fijar un nivel de referencia bajo. Si se busca un ensayo con una especificidad elevada, se puede fijar un nivel de referencia más alto.

Dependiendo de la prevalencia o prevalencia esperada de la presencia de la enfermedad, se puede ajustar el nivel de referencia para obtener tan pocos resultados que sean falsos positivos o tan pocos resultados que sean falsos negativos como se quiera, dependiendo de la gravedad de la enfermedad y de las consecuencias de determinar si el paciente ha dado positivo para el ensayo o ha dado negativo para el ensayo.

Los métodos para cuantificar la metilación, la parte escogida de las secuencias de ácido nucleico que comprenden la región promotora de los genes marcadores u otros parámetros, darán como resultado otros valores de referencia, los cuales pueden determinarse de acuerdo con los contenidos de la presente.

El nivel de referencia puede ser diferente si se va a diagnosticar un único paciente con síntomas o si el ensayo se va a utilizar en un cribado de un número mayor de individuos en una población.

El nivel de referencia se puede basar en determinaciones del nivel o estado de metilación combinado de diferentes marcadores tales como, sin carácter limitante, CNRIP1, SPG20, FBN1, SNCA, INA, MAL, ADAMTS1, VIM, SFRP1

y/o SFRP2. El nivel de referencia de un compuesto puede dar como resultado otros valores, los cuales pueden determinarse de acuerdo con los contenidos de la presente descripción.

Se compara el nivel de metilación con un conjunto de datos de referencia o un nivel de referencia, tal como el valor de corte, para determinar si el individuo tiene un riesgo o probabilidad mayor de padecer cáncer.

5 *Especificidad y sensibilidad*

La sensibilidad de un ensayo de cribado determinado cualquiera se refiere a la proporción de individuos con la afección que el ensayo identifica o diagnostica correctamente, p. ej., una sensibilidad es de un 100% si todos los individuos con una afección determinada dan positivo en el ensayo. La especificidad de un ensayo de cribado determinado es la proporción de individuos sin la afección que el ensayo identifica o diagnostica correctamente, p. ej., una especificidad de un 100% se refiere a que todos los individuos sin la afección tienen un resultado de ensayo negativo.

**Por lo tanto la sensibilidad se define como el (número de resultados de ensayo positivos verdaderos)/(número de resultados positivos verdaderos + número de resultados de ensayo negativos falsos).**

15 **La especificidad se define como el (número de resultados negativos verdaderos)/(número de resultados negativos verdaderos + número de resultados positivos falsos).**

Los genes de acuerdo con la presente solicitud se caracterizan por tener una sensibilidad elevada (la cantidad relativa de muestras que comprenden el gen metilado de interés procedente de individuos con cáncer es elevada) y una especificidad elevada (la cantidad relativa de muestras que comprenden el gen metilado de interés procedente de individuos sin cáncer es baja).

Un buen marcador para el cáncer es un gen el cual está metilado en casi todas las muestras cuando un individuo tiene cáncer y no metilado en muestras que proceden de un individuo que no tiene cáncer.

La especificidad del método de acuerdo con la presente descripción está comprendida preferentemente entre un 70% y un 100%, tal como entre un 75% y un 100%, más preferentemente entre un 80% y un 100%, más preferentemente entre 90% y un 100%. Por lo tanto, en una realización de la presente descripción, la especificidad de la descripción es de un 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100%.

La sensibilidad del método de acuerdo con la presente descripción está comprendida preferentemente entre un 80% y un 100%, más preferentemente entre un 85% y un 100%, más preferentemente entre un 90% y un 100%. Por lo tanto, en una realización de la presente descripción la sensibilidad de la descripción es de un 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100%.

Se sobreentenderá que los marcadores de acuerdo con la descripción se pueden utilizar combinados en los métodos de acuerdo con la descripción. Utilizar varios marcadores combinados probablemente incrementará la especificidad y/o sensibilidad del ensayo en comparación con un ensayo que implique la utilización de un único marcador. Por lo tanto, cuando se utilizan varios marcadores combinados puede ser aceptable que la especificidad y sensibilidad de cada marcador sean inferiores de lo que se ha especificado anteriormente.

A modo de ejemplo ilustrado en la tabla 3, el gen CNRIP1 está metilado en 20 de las 20 muestras procedentes de una línea celular de cáncer de colon (100%) y en 45 de las 48 (94%) muestras de adenoma – es decir, este gen tiene una sensibilidad elevada y la probabilidad de detectar la enfermedad es de un 100 y un 94% a partir de las muestras respectivas. Si se considera que la metilación de los mismos genes en muestras procedentes de tejido normal es 0 de 21, el gen tiene, por tanto, una especificidad elevada, la probabilidad de detectar un positivo falso es 0 y se detectan todos los individuos exentos de cáncer. Las muestras procedentes de mucosa normal de pacientes oncológicos mostraron metilación en 9 de las 21 muestras, lo que indica que puede detectarse el cáncer lejos de un tumor.

45 *Rendimiento diagnóstico*

La exactitud de un ensayo diagnóstico se describe mejor mediante su rendimiento diagnóstico (ROC, por sus siglas en inglés) (remítase a Zweig, M. H. y Campbell, G., *Clin. Chem.* 39 (1993) 561-577). El gráfico ROC es una representación de todas las parejas sensibilidad/especificidad resultantes de variar continuamente el nivel de referencia a lo largo del intervalo completo de datos observados.

50 El rendimiento clínico de un ensayo de laboratorio depende de su exactitud diagnóstica, o de la capacidad de clasificar correctamente individuos en subgrupos clínicamente relevantes. La exactitud diagnóstica cuantifica la capacidad del ensayo para distinguir correctamente dos condiciones diferentes en los individuos investigados. Las condiciones de este tipo son, por ejemplo, enfermedad y salud, infección latente o reciente *versus* no infección, o

enfermedad benigna *versus* maligna.

En cada caso, la curva ROC representa el solapamiento entre dos distribuciones mediante la representación de la sensibilidad *versus* 1 – la especificidad para el intervalo completo de umbrales de decisión. En el eje y se representa la sensibilidad, la cual se calcula enteramente a partir del subgrupo afectado. En el eje x se representa la fracción de positivos falsos, o 1 – la especificidad, la cual se calcula enteramente a partir de grupo que no está afectado.

Debido a que la sensibilidad y la especificidad se calculan totalmente por separado, mediante la utilización de los resultados de ensayo procedentes de dos subgrupos diferentes, la curva ROC es independiente de la prevalencia de la enfermedad en la muestra. Cada punto de la curva ROC representa una pareja sensibilidad/especificidad correspondiente a un umbral de decisión particular. Un ensayo con una discriminación perfecta (sin solapamiento entre las dos distribuciones de resultados) tiene una curva ROC que pasa a través de la esquina superior izquierda, donde la fracción de positivos verdaderos es 1.0 o de un 100% (sensibilidad perfecta) y la fracción de positivos falsos es 0 (especificidad perfecta). El gráfico teórico para un ensayo sin discriminación (distribuciones de resultados idénticas para los dos grupos) es una diagonal de 45° desde la esquina inferior izquierda hasta la esquina superior derecha. La mayoría de los gráficos están entre estos dos extremos. (Si la curva ROC está completamente por debajo de la diagonal de 45°, esto se soluciona fácilmente invirtiendo el criterio de “positividad” de “más que” a “menos que” o viceversa). Cualitativamente, cuanto más cerca está la curva de la esquina superior izquierda, mayor será la exactitud global del ensayo.

Un objetivo conveniente para cuantificar la exactitud diagnóstica de un ensayo de laboratorio es expresar su rendimiento mediante un único número. La medida global más común es el área bajo la curva ROC. Por convenio, esta área es siempre  $\geq 0.5$  (si no lo es, se puede invertir la regla de decisión para hacer que lo sea). Los valores están comprendidos en el intervalo entre 1.0 (separación perfecta entre los valores de ensayo de ambos grupos) y 0.5 (sin diferencia distribucional aparente entre ambos grupos de valores de ensayo). El área no depende únicamente de una porción particular de la curva tal como el punto más cercano a la diagonal o la sensibilidad a una especificidad de un 90%, sino de la gráfica completa. Esta es una expresión descriptiva y cuantitativa de lo cerca que está la curva ROC de la curva perfecta (área = 1.0).

La utilidad clínica de los genes marcadores del cáncer novedosos se puede evaluar mediante la comparación con otros marcadores para un cáncer concreto y en combinación con estos, p. ej., la utilidad clínica de los marcadores del cáncer novedosos CNRIP1, SPG20, FBN1, SNCA, INA y MAL se evalúa por comparación con: herramientas establecidas de diagnóstico, p. ej., la medida del nivel de expresión de los marcadores de metilación correspondientes o establecidos tales como, sin carácter limitante, ADAMTS1, VIM, SFRP1, SFRP2 y CRABP1.

#### *Evaluación del riesgo y valor de corte*

Para determinar si un sujeto corre un riesgo mayor de desarrollar, p. ej., cáncer, se debe establecer un límite para el valor de corte para el ensayo positivo. Este valor de corte lo puede establecer el laboratorio, el médico, o caso por caso cada individuo.

De manera alternativa, se puede determinar un valor de corte como el promedio, mediana o promedio geométrico del grupo de control negativo (p. ej., que no padece cáncer) +/- una o más desviaciones estándar o un valor derivado de la desviación estándar.

El valor de corte límite para los resultados de ensayo positivos de acuerdo con la invención es el nivel o estado de metilación para el cual la metilación es un indicador de cáncer.

Otro valor de corte puede ser la cantidad de sitios CpG que han de estar metilados para que se considere que un gen está metilado.

Los presentes inventores han identificado con éxito nuevos marcadores para el cáncer. El estado de metilación de los sitios CpG o el nivel de metilación en la región promotora de la secuencia de ácido nucleico de un gen seleccionado a partir de los siguientes: CNRIP1, SPG20, FBN1, SNCA, INA y MAL aumenta en individuos con cáncer y, por tanto, estos genes son marcadores eficaces para la detección de, p. ej., cáncer.

Los valores de corte pueden variar en función de condiciones específicas del individuo sometido a ensayo tales como, sin carácter limitante, el riesgo de padecer la enfermedad, ocupación, residencia o exposición geográficas.

Los valores de corte pueden variar en función de condiciones específicas del individuo sometido a ensayo tales como, sin carácter limitante, la edad, sexo, antecedentes genéticos (p. ej., tipo de HLA), función inmunitaria comprometida heredada o adquirida (p. ej., infección por VIH, diabetes, pacientes con insuficiencia renal o hepática, pacientes en tratamiento con fármacos modificadores de la inmunidad tales como, sin carácter limitante, corticoesteroides, quimioterapia, bloqueantes del TNF- $\alpha$  e inhibidores de la mitosis).

Ajustando el límite de corte o de la decisión se determinará, por tanto, la sensibilidad del ensayo para detectar el

cáncer, si está presente, o su especificidad para excluir el cáncer o la enfermedad si está por debajo de este límite. En consecuencia, el principio es que un valor por encima del valor de corte indica un riesgo mayor y un valor por debajo del valor de corte indica un riesgo reducido.

*Expresión*

5 Se ha detectado la inactivación de varios genes supresores de tumores mediante la metilación del promotor de las islas de CpG. A modo de ejemplo, el gen *MLH1* en el cual la hipermetilación de un número limitado de sitios CpG a aproximadamente 200 pares de bases en dirección 5' del sitio de inicio transcripcional se correlaciona invariablemente con la ausencia de expresión génica.

10 Los presentes análisis de líneas celulares de cáncer procedentes de varios tejidos indican que la hipermetilación de un área limitada en la proximidad del sitio de inicio transcripcional de MAL se asocia con la pérdida o reducción de la expresión génica. Los resultados cuantitativos de la expresión génica a partir de líneas celulares de cáncer de colon analizadas antes y después del tratamiento farmacológico epigenético indican que este también es el caso de SPG20, INA y CNRIP1.

15 Por lo tanto, la medida del nivel o estado de metilación y, además, la expresión del gen de interés aumentan la especificidad y sensibilidad del método.

**Métodos**

20 La presente descripción se basa en el descubrimiento de genes, los cuales están hipermetilados con una frecuencia excepcionalmente alta en diferentes tipos de cáncer como el cáncer colorrectal, comprendidos en un subconjunto particular de 6 genes seleccionados a partir de los 21 genes descritos previamente por Und *et al.* Estos marcadores de hipermetilación sumamente adecuados incluyen CNRIP1, SPG20, FBN1, SNCA, INA y MAL. Los resultados sobre, p. ej., MAL contradicen a los informes anteriores en los que se había observado hipermetilación de MAL únicamente con una frecuencia baja en el cáncer colorrectal.

25 En un primer aspecto, la presente invención proporciona un método para determinar si un sujeto ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras un tratamiento contra el cáncer, que comprende los siguientes pasos:

- a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora, primer exón o intrón, de al menos un gen en una muestra, obtenida a partir de dicho individuo, donde dicho gen es:

30 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927

y donde dicha secuencia comprende una secuencia de ácido nucleico constituida por la SEQ ID NO:7. En una realización el cáncer es un tumor tal como un tumor en el sistema aereodigestivo (benigno o maligno). El método además puede comprender los siguientes pasos:

- 35 b) comparar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG respecto a una referencia; e
- 40 c) identificar si es probable que dicho individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar cáncer, o esté recayendo tras el tratamiento contra el cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG es más elevado que el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de la referencia e identificar si es poco probable que un individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar cáncer, o esté recayendo tras el tratamiento contra el cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG está por debajo del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de dicha referencia donde el cáncer es un
- 45 cáncer del sistema aéreo-digestivo.

En otro aspecto la presente invención proporciona un método para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer, que comprende los siguientes pasos:

- 50 a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una muestra de un individuo;
- b) construir un gráfico del percentil del nivel de metilación, el número de sitios CpG metilados o el estado de

metilación de los sitios CpG de dicho gen obtenido a partir de una muestra de una población sana;

- 5 c) construir una curva ROC (curva de rendimiento diagnóstico) basándose en el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG determinados en la población sana y en el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG determinados en una población con cáncer;
- d) seleccionar la combinación deseada de sensibilidad y especificidad a partir de la curva ROC;
- e) determinar a partir del gráfico del percentil el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG correspondientes a la especificidad escogida o determinada; y
- 10 f) predecir que el individuo probablemente padecerá cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de dicho gen de la muestra es igual o superior a dicho nivel de metilación, número de sitios CpG metilados o estado de metilación de los sitios CpG correspondientes a la combinación deseada de sensibilidad/especificidad, y predecir si el individuo probablemente padecerá cáncer si o no, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de la muestra es menor que dicho nivel de metilación, número de sitios CpG metilados o estado de metilación de los sitios CpG correspondientes a la combinación deseada de sensibilidad/especificidad.
- 15

En particular, el método tal como se describe anteriormente, comprende un método que comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:

- 20 A) Una secuencia de ácido nucleico tal como la define la SEQ ID NO.:7,  
 B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);  
 C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);  
 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

25 El método tal como se describe anteriormente también puede comprender determinar el estado de metilación de los sitios CpG en una secuencia de ácido nucleico de los genes adicionales seleccionados a partir del grupo constituido por:

- 30 A) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes: la SEQ ID NO.: 1, SEQ ID NO.: 2, SEQ ID NO.: 3, SEQ ID NO.: 4, SEQ ID NO.: 5, SEQ ID NO.: 8, SEQ ID NO.: 10, SEQ ID NO.: 11, SEQ ID NO.: 12, SEQ ID NO.: 15 y la SEQ ID NO.: 16;  
 B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);  
 C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);  
 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

35 En otra realización el método de la descripción comprende determinar el estado de metilación de los sitios CpG en la región promotora de MAL. De acuerdo con esta realización la secuencia de ácido nucleico es una de las siguientes:

- 40 A) Una secuencia de ácido nucleico tal como la define la SEQ ID NO.:1,  
 B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);  
 C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);  
 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

45 Los identificadores de secuencias 1-16 representan secuencias de ácido nucleico de los genes mencionados anteriormente. Tal como comprenderá el experto en la técnica, el análisis del estado de metilación de los sitios CpG comprendidos en estas secuencias así como en sus secuencias complementarias está dentro del alcance de la presente descripción. La siguiente tabla enumera los genes de acuerdo con la descripción y los correspondientes números de id, identificadores de secuencia y alias.

Nombre HGNC	Entrada	Ensemb	SEQ ID NO.	Alias	Nombre aprobado
MAL	4118	ENSG00000172005	1-4		proteína de diferenciación de linfocitos T

ES 2 443 230 T3

Nombre HGNC	Entrada	Ensemb	SEQ ID NO.	Alias	Nombre aprobado
FBN1	2200	ENSG00000166147	6	FBN; SGS; WMS; MASS; MFS1; OCTD	fibrilina 1
CNRIP1	25927	ENSG00000119865	7	DKFZP566K1924, CRIP1, CRIP1a, CRIP1b, marco de lectura abierto 32 del cromosoma 2, C2orf32	proteína 1 que interacciona con el receptor canabinoide
SPG20	23111	ENSG00000133104	9	SPARTIN; TAHCCP1; KIAA0610	paraplejia espástica 20
SNCA	6622	ENSG00000145335	13, 14	NACP, PD1, PARK1, PARK4, MGC110988	sinucleína, alfa (componente no A4 del precursor amiloide)
INA	9118	ENSG00000148798	16	NEFS; NF-66; TXBP-1; MGC12702	proteína del filamento intermedio neuronal, internexina alfa

Las secuencias de ácido nucleico de acuerdo con la descripción se enuncian en la lista de secuencias. Cada secuencia comprende en el orden en que se menciona y en la orientación de 5' a 3' lo siguiente: una secuencia consecutiva de residuos de ácidos nucleicos localizados dentro de la región de 1000 pb en dirección 5' respecto al sitio de inicio transcripcional (indicado en letras minúsculas) seguida de una secuencia consecutiva de residuos de ácido nucleico localizados en dirección 3' respecto al sitio de inicio transcripcional (indicado en letras mayúsculas) y de la secuencia intrónica de residuos de ácidos nucleicos (indicados en letras minúsculas).

El método de la descripción puede tratar sobre el análisis de subsecuencias particulares tal como se sugiere en el apartado C) anterior. De acuerdo con estas realizaciones, la subsecuencia en C) tiene una longitud de al menos 8 residuos de ácidos nucleicos, tal como una longitud de al menos 9 residuos de ácidos nucleicos, al menos 10 residuos de ácidos nucleicos, al menos 11 residuos de ácidos nucleicos, al menos 12 residuos de ácidos nucleicos, al menos 13 residuos de ácidos nucleicos, al menos 14 residuos de ácidos nucleicos, al menos 15 residuos de ácidos nucleicos, al menos 20 residuos de ácidos nucleicos, al menos 25 residuos de ácidos nucleicos, al menos 30 residuos de ácidos nucleicos, al menos 35 residuos de ácidos nucleicos, al menos 40 residuos de ácidos nucleicos, al menos 45 residuos de ácidos nucleicos, al menos 50 residuos de ácidos nucleicos, al menos 70 residuos de ácidos nucleicos, o tal como una longitud de al menos 90 residuos de ácidos nucleicos. Normalmente se desea que el análisis se realice sobre secuencias de una cierta longitud a fin de garantizar que el método sea lo suficientemente sensible.

A efectos prácticos también puede ser deseable minimizar la longitud de las subsecuencias que son objeto de los estudios de metilación en el método de la invención. En consecuencia, puede ser deseable que dicha subsecuencia en C) tenga una longitud de 10 residuos de ácidos nucleicos como máximo, tal como 13 residuos de ácidos nucleicos como máximo, 14 residuos de ácidos nucleicos como máximo, 15 residuos de ácidos nucleicos como máximo, 20 residuos de ácidos nucleicos como máximo, 25 residuos de ácidos nucleicos como máximo, 30 residuos de ácidos nucleicos como máximo, 35 residuos de ácidos nucleicos como máximo, 40 residuos de ácidos nucleicos como máximo, 45 residuos de ácidos nucleicos como máximo, 50 residuos de ácidos nucleicos como máximo, 70 residuos de ácidos nucleicos como máximo, 90 residuos de ácidos nucleicos como máximo, 110 residuos de ácidos nucleicos como máximo, 150 residuos de ácidos nucleicos como máximo, o tal como 200 residuos de ácidos nucleicos como máximo.

Más concretamente, puede ser deseable que la subsecuencia en C) tenga una longitud de entre 8 y 200 residuos de ácidos nucleicos, tal como una longitud de entre 8 y 150 residuos de ácidos nucleicos, de entre 8 y 100 residuos de ácidos nucleicos, de entre 8 y 75 residuos de ácidos nucleicos, de entre 8 y 50 residuos de ácidos nucleicos, de entre 9 y 200 residuos de ácidos nucleicos, tal como una longitud de entre 9 y 150 residuos de ácidos nucleicos, de entre 9 y 100 residuos de ácidos nucleicos, de entre 9 y 75 residuos de ácidos nucleicos, de entre 9 y 50 residuos de ácidos nucleicos, tal como una longitud de entre 10 y 200 residuos de ácidos nucleicos, de entre 10 y 150 residuos de ácidos nucleicos, de entre 10 y 100 residuos de ácidos nucleicos, de entre 10 y 75 residuos de ácidos nucleicos, de entre 10 y 50 residuos de ácidos nucleicos, tal como una longitud de entre 11 y 200 residuos de ácidos nucleicos, de entre 11 y 150 residuos de ácidos nucleicos, de entre 11 y 100 residuos de ácidos nucleicos, de entre 11 y 75 residuos de ácidos nucleicos, de entre 11 y 50 residuos de ácidos nucleicos, o tal como una longitud de entre 12 y 200 residuos de ácidos nucleicos, tal como una longitud de entre 12 y 150 residuos de ácidos nucleicos, de entre 12 y 100 residuos de ácidos nucleicos, de entre 12 y 75 residuos de ácidos nucleicos, o tal como una longitud de entre 12 y 50 residuos de ácidos nucleicos.

## ES 2 443 230 T3

Las regiones promotoras de los genes de acuerdo con la descripción se enumeran en la siguiente tabla:

Nombre HGNC	Entrada	Ensemb	SEQ ID NO.
MAL	4118	ENSG00000172005	17-20
FBN1	2200	ENSG00000166147	22
CNRIP1	25927	ENSG00000119865	23
SPG20	23111	ENSG00000133104	25
SNCA	6622	ENSG00000145335	29, 30
INA	9118	ENSG00000148798	32

- 5 Para cada uno de los genes mencionados anteriormente, los inventores han identificado subsecuencias que son particularmente útiles en el método de la descripción. En el caso de MAL, la subsecuencia en C) puede, en consecuencia, seleccionarse a partir del grupo de secuencias constituido por la secuencia especificada por la SEQ ID NO.:17 y su secuencia complementaria, la secuencia especificada por la SEQ ID NO.:18 y su secuencia complementaria, la secuencia especificada por la SEQ ID NO.:19 y su secuencia complementaria, la secuencia especificada por la SEQ ID NO.:20 y su secuencia complementaria y las subsecuencias de cualquiera de estas secuencias.
- 10 En el caso del gen de la fibrilina 1, la subsecuencia en C) es preferentemente la secuencia especificada por la SEQ ID NO.:22, o su secuencia complementaria, o una subsecuencia de cualquiera de estas.
- En el caso del marco de lectura abierto 32 del cromosoma 2, (CNRIP1), la subsecuencia en C) es preferentemente la secuencia especificada por la SEQ ID NO.:23 o su secuencia complementaria o una subsecuencia de cualquiera de estas.
- 15 En el caso de la paraplejia espástica 20, espartina (síndrome de Troyer), la subsecuencia en C) es preferentemente la secuencia especificada por la SEQ ID NO.: 25, o su secuencia complementaria o una subsecuencia de una de estas.
- 20 En el caso de la sinucleína alfa (componente no A4 del precursor amiloide), la subsecuencia en C) se selecciona preferentemente a partir del grupo de secuencias constituido por la secuencia especificada por la SEQ ID NO.:29 y su secuencia complementaria, la secuencia especificada por la SEQ ID NO.: 30 y su secuencia complementaria y las subsecuencias de cualquiera de estas secuencias.
- En el caso de la proteína del filamento del intermedio neuronal internexina alfa, la subsecuencia en C) es preferentemente la secuencia especificada por la SEQ ID NO.: 32, o su secuencia complementaria o una subsecuencia de cualquiera de estas.
- 25 También pueden ser útiles en la presente descripción los ácidos nucleicos que comprenden la región promotora de genes adicionales seleccionados, sin carácter limitante, a partir del grupo del factor potenciador de miocito 2C (SEQ ID NO.: 24), C3orf14/14HT021 (SEQ ID NO.: 21), ubiquitina-proteína-ligasa E3A (SEQ ID NO.: 26, 27 y 28), expresada en el cerebro, ligada a X 1 (SEQ ID NO.: 31) o su secuencia complementaria, o una subsecuencia de una de estas.
- 30 Hasta la fecha pocos genes han demostrado ser útiles en la detección precoz del cáncer en función de los eventos de metilación en las regiones promotoras. Sin embargo, se incluye en el alcance de la presente descripción incluir en el método análisis del estado o nivel de metilación de las regiones promotoras de marcadores conocidos de la hipermetilación en el cáncer. En realizaciones adicionales de la descripción, por lo tanto, el método comprende determinar el estado o nivel de metilación de los sitios CpG en la secuencia/región promotora de uno o más genes,
- 35 donde dicho gen o genes se seleccionan a partir del grupo constituido por:
- Metalopeptidasa ADAM con motivo de trombospondina de tipo 1 (ADAMTS1, C3-C5, KIAA1346, METH1)
  - Vimentina (VIM)
  - Proteína 1 secretada relacionada con Frizzled (SFRP1, por sus siglas en inglés); y
  - Proteína 2 secretada relacionada con Frizzled (SFRP2, por sus siglas en inglés).
- 40 Las regiones promotoras de estos genes se representan mediante los identificadores de secuencia 35-38. En consecuencia, el método de la descripción puede comprender determinar el estado de metilación de los sitios CpG en una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:



- 5
- i) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes secuencias: SEQ ID NO.: 33, SEQ ID NO.: 34, SEQ ID NO.: 35 y SEQ ID NO.: 36;
  - ii) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en i)
  - iii) Una subsecuencia de una secuencia de ácido nucleico tal como se define en i) o ii);
  - iv) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75%, tal como en al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 98% o al menos un 99% a la secuencia tal como se define en i), ii) o iii).

10 El experto en la técnica además comprenderá que las diferentes regiones promotoras mostrarán un cierto grado de degeneración. En consecuencia, tal como se indica en el apartado D) anterior, la secuencia promotora de cualquiera de los genes concretos puede ser una que no sea completamente idéntica a una de las secuencias representadas por los identificadores de secuencias 1-16. En realizaciones particulares, la secuencia de ácido nucleico en D) es idéntica en al menos un 80% a la secuencia tal como se define en A), B) o C), tal como idéntica en al menos un 85%, al menos un 90%, al menos un 95%, al menos un 98%, al menos un 99% o tal como en al menos un 99,5% a la secuencia tal como se define en A), B) o C).

15 La especificidad y sensibilidad de los genes de acuerdo con la descripción son muy altas y cada uno de los genes puede estar comprendido en los métodos de la descripción.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 20
- I) una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por: las SEQ ID NO.s: 1-4, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 25
- I) una secuencia de ácido nucleico que comprende una secuencia constituida por: la SEQ ID NO.: 6, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 30
- I) una secuencia de ácido nucleico que comprende una secuencia constituida por: la SEQ ID NO.: 7, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 35
- I) una secuencia de ácido nucleico que comprende una secuencia constituida por: la SEQ ID NO.: 9, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 40
- I) una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por: las SEQ ID NO.: 13-14, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

En una realización, el método comprende determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora en:

- 45
- I) una secuencia de ácido nucleico que comprende una secuencia constituida por: la SEQ ID NO.: 16, y sus secuencias relacionadas, y
  - II) 1, 2 o 3 secuencias de ácido nucleico tal como se definen en el párrafo precedente.

50 Los métodos de acuerdo con la invención tienen como objetivo la detección o el diagnóstico del cáncer, tal como un tumor, en el sistema aéreo-digestivo. El "sistema aéreo-digestivo" o "tracto aéreo-digestivo" incluye los pulmones y el tracto gastrointestinal: esófago, estómago, páncreas, hígado, vesícula biliar/conducto biliar, intestino delgado e intestino grueso, incluido el colon y recto. En particular, el tumor puede seleccionarse a partir del grupo constituido por: tumores colorrectales, tumores pulmonares (incluidos el carcinoma microcítico de pulmón y/o carcinomas no microcíticos de pulmón), tumores esofágicos, tumores gástricos, tumores pancreáticos, tumores hepáticos, tumores

de la vesícula biliar y/o del conducto biliar, tumores del intestino delgado y tumores del intestino grueso.

Por lo tanto, en una realización de la invención el cáncer se selecciona a partir del grupo constituido por: tumores colorrectales, tumores pulmonares (incluido el carcinoma microcítico de pulmón y/o el carcinoma no microcítico de pulmón), tumores esofágicos, tumores gástricos, tumores pancreáticos, tumores hepáticos, tumores de la vesícula biliar y/o conducto biliar, tumores del intestino delgado y tumores del intestino grueso.

5 A fin de determinar el número de sitios CpG metilados, el estado de metilación de los sitios CpG o el nivel de metilación en dicha secuencia/región promotora, el método de acuerdo con la descripción requiere que se aisle una cantidad suficiente de ADN a partir de un individuo concreto. El experto en la técnica conocerá suficientes técnicas para aislar y purificar ADN en la cantidad y de la calidad requeridas. Para la mayoría de los propósitos, el ADN puede aislarse a partir de una muestra sanguínea, una muestra fecal, una muestra tisular o una muestra de moco procedente de los pulmones de dicho individuo. En general, es deseable llevar a cabo el método de la invención de manera no invasiva siempre que esto sea posible: para el cáncer gastrointestinal, recoger ADN a partir de muestras fecales será a menudo práctico y conveniente. En relación a los tumores pulmonares, aislar ADN a partir de muestras de moco procedentes del pulmón puede ofrecer un método conveniente para la recogida no invasiva de ADN. Para otros tumores, incluidos los tumores en el hígado y páncreas, puede ser preferible recoger muestras tisulares para el aislamiento posterior de ADN.

Por lo tanto, en una realización se obtiene la muestra a partir de sangre, excrementos, orina, líquido pleural, bilis, líquido bronquial, enjuagues bucales, piezas histológicas, ascitis, pus, líquido cefalorraquídeo, muestras obtenidas por aspiración, líquido folicular, tejido o moco.

20 Cuando se utiliza el método de acuerdo con la descripción con el propósito de determinar únicamente la presencia de cáncer o un tumor en el sistema aéreo-digestivo, en particular cuando se requiere un resultado del tipo "sí/no", es deseable que el método pueda limitarse al análisis del nivel de metilación o estado de metilación de los sitios CpG en las regiones promotoras de 2-4 genes. Esto requiere claramente que los genes tengan una frecuencia extremadamente alta de hipermetilación durante el desarrollo y evolución del cáncer.

25 Por lo tanto, para propósitos meramente diagnósticos, se prefiere sobre todo limitar el análisis a las regiones promotoras en muy pocos genes. Como se ha mencionado anteriormente, esto requiere la disponibilidad de un conjunto de marcadores para la hipermetilación, donde cada marcador tenga una sensibilidad y especificidad elevadas. Para propósitos más sutiles, sin embargo, puede ser necesario analizar el estado o nivel de metilación de las regiones promotoras en un número mayor de genes marcadores. Los métodos de acuerdo con la invención, por tanto, pueden comprender la determinación del estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la secuencia/región promotora de al menos 2 genes, tal como al menos 3 genes, tal como al menos 4 genes, tal como al menos 5 genes, tal como al menos 7 genes, al menos 8 genes, al menos 9 genes, al menos 10 genes, al menos 11 genes, al menos 12 genes, al menos 13 genes, al menos 14 genes, al menos 15 genes, al menos 16 genes, al menos 17 genes, al menos 18 genes, al menos 19 genes o al menos 20 genes, incluido al menos 1 gen tal como se define en la reivindicación 1 a fin de determinar el nivel de riesgo de inicio y/o evolución tumoral en un individuo.

40 De acuerdo con lo que se explica anteriormente, el método de la descripción además comprende determinar el estado de metilación de los sitios CpG en al menos una, tal como al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19 o tal como al menos 20 secuencias de ácido nucleico adicionales tal como se definen anteriormente o sus secuencias relacionadas.

45 Por lo tanto, para la mayoría de los propósitos será insuficiente analizar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en la región promotora de un único gen. Los métodos de acuerdo con la descripción, por tanto, pueden comprender la determinación del estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la secuencia/región promotora de al menos 2 genes, tal como al menos 3 genes, tal como al menos 4 genes, tal como al menos 5 genes, tal como al menos 7 genes, al menos 8 genes, al menos 9 genes, al menos 10 genes, al menos 11 genes, al menos 12 genes, al menos 13 genes, al menos 14 genes, al menos 15 genes, al menos 16 genes, al menos 17 genes, al menos 18 genes, al menos 19 genes o al menos 20 genes, donde se selecciona al menos un gen a partir del grupo de genes definido anteriormente.

50 Por lo tanto, en otro aspecto la descripción trata sobre un método donde se determina el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional.

Donde al menos un marcador adicional se selecciona a partir del grupo constituido por:

55 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927  
SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111

- FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200  
 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622;  
 5 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118; y  
 MAL, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000172005.

*Combinación de marcadores génicos*

10 Por lo tanto, por las razones explicadas anteriormente, el nivel de metilación o el estado de metilación puede combinarse con determinaciones de uno o más marcadores diferentes, y compararse con un nivel de referencia combinado. Los niveles determinados de marcadores se pueden combinar mediante operaciones aritméticas tales como adición, sustracción, multiplicación y manipulaciones aritméticas de los porcentajes, raíz cuadrada, exponenciación y funciones logarítmicas. También pueden combinarse los niveles tras las manipulaciones utilizando  
 15 varios modelos, p. ej., regresión logística y estimaciones de la máxima verosimilitud. Se pueden llevar a cabo varias combinaciones de biomarcadores y varios medios para calcular el valor de referencia combinado mediante medios conocidos por el experto en la técnica.

Por lo tanto, otra realización de la descripción trata sobre un método de acuerdo con cualquiera de las reivindicaciones precedentes, donde se determina el nivel de metilación o estado de metilación de al menos un marcador adicional.

20 Dicho marcador adicional puede ser, sin carácter limitante, CNRIP1, SPG20, FBN1, SNCA, INA, MAL, ADAMTS1, VIM, SFRP1, SFRP2 o CRABP1. Los marcadores pueden compararse con un conjunto de datos de referencia para determinar si el individuo padece cáncer o tiene un riesgo mayor de desarrollar un cáncer.

Se puede lograr un método para construir un ensayo diagnóstico basado en un marcador combinado mediante la combinación de los niveles de metilación o el estado de metilación (o un valor derivado de estos) de dos o más  
 25 marcadores individuales mediante manipulación aritmética (p. ej., adición). Como puede haber una variedad en el nivel de metilación o estado de metilación de los diferentes marcadores es relevante ponderar las medidas a fin de conseguir una combinación independiente de las diferencias en, p. ej., el nivel o estado de metilación. Esto puede llevarse a cabo mediante una simple normalización a partir de una mediana o un promedio a partir de material estándar.

30 *Sinergia*

Combinando los diferentes marcadores de genes de acuerdo con la descripción puede conseguirse un efecto sinérgico.

Específicamente, tal como se utiliza en la presente, sinergia se refiere al fenómeno en el cual varios marcadores que actúan juntos crean una "señal marcadora combinada" con una sensibilidad o especificidad mayores para el  
 35 diagnóstico, que la predicha conociendo únicamente de la sensibilidad o especificidad separadas de los marcadores.

Por lo tanto, en una realización de la presente descripción el uso combinado de al menos un marcador adicional (p. ej., CNRIP1, SPG20, FBN1, SNCA, INA, MAL) proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

40 En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1 e INA proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1 y SNCA proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1 y FBN1 proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

45 En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1 y SPG20 proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

En otra realización de la presente descripción, el uso combinado de los marcadores INA y SNCA proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

50 En otra realización de la presente descripción, el uso combinado de los marcadores INA y FBN1 proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

En otra realización de la presente descripción, el uso combinado de los marcadores INA y SPG20 proporciona un





En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1, SPG20, FBN1, MAL e INA proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

- 5 En otra realización de la presente descripción, el uso combinado de los marcadores CNRIP1, SPG20, FBN1, SNCA y MAL proporciona un efecto sinérgico en relación con la sensibilidad y/o especificidad.

10 Por lo tanto, en una realización los métodos de acuerdo con la descripción comprenden determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de CNRIP1 junto con la determinación del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional seleccionado a partir del grupo constituido por:

- 15 SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111  
 FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200  
 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e  
 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118.

20 Por lo tanto, en una realización los métodos de acuerdo con la descripción comprenden determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de SPG20 junto con la determinación del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional seleccionado a partir del grupo constituido por:

- 25 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927  
 FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200  
 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e  
 30 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118.

35 Por lo tanto, en una realización los métodos de acuerdo con la descripción comprenden determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de FBN1 junto con la determinación del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional seleccionado a partir del grupo constituido por:

- CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927  
 SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111  
 40 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e  
 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118.

45 Por lo tanto, en una realización los métodos de acuerdo con la descripción comprenden determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de SNCA junto con la determinación del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional seleccionado a partir del grupo constituido por:

- CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927  
 50 SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111  
 FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200; e  
 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118.  
 55

Por lo tanto, en una realización los métodos de acuerdo con la descripción comprenden determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de INA junto con la

determinación del nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG de al menos un marcador adicional seleccionado a partir del grupo constituido por:

- 5 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927
- 5 SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111
- FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200; y
- 10 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622.

*Tratamiento con bisulfito y reacción en cadena de la polimerasa específica de metilación*

15 La invención no está limitada por los tipos de ensayos utilizados para determinar el estado de metilación de los miembros del gen o del conjunto de genes. De hecho, cualquier ensayo que pueda emplearse para determinar el nivel o estado de metilación del gen o conjunto de genes debería bastar para los propósitos de la presente invención.

20 Una estrategia práctica para determinar el estado de metilación de las islas de CpG en una región promotora puede comprender un paso de tratamiento de la secuencia promotora con bisulfito. El tratamiento del ADN con bisulfito da lugar a variaciones en la secuencia ya que convierte las citosinas no metiladas, pero no las metiladas, en uracilo. El tratamiento con bisulfito seguido del análisis de la secuencia permite una presentación positiva de 5-metilcitosinas en el promotor génico tras la modificación con bisulfito ya que las citosinas no metiladas aparecen como timidinas, mientras que las 5-metilcitosinas aparecen como citosinas en la secuencia final. En realizaciones particulares de la invención, el estado de metilación de dicha secuencia/región promotora se determina, por tanto, mediante la secuenciación del ácido nucleico (secuenciación con bisulfito).

25 En realizaciones adicionales de la descripción, se determina el número de sitios CpG metilados, el estado de metilación de los sitios CpG o el nivel de metilación de dicha región/secuencia promotora mediante PCR específica de metilación. En los ejemplos de la presente solicitud se proporcionan un conjunto de diseños de cebadores y condiciones de PCR adecuados. En general, sin embargo, el experto en la técnica también poseerá el conocimiento requerido para ser capaz de determinar los diseños de cebadores y condiciones apropiados para los análisis por PCR.

30 Como sabrá el experto en la técnica, la fluorescencia en tiempo real ofrece una estrategia rápida y conveniente para la detección de productos de PCR y se puede aplicar fácilmente a procedimientos diagnósticos donde se requiere un procesamiento ultrarrápido. En la realización de la invención actualmente preferida dicha PCR específica de metilación comprende, por tanto, la detección fluorescente en tiempo real de los productos de PCR.

35 Al igual que para la mayoría de los otros procedimientos de PCR, el método de la descripción puede comprender un paso de separación de los productos de acuerdo con el tamaño. En particular, los métodos de la descripción pueden comprender un paso de separación de los productos de PCR resultantes mediante electroforesis capilar o en gel.

40 Como parte del análisis, los productos de PCR resultantes pueden detectarse mediante la utilización de un marcador seleccionado a partir del grupo constituido por marcadores fluorescentes, marcadores quimioluminiscentes y marcadores radiactivos. Por motivos prácticos y de seguridad, para la mayoría de los casos se prefieren los marcadores no radiactivos.

El nivel o estado de metilación de dicha secuencia/región promotora también puede determinarse mediante pirosecuenciación, espectrometría de masas o mediante el uso de enzimas de restricción específicas de metilación.

45 El nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG se determina, sin carácter limitante, mediante secuenciación con bisulfito, reacción en cadena de la polimerasa específica de metilación (MSP) cualitativa y/o cuantitativa, pirosecuenciación, transferencia de Southern, barrido del genoma de los puntos de referencia de restricción (RLGS, por sus siglas en inglés), extensión del cebador en un único nucleótido, micromatriz de islas de CpG, SNUPE, COBRA, espectrometría de masas, mediante la utilización de enzimas de restricción específicas de metilación, mediante la determinación del nivel de expresión de dichos genes  
50 o una combinación de estos.

55 En realizaciones preferidas, la PCR específica de metilación utilizada en el método de la invención comprende la utilización de cebadores de ácidos nucleicos los cuales son capaces de hibridarse con una secuencia de ácido nucleico que comprende 2 sitios CpG y un residuo de citosina el cual no está en un sitio CpG. Es deseable la inclusión de un residuo de citosina de este tipo, la cual no está metilada, a fin de distinguir mejor ADN modificado con bisulfito del ADN no modificado con bisulfito. Los cebadores para las secuencias metiladas siempre se unirán a

5 los sitios CpG metilados, los cuales son sitios que se mantienen como CpG tras la conversión con bisulfito. En presencia de ADN no modificado, este contendrá sitios CpG independientemente del estado de metilación y los cebadores específicos de metilación se unirán entonces al ADN no modificado, lo que originará positivos falsos. La inclusión de una "C", la cual no está en un sitio CpG en el área atacada por el cebador evitará que el cebador se una a ADN no modificado ya que este ADN contendrá "C" mientras que el ADN modificado contendrá "T" en la misma posición.

En otras realizaciones adicionales, la PCR específica de metilación comprende la utilización de cebadores de ácidos nucleicos los cuales son capaces de hibridarse con una secuencia de ácido nucleico que comprende 2 sitios CpG y un residuo de citosina el cual no está en un sitio CpG.

10 Los métodos de acuerdo con la descripción pueden combinarse con cualquier otro parámetro conocido del cáncer. Por lo tanto, el estado o nivel de metilación de un gen de la descripción puede combinarse, sin carácter limitante, con cualquiera de los siguientes parámetros del cáncer: un ensayo de la integridad del ADN genético, ploidía, estado de mutación de los genes, cambios genómicos, genes de fusión, variantes de corte y empalme, diferencias en la expresión, miARN.

15 Utilización

Debido a su sensibilidad y especificidad elevadas, los marcadores de acuerdo con la descripción son muy adecuados para su utilización como marcadores para el cáncer. Por lo tanto, otro aspecto de la invención trata sobre la utilización del gen

20 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927

en un ensayo diagnóstico donde se evalúa el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG como un indicador de si el individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras el tratamiento contra el cáncer en el sistema aéreo-digestivo.

25 Otra realización trata sobre la utilización de una secuencia de ácido nucleico, donde dicho ácido nucleico comprende una secuencia de ácido nucleico seleccionado a partir de grupo constituido por:

- A) Una secuencia de ácido nucleico tal como la define la SEQ ID NO:7,
- B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);
- C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);
- 30 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

35 En un ensayo diagnóstico donde el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG se evalúa como un indicador de si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer o si un individuo está recayendo tras el tratamiento contra el cáncer en el sistema aéreo-digestivo.

Anticuerpo

La descripción también se refiere a un anticuerpo contra las secuencias metiladas. Por lo tanto, otra realización de la descripción se refiere a un anticuerpo que reconoce secuencias de ácido nucleico metilado seleccionadas a partir del grupo constituido por:

- 40 A) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes: la SEQ ID NO.: 6, SEQ ID NO.: 7, SEQ ID NO.: 9, SEQ ID NO.: 13, SEQ ID NO.: 14 y la SEQ ID NO.: 16;
- B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);
- C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);
- 45 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

Kit de diagnóstico

50 En un aspecto preferido la descripción proporciona un kit de diagnóstico para la determinación del cáncer que comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario a una secuencia de ácido nucleico de los genes seleccionados a partir de:

CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927



SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111

FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200

5 SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e

INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118

10 Un segundo aspecto de la descripción proporciona un kit de diagnóstico que comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores, tales como 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 11 o más, 12 o más, 13 o más, 14 o más, 15 o más, 16 o más, 17 o más, 18 o más, 19 o más o tales como 20 o más oligonucleótidos que actúan de cebadores o conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico en la región/secuencia promotora de uno o más genes, donde dicho gen o genes se seleccionan a partir del grupo constituido por:

CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927

SPG20, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000133104, identificación de entrada 23111

20 FBN1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000166147, identificación de entrada 2200

SNCA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000145335, identificación de entrada 6622; e

25 INA, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000148798, identificación de entrada 9118

En particular, el kit de acuerdo con este aspecto de la descripción comprende uno o más oligonucleótidos que actúan de cebadores o conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:

- 30 A) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes: la SEQ ID NO.: 6, SEQ ID NO.: 7, SEQ ID NO.: 9, SEQ ID NO.: 13, SEQ ID NO.: 14 y la SEQ ID NO.: 16;  
 B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);  
 C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B); Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

35 En realizaciones particulares el kit comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico en la región/secuencia promotora de un gen que se selecciona a partir del grupo constituido por:

40 Metalopeptidasa ADAM con motivo de trombospondina de tipo 1 (ADAMTS1, C3-C5, KIAA1346, METH1)  
 Vimentina (VIM)

Proteína 1 secretada relacionada con Frizzled (SFRP1); y

Proteína 2 secretada relacionada con Frizzled (SFRP2)

MAL (proteína de diferenciación de linfocitos T)

45 marco de lectura abierto del cromosoma 3 (C3orf14/14HT021),  
 ubiquitina-proteína-ligasa E3A (UBE3A, AS, ANCR, E6-AP, FLJ26981), expresada en el cerebro, ligada a X 1 (BEX1),  
 factor potenciador de miocito 2C, MEF2c.

50 En otro aspecto de la descripción, el kit comprende uno o más oligonucleótidos que actúan de cebadores o conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:

- 55 A) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes: la SEQ ID NO.: 1, SEQ ID NO.: 2, SEQ ID NO.: 3, SEQ ID NO.: 4, SEQ ID NO.: 5, SEQ ID NO.: 8, SEQ ID NO.: 10, SEQ ID NO.: 11, SEQ ID NO.: 12 y la SEQ ID NO.: 15;  
 B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);  
 C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);  
 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se

define en A), B) o C).

De acuerdo con estas realizaciones el kit comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores, donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:

- 5
- i) Una secuencia de ácido nucleico tal como la define cualquiera de las siguientes: la SEQ ID NO.: 33, SEQ ID NO.: 34, SEQ ID NO.: 35 y la SEQ ID NO.: 36;
  - ii) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en i);
  - iii) Una subsecuencia de una secuencia de ácido nucleico tal como se define en i) o ii);
  - 10 iv) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75%, tal como al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 98% o al menos un 99%, a una secuencia tal como se define en i), ii) o iii).

A la vista de la especificidad extremadamente elevada de MAL como marcador para el cáncer, el kit comprende uno o más oligonucleótidos que actúan de cebadores o uno o más conjuntos de oligonucleótidos que actúan de cebadores donde cada uno de los cuales es complementario/capaz de hibridarse con una secuencia de ácido nucleico seleccionada a partir del grupo constituido por:

- 15
- A) Una secuencia de ácido nucleico tal como la define la SEQ ID NO.: 1;
  - B) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en A);
  - C) Una subsecuencia de una secuencia de ácido nucleico tal como se define en A) o B);
  - 20 D) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en A), B) o C).

Se pueden considerar varios diseños para el kit de la descripción. En una realización cada uno de dichos cebadores o conjuntos de cebadores está en recipientes separados. Esto permitirá al usuario final del kit preparar diferentes mezclas de cebadores para diferentes objetivos. De acuerdo con otras realizaciones, sin embargo, los cebadores o conjuntos de cebadores pueden suministrarse como una mezcla.

25

Para ciertos usos, tales como a efectos diagnósticos tradicionales, el kit de diagnóstico puede incluir pocos cebadores o conjuntos de cebadores. En particular, esto es relevante cuando el kit se ha de utilizar en una aplicación donde se requiere un resultado del tipo "sí/no", tal como cuando el kit se utiliza simplemente a fin de determinar si se está desarrollando un tumor o carcinoma en el sistema aéreo-digestivo. A tales efectos el número de cebadores o conjuntos de cebadores en el kit puede limitarse a 2-4, donde al menos un cebador o un conjunto de cebadores, tal como al menos 2, al menos 3 o al menos 4 cebadores o conjuntos de cebadores, es complementario/capaz de hibridarse con una secuencia de ácido nucleico de acuerdo con las SEQ ID NO.: 1-16 o secuencias que son complementarias o idénticas en parte a estas tal como se ha definido anteriormente en los apartados C) y D).

35

Sin embargo, tal como se ha discutido anteriormente en relación con el método de la descripción, el estado de metilación de las islas de CpG de los genes marcadores de la descripción también pueden utilizarse para análisis más complejos, tal como a fin de determinar el nivel de riesgo del inicio de un tumor y/o evolución en un individuo.

40

En tales realizaciones el kit de diagnóstico de la descripción contendrá típicamente cebadores o conjuntos de cebadores que son capaces de atacar un mayor número de genes marcadores. Un kit con tales propósitos deberá incluir normalmente 5 o más cebadores o conjuntos de cebadores, donde al menos un cebador o un conjunto de cebadores, tal como al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 12, al menos 13, al menos 14, al menos 15, o tal como al menos 16 cebadores o un conjunto de cebadores, es complementario/capaz de hibridarse con una secuencia de ácido nucleico de un gen marcador de acuerdo con la descripción.

45

El kit de diagnóstico de acuerdo con la descripción puede además comprender cualquier reactivo o medio necesario a fin de llevar a cabo los análisis requeridos, tales como análisis por PCR, tales como análisis de secuencia mediante la reacción en cadena de la polimerasa específica de metilación (MSP), tratamiento con bisulfito, secuenciación con bisulfito, electroforesis, pirosecuenciación, espectrometría de masas y análisis de secuencias mediante digestión restrictiva, metilación cualitativa y/o cuantitativa, pirosecuenciación, transferencia de Southern, barrido del genoma de los puntos de referencia de restricción (RLGS), extensión del cebador en un único nucleótido, micromatriz de islas de CpG, SNUPE, COBRA, espectrometría de masas, mediante la utilización de enzimas de restricción específicas de metilación o mediante la determinación del nivel de expresión de dichos genes. En particular el kit puede comprender además uno o más componentes seleccionados a partir del grupo constituido por:

50

55

desoxirribonucleósidos trifosfatos, tampones, estabilizantes, ADN-polimerasas termoestables, endonucleasas de restricción (incluidas las endonucleasas específicas de metilación) y marcadores (incluidos marcadores fluorescentes, quimioluminiscentes y radioactivos). El ensayo diagnóstico de acuerdo con la descripción puede

además comprender uno o más reactivos requeridos para el aislamiento de ADN.

Cabe destacar que las realizaciones y características descritas en el contexto de uno de los aspectos de la presente descripción también son aplicables a otros aspectos de la descripción.

5 Cuando se hace referencia en singular a un objeto de acuerdo con la presente invención o a uno de sus distintivos o características, esto también se refiere al objeto o a sus características en plural. A modo de ejemplo, cuando se hace referencia a “un polipéptido” debe sobreentenderse que se hace referencia a uno o más polipéptidos.

10 A lo largo de la presente memoria descriptiva se sobreentenderá que la palabra “comprender” o variaciones tales como “comprende” o “que comprende” implica la inclusión de un elemento, número entero o paso enunciados, o grupo de elementos, números enteros o pasos, pero no la exclusión de cualquier otro elemento, número entero o paso, o grupo de elementos, números enteros o pasos.

La invención se describirá a continuación más detalladamente en los siguientes ejemplos no limitantes.

### Ejemplos

Ejemplo 1:

#### *Tratamiento con bisulfito y PCR específica de metilación*

15 Se trató ADN procedente de líneas celulares y carcinomas colorrectales con bisulfito tal como se ha descrito previamente (Grunau *et al.* y Fraga *et al.*). Mientras que el ADN procedente de adenomas se trató con bisulfito de acuerdo con el protocolo del CpGenome™ DNA modification kit (Intergen Boston, MA) (Smith-Sørensen *et al.*). El estado de metilación del promotor de MAL, C3orf14, FBN1, MEF2c, CNRIP1, SPG20, UBE3A, SNCA, BEX e INA se analizó subsiguientemente mediante PCR específica de metilación (MSP), un método que permite la distinción entre  
20 alelos metilados y no metilados (Herman *et al.* y Derks *et al.*). Se diseñaron todos los cebadores con MethPrimer (U y Dahiya) o Methyl Primer Express (Applied Biosystems). Sus secuencias se enumeran en la Tabla 1, junto con la longitud del fragmento que es el producto, localización del cebador y temperatura de hibridación para cada PCR. Se amplificaron los fragmentos utilizando la ADN-polimerasa HotStarTaq (QIAGEN Inc., Valencia, CA) y se confirmaron todos los resultados con una segunda serie independiente de MSP.

Tabla 1: Cebadores para la PCR específica de metilación

Gen	Secuencia metilada (secuencia del cebador directo/inverso)	SEQ ID NO.	Secuencia no metilada (secuencia del cebador directo/inverso)	SEQ ID NO.
MAL	TTCCGGGTTTTTTGTTTTAATTC/ GAAAACCATACGACCGTACTAACGT	37/38	TTTTGGGTTTTTTGTTTTAATTT/ ACAAAACCCATAACACATACTAATC	39/40
C3orf14	GTAATTTAGATTTCCGGAGGGC/ CGACCAAAAAAAAAACGAAAA	41/42	TTTGTAATTTAGATTTTGGAGGGT/ CCAACCAAAAAAAAAACAAAAACA	43/44
FBN1	GTATTTTTTTCGCGAGAAATC/ AATCGTAACCGCTACAACC	45/46	AAAGTATTTTTTTTGTGAGAAATT/ CCCAATCATAAACCACCTACAACC	47/48
MEF2c	GTTATTTTTTAATTCGATCGGTC/ AAACCGCTCGAAAAAAA	49/50	TTGGTTATTTTTTAATTTTGATTTGGTT/ CCAAAACCCACTCAAAAAAAA	51/52
CNRIPI	TCGTTTTTTTGGTATAGTGGTC/ CAAAATCCGGCAACTAAA	53/54	GTTTTGTTTTTTGGTATAGTGGTT/ CAAAATCCACACAACATAAAAAAC	55/56
SPG20	TGGAACGTTTTGGTTGTTAC/ TACCCTCGAAAACCTCCCTACG	57/58	GTGGAATGTTTTGGTTGTTAT/ TTACCTCAAAAAACTCCCTACA	59/60
UBE3A	CGTTGTTTTCGCGGATATTC/ CCCGTCGTCCTCTATAATCA	61/62	GTGTTGTTTGTGGGATATTT/ CCCCATCATCTCCTATAATCA	63/64
SNCA	CGGGTTGTAGCGTAGATTTTC/ CGTCGAATTAACCACTCCC	65/66	GTGTGGGTTGTAGTGTAGATTTT/ TCATCAAAATAACCACTCCCAA	67/68

Gen	Secuencia metilada (secuencia del cebador directo/inverso)	SEQ ID NO.	Secuencia no metilada (secuencia del cebador directo/inverso)	SEQ ID NO.
BEX1	AGTTAATTGGTCGGTTC / CGAATAACGACTACACCGAA	69/70	ATTAGTTAATTGGTTGTTGGTTT / ACACAAATTAACAACACTACACCAAA	71/72
INA	AGGAGTTTCGTTTTTAGCGC / ACGACTTCACCGGCACTAC	73/74	AGTAGGAGTTTTGTTTTTAGTGT / ACAACCTCACACAAACTACAAA	75/76

*Secuenciación con bisulfito*

5 Se amplificaron todos los fragmentos con la ADN-polimerasa HotStarTaq y se eluyeron desde un gel de agarosa al 2% mediante el MinElute™ Gel Extraction kit (QIAGEN). Se secuenciaron subsiguientemente las muestras utilizando el dGTP BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) en un secuenciador 3730 (Applied Biosystems). Se calculó la cantidad aproximada de metilcitosina de cada sitio CpG en los diferentes fragmentos comparando la altura del pico de la señal de citosina con la suma de la altura de los picos de las señales de citosina y timina, tal como describieron previamente Melki *et al.*

Tabla 2: Cebadores para la secuenciación con bisulfito

Gene	Secuencia del cebador (directo/inverso)	SEQ ID NO.
<i>MAL</i>	GGGTTTTTTTTGTTTTTAATT/ACCAAAAACCACTCACAAACTC	77/78
<i>C3orf14</i>	GGAGGGTAGATGATTTTTGAGAA/CTTCCCCTTCCCCTAACTACTA	79/80
<i>FBN1</i>	AGGGGGTGTTATTTTTTTTTTTTT/CCCAATCCCTATCCCTACC	81/82
<i>MEF2c</i>	TTTTTGAYGAGTTTGTTATT/CCACCTAATTCAAACATACAACC	83/84
<i>CNRIP1</i>	TTTTAYGTAGTTGGTYGAGG/CTCCTTAAACTATAACCCCCCT	85/86
<i>SPG20</i>	ATTTAGTTTGAGTAGGTYGGTG/CTCCATCCTAACAATCCATAAA	87/88
<i>UBE3A</i>	GGGGGGTGTTTAGAGGG/CCTCCTACCAAAAACCTACAACC	89/90
<i>SNCA</i>	AGAAGGGGTTTAAGAGAGG/ACTATCCCCAAAAAAAACC	91/92
<i>BEX1</i>	ATTTGTGGTTTTTTAGATTGGA/CAAAAAAACCACTATATTCCCA	93/94
<i>INA</i>	GATGTAGATGGTTTTGTTYGG/CAAACRAAAACCATCCCC	95/96

10 Cuando el análisis se llevó a cabo mediante la reacción en cadena de la polimerasa específica de metilación (MSP) se observó hipermetilación de *MAL* con una frecuencia excepcionalmente alta en los tumores malignos (83%, 40 de 48 carcinomas) así como en los tumores benignos del intestino grueso (73%, 43 de 59 adenomas) (Figura 1).

**Ejemplo 2**

15 *Se llevó a cabo la reacción en cadena de la polimerasa específica de metilación (MSP) para los siguientes genes: MAL, C3orf14, FBN1, SPG20, SNCA, BEX1, INA, CNRIP1, UBE3A, MEF2C*

20 Para cada muestra (líneas celulares de cáncer de colon, carcinomas colorrectales, adenomas y mucosa normal), se trataron con bisulfito 1,3 µg de ADN utilizando el EpiTect bisulphite kit (Qiagen Inc., Valencia, CA) siguiendo el protocolo de los fabricantes. El ADN modificado se eluyó en 40 µL de tampón de elución (incluido en el kit). Debido a que la modificación con bisulfito da lugar a diferencias en la secuencia, se utilizaron dos pares de cebadores para amplificar cada gen (remítase a la lista de cebadores del Ejemplo 1), uno específico para la plantilla no metilada y el otro específico para la plantilla metilada. La mezcla de PCR de 25 µL contuvo 1 x tampón de PCR, 0,75 µL de la plantilla tratada con bisulfito, MgCl<sub>2</sub> 1,5-2,0 mM, 20 pmol de cada cebador, dNTP 200 µM y 0,625-1 U de la ADN-polimerasa HotStarTaq (Qiagen). Se utilizó ADN de placenta humana (Sigma Chemical Co., San Luis, MO, EE. UU.) tratado *in vitro* con metiltransferasa SssI (New England Biolabs Inc., Beverly, MA, EE. UU.) como control positivo para la reacción de MSP metilada, mientras que se utilizó ADN procedente de linfocitos normales como control positivo para los alelos no metilados. Se utilizó agua como control de PCR negativo en ambas reacciones.

El programa de PCR consistió en 15 min de desnaturalización a 95 °C, seguido de 35 ciclos de 30 s a 95 °C, 30 s a la temperatura de hibridación y 30 s a 72 °C. Se llevó a cabo una elongación final a 72 °C en 7 minutos.

Temperatura de hibridación y contenido de MgCl<sub>2</sub> para los respectivos genes evaluados hasta el momento:

30 *MAL*: 56 °C, MgCl<sub>2</sub> 1,5 mM  
*C3orf14*: 53 °C, MgCl<sub>2</sub> 1,5 mM  
*FBN1*: 48 °C, MgCl<sub>2</sub> 1,7 mM para la reacción no metilada y 2,0 para la reacción metilada  
*SPG20*: 56 °C, MgCl<sub>2</sub> 1,5 mM  
*SNCA*: 53 °C, MgCl<sub>2</sub> 1,5 M  
35 *BEX1*: 51 °C, MgCl<sub>2</sub> 1,5 M  
*INA*: 55 °C, MgCl<sub>2</sub> 1,5 mM

CNRIP1: 52°C, MgCl<sub>2</sub> 1,5 mM

Resultados: Metilación de MAL, UBE3A, MEF2C, FBN1, C3orf14, BEX1, INA, SNCA, SPG20 y CNRIP1.

Los resultados obtenidos en la reacción en cadena de la polimerasa específica de metilación (MSP) se presentan a continuación en la Tabla 3:

5 Tabla 3

	MAL	UBE3A	MEF2C	FBN1	c3orf14
<b>Líneas celulares de cáncer de colon</b>	19/20 (95%)	0/20 (0%)	4/20 (20%)	18/20 (90%)	18/20 (90%)
<b>Carcinomas colorrectales</b>	49/61 (80%)			49/49 (82%)	27/49 (55%)
<b>Adenomas</b>	45/63 (71%)			34/59 (58%)	33/59 (56%)
<b>Mucosa (cáncer normal)</b>	2/21 (10%)			2/21 (10%)	9/21 (43%)
<b>Mucosa (normal normal)</b>	1/23 (4%)			1/19 (5%)	5/21 (24%)
	BEX1	INA	SHCA	SPG20	CNRIP1
<b>Líneas celulares de cáncer de colon</b>	18/19 (95%)	19/20 (95%)	19/20 (95%)	20/20 (100%)	20/20 (100%)
<b>Carcinomas colorrectales</b>	45/49 (92%)	33/48 (69%)	37/48 (77%)	44/49 (90%)	45/48 (94%)
<b>Adenomas</b>	52/58 (90%)	31/59 (53%)	42/61 (69%)	48/58 (83%)	53/59 (88%)
<b>Mucosa (cáncer normal)</b>	15/21 (71%)	2/20 (10%)	14/21 (67%)		9/21 (43%)
<b>Mucosa (normal normal)</b>	9/22(45%)	0/21 (0%)	2/21 (10%)	1/20 (5%)	0/21 (0%)

10 En general, los genes marcadores analizados aquí están metilados con una frecuencia extremadamente alta en líneas celulares de cáncer colorrectal y carcinomas colorrectales mientras que la frecuencia de metilación es baja en la mucosa normal de donantes no cancerosos. Estos resultados confirman la utilidad de los genes marcadores en el diagnóstico de tumores en el sistema aéreo-digestivo y en la monitorización del desarrollo tumoral.

15 En particular, MAL está metilado en 1/23 (4%) de las muestras de mucosa normal procedente de donantes no cancerosos, en 2/21 (10%) de muestras de mucosa normal tomadas lejos del tumor primario, en 45/63 (71%) de los adenomas, en 49/61 (80%) de los carcinomas y en 19/20 (95%) de las líneas celulares de cáncer de colon. Cabe destacar que las frecuencias de metilación observadas para MAL se desvían ligeramente de aquellas observadas en el Ejemplo 1. Esta desviación se debe principalmente al hecho de que el conjunto de muestras analizadas se ha expandido.

20 De igual modo, FBN1 e INA están metilados muy raramente en muestras de mucosa normal procedentes tanto de donantes no cancerosos como en muestras de mucosa normal tomadas lejos del tumor primario (1/19, 5% y 2/21, 10%, respectivamente para FBN1; 0/21, 0% y 2/20, 10%, respectivamente para INA). Simultáneamente, tanto FBN1 como INA están metilados frecuentemente en carcinomas (40/49, 82% y 33/48, 69% respectivamente) así como en adenomas (37/59, 58% y 31/59, 53%, respectivamente). Las frecuencias de metilación bajas en ambas cohortes de muestras de mucosa normal y las frecuencias de metilación elevadas tanto en los tumores malignos como benignos indican que FBN1 e INA son particularmente prometedores para la detección de tumores en una etapa temprana del desarrollo. Simultáneamente, un ensayo que comprenda estos dos marcadores tendrá, con toda probabilidad, una especificidad elevada.

30 En general, SNCA, SPG20 y CNRIP1 tienen frecuencias de metilación más elevadas en carcinomas que el último grupo (37/48 (77%), 44/49 (90%) y 45/48 (94%), respectivamente), así como en adenomas (42/61 (69%), 48/58 (83%) y 52/59 (88%), respectivamente). Incluyendo estos marcadores en un ensayo no invasivo, es probable que aumente la sensibilidad. Además de tener frecuencias de metilación bajas en muestras de mucosa normales procedentes de donantes no cancerosos, estos marcadores tienen frecuencias de metilación relativamente elevadas en muestras de mucosa normal tomadas lejos del tumor primario (14/21 (67%), (29-90%) y 9/21 (43%) respectivamente). Esto puede indicar un "efecto de campo" alrededor del tumor, donde las células aparentemente normales en las inmediaciones del tumor también presentan metilaciones de estos tres genes. La presencia de un efecto de campo de este tipo podría aumentar la sensibilidad de un ensayo no invasivo a partir de heces, ya que se verterían más células que presentan metilaciones de SNCA, SPG20 y/o CNRIP1 en la luz del colon y se excretarían con las heces.

Ejemplo 3

*Metilación de los genes marcadores en diferentes muestras*

Se purificó ADN a partir de excrementos y se llevó a cabo la MSP para los siguientes genes: MAL, FBN1, CNRIP1, INA, SPG20 y SNCA tal como se ha descrito en los ejemplos anteriores.

5 Se analizaron los seis genes en el ADN purificado procedente de 10 muestras de excrementos. Se averiguó el estado de metilación del tumor primario correspondiente a partir de 4-5 de los tumores correspondientes.

Además, en nueve de los 10 pacientes a partir de los cuales se obtuvieron las muestras de excrementos, se extrajeron también muestras sanguíneas y se comparan los resultados en la tabla 5. Se aisló el ADN a partir de 250 mg de heces utilizando el QIAamp DNA stool kit (QIAGEN).

10 Los resultados del estado de metilación en genes procedentes de diferentes muestras se presentan a continuación en la tabla 4:

Tejido/gen	MAL	CNRIP1	INA	FBN1	SPG20	SNCA
Excrementos	0/3	1/4	2/6	0/4	2/5	3/8
Sangre	3/9	8/9	3/9	0/9	6/9	9/9

15 Con la excepción de MAL y FBN1 se detectó metilación en todos los marcadores en las muestras procedentes de excrementos. En las muestras sanguíneas se detectó metilación en todos los genes excepto en FBN1. En general, la frecuencia de metilación de las muestras fue elevada para todos los genes procedentes de muestras sanguíneas. Especialmente, la frecuencia de metilación de las muestras en las muestras sanguíneas que comprendían SNCA y CNRIP1 fue muy elevada. Estos marcadores parecen especialmente adecuados como marcadores del cáncer ya que pueden evaluarse con un procedimiento no invasivo, tal como en muestras sanguíneas.

20 Se purificó el ADN a partir de sangre (utilizando un método fenol/cloroformo estándar) y se llevó a cabo MSP para los siguientes genes: MAL, CNRIP1, INA, FBN1, SPG20 y SNCA tal como se ha descrito en los ejemplos anteriores.

Se purificó ADN a partir de 14 muestras sanguíneas procedentes de pacientes con un tumor primario correspondiente el cual estuvo metilado en los seis genes.

Tabla 5 conjunto de muestra expandido

Tejido/gen	MAL	CNRIP1	INA	FBN1	SPG20	SNCA
Sangre	4/13	11/13	6/13	3/13	12/13	12/13

25 Este ejemplo confirma la elevada frecuencia de metilación de las muestras de los genes en las muestras sanguíneas y especialmente CNRIP1 y SNCA son marcadores sumamente adecuados para el diagnóstico y/o cribado del cáncer o desarrollo del cáncer. Igualmente, SPG20 parece un marcador muy prometedor para el cribado de muestras sanguíneas.

Ejemplo 4

30 *Frecuencias de metilación histoespecíficas en muestras de INA, SNCA, CNRIP1, SPG20 o FBN1 en diferentes líneas celulares*

Para cada muestra (líneas celulares procedentes de mama, riñón, ovario, páncreas, próstata, útero y gástricas) se trató el ADN con bisulfito antes de la MSP tal como se ha descrito en el ejemplo 2.

35 Los resultados de la frecuencia de metilación de las muestras de genes en diferentes muestras se presentan a continuación en la tabla 6:

	INA	SNCA	CNRIP1	SPG20	FBN1
Mama	4/6	6/7	2/6	2/6	4/6
Riñón	2/4	1/4	0/4	0/4	0/4
Ovario	2/4	0/4	0/4	1/4	1/4



	INA	SNCA	CNRIP1	SPG20	FBN1
Páncreas	3/6	4/6	5/6	4/6	2/6
Próstata	1/1	1/1	0/1	0/1	0/1
Útero	2/4	2/4	1/3	2/4	2/4
Gástricas	3/3	3/3	3/3	3/3	3/3

Se analizó el estado de metilación del promotor de INA, SNCA, CNRIP1, SPG20 y FBN1 con MSP. En todas las muestras procedentes de líneas celulares gástricas los genes ensayados estuvieron metilados y, por tanto, la frecuencia de metilación fue (de un 100%) para todos los genes ensayados. En general, se metiló INA en al menos una muestra de todos los tejidos ensayados. Se observó la mayor frecuencia de metilación de las muestras para este gen en líneas celulares procedentes de la próstata 1/1 (100%), mama 4/6 (66%) y gástricas 3/3 (100%). En el caso las líneas celulares procedentes de todos los demás tejidos la frecuencia de metilación de las muestras fue de un 50%. Se metiló SNCA en las líneas celulares procedentes de todos los tejidos ensayados, excepto en el caso del ovario. La mayor frecuencia de metilación de las muestras tuvo lugar en las líneas celulares procedentes de la próstata 1/1 (100%), páncreas 4/6 (66%), mama 6/7 (85%) y gástricas 3/3 (100%). En el caso del útero, la frecuencia de metilación de las muestras fue de un 50%. La frecuencia de metilación de CNRIP1 fue elevada en muestras gástricas 3/3 (100%) y de páncreas (83%) donde 5 de las 6 muestras estuvieron metiladas. Se metiló SPG20 en 4 de las 6 (66%) líneas celulares pancreáticas y en 2 de las 4 líneas celulares uterinas (50%). Se metiló FBN1 en 4 de las 6 líneas celulares de cáncer de mama y se metilaron 2 de las 4 (50%) muestras en líneas celulares procedentes del útero.

En general estuvieron metilados todos los genes en muestras procedentes de líneas celulares gástricas, por lo tanto, la frecuencia de metilación de las muestras fue elevada para todos los genes. Además, estuvieron metilados todos los genes en líneas celulares de mama aunque las frecuencias variaron entre los genes. Igualmente, estuvieron metilados todos los genes en las muestras procedentes del páncreas y para todos los genes, excepto para FBN1 (33%), la frecuencia fue un 50% o superior.

Este experimento indica claramente que los genes de acuerdo con la descripción están metilados en líneas celulares procedentes de diferentes tejidos cancerosos y, por lo tanto, podrían utilizarse como marcadores del cáncer para varios tipos de cáncer. Resultará obvio para el experto en la técnica que cada uno de los genes de acuerdo con la descripción puede combinarse de manera diferente dependiendo del tipo de cáncer que se quiere detectar. Por lo tanto, los genes que muestran mejores resultados en líneas celulares de mama se seleccionarían como marcadores para detectar cáncer de mama.

El resultado de la frecuencia de metilación histo-específica de las muestras de MAL se enumera en la tabla 8.

#### Ejemplo 5

*Se llevaron a cabo análisis cuantitativos de la expresión génica para los siguientes genes: SPG20, INA y CNRIP1.*

Se cuantificó la expresión génica en 6 líneas celulares de cáncer de colon antes y después del tratamiento con fármacos epigenéticos. Se cuantificaron los niveles de expresión relativa de SPG20, INA y CNRIP1 en las líneas celulares de cáncer de colon (n = 6). Se presentan los niveles de expresión como proporción de aumento calculada mediante el método  $\Delta\Delta\text{CT}$  utilizando la muestra no tratada como un calibrador. Se utilizó la expresión media de ACTB y GUSB como control endógeno.

Se utilizó la detección fluorescente en tiempo real TaqMan (Applied Biosystems, Foster city, CA) para cuantificar los niveles de ARNm en las líneas celulares de cáncer de colon, como se ha descrito previamente [Gibson *et al.* y Heid *et al.*]. Se generó ADNc a partir de cinco  $\mu\text{g}$  de ARN total utilizando el High-Capacity cDNA Archive kit (Applied Biosystems), que incluye cebadores aleatorios de acuerdo con el protocolo del fabricante. Se amplificaron de manera separada el ADNc procedente de los genes de interés (SPG20, INA y CNRIP1) y los controles endógenos (ACTB y GUSB) mediante el Sistema de Detección de Secuencias 7900HT (Applied Biosystems) siguiendo el protocolo recomendado por el proveedor. Se analizaron todas las muestras en triplicados. Se calcularon los niveles de expresión como proporción de aumento utilizando el método  $\Delta\Delta\text{CT}$  y la muestra no tratada como un calibrador. A fin de ajustarse a las posibles variaciones en la cantidad de ADNc que se aporta a cada PCR, se normalizó la cantidad de expresión de los genes diana con los genes constitutivos ACTB y GUSB.

En el caso de SPG20, INA y CNRIP1, la expresión génica aumentó significativamente en la mayoría de las líneas celulares de cáncer de colon metiladas inicialmente tras la desmetilación del promotor inducida por el tratamiento combinado de 5-aza-2'-desoxicitidina y tricostatina A (Figuras 2, 3 y 4). En el caso de INA y CNRIP1, el tratamiento combinado resultó ser más efectivo que el tratamiento individual con 5-aza-2'-desoxicitidina sola y tricostatina A sola. El tratamiento combinado también incrementó la expresión de SPG20, sin embargo, se puede conseguir una

reactivación similar o mayor mediante el tratamiento con 5-aza-2'-desoxicitidina sola. El tratamiento con el inhibidor de desacetilasas tricostatina A solo no incrementó la expresión génica ni de SPG20 ni de INA ni de CNRIP1. Las dos dosis de 5-aza-2'-desoxicitidina ensayadas aquí (1 µM y 10 µM) proporcionaron efectos comparables. Esto significa que la desmetilación de las líneas celulares puede conseguirse cultivándolas en presencia de dosis bajas de 5-aza-2'-desoxicitidina, lo cual es una ventaja considerando la citotoxicidad de este fármaco.

Existe una clara relación entre el estado de metilación y la expresión de SPG20, INA y CNRIP1. Por lo tanto cuantificar el nivel o estado de metilación mediante, p. ej., MSP y combinar el resultado con el nivel de expresión del gen correspondiente podría incrementar la sensibilidad y especificidad de un método de la descripción.

Ejemplo 6

*Hipermetilación de MAL*

Se sometió ADN de líneas celulares y pacientes procedente de 218 muestras congeladas en fresco a análisis de metilación, incluidos 65 carcinomas colorrectales (36 con estabilidad de microsatélites; MSS, por sus siglas en inglés, y 29 con inestabilidad de microsatélites; MSI, por sus siglas en inglés) procedentes de 64 pacientes, 63 adenomas, tamaño de la mediana 8 mm, intervalo 5-50 mm (61 MSS y 2 MSI) procedentes de 52 pacientes, 21 muestras de mucosa normal procedentes de 21 pacientes con cáncer colorrectal (tomadas en sitios lejanos respecto al carcinoma primario) y otras 23 muestras de mucosa colorrectal normal procedentes de 22 individuos sin cáncer, junto con 20 líneas celulares de cáncer de colon (11 MSS y 9 MSI), y 29 líneas celulares de cáncer procedentes de varios tejidos (mama, riñón, ovario, páncreas, próstata, útero y gástricas; Tabla 9). La edad media en el diagnóstico fue de 70 años (intervalo de 33 a 92) en el caso de pacientes con carcinoma, 67 años (intervalo de 62 a 72) en el caso de personas con adenomas, 64 años (varía entre 24 y 89) en el caso del primer grupo de donantes de mucosa normal, y 54 años (varía entre 33 y 86) en el caso del segundo grupo de donantes de mucosa normal. Los carcinomas colorrectales y las muestras normales procedentes de pacientes oncológicos se obtuvieron a partir de una serie prospectiva no seleccionada recogida a partir de siete hospitales ubicados en la región suroeste de Noruega. Se obtuvieron los adenomas a partir de individuos según un programa de cribado sigmoidoscópico del cáncer colorrectal en función de la población. Las muestras de mucosa normal procedentes de individuos sin cáncer se obtuvieron a partir de personas fallecidas, y la mayoría del conjunto total de muestras normales (27/44) consistió únicamente de mucosa, mientras que las muestras restantes se tomaron a partir de la pared intestinal. Los datos clínico-patológicos adicionales de la serie tumoral actual incluyen el género y la localización tumoral, así como el tamaño del pólipo y el número total de pólipos por individuo para la serie de adenomas.

Todas las muestras pertenecen a biobancos de investigación aprobados y son parte de proyectos de investigación aprobados de acuerdo con las directrices nacionales (Biobanco; registrado en el Instituto Noruego de Salud Pública; Proyectos: Comité de Ética Regional y Cuerpo de Inspectores de Datos Nacional).

Se sometieron seis líneas celulares de cáncer de colon, HCT15, HT29, SW48, SW480, RKO y LS1034, a tratamiento con el fármaco desmetilante 5-aza-2'-desoxicitidina (1 µM durante 72 h y 10 µM durante 72 h), el inhibidor de desacetilasas de histonas tricostatina A (0,5 µM durante 12 h) y una combinación de ambos (5-aza-2'-desoxicitidina 1 µM durante 72 h, tricostatina A 0,5 µM añadida las últimas 12 h).

*Tratamiento con bisulfito y reacción en cadena de la polimerasa específica de metilación (MSP)*

Se trató el ADN procedente de muestras de tumores primarios y de mucosa normal con bisulfito tal como se ha descrito previamente. El ADN procedente líneas celulares de cáncer colorrectal se trató con bisulfito utilizando el EpiTect bisulphite kit (Qiagen Inc., Valencia, CA, EE. UU.). Se analizó el estado de metilación del promotor de todos los genes mediante la reacción en cadena de la polimerasa específica de metilación (MSP) utilizando la ADN-polimerasa HotStarTaq (Qiagen). Se confirmaron todos los resultados con una segunda serie independiente de MSP. Se utilizó ADN de placenta humana (Sigma Chemical Co., San Luis, MO, EE. UU.) tratado *in vitro* con metiltransferasa Sss1 (New England Biolabs Inc., Beverly, MA, EE. UU.) como control positivo para la reacción de MSP metilada, mientras que se utilizó ADN procedente de linfocitos normales como control positivo para los alelos no metilados. Se utilizó agua como control negativo en ambas reacciones. Se diseñaron los cebadores con MethPrimer y Methyl Primer Express y se enumeran sus secuencias en la Tabla 7.

Conjunto de cebadores	Cebador sentido/ antisentido	Cebador	Tamaño frag. pb	Temp. hibridación	Ubicación del fragmento	SEQ ID NO.
MAL MSP-M	<b>TTCGGGTTTTTTGTTTTAATTC/ GAAAACCATAACGACGTAACGT</b>		139	56	-71 to 68	37/38
MAL MSP-U	<b>TTTTGGGTTTTTTGTTTTAATT/ ACAAAACCATAACAACATACTAACATC</b>		142	56	-72 to 70	39/40

Conjunto de cebadores	Cebador sentido/ antisentido	Cebador	Tamaño frag. pb	Temp. hibridación	Ubicación del fragmento	SEQ ID NO.
MAL BS_A	<b>GGGTTTTTTTGTTTTAATT/ ACCAAAAACCACTCACAAACTC</b>		236	53	-68 to 168	97/98
MAL BS_B	<b>GGAAAAATGAAGGAGATTAAATTT/ AATAACCTAAACRCCCCC</b>		404	50	-427 to -23	99/100

Abreviaturas: MSP, reacción en cadena de la polimerasa específica de metilación; BS, secuenciación con bisulfito; M, cebadores específicos para metilados; U, cebadores específicos para no metilados; Tamaño frag., tamaño del fragmento; Temp. hib., temperatura de hibridación (en grados celsius). La ubicación del fragmento señala el punto de inicio y el punto de terminación (en pares de bases) de cada fragmento respecto al punto de inicio de la transcripción proporcionado por NCBI (ID de RefSeq NM\_002371).

#### Secuenciación con bisulfito

Se sometieron todas las líneas celulares de cáncer de colon (n = 20) a secuenciación directa con bisulfito del promotor de MAL. Se amplificaron dos fragmentos: el fragmento A, que abarcaba desde la base -68 hasta la 168 respecto al punto de inicio transcripcional (que se solapa con el producto de la MSP de los inventores) y el fragmento B que abarca desde la base -427 hasta la -23. El fragmento A también abarcó 24 sitios de CpG y se amplificó utilizando la ADN-polimerasa HotStarTaq y 35 ciclos de PCR. El fragmento B abarcó en total 32 sitios de CpG y se amplificó utilizando la misma polimerasa y 36 ciclos de PCR. Las secuencias de los cebadores se enumeran en la Tabla 8. El cebador y los nucleótidos en exceso se eliminaron mediante el tratamiento ExoSAP-IT siguiendo el protocolo del fabricante (GE Healthcare, USB Corporation, Ohio, EE. UU.). Los productos purificados se secuenciaron posteriormente utilizando el dGTP BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA, EE. UU.) en un secuenciador 3730 AB Prism (Applied Biosystems). Se calculó la cantidad aproximada de metilcitosina de cada sitio CpG comparando la altura del pico de la señal de citosina con la suma de la altura de los picos de las señales de citosina y timina, tal como se ha descrito previamente. Los sitios de CpG con ratios comprendidos en el intervalo 0 – 0,20 se clasificaron como no metilados, los sitios de CpG comprendidos en el intervalo 0,21 – 0,80 se clasificaron como parcialmente metilados y los sitios de CpG comprendidos en el intervalo 0,81 – 1,0 se clasificaron como hipermetilados.

#### Preparación de ADNc y expresión génica cuantitativa en tiempo real

Se extrajo el ARN total a partir de líneas celulares (n = 46), tumores (n = 16) y tejido normal (n = 3) utilizando Trizol (Invitrogen, Carlsbad, CA, EE. UU.) y se determinó la concentración de ARN utilizando un Nanodrop ND-1000 (NanoDrop Technologies, Wilmington, DE, EE. UU.). Para cada muestra, se convirtió el ARN total en ADNc utilizando un High-Capacity cDNA Archive kit (Applied Biosystems), que incluía cebadores aleatorios. Se amplificaron de manera separada MAL (Hs00242749\_m1 y Hs00360838\_m1) y los controles endógenos ACTB (Hs99999903\_m1) y GUSB (Hs99999908\_m1) en placas de 96 pocillos Fast siguiendo el protocolo recomendado (Applied Biosystems) y se cuantificó la expresión génica cuantitativa en tiempo real mediante el Sistema de Detección de Secuencias 7900HT (Applied Biosystems). Se analizaron todas las muestras por triplicado y se utilizó el valor de la mediana para el análisis de los datos. Se utilizó el ARN de referencia universal humano (que contenía una mezcla de ARN procedente de diez líneas celulares diferentes; Stratagene) para generar una curva estándar y se normalizaron los niveles de expresión cuantitativa de MAL resultantes respecto al valor medio de los dos controles endógenos.

#### Micromatriz tisular

Se construyó una micromatriz tisular (TMA, por sus siglas en inglés), basada en la tecnología descrita previamente para la detección *in situ* de la expresión proteica en el cáncer colorrectal. En la TMA se incluyeron 292 núcleos tisulares cilíndricos (diámetro de 0,6 mm) procedentes de muestras tumorales fijadas con etanol e incluidas en parafina derivadas de 281 individuos. Se examinaron las muestras procedentes de la misma serie de pacientes para diferentes variables biológicas y criterios clínicos de valoración. Además, la matriz contuvo tejidos normales procedentes de riñón, hígado, bazo y corazón como controles. Los tejidos de colon normales fijados con etanol procedentes de cuatro personas sin antecedentes conocidos de cáncer colorrectal se obtuvieron de manera separada.

#### Análisis de la expresión proteica inmunohistoquímica *in situ*

Se transfirieron cortes de cinco µm de grosor de los bloques de la TMA a portaobjetos de vidrio para los análisis inmunohistoquímicos. Se eliminó la parafina de los cortes en un baño de xileno durante 10 minutos y se rehidrataron

5 mediante una serie de baños de etanol graduado. Se llevó a cabo la recuperación del epítipo inducida por el calor calentando en un horno microondas a máxima potencia (850 W) durante 5 minutos y a continuación se sumergió 15 minutos a 100 W en tampón de citrato 10 mM a pH 6,0 que contenía un 0,05% de Tween-20. Después de enfriar hasta temperatura ambiente, se llevó a cabo la tinción inmunohistoquímica de acuerdo con el protocolo del kit Envision+™ K5007 de DAKO (Dako, Glostrup, Dinamarca). El anticuerpo primario, clon 6D9 anti-MAL de ratón, se utilizó con una dilución de 1:5000, lo que permitió la tinción de los túbulos renales como control positivo, mientras que el tejido muscular cardíaco permaneció sin teñirse como control negativo. Los portaobjetos se sometieron a una contratinción con hematoxilina durante 2 minutos y a continuación se deshidrataron en etanol con una graduación creciente y finalmente en xileno. Se obtuvieron los resultados de la inmunohistoquímica mediante una puntuación independiente por parte de uno de los autores y un patólogo de referencia.

*Datos estadísticos*

15 Se obtuvieron todos los valores P a partir de pruebas estadísticas bilaterales utilizando el software SPSS 13.0 (SPSS, Chicago, IL, EE. UU.). Se utilizó la prueba exacta de Fisher para analizar tablas de contingencia 2 x 2. Se utilizaron una prueba de Chi cuadrado y una tabla 2 x 3 para analizar la asociación potencial entre la expresión génica cuantitativa de MAL y el estado de metilación del promotor. Se dividieron las muestras en dos categorías de acuerdo con sus niveles de expresión génica: la expresión baja incluyó muestras con una expresión génica igual o inferior al valor de la mediana en todas las células, líneas o todos los tumores, la expresión elevada incluyó muestras con una expresión génica superior a la mediana. El estado de metilación se dividió en tres categorías: no metilado, metilación parcial e hipermetilado.

20 *Estado de metilación del promotor de MAL en tejidos y líneas celulares*

25 Se analizó el estado de metilación del promotor de MAL con MSP (Figura 5). Una de las 23 (4%) muestras de mucosa normal procedentes de pacientes no cancerosos y dos de las 21 (10%) muestras de mucosa normal tomadas lejos del tumor primario estuvieron metiladas pero presentaron únicamente una banda de intensidad baja en comparación con el control positivo tras la electroforesis en gel. Cuarenta y cinco de los 63 (71%) adenomas y 49/61 (80%) carcinomas mostraron hipermetilación del promotor. Diecinueve de las veinte líneas celulares de cáncer de colon (95%) y 15/26 (58%) de las líneas celulares cancerosas procedentes de varios tejidos (mama, riñón, ovario, páncreas, próstata y útero) estuvieron hipermetiladas (la Tabla 9 enumera las frecuencias histoespecíficas).

Tabla 8: Estado de metilación del promotor de MAL en líneas celulares de varios tejidos

Línea celular	Tejido	Estado de metilación del promotor	Frecuencia de metilación
BT-20	Mama	M	
BT-474	Mama	U/M	
Hs 578T	Mama	U	
SK-BR-3	Mama	U	57%
T-47D	Mama	U/M	
ZR-75-1	Mama	U	
ZR-75-38	Mama	M	
Co115	Colon	M	
HCT15	Colon	M	
HCT116	Colon	M	
LoVo	Colon	M	
LS174T	Colon	M	
RKO	Colon	M	
SW48	Colon	M	
TC7	Colon	M	
TC71	Colon	M	
ALA	Colon	M	95%
Colo320	Colon	M	
EB	Colon	M	

Línea celular	Tejido	Estado de metilación del promotor	Frecuencia de metilación
FRI	Colon	U/M	
HT29	Colon	M	
IS1	Colon	M	
IS2	Colon	M	
IS3	Colon	M	
LS1034	Colon	M	
SW480	Colon	M	
V9P	Colon	U	
ACHN	Riñón	U	
Caki-1	Riñón	U	50%
Caki-2	Riñón	M	
786-O	Riñón	U/M	
ES-2	Ovario	U/M	
OV-90	Ovario	U/M	50%
Ovcar-3	Ovario	U	
SK-OV-3	Ovario	U	
AsPC-1	Páncreas	M	
BxPC-3	Páncreas	U	
CFPAC-1	Páncreas	U	67%
HPAF-II	Páncreas	M	
PaCa-2	Páncreas	M	
Panc-1	Páncreas	U/M	
LNCaP	Próstata	U	0%
AN3 CA	Útero	U/M	
HEC-1-A	Útero	M	75%
KLE	Útero	U	
RL95-2	Útero	M	

Se evaluó el estado de metilación del promotor de las líneas celulares individuales mediante la reacción en cadena de la polimerasa específica de metilación (MSP). La frecuencia de metilación refleja el número de muestras metiladas (M y U/M) procedentes de cada tejido. Abreviaturas: U, no metilado; M, metilado.

- 5 La frecuencia de hipermetilación observada en muestras normales fue significativamente menor que en los adenomas ( $P < 0,0001$ ) y carcinomas ( $P < 0,0001$ ). La hipermetilación del promotor de *MAL* no se asocia con el estado MSI, género ni edad, ni en tumores benignos ni en los malignos. Entre los carcinomas, los tumores con una ubicación distal en el intestino (lateral izquierdo y recto) estuvieron hipermetilados con una frecuencia mayor que la de los tumores con una ubicación proximal, aunque no fue estadísticamente significativa ( $P = 0,088$ ). Entre los adenomas, no se observó una asociación significativa entre el estado de metilación del promotor de *MAL* y el número o tamaño de los pólipos.

*Verificación mediante la secuenciación con bisulfito del estado de metilación del promotor de MAL*

- 15 Se secuenciaron con bisulfito dos fragmentos solapantes del promotor de *MAL* en 20 líneas celulares de cáncer de colon. Los resultados se resumen en la Figura 6 y los datos no tratados representativos se pueden observar en la Figura 7. Se apreció una buena asociación entre el estado de metilación, tal como se evaluó mediante MSP, y las secuencias con bisulfito del fragmento solapante A. Sin embargo, en el fragmento B hubo una asociación poco

satisfactoria con los datos de MSP. En el caso de este fragmento, el cual está ubicado más allá del sitio de inicio transcripcional en dirección 5', varios sitios CpG consecutivos estuvieron frecuentemente no metilados y/o parcialmente metilados. Este fue también el caso en líneas celulares en las que se había observado una metilación muy elevada alrededor del sitio de inicio transcripcional (fragmento A; Figura 6).

#### 5 *Expresión génica cuantitativa en tiempo real*

Se evaluó el nivel de la expresión de ARNm de MAL en líneas celulares (n = 46), carcinomas colorrectales primarios (n = 16) y mucosa normal (n = 3) mediante PCR cuantitativa en tiempo real. Hubo una estrecha correlación entre la hipermetilación del promotor de MAL y una expresión génica reducida o ausente entre las líneas celulares (P = 0,041; Figura 8). Además, hubo un incremento en la expresión génica de MAL en las líneas celulares de cáncer de colon tras la desmetilación del promotor inducida por el tratamiento combinado con 5-aza-2'-desoxicitidina y tricostatina A (Figura 9). El tratamiento con el inhibidor de desacetilasas tricostatina A solo no aumentó la expresión de MAL, mientras que el tratamiento con el compuesto desmetilante del ADN 5-aza-2'-desoxicitidina conllevó una expresión elevada en células HT29, pero niveles más moderados en células HCT15 (Figura 9). Entre los carcinomas colorrectales primarios, aquellos que presentaban hipermetilación del promotor de MAL (n = 13) expresaron niveles de ARNm de MAL un poco menores en comparación con los tumores no metilados (n = 3), aunque no de manera estadísticamente significativa (Figura 8).

#### *La expresión de la proteína MAL está ausente en carcinomas colorrectales*

Para evaluar los análisis inmunohistoquímicos de MAL, se incluyeron tejidos musculares cardíacos y renales como controles negativos y positivos respectivamente (Figura 10 A-B). A partir de los 231 núcleos de tejido colorrectal puntuables, es decir, aquellos que contenían tejido epitelial colorrectal maligno, 198 presentaron una tinción negativa para MAL (Figura 10 C-D). Veintinueve de estos presentaron una tinción positiva en componentes tisulares no epiteliales comprendidos en los mismos núcleos tisulares, principalmente en neuronas y vasos sanguíneos (no se muestran). En comparación, todos los cortes de tejido de colon normal presentaron una tinción positiva para MAL en las células epiteliales (Figura 10 E-F).

Estos experimentos concluyen que el promotor de MAL cercano al inicio transcripcional está hipermetilado en la gran mayoría de los tumores colorrectales malignos, así como en los benignos, a diferencia de las muestras de mucosa de colon normal las cuales no están metiladas, y los autores sostienen que MAL sigue siendo un biomarcador diagnóstico prometedor para la tumorigénesis colorrectal incipiente. Además, no se observó MAL hipermetilado en líneas celulares cancerosas procedentes de mama, riñón, ovario y útero.

Otros han demostrado previamente, mediante la reacción en cadena de la polimerasa específica de metilación (MSP), que la hipermetilación de MAL está presente únicamente en una pequeña fracción (6%, 2/34) de los carcinomas de colon (Mori *et al.*). En cambio, los solicitantes demuestran aquí una frecuencia de metilación de MAL significativamente mayor tanto en los tumores colorrectales malignos como en los benignos (71% en los adenomas y 80% en los carcinomas). La discrepancia en las frecuencias de metilación entre el presente informe y el estudio previo de Mori *et al.* es probablemente consecuencia del diseño del estudio. A partir de la secuenciación directa con bisulfito de las líneas celulares de cáncer de colon, los autores han demostrado que la metilación del ADN de MAL está distribuida de manera desigual en las islas de CpG de su promotor (Figura 6). Las islas de CpG a menudo abarcan más de una kilobase del promotor génico, y a veces se asume erróneamente que el estado de metilación en esta región está distribuido por igual. Debido a que los resultados del análisis de MSP dependen del emparejamiento o emparejamiento incorrecto de las secuencias de los cebadores metilados y no metilados con el ADN tratado con bisulfito, se debe garantizar que los cebadores se hibridan con sitios CpG relevantes en el promotor génico. En el presente estudio, los solicitantes diseñaron cebadores de MSP cercanos al punto de inicio transcripcional del gen (de -72 a +70) y descubrieron, mediante la secuenciación con bisulfito, concordancia entre el estado de metilación global de MAL tal como se evaluó mediante MSP y el estado de metilación de los sitios CpG individuales cubiertos por el primer conjunto de cebadores de MSP de los autores (Figura 6). Esta parte de la isla de CpG estuvo hipermetilada en la mayoría de las líneas celulares de cáncer de colon (95%). Los autores también observaron a partir de los análisis cuantitativos en tiempo real que estas líneas celulares, así como las de otros tejidos, presentaban pérdida de la expresión del ARN de MAL, y que la eliminación de la hipermetilación de ADN mediante el tratamiento combinado con 5-aza-2'-desoxicitidina y tricostatina A reinducía la expresión de MAL en las líneas celulares de cáncer de colon (Figura 9). Además, analizando una larga serie de muestras clínicamente representativas mediante inmunohistoquímica proteica confirmaron que se perdía la expresión de MAL en células epiteliales colorrectales malignas en comparación con la mucosa normal.

Los inventores han analizado además la misma región del promotor de MAL al igual que Mori *et al.*, la cual está localizada de -206 a -126 pares de bases en dirección 5' respecto al punto de inicio transcripcional. Han demostrado, mediante secuenciación con bisulfito, que únicamente una minoría de los sitios CpG cubiertos por el cebador antisentido de Mori estaban metilados en las 19 líneas celulares de cáncer de colon con una metilación muy elevada alrededor del punto de inicio transcripcional (Figura 6). Por lo tanto, los autores concluyen que la frecuencia de metilación extremadamente baja (seis por ciento) descrita inicialmente para MAL en los carcinomas de colon (Mori *et*

al.) es muy probablemente consecuencia del diseño del cebador y elección de los sitios CpG que debían examinarse.

5 La inactivación de la hipermetilación del promotor de MAL también podría ser prevalente en otros tipos de cáncer. En el presente estudio, se detectó MAL hipermetilado en líneas celulares cancerosas procedentes de mama, riñón, ovario y útero.

Los presente análisis de líneas celulares cancerosas procedentes de varios tejidos indican que la hipermetilación de un área limitada en las proximidades del punto de inicio transcripcional de MAL está asociada con una expresión génica reducida o ausente.

10 Una estrategia de cribado no invasiva y sensible para el cáncer colorrectal podría mejorar notablemente el desenlace clínico del paciente. Un ensayo diagnóstico de este tipo podría en principio cuantificar el estado de un único biomarcador.

15 La hipermetilación del promotor de MAL representa un gen hipermetilado frecuentemente entre las lesiones colorrectales premalignas, y viene acompañado por frecuencias de metilación bajas en la mucosa de colon normal. La presencia de este tipo de cambios epigenéticos en tejidos premalignos también podría tener implicaciones en la quimioprevención del cáncer. Inhibiendo o invirtiendo estas alteraciones epigenéticas, podría evitarse la evolución a un fenotipo maligno (Kopelovich *et al.*). La hipermetilación del promotor de MAL sigue siendo uno de los biomarcadores diagnósticos más prometedores para la detección precoz de los tumores colorrectales.

### Referencias

- 20 1. Mori Y, Cai K, Cheng Y, Wang S, Paun B, Hamilton JP, Jin Z, Sato F, Berki AT, Kan T, Ito T, Mantzur C, Abraham JM, Meltzer SJ. A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer. *Gastroenterology* 2006; 131:797-808.
2. Lind GE, Kleivi K, Meling GI, Teixeira MR, Thiis-Evensen E, Rognum TO, Lothe RA. ADAMTS1, CRABP1, and NR3C1 Identified as Epigenetically Deregulated genes in Colorectal Tumourigenesis. *Cell Oncol* 2006; 28:259-272.
- 25 3. Laird PW. The power and the promise of DNA methylation markers. *Nat Rev Cancer* 2003; 3:253-266.
4. Muller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, Muhlthaler M, Ofner D, Margreiter R, Widschwendter M. Methylation changes in faecal DNA: a marker for colorectal cancer screening? *Lancet* 2004; 363:1283-1285.
- 30 5. Chen WD, Han ZJ, Skoletsy J, Olson J, Sah J, Myeroff L, Platzer P, Lu S, Dawson D, Willis J, Pretlow TP, Lutterbaugh J, Kasturi L, Willson JK, Rao JS, Shuber A, Markowitz SD. Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 2005; 97:1124-1132.
6. C.Grunau, S.J.Clark y A.Rosenthal, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters, *Nucleic Acids Res.* 29 (2001), E65.
- 35 7. M.F.Fraga y M.Esteller, DNA methylation: a profile of methods and applications, *Biotechniques* 33 (2002), 632-649.
8. B.Smith-Sørensen, G.E.Lind, R.I.Skotheim, S.D.Fosså, Ø.Fodstad, A.E.Stenwig, K.S.Jakobsen y R.A.Lothe, Frequent promoter hypermethylation of the 06-Methylguanine-DNA Methyltransferase (MGMT) gene in testicular cancer, *Oncogene* 21 (2002), 8878-8884.
- 40 9. J.G.Herman, J.R.Graff, S.Myöhänen, B.D.Nelkin y S.B.Baylin, Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands, *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996), 9821-9826.
10. S.Derks, M.H.Lentjes, D.M.Hellebrekers, A.P.de Bruine, J.G.Herman y E.M.van, Methylation-specific PCR unraveled, *Cell Oncol.* 26 (2004), 291-299.
11. L.C.Li y R.Dahiya, MethPrimer: designing primers for methylation PCRs, *Bioinformatics.* 18 (2002), 1427-1431.

12. J.R.Melki, P.C.Vincent y S.J.Clark, Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia, *Cancer Res.* 59 (1999), 3730-3740.

13. Zweig, M. H., y Campbell, G., *Clin. Chem.* 39 (1993) 561-577.

LISTADO DE SECUENCIAS

5 <110> Rikshospitalet-Radiumhospitalet HF  
Lothe, Ragnhild A.  
Lind, Guro E.  
Ahlquist, Terje  
Skotheim, Rolf I

10 <120> Nuevos marcadores para el cáncer

<130> 42139pc01

<150> PA 2007 00273

<151> 27-02-2007

15 <150> PA 2007 00601

<151> 24-04-2007

<160> 100

<170> FastSEQ para Windows Versión 4.0

<210> 1

20 <211> 1300

<212> ADN

<213> *Homo Sapiens*

<400> 1

```

gactcgggcg gatttcaggc ttcagtgttt gtaggaggaa acacagcaat cacactatta 60
atagttaaatt aaaataaatg ggcaactgct gcatggtaat actttttttt ttaaggcaaa 120
aaataaaaaa tagtgaaaca gagaacaaa acatgaaaca ccggcagtcac acaggcaggc 180
aaagaacctg ggggtggggg tagcagcggc cccaccctca aaaggcccgg gctgcccaga 240
ccaagagaaa gcgatgaatc tcttctggta acgtcccttc ctgtcgcatt gattcaagc 300
cgacctgcc cagcaccacc accagcagcc ttctgctggg gccggcacag ctgggagcaa 360
cctctactc tcaggcagac gcgcagcacc aagcagagag gcccggtgca ggatcccagc 420
gccgaaccag cgccggctca gtggacggc aaggggcccgg cggccgcccgc cgggtcccac 480
ccccactgca gacccccagc ctgtggcggc ggtccagttc cggcaggaaa ccgcccgcctg 540
gagctgtggg tcgcgcacat taacgcatcc agcggaaaaa tgaaggagac ccaaattcaa 600
agttaaagta atggtgaccc gagaggtgcc ttgatgagaa ggtttggggc cccggttact 660
gatggttacc attcttacga gatgctggc acctacgaag ggagaaaggc acgaggagcg 720
cctgacaaa gtggttttgc cctgcttccc gcaagaggtg gcacccacgg ctggaacgca 780
ggagtcagac ccacagtcac cagctctgga cgcccgcagc ggggcctcga agaggttcag 840
ggcggtgccc gcggcgctcg ggcccgggtc cccggggcgt ggggcggggg gcggggttcg 900
gcggcggcgg gggctcctcc ctcttctgcc ccgggtccc ctgctcttaa cccgcgcgcg 960

```

25



ES 2 443 230 T3

```

ggggcgccca ggccactggg ctccgcggag ccagcgagag gtctgcgcgg agtctgagcg 1020
gcgctcgtec cgteccaagg ccgacgccag cacgccgta tggccccgc agcggcgacg 1080
ggggcgagca ccctgcccag tggttctcgt gtcttcacca ccttgcccga cttgctcttc 1140
atctttgagt ttgtgagtgg ctctggccg gggaaaggac ggggtgggct gagccgtgcg 1200
ctctctcggg cgcccagcac agctgtcgga cgggatccgc tagctgcgca ggttctggga 1260
gcatcggggc agcagggcga gggcggggac taagccaggg 1300

```

```

<210> 2
<211> 1300
<212> ADN
<213> Homo Sapiens

```

5

<400> 2

```

gactcggggc gatttcaggc ttcagtgttt gtaggaggaa acacagcaat cacactatta 60
atagtaaatt aaaataaatg ggcaactgct gcatggtaat actttttttt ttaaggcaaa 120
aaataaaaaa tagtgaaaca gagaaacaaa acatgaaaca ccggcagtcg acaggcaggc 180
aaagaacctg ggggtggggg tagcagcggg cccaccctca aaaggcccgg gctgcccaga 240
ccaagagaaa gcgatgaatc tcttctggta acgtcccttc ctgtcgcgatg gattcaaggc 300
cgacctgccc cagcaccacc accagcagcc ttctgctggg gcccgccacag ctgggagcaa 360
cctcctactc tcaggcagac gcgcagcacc aagcagagag gcccggtgca ggatcccagc 420
gccgaaccag cgccggctca gtggacgcgg aaggggccgg cggccgcggc cggtcccatac 480
ccccactgca gacccccagc ctgtggcggg ggtccagttc cggcaggaaa ccgcccctg 540
gagctgtggg tcgcgcacat taacgcaccc agcggaaaaa tgaaggagac ccaaattcaa 600
agttaaagta atggtgacct gagagggtcc ttgatgagaa ggtttggggg cccggttact 660
gatggttatc attcttacga gatgctggtc acctacgaag ggagaaaggc acgaggagcg 720
cctgacaaaa gtggttttgc cctgcttccc gcaagaggtg gcacccacgg ctggaacgca 780
ggagtacagc ccacagtccc cagctctgga cggccgcagc ggggcctcga agaggttcag 840
ggcgggtgcc gcgcgcgctc ggcggggtct cccggggcgt ggggcggggg gcgggggttg 900
gcgcggcgcg gggctcctcc ctcttctgcc ccgggctccc ctgctcttaa cccgcgcgcg 960
ggggcgccca ggccactggg ctccgcggag ccagcgagag gtctgcgcgg agtctgagcg 1020
gcgctcgtec cgteccaagg ccgacgccag cacgccgta tggccccgc agcggcgacg 1080
ggggcgagca ccctgcccag tggttctcgt gtcttcacca ccttgcccga cttgctcttc 1140
atctttgagt ttgtgagtgg ctctggccg gggaaaggac ggggtgggct gagccgtgcg 1200
ctctctcggg cgcccagcac agctgtcgga cgggatccgc tagctgcgca ggttctggga 1260
gcatcggggc agcagggcga gggcggggac taagccaggg 1300

```

```

<210> 3
<211> 1300
<212> ADN
<213> Homo Sapiens

```

10

<400> 3

```

gactcggggc gatttcaggc ttcagtgttt gtaggaggaa acacagcaat cacactatta 60
atagtaaatt aaaataaatg ggcaactgct gcatggtaat actttttttt ttaaggcaaa 120
aaataaaaaa tagtgaaaca gagaaacaaa acatgaaaca ccggcagtcg acaggcaggc 180

```

15

aaagaacctg ggggtggggg tagcagcggg cccaccctca aaaggcccgg gctgcccaga 240  
 ccaagagaaa gcgatgaatc tcttctggta acgtcccttc ctgtcgcag gattcaaggc 300  
 cgacctgccc cagcaccacc accagcagcc ttctgctggg gccggcacag ctgggagcaa 360  
 cctcctactc tcaggcagac gcgcagcacc aagcagagag gcccggtgca ggatcccagc 420  
 gccgaaccag cgccggctca gtggacgcgg aaggggccgg cgccgcggc cggccccatc 480  
 cccactgca gacccccagc ctgtggcggg ggtccagttc cgccaggaaa ccgccgcctg 540  
 gagctgtggg tcgcgcacat taacgcatcc agcggaaaaa tgaaggagac ccaaattcaa 600  
 agttaagta atggtgaccg gagaggtgcc ttgatgagaa ggtttggggg cccggttact 660  
 gatggttatic attcttacga gatgctggtc acctacgaag ggagaaagc acgaggagcg 720  
 cctgaccaa gtggttttgc cctgcttccc gcaagaggtg gacccacgg ctggaacgca 780  
 ggagtcagac ccacagtccc cagctctgga cgcccgcagc ggggcctcga agaggttcag 840  
 ggcggtgccc gcgcgcctcg ggcgggtct cccggggcgt ggggcggggg gcggggttg 900  
 gcgcggccgg gggctcctcc ctcttctgcc cccggctccc ctgctcttaa cccgcgcgcg 960  
 ggggcgcccc ggccactggg ctccgcggag ccagcgagag gtctgcggc agtctgagcg 1020  
 gcgctcgtcc cgtcccaagg ccgacgccag cacgccgtca tggccccgc agcggcgacg 1080  
 gggggcagca ccctgcccag tggcttctcg gtcttcacca ccttggccga cttgctcttc 1140  
 atctttgagt ttgtgagtgg ctctggccg gggaaaggac ggggtgggct gagccgtgcg 1200  
 ctctctcggg cgcccagcac agctgtcggg cgggatccgc tagctgcgca ggttctggga 1260  
 gcatcggggc agcagggcga gggcggggac taagccaggg 1300

<210> 4  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 4

gactcggggc gatttcagc ttcagtgtt gtaggaggaa acacagcaat cactactata 60  
 atagtaaat aaaaataatg ggcaactgct gcatggtaat acttttttt ttaaggcaa 120  
 aaataaaaa tagtgaaaca gagaacaaa acatgaaaca ccggcagtc acaggcaggc 180  
 aaagaacctg ggggtggggg tagcagcggg cccaccctca aaaggcccgg gctgcccaga 240  
 ccaagagaaa gcgatgaatc tcttctggta acgtcccttc ctgtcgcag gattcaaggc 300  
 cgacctgccc cagcaccacc accagcagcc ttctgctggg gccggcacag ctgggagcaa 360  
 cctcctactc tcaggcagac gcgcagcacc aagcagagag gcccggtgca ggatcccagc 420  
 gccgaaccag cgccggctca gtggacgcgg aaggggccgg cgccgcggc cggccccatc 480  
 cccactgca gacccccagc ctgtggcggg ggtccagttc cgccaggaaa ccgccgcctg 540  
 gagctgtggg tcgcgcacat taacgcatcc agcggaaaaa tgaaggagac ccaaattcaa 600  
 agttaagta atggtgaccg gagaggtgcc ttgatgagaa ggtttggggg cccggttact 660  
 gatggttatic attcttacga gatgctggtc acctacgaag ggagaaagc acgaggagcg 720  
 cctgaccaa gtggttttgc cctgcttccc gcaagaggtg gacccacgg ctggaacgca 780  
 ggagtcagac ccacagtccc cagctctgga cgcccgcagc ggggcctcga agaggttcag 840  
 ggcggtgccc gcgcgcctcg ggcgggtct cccggggcgt ggggcggggg gcggggttg 900  
 gcggcggccg gggctcctcc ctcttctgcc cccggctccc ctgctcttaa cccgcgcgcg 960  
 ggggcgcccc ggccactggg ctccgcggag ccagcgagag gtctgcggc agtctgagcg 1020  
 gcgctcgtcc cgtcccaagg ccgacgccag cacgccgtca tggccccgc agcggcgacg 1080  
 gggggcagca ccctgcccag tggcttctcg gtcttcacca ccttggccga cttgctcttc 1140  
 atctttgagt ttgtgagtgg ctctggccg gggaaaggac ggggtgggct gagccgtgcg 1200  
 ctctctcggg cgcccagcac agctgtcggg cgggatccgc tagctgcgca ggttctggga 1260  
 gcatcggggc agcagggcga gggcggggac taagccaggg 1300

<210> 5  
 <211> 1300  
 <212> ADN

10

<213> *Homo Sapiens*

<400> 5

```

gcttcactct ctgaaactct ttcaacatat gcaggtctca ctttaaag tcttgaatg 60
gagttcctac cagacggcct tccctaggg ttcataaaga ccctttcaag ccaccctccc 120
gcgcgccgcy ttaccacacg cccacctcct ggttgccgcy actacacctc acgaactttt 180
cgcgccccaa ggcggccctt tataaagatg cggtgacgtc accgcgcgcc gcgcgtgcgc 240
aaatatggta acgtgaccac gccctggcct gcgctaaggc cccggcgctc gcgcagtagc 300
tgggtccttc ccggaagggc tcagagggcg gctcgggtgag tggagtcgga cagtgggtgg 360
gtgcctgac tgcctgacgg ccgcgcaggg ccttgcagga ccgctgggga ggatgttgtt 420
aacgtaaaga gccaggctga tgaggaaggt tcatgagtcc tttgggtata agcgggagtt 480
ggggcgccg gagggcagat gactctgaga aggtgcggat cttcgattct aaggggagaa 540
tgtgatttgg ttttgtctaa agtgctgata gtggcctca cagaaaagtt accatgaacc 600
gactttccgg gattcccttg gcgctgcatt aattggcctc gttgcaaaaag gaaaccgaga 660
ggaagcgcg aagacttggc ctgcgggtga tgtagcggga aaatatctgt aatttagacc 720
ccgagggcg ggcctgtgag cgattgctgt ccgcgggtgct gaaggaaag cagccacgcc 780
cagagggaga aacctaata gggcagggcg gtgtttccgt tttccttggc cgggttcagc 840
accctacact caataagcgt gggctttcta gcaggactgg acaactcagc ggtgtgtcaa 900
tctttagaca atctttatag agtgcataga tgtttaattt cataaaattt aggaaagctt 960
catttcagga aggaatttta ttctctttt tctcccttag ggcaagcact ttaacctttt 1020
aagccaacc agatgagttg cctgcagttt tggaggcctt cagagcattt cactagacct 1080
ctgtctgtgt cgggtccagt tctttagcca aggtgagctt gctcgtggtg tgaagtgtgc 1140
cacgaaaatg ggctttcctg cactccacaa atgagagtgc ttagcagcca ggggaagggg 1200
aaggtatcta tctaattgtg ctttctctcg cagttacaca tgaaaagtac tttcctattt 1260
agaagagaga tggtagaact gggcactcct ccaataaagg 1300
    
```

5

<210> 6

<211> 1300

<212> ADN

<213> *Homo Sapiens*

<400> 6

10

```

aggccttctt aaactgttac aggggtgcagg catcattggg cattatccat ataaggatgc 60
cacaataact atacctttca aagccactaa attcactgtg ccgcgcataat cctggggaga 120
ggcagctgtg tgtgggggag ggcgagggga ccgggagggg gtaagtagag .tctccagat 180
gaacccagct cttaatctgg gctgggcaag aggaccatcc cttcaccgct cccatcctac 240
atcctagggt ccaacattta caggccctg ggttgacac cgacttagag agagcaccct 300
ggattcaggc ctgggtggct tgaattgagc ctcagaagag ccgcgtctgg agtgggctct 360
    
```

ES 2 443 230 T3

```

cgacacccag ggcaagtggg ggcggcagag ccctctctc ggteggcaca gcagcctctg 420
ccgcggtecc ggectgcgac gcgccagtc ttagcctccc ggctccggc gtctgtgag 480
tgtccggcgg gagaggcgca gggagcgcgc taccgggagg cgcgggcagc ggggactgg 540
tttctctcgg gccaggcct ccggggcaac cgtctccagc gcgcattctt ggtgcaggtg 600
gaacagcttt ctgctccggt agggcttcac ctatcgcggg agaggtaat ctcggatcta 660
aacctcgag ccgcagagcg ggctaaaacc gctactccac ctcttcccat ttctcccctc 720
cccacctcaa gacaaaagt cccaggccgg gcaggacctg atcacctctg cctcctccca 780
ctgcgcta at cctgagagcg agaggcccc caccgagcgg gaggctggca aaggggagtg 840
gaaagggagg atgatgggg ccgggggggtg ggggtgggat gagggcgagc aagggagggg 900
tgtcattttc tttttcttcc tttttttaa aaaagtattt ctctcgcgag aaaccgctgc 960
gcggagcagata cttgaagagg tgggaaagg agggggctgc gggagcccg gcagagactg 1020
tgggtgccac aagcggacag gagccacagc tgggacagct gcgagcggag ccgagcagtg 1080
gctgtagcgg ccacgactgg gagcagccgc cgcgcctcc tcgggagtcg gagccgccgc 1140
ttctccagtg ggtcagccg gggctcccagc gggggctcgg cggccaccgg ggtggagct 1200
gcggccacgg aggttttgc gtttgcggc cgcagagggc agggacaggg actggggtga 1260
ggggctgtcc cggaacgtcc acagctggcg ctggccctcc 1300

```

<210> 7  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 7

```

acatatatat tatatgtata gcttattatt tcttaaaaa gagcagtcta acttgggtaa 60
acaaaaactg ccaaggaaac gcateccctac ccatttgttc tcaactagccc acttcatttc 120
ccagggtgtgc tccctggctc tggtaactat catctgggtg gtgctgaagc aggtgctctc 180
tctgctccgg gaagccattc ccagtatagt cctcaccccc tgagtctccc ctgcttcca 240
aaagtcagct gtctaccgac ctctgccagc aaaactgtct aatcacgctt aatattgcaa 300
tctaaactct gcggtgtcgc atatggcccc ctgccaataa tcctttgacc catcccact 360
cttggccatt ttgtcaccag aatgccttag gctcatgtca gttgccttt acagggttaa 420
tagctcacc tccaaaatct gtgtaatat atgtgattca ttacatata ttatgtatgt 480
gtcatttttg ggacgagcat acgtataatg gaaatgtaca cccacaagaa gatgtagctc 540
cttgaatatg gttattttct tctcttga tccattctc atccagcgtg gctccatcac 600
gtggaattgc gccttctc ggccatttcc cctaagcttt ggaagactc atatttctg 660
aaatttaaat caccggctag cggtttctg tcccccttc ccgctccct ttaccccc 720
tcagcctcct cccgcctggc tggcgccgga tccgcgagca gcagtgtcca cctcacgtag 780
ctggccgagg cggtatccag attccgggg tctcgtcct tggcatagtg gtcgctaac 840
tccagcgcct tcccagtgcc tcccaaact ctctcctcc tgcgggggc tgttctagga 900
ggagcctaaa tgtccgcttc cccagctgaa ggggattgtg gtgcgtaaac aggattaact 960
ccctcgcccc cagctgcgcg gatctgcgcg gcggcctcgc gccagagcg agccttcggg 1020
tgcgtggagg gcgcagcact ggagcgggg atggggaagc gcgcgccct ggagccgtct 1080
cttgggcccc ctctctcgc gagtggggag gggggccaca gcctaaggag ctgctcctca 1140
aaagcccgg cggaaggcgg cgacggcgc ggttcttagc gccagcctca gctccaccac 1200
ctccctcacc ctccccgc tctctctcc cccctcgc tccagccact cggccgcagc 1260
cggcgtgtc ctccgcccc cggaagcgc gcgccagacc 1300

```

<210> 8  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

10

<400> 8

cccgacttta gaaggacttt cctgctgget ggctcgcage gtccccggag agcaagcgag 60  
 cgcgccgcag cgggagacta ggttctctcc gagcgtcctc cgcagaggcg ccgcgagagg 120  
 agcaggggagc cgcagctcgc cgtgttctct tcatttctgt caataaagaa ttgtcattag 180  
 gtttgcgtat ggcatgtgct attacccac cgtaaaacta aaattagcaa aatgtcagga 240  
 atgaaagat tgattcacca agatgcaatt atcatttaa agtgcttgat tgaggactg 300  
 atgttcagt atttattctg cataccatat acataattaa agtagttag taggagtaatt 360  
 tatcaatcta gttgagactg gaggggtgag gagggagctg ctgtatgttt gttttattaa 420  
 aatgctccga ggtctagtcc cgcctccctt ttgcaagagt gaaactgat atttctccag 480  
 ctcgcgagga aagagtcaac ggtttgggat tgtgggggag agagagacgg agagaaaaag 540  
 gcagcgcgaa gcaaggcaa ggacaaaatt aaataaagg ggaaaaaagg aggcaaaaaga 600  
 catttcatcg gacgtgctgc ttagaacccc aaccattcgt gctccgtctt ccctaccacc 660  
 cccgcctccc ctcccagat ccttcaatcc ccccccccc caccaccccc agtctttttt 720  
 acgcgatgtt tcaaacgctg tgagctgttc tcttttccc attcgtcttc tgtcacttcc 780  
 ttcttgagc cagttttctg gacgagtctg gttactttta atccgaccgg ccgctgagag 840  
 ccactttctc ctctctctcc tctctctct tctctctctc ctctctctc ctctctctc 900  
 tctctctcc agcggcctcg gcgcgcgca atgcgcggcc cgcgcctccc cccctcgcgc 960  
 gcgctccct cgcgcgcgcg cacacacgca cacatcgtct ccagctctct gctcgtctg 1020  
 ctcgcagtca cagacacttg agcacacgcg tacacccaga catctcggg ctgctattgg 1080  
 attgactttg aagttctgt gtgggtcgcg gtggctgcat gtttgaatca ggtggagaag 1140  
 cacttcaacg ctggacgaag taaagattat tgttgttatt tttttttct ctctctctct 1200  
 ctcttaagaa aggaaaatat cccaaggact aatctgatcg ggtcttctt catgtaagta 1260  
 ccctgatat ttctcgagga aatagaaaac ccgggtattg 1300

<210> 9  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 9

aacctgggaa gcagaggttg cagtgagccg agatcatgcc actgcactcc agcctgggca 60  
 acatagtaag actctgtctc aaaacaaaac aaaaacacaa caacaacaaa actggttcta 120  
 aatgctccgg gcaagtacat actgcttttt cttttttctg ttttttttc cctaaggcag 180  
 ccctattgaa gagtctgtca tctacaatat tttactaatt taaaaacact tgtcaagaag 240  
 ggtaaagtat gttcatcttg attttactct ttgtatcccg gaatcaggaa ctaacagttt 300  
 tatgatagt atatagactt tacacacaca ccaagactta aacacgtgg tgcaaggcgg 360  
 cttttgaaa ttaactgtgc tctggataca gaagtcacaa agctcaggct gcagagacat 420  
 tgaggggata atcagcctt attgtctctc gtgcactgtg cccagcctc cctctagcgg 480  
 ctacaagcgc tcaactgcca cactctctac taggctctcc ggaggcctcg gtttcgctgg 540  
  
 agagaaagag attgcgcagg cgcagcagc cgcagctggc cgcaggggac cgtgcctggg 600  
 ggggtctgag cggggccgct gactcgcgag gttctctgtc gectgccgcc ctgcctcctt 660  
 ttctcccctt ctagagtgtt acggtgactc ttgggtccga aggggacgga cgtcctccca 720  
 cgtaccattt agaaggtcag agtgggcaga acctagtctg agcaggccgg tgatgaatcc 780  
 gctggcaaga cgcgcgcgtg ggtactcacg tgacctaggc tgcgccgagc gcgtgctgag 840  
 ggctctgtgg cgggagcag gccgacgggc ggggcccgcg gccccgctga cgcgaagcgt 900  
 tcgagagcgc gcgtcgtgga acgtcttggg tgccacggca agcgcgcgcg cgaggccttg 960  
 ggaacctcgg gaccggcccc cggcagcgc agcggcggcc agtataagg agtgggagct 1020  
 ggtccgtgcc gcgcgcggcg cgcagggagc tctcagggca acgccggggc gcccgaggtc 1080  
 tggaaaggc aggtggggcg cgggggagtt caccctcgc cggctccctg ggtgggaggg 1140  
 gtgggcgcc caggccccgc ggactcggct gggaggtgga gcgggcgac gtcacgggac 1200  
 aagggtggag ggggtgggga cctcatccga gtctcagaga tggaaatttc gaggaaggga 1260  
 gcgatttcga gagagaatag tttgagaagt tgttctcagt 1300

10

<210> 10

<211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

<400> 10

5

```

gttgctgatg ctgtttataa aagtgaatag atagggaaat actgattcat atctcgtgaa 60
atgaaaatga aaggcttttt agtgacttag aattttaa atttctacat gagagagcag 120
tagtataat agaaataaca aagtaactgg caactgttta aaactgaagt taattcacag 180
ctatccagtg caaaacttca cctcaggtga tacacttttg acaggttaata catacagtaa 240
gtgtattttt agggaaacag tttcattggt gaaccaagat aatcatcatt agaatggtgt 300
atctgattta agtgctctta aacttaccaa ggtatttagat tttagtttga attgtctgga 360
gtagcggtag cggttggcac atttgttctt aaagggccag aaaataaata ttttaggctt 420
tatgggcat gtggtttctg ccacaagtcc tcagctctgt tcttttagtg tgaaagcagc 480
catagatgat acctaaatga atgagtatgg ctgcatocca ataaaacttc atttcaaaa 540
ctacatggct ggcctaagct ttagtctgtc tgcctaggga gtagtttact gagccactaa 600
tctaaagtgt aatactgtga gtgaatacca gtgagtacct ttgttaatgt ggataaccaa 660
tacttggtta taggaagttt tttagttgtg tgtttatta cacgtatttg actttgtgaa 720
taattatggc ttataatggc ttgtctgttg gtatctatgt atagcgttta cagtttctt 780
taaaaaacat gcattgagtt ttttaatagt ccaaccotta aaataaatgt gttgtatggc 840
cacctgatct gaccactttc tttcatgttg acatctttaa ttttaaaact gttttattta 900
gtgcttaaat cttgtttaca aaattgtctt cctaagtaat atgtctacct tttttttgg 960
aatatggaat attttgctaa ctgtttctca attgcatttt acagatcagg agaacctcag 1020
tctgacgaca ttgaagctag ccgaatgtaa gtgtaacttg gttgagactg tggttcttat 1080
tttgagttgc cctagactgc tttaaattac gtcacattat ttggaataa tttctggtta 1140
aaagaaagga atcatttagc agtaaatggg agataggaac atacctactt ttttctctat 1200
cagataactc taaacctcgg taacagttta ctaggtttct actactagat agataaatgc 1260
acacgcctaa attcttagtc tttttgcttc cctggtagca 1300
    
```

<210> 11  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

10

<400> 11

caaaaaate cateccttga taggaagtca gtatatctgg gttttagtca gtccttcaac 60  
 actagtcca aaatcctggg caagtcagtt aatcactcgg aggctgctta atttttttt 120  
 ttttaattc ttagtatatt atctgtcttg attaggtaaa gccaaagttg ttataagact 180  
 caagtgacta ttgaccaaaa tgcttcagaa actataacat tacataaaag tacaaggaat 240  
 tattacattt tctcaaatga agagatcatt taagaaaaac aaaatgctgg aaaatagtcc 300  
 ttaaaaactt ttctataata attactcttc tgatgaattt tacattgaca cctaatttga 360  
 agccttgaat attatTTTT gaacaatgaa ttgggtttat atattaaagt gttctaacca 420  
 aaggttgtt taatttatta gagttattaa gcctacgctc agatcaaggt agcagctaga 480  
 ctggtgtgac aacctgtttt taatcagtga ctcaaagctg tgatcacctc gatgtcaccg 540  
 aatggccaca gcttgtaaaa ggtaattttt aattatTTTta cagcctttaa aaggctgtca 600  
 ttgtaaagt aaatacatac tgtaaaaatg aagacaatgt atagttggca ggaatactgc 660  
 taagaatagt gggcctgaga gttgacctac ttcatttaat ctccttaaaa tatttttagt 720  
 tcttttttct cattaataaa tatcagtgc cactatcaaa gacctatctg tatgtataaa 780  
 caggaaagta tttctcaaaa aaacttcaag gttttaaaca ttttactata aacattatag 840  
 aaaacattaa aaatttctta atatgcattt taaagtatgc cttttatgtg gtgaacgttt 900  
 taactaaaca tttctctaga agtttttata acattaataa aactagtata cttttctctc 960  
 ctagagagtt acagtggagg taaaaggagt ggcttgcagg atggagaagc tgcaccagt 1020  
 ttattggaag tgagccacca tttgaatttg ctagctcatg ctgcagtatt cagattagt 1080  
 ggtgttttgt gatcatattt tctgcgattc aaatcaagac ctaaaattga ttctcattct 1140  
 gagatttaca ttttattatc caatgttatt tgggtattgt ttcagtttta aaatttctga 1200  
 cttgttattc tgacacactt taggataacc taagtcaata gtttaataca aaaatttccc 1260  
 tttggttata tgttttgaat taggttatat gttctgcatt 1300

<210> 12  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 12

agtacttcaa gggcttatca acaaagttgc aagttgtaac actatgaatt gttagtata 60  
 ctcttttggc tttgctagca agtgttgtaa agctatacac acacacacac acacacacac 120  
 acatacacac acacaccttt taaaatgggtg accctggtag caaatatgac tttaaatgga 180  
 tttaatTTTta atggctttaa ctacgttcag ctgtcatatg gatcaaaatt agcctctatc 240  
 cagctggggg caaccaggga gccacttttc ttaaccgagc acctactgaa cgtcaacaac 300  
 tgcaggagac gggactttac cttcgtctct ggtaaaactag ttgacacatc ctgtgttggc 360  
 aagaggccta agtagatgac cttggtcctc taaaatctgg cctgcactct cggggcacc 420  
 ctgcaacatc tacaaggca gctccagata gaaaaggggt ggggtcgaaa agccaataac 480  
 ggcaggcacc tgccccgcct cggggctggg gggctattcc agcggcttca gctaactttc 540  
 agagccattc gtttcccaac aaagtctgag gcgttctctc gctgggtaca ccaaggggct 600  
 ctgcaaccct cctggggggg ggggtgccca gagggcttcc ggaagtccca ggtttattct 660  
 ttcgggtcac agacagcaga aactaaaaag agggattacc ctttctgtcc agtcgcaaga 720  
 tggcgaccga gcctggtggg actccgaggg gccgcaggcc acctcctctt cccaatggcc 780  
 cgtgcgccgg cggcgacggc aagcgggagg gaggcggggc cggcgaagga aggagggggc 840  
 gagcggggcg ccctcccgcg cgtcttggcc ccgccccacg tccccgcgtc ccggcctgga 900  
 gccctgcgcc gggcggggcg cgcgcgctgc ctgccgggat actcggcccc cccagccagt 960  
 cctcccgctt tgcgcgcggg ccgcgagatc cgtgtgtctc ccaagatggt ggcgctgggc 1020  
 tcggggtgac tacaggagac gacggggcct tttcccttcg ccaggacccg acacaccagg 1080  
 cttcgcctgc tgcgcacccc ctccgcgcg tagccatccg ccagcggggg cggccgccat 1140  
 ccgcgccta cttacgctc acctctgccg acccggcgcg ctcggctgcg ggcggcgcg 1200  
 cctccttcg ctctcctcga gaatagctcg cggcctgtag cccctggcag gagggcccct 1260  
 cagccccccg gtgtggacag gcagcggcg ctggcgacga 1300

10

<210> 13

ES 2 443 230 T3

<211> 1300  
<212> ADN  
<213> *Homo Sapiens*

<400> 13

5

```
gCGAAAAGCC ttaggaccgc ttgttttaga cggctgggga atatttattc cttgttccac 60
tgatgggaaa atcagcgtct ggcaggcgt gattggtgga aaggaaaatg gtgatagtgg 120
cgtggaaaga ggatttgcgt agccttctcc tgccctctca acctgtgact cttccttagt 180
agtctccctt tcaccctcag gaccctttcc ggctcttccct agattaagag caaacgaaaa 240
ccttgaagat atttgaacta aagcgacccc taacgttcta acctgtgacc gtgattaaat 300
ttcagcgatg cgagggcaaa gcgctctcgg cggtagcgtg tgagccacct cccggcgtg 360
cctgtctcct ccagcagctc cccaagggat aggcctctgc cttggtggtc gaccctcagg 420
ccctcggctc tcccagggcg actctgacga gggtagggg gtggtccccg ggaggacca 480
gaggaaagc ggggacaaga agggagggga aggggaaaga ggaagaggca tcatccctag 540
cccaaccgct cccgatctcc acaagagtgc tcgtgaccct aaacttaacg tgaggcgcaa 600
aagcgcccc actttcccgc cttgcgcggc caggcaggcg gctggagtgt atggctcacc 660
ccgcgcccc tgccccatcc ccatccgaga tagggacgag gacacgctg cagggaaagc 720
agcagcgcgc gggagagggg cgggcagaaq cgctgacaaa tcagcgtggt gggcggagag 780
ccgaggagaa ggagaaggag gaggactagg aggaggagga cggcgacgac cagaaggggc 840
ccaagagagg gggcgagcga ccgagcgcgc cgacgcggaa gtgaggtgct tgcgggctgc 900
agcgcagacc cccgcccggc ccctccgaga gcgtcctggg cgctccctca cgccttgcct 960
tcaagccttc tgcctttcca ccctcgtgag cggagaactg ggagtggcca ttcgacgaca 1020
ggttagcggg tttgcctccc actccccag cctcgcgtcg ccggtcaca gcggcctcct 1080
ctggggacag tcccccccg gtgcgcctc cgcccttccct gtgcgctcct tttccttctt 1140
ctttcctatt aaatattatt tgggaattgt ttaaattttt tttttaaaaa aagagagag 1200
cggggaggag tcggagtgtt ggagaagcag agggactcag gtaagtacct gtggatctaa 1260
acggcgtctc ttggaatcc tggagaacgc cggatgggag 1300
```

<210> 14  
<211> 1300  
<212> ADN  
<213> *Homo Sapiens*

10

<400> 14



gaatggtcgt gggcaccggg aggggggtggt gctgccatga ggaccgcgtg ggccaggtct 60  
 ctgggaggtg agtacttgtc cctttgggga gcctaaggaa agagacttga cctggctttc 120  
 gtcctgcttc tgatattccc ttctccacaa gggctgagag attagctgc ttctccggga 180  
 tccgcttttc cccgggaaac gcgaggatgc tccatggagc gtgagcatcc aacttttctc 240  
 tcacataaaa tctgtctgcc cgctctcttg gtttttctct gtaaagtaag caagctgcgt 300  
 ttggcaataa atgaaatgga agtgcaagga ggccaagtca acaggtggta acgggttaac 360  
 aagtgtctgc gcggggtccg ctagggtgga ggctgagaac gccccctcgg gtggctggcg 420  
 cggggttggg gacggcccgc gagtgtgagc ggcgctgct cagggtagat agctgagggc 480  
 gggggtggat gttggatgga ttagaacat cacacttggg cctgctgttt gcctgagttt 540  
 gaaccacacc ccgagtgagc agttagtctt gttgcctacg cctttccacc atcaacctgt 600  
 tagccttctt ctgggattca tgtaaggat acccctgacc ctaagcctcc agcttccatg 660  
 cttctaactc atactgttac cctttagacc ccgggaattt aaaaaagggg ttaatctttt 720  
 catgcaactc cacttctgaa atgcagtaat aacaactcag aggattcatc ctaatccgtg 780  
 gttagggtgc tagactttta ctagccaaga tggatgggag atgctaaatt tttaatgcca 840  
 gagctaaaaa tgtctgcttt gtccaatggt taaatgagtg tacacttaa agagtctcac 900  
 actttggagg gtttctcatg atttttcagt gttttttgtt tatttttccc cgaaagtctc 960  
 cattcaaagt gtattttatg ttttccagtg tgggtgaaag gaattcatta gccatggatg 1020  
 tattcatgaa aggactttca aaggccaagg agggagtgtt ggctgctgct gaaaaacca 1080  
 aacaggggtg ggcagaagca gcaggaaaga caaaagaggg tgttctctat gtaggtaggt 1140  
 aaaccccaaa tgtcagtttg gtgcttgttc atgagtgatg ggttaggata atcaatactc 1200  
 taaatgctgg tagttctctc tcttgattca tttttgcac attgcttgtc aaaaaggtgg 1260  
 actgagtcag aggtatgtgt aggtaggtga atgtgaacgt 1300

<210> 15  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 15

gccaaagtga tatattaagc tcaccacctt ggtcacctgt ccatctatat ttttgagta 60  
 agcaacagaa ttaggagtct gaacatagga attaagaagc tctcgaatat aacagtggaa 120  
 gaccagcacc ttccaacct cttgtcaggg cagggatact ctatctctgc gtcactcaac 180  
 ctgccttggc ctcacagctc ttcttttgc ttaagtcttc aaaccatcaa cagtgatagc 240  
 cactaatttt agcaggtggc attagttgtc agcaaatgtt ttttcatatt taatgaatgt 300  
 ttattataat agaataaaat agaaatgttt attaataaaa cttgaaattt gtcccagaag 360  
 gataaaaacg tagattatft acataaaatg gccaaagatct gacaacagaa ctgtttcctg 420  
 aagatttact tcacagatta ttttgaaaac accaactgtg ttcaggaaaa ttttctctcc 480  
 ttttttctat aaaattttca aacaagtgag aatgctctaa aaatgatggt gtggagccta 540  
 gagaaaaatg tttctgtgtg ctatagatga acataaacta tatatatata ttactttccc 600  
 aggttgaagt gtttctttca aaaataaaat ttgattatft tcaatttttt gagggatgct 660  
 gatttttaaa aaggaaataa agagaaatta tttaggaagca tatacacagc cagtgagaag 720  
 aggaaagaag aaaaggccaa ggccagagga agggggtttg ggggtggggg aaggaaaggt 780  
 acagaccctc ccaataaacc gaagatgggt aaactccaac cgcaaagcga tgatgccact 840  
 tgtgggccct cagattggag aggacggaga tgagtgcagc ggctgtgagc actggggcac 900  
  
 cgcgcacaga cgggaggagg ggggtgtctg tgtgttccaa aggggggagt accagttaac 960  
 tggccgcggg cccgtggggg ttggtgagaa ggagggtgag cttggcggtg acgcacggcc 1020  
 ctcacgtgac cgggagctgc agagctacgc agcctcgggt gcagtcgtca ctctgtctc 1080  
 gctaccagct ccccgtgcc ctgcctcggc cgggctggca tccggcccgg gggaaagcgg 1140  
 accaggtaag ggccccgggg cccgcgcgcc ctggctagge cagcagagge tcaccgacc 1200  
 ccccagcagc cgcagaacag cgcaggggat ccccgccagc tttctccaca cccaacatt 1260  
 cttgggaaca tagtggctcc ctggggtggg gaagcagtg 1300

10

<210> 16

<211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

<400> 16

5

```
actgtagtga acaggagagc acacttctta tacaatggc aagggaact cacctctgcc 60
actgtgggca tcccagctct tcccattttc agaagaagcc agaaatccaa tttctttaca 120
ggtgaaatct cctggtttaa aaatgctcct ttcgaaatat tttaaaatat aatgctctaa 180
acacaaaagt ctgagcccc ctttccagca gcttgtggtc ctggcctgtg ccatgcttta 240
gtgccctaac gatgagcccc agcaattccc agcgcctcct gcttctccaa gtctggcctc 300
ttttcccctc tgtaccgccc cttctcccga ctctccgcag ctccgggtca ctttctttac 360
tgatctcccc tctcttctgg gagaggcact tgctgaacat ctttccgccc ggagcctccc 420
caggagctgc agaggggagc ctggagagat gctttctgac gtatcatctc ccacctttct 480
ccttggggac cctgaggaag gggacaaggc ctctctctgc agtccctcac tgtaactgga 540
aaaacacgac ctcccccctc ggaaggcttt ctgtgcgctc cacctcagga tgaggggagg 600
tgtaggggac acctcccaga aaccctaac ctcccagctc gttaaagaag aggggatagg 660
gtcaagggat gcgacagagc tgtgtggttt ccgatggga aacctcagtc gtttaggcac 720
ccctccgctc gactcacttc cgaagcagtc gattcttggg gagaagcgtc gcggaaggag 780
gcgactccga tgcagatggc cctgtcccgc cgcccaggtc cgtcgcgcgc gcagctgcgg 840
tagtcactgc gcctcccgc cccactcctt ggatgcccc cttccctctc ccggccagac 900
tctgagcagc agctccgccc ccagcgcgccc gccccagccc cggcgcctta aaagccgggc 960
gcaccgcccc gcgcccgcct gctgcccga cctctccttt cttctgtagc tcgcttgtaa 1020
gccgacgctc cggccccgat cccggcacca tgagcttcgg ctccggagcac tacctgtgct 1080
cctctcctc ctaccgcaag gtgttcgggg atggtctctc cctgtccgcc cgcctctctg 1140
gggcccggcg cgcggcgcgc ttccgctcgc agtctctgtc ccgagcaat gtggcctcct 1200
cggccgctgc ctctcggcc tcgtcgtctc gcctcggcct ggcctatcgc cggccgcccg 1260
cgtccgacgg gctggacctg agccaggcgg cggcgcgcac 1300
```

<210> 17  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

10

<400> 17

```
cagctctgga cggccgagc ggggcctcga agaggttcag ggcggtgcc gcggcgctcg 60
ggccgggtct cccggggcgt ggggcggggg gcgggggttg gcggcgcccg gggctcctc 120

ctctctgccc ccgggctccc ctgctcttaa cccgcgcgcg ggggcgccc ggccactggg 180
ctccgaggag ccagcgagag gtctgcgagg agtctgagcg gcgctcgtcc cgtcccaagg 240
ccgacgccag cacgccgca tggccccgc agcggcgagc gggggcagca ccctgcccag 300
```

15

<210> 18  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

20

<400> 18

ES 2 443 230 T3

cagctctgga cgccccgagc ggggcctcga agaggttcag ggcggtgcc ggcgcgctcg 60  
ggccgggtct cccggggcgt ggggcggggg gcggggttg gcggcggcc gggctcctcc 120  
ctcttctgcc ccgggtccc ctgctcttaa cccgcgcgcg ggggcgcca ggccactggg 180  
ctccgaggag ccagcgagag gtctgcgcgg agtctgagcg gcgctcgtcc cgtcccaagg 240  
ccgacgccag cacgccgtca tggccccgc agcggcgacg gggggcagca ccctgccag 300

<210> 19  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

5

<400> 19

cagctctgga cgccccgagc ggggcctcga agaggttcag ggcggtgcc ggcgcgctcg 60  
ggccgggtct cccggggcgt ggggcggggg gcggggttg gcggcggcc gggctcctcc 120  
ctcttctgcc ccgggtccc ctgctcttaa cccgcgcgcg ggggcgcca ggccactggg 180  
ctccgaggag ccagcgagag gtctgcgcgg agtctgagcg gcgctcgtcc cgtcccaagg 240  
ccgacgccag cacgccgtca tggccccgc agcggcgacg gggggcagca ccctgccag 300

<210> 20  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

10

<400> 20

cagctctgga cgccccgagc ggggcctcga agaggttcag ggcggtgcc ggcgcgctcg 60  
ggccgggtct cccggggcgt ggggcggggg gcggggttg gcggcggcc gggctcctcc 120  
ctcttctgcc ccgggtccc ctgctcttaa cccgcgcgcg ggggcgcca ggccactggg 180  
ctccgaggag ccagcgagag gtctgcgcgg agtctgagcg gcgctcgtcc cgtcccaagg 240  
ccgacgccag cacgccgtca tggccccgc agcggcgacg gggggcagca ccctgccag 300

15

<210> 21  
<211> 300  
<212> DNA  
<213> *Homo Sapiens*

20

<400> 21

gggcagggcg gtgtttcgt tttccttggc cgggttcagc accctacact caataagcgt 60  
gggctttcta gcaggactgg aactcaggc ggtgtgtcaa tctttagaca atctttatag 120  
agtgcataga tgtttaattt cataaaattt aggaaagctt catttcagga aggaatttta 180  
ttcctctttt tctcccttag ggcaagcact ttaacctttt aagccaacc agatgagttg 240  
cctgcagttt tggaggcctt cagagcattt cactagacct ctgtctgtgt cgggccagtg 300

<210> 22  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

25

<400> 22

ES 2 443 230 T3

agaggcccc caccgaggcg gaggtggca aaggggagtg gaaagggagg atggatgggg 60  
ccgggggggtg ggggtggtgat gagggcgacg aaggaggggg tgtcattttc tttttcttc 120  
tttttttaa aaaagtattt ctctcgcgag aaaccgctgc gcggacgata cttgaagagg 180  
tggggaaagg agggggctgc gggagccgcg gcagagactg tgggtgccac aagcggacag 240  
gagccacagc tgggacagct gcgagcggag ccgagcagtg gctgtagcgg ccacgactgg 300

<210> 23  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

5

<400> 23

attccggggg tctcgtcct tggcatagt gtcgctaac tccagcgcct tcccagtgcc 60  
tcccaaacet ctctcctcc tgcccgggce tgttctagga ggagcctaaa tgtccgcttc 120  
cccagctgaa ggggattgtg gtgcgtaaac aggattaact ccctcgcccc cagctgcgag 180  
gatctgcgcg gcggcctgcg ggcagagcgc agccttcggg tgcgtggagg gcgcagcact 240  
ggagcggggg atggggaagc gcgcgcctt ggagccgtct cttgggcccg cttcctcgcc 300

<210> 24  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

10

<400> 24

gacgagtctg gttactttta atccgaccgg ccgctgagag ccactttctc ctctcctcc 60  
tctcctcct tctcttctc ctctctctc ctctcctcc tcctcttccg agcggcctcg 120  
gcgcgcgca atgcgaggcc ccgcccccc cccctcgcgc gcgctcccct cgcgcgagcg 180  
cacacacgca cacatcgtct ccagctctct gctcgtctg ctcgcagtca cagacacttg 240  
agcacacgcg tacaccaga catctcggg ctgctattgg attgactttg aaggttctgt 300

<210> 25  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

20

<400> 25

ggtactcacg tgacctaggc tgcggcggcg gcgtgctgcg ggctctgtgg cgggagcag 60  
gccgacgggc gggccgagc ggccgcgtga cgcgaagcgt tcgagagcgc gcgtcgtgga 120  
acgtcttggg tgccacggca agcgcgagcg cgaggccttg ggaacctcgg gaccggcccc 180  
cggcagagcgc agcggcggcc agtctaaggg agtgggagct ggtccgtgcc gcggcggccg 240  
cgcaggggagc tctcagggca acgcccgggc gcccgaggtc tggaaaggcgc aggtggggcg 300

<210> 26  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

25

<400> 26

ES 2 443 230 T3

ttttaatagt ccaaccotta aataaatgt gttgtatggc cacctgatct gaccactttc 60  
 tttcatgttg acatctttaa ttttaaaact gttttattta gtgcttaaat cttgtttaca 120  
 aaattgtcct cctaagtaat atgtctacct ttttttttgg aatatggaat attttgctaa 180  
 ctgtttctca attgcatttt acagatcagg agaacctcag tctgacgaca ttgaagctag 240  
 ccgaatgtaa gtgtaacttg gttgagactg tggttcttat tttgagttgc cctagactgc 300

<210> 27  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 27

aaacttcaag gttttaaaca ttttactata aacattatag aaaacattaa aaatttctta 60  
 atatgcattt taaagtatgc cttttatgtg gtgaacgttt taactaaaca tttctctaga 120  
 agtttttata acattaataa aactagtata cttttctctc ctgagagatt acagtggagg 180  
 taaaaggagt ggcttgagg atggagaagc tgcaccagtg ttattggaag tgagccacca 240  
 tttgaatttg ctgagctcatg ctgcagtatt cagattagtg ggtgttttgt gatcatattt 300

<210> 28  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

10

<400> 28

aagcgggagg gagcgggggc cggcgaagga aggagggggc gagcgcggcg ccctcccgcg 60  
 cgtcttggcc ccgccccacg tccccgcgtc ccggcctgga gccctcgecc ggccggggcg 120  
 cgcgcgctgc ctgcgggat actcggcccc ccagccagt cctcccgtct tgcgcccgcg 180  
 ccgcgagatc cgtgtgtctc ccaagatggt ggcgctgggc tcgggggtgac tacaggagac 240  
 gacggggcct tttcccttcg ccaggaccog acacaccagg cttcgcctgc tcgqcaccc 300

15

<210> 29  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

20

<400> 29

gaggactagg aggaggagga cggcgcgac cagaaggggc ccaagagagg gggcgcgca 60  
 ccgagcgccg cgacgcggaa gtgaggtgctg tgcgggctgc agcgcagacc ccggccccgc 120  
 ccctccgaga gcgtcctggg cgctccctca cgccttgctt tcaagccttc tgcctttcca 180  
 ccctcgtgag cggagaactg ggagtggcca ttcgacgaca ggttagcggg tttgcctccc 240  
 actccccag cctcgcgctc cggctcaca gcggcctcct ctggggacag tcccccccg 300

25

<210> 30  
 <211> 300  
 <212> ADN  
 <213> *Homo Sapiens*

<400> 30

ES 2 443 230 T3

ctagccaaga tggatgggag atgctaaatt ttaaatgcca gagctaaaaa tgtctgcttt 60  
gtccaatggt taaatgagtg tacacttaaa agagtctcac actttggagg gtttctcatg 120  
atttttcagt gttttttgtt tatttttccc cgaaagttct cattcaaagt gtattttatg 180  
ttttccagtg tgggtgaaag gaattcatta gccatggatg tattcatgaa aggactttca 240  
aaggccaagg agggagttgt ggctgctgct gagaaaacca aacagggtgt ggcagaagca 300

<210> 31  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

5

<400> 31

gaagatgggt aaactccaac cgcaaagcga tgatgccact tgtgggccct cagattggag 60  
aggacggaga tgagtgacgg ggetgtgcgc actggggcac cgcgcacaga cgggaggagg 120  
ggggtgtctg tgtgttccaa aggggggagt accagttaac tggccgccgg cccgtggggg 180  
ttggtgagaa ggaggggtgag cttggcggtg acgcacggcc ctcacgtgac cgggagctgc 240  
agagctacgc agccttcggt gcagtcgtca ctcgtgtctc gctaccagct ccccgctgcc 300

<210> 32  
<211> 300  
<212> ADN  
<213> *Homo Sapiens*

10

<400> 32

ctgtcccgg cgcgccaggt cgtcgcgcgc gcagctgagg tagtcaactgc gcctccccgc 60  
ccccactcct ggatgcccc cttccctctc ccggccagac tctgagcagg agctccgccc 120  
ccagcgcgcc gccccagccc cggcgctta aaagccgggc gcaccgccc gccgcgccc 180  
gectgcega cctctcctt cttctgtagc tcgcgttgaa gccgcacgtc cggccccgat 240  
cccgccacca tgagcttcgg ctcggagcac tacctgtgct cctcctcctc ctaccgcaag 300

15

<210> 33  
<211> 1300  
<212> ADN  
<213> *Homo Sapiens*

20

<400> 33

attcctctt tccccctct gcacgcttgc tagccccagc gatcgctgct ggccccggg 60  
 taggaaagt gggttcctgg ccgtttctgc gacgctggcc tagggcttgc agctgctgtt 120  
 gagtgaagc acgcagactg gcgggagccg atcattttctc gaatgaagaa gaaaaagcgc 180  
 aattccctcc ttatgctcta ggggaattgag ccgcttccca gatcacccat tccagaaatg 240  
 tgaaccggg ccctcacaaa gtcgtctctg gtgaagaggt ggcgtgctgg gtgggggtt 300  
 gtggagggtg aaggcataag caaacatatt ttaaaatcca gatcgtagga agtgtcacct 360  
 ggccccctac ccaggcatgc tttctggggg aagcgcaggg ccaagctttc cctagaaaag 420  
 ctggggcgaa gagagagcag gcggcggcta aggagctcct ggcaggtctg gaaggtggag 480  
 aagtgggggtg aggtatTTTT ctagaagtg tagccctagc tcatctccta gattggggaa 540  
 gagggaaactg agggaggagg gaaggagacc cagggcagct ccaggatagg gaaatgttga 600  
 agaagggact gcgttctcca accgaaccct ccctcctggg aaccgcagcc cagcgcggtg 660  
 actgagttac cgcaaccggg cgggtggggag gaagggtggt ccaggaaacc ggcgagggag 720  
 aaaagcgggtg gaagggagag tcttctccct ggagcggccc cagcagtaca aagtgtctgt 780  
 cacagcggcc cttccgcccc tagattgacg agcagtggcg tggagccagc gcggaggtctg 840  
 cccccctccc ctcccagacc cgcagcggcg agcgcggtt agcaccaacg gagccggggg 900  
 cggcgtcttt gggatgaaa agggccaaag gggaggagtg ggggtgggggt gggggtttca 960  
 ctggtccact ataaaaggac cgctcggctg cccggttctt gcaactcgtg gaaagcggct 1020  
 ccgagccaggg ggtattgca aagccagggg gcgctaccgg accgagaggg gagagccctg 1080  
 agcagagtga gcaacatcgc agccaaggcg gaggccgaag aggggcgcca ggcaccaatc 1140  
  
 tcccgcttgc ctcagccccg gaggcgcccc agagcgcttc ttgtcccagc agagccactc 1200  
 tgcctgcgcc tgcctctcag tgtctccaac tttgcgctgg aagaaaaact tcccgcgcgc 1260  
 cggcagaact gcagcgcctc ctttagtga ctccgggagc 1300

<210> 34  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

5

<400> 34

tccgcacatt cgagcaaaga cagcctttag cgagttatta aaaacttagg ggcgctcttg 60  
 tccccacag gggccgaccg cacacagcaa ggcgatggcc cagctgtaag ttggtagcac 120  
 tgagaactag cagcgcgcgc ggagcccgtg gagactttaa tcaatctggt ctaacggttt 180  
 cccctaacc gctaggagcc ctcaatcggc gggacagcag ggcgcggtga gtcaccgccc 240  
 gtgactaagc gacccaccct ctctccctcg ggctttctc tggcaccgcc gtctcgcaac 300  
 tcccggctc cgaagctgga ctgagcccgt taggtccctc gacagaacct cccctcccc 360  
 caacatctct ccgccaaggc aagtcgatgg acagagggcg gggccggagc agccccctt 420  
 tccaagcggg cggcgcgga ggctcggcg aggcctgagc cctgcgttcc tgcgctgtgc 480  
 gcgccccac cccgcttcc aatctcaggc gctctttgtt tcttctccg cgacttcaga 540  
 tctgagggat tcctactct tctctctcc cgctccttg cccgcggtc tcccgcctg 600  
 accgcagccc cgagaccgcc gcgcacctcc tcccagccc ctttggcgtg gtgccaccgg 660  
 accctcttg ttcagtcca ggcggacccc cccctaccg cgcgacccc ctttttctcag 720  
 caccacaggg tgagcccagc tcagactatc atccgaaaag cccccaaaag tcccagccc 780  
 gcgctgaagt aacgggacca tgcccagtc caggccccgg agcaggaagg ctgagggcg 840  
 cccccacccc accgcccac cctcccgcct tctcgtagg tccctattgg ctggcgcgt 900  
 ccgcgctg gatggcagtg ggaggggacc ctcttctca acggggttat aaaaacagc 960  
 cctcggcgg ggtccagtc tctgcccactc tcgctccgag gtcccgcgc cagagacgca 1020  
 gccgctcc caccaccac acccaccgcg ccctcgttcg cctcttctcc gggagccagt 1080  
 ccgcccacc gccgcccgc aggcacatcgc caccctcgc agccatgtcc accaggtccg 1140  
 tgtctcgtc ctctaccgc aggatgttcg gcgcccggg caccgcgagc cggccgagct 1200  
 ccagcgggag ctacgtgact acgtccacc gcacctacag cctgggcagc gcgctgcgcc 1260  
 ccagcaccag ccgagcctc tacgctcgt ccccgggagc 1300

<210> 35  
 <211> 1300

10

<212> ADN  
 <213> *Homo Sapiens*

<400> 35

5

```

cttcctctac ctcgctccct cccagcctgc ccatcttttg tgcccgtctt ttgtgccggc 60
caccagtctt tacaacagc gccctggggg cccatactcc agctttccag agcgcacagc 120
tgcaaggcg gctgtgtgtt agagcccgtg atgccttga gtattgagtc atgccattca 180
tttaagaagg aaaaatctct ccaaaacggg gaaatggtgg gcagtcctgt gtgactggtg 240
agactcttgt aggggctgtt ctacaacgac gaaacccttc ctaggcactc actccaacag 300
aataacaage ccattttatt agtatttcgt tttccatgta aagtctctgt catacgaata 360

tatttataat tctgattttt ttacggcatt ggggagcaca ccgacaggct gctgaacggc 420
ggctggagat tcgaggaaa acgaagtctg ccgagggcggc ctcggggcggg cagggtcccgg 480
gctccatcac agggcacacg cggtaccag ggacgcagcc cccaacaca cacacacaca 540
cacacacaca cacacacaca cacacacaca ccctctcca ctcatgctg gcaaccagc 600
agaaactctg gactggggca aaacaagccc gggccccggc ggcacgcggg gctagggcgg 660
ttcccgcag tacctggtcg cgaggccgct cgcggggtgc cctgcgtgcc ccccactccc 720
gcagcccgcg ccctgctcgc tcaactgtgg ggcgcagcgg ccaggcttct ctgtttgttg 780
tttaagaaa tcctagggcg ggcgagcggc ggcacttagg ggagggggcg cagccagaat 840
tccctccag caagcgcgtg aggggcattc tcaacgaaa accagacca gaaagtagtg 900
accagccctc ctcggattac ccttcattgg ctctccctt gctccccca ccctccagat 960
ttgcataaaa aaggccaaga aaactctggc tgtgccccag caacggctca ttctgetccc 1020
ccgggtcgga gcccccgga gctgcgcgcg ggettgagc gcctcgcccg cgtgtctctc 1080
ccgggtgccc gcttctcgc gccccagccg ccggtgcca gcttttcggg gccccgagtc 1140
gcaccagcg aagagagcgg gccccggaca agctcgaact ccggccgct cgccttccc 1200
cggtccgct ccctctgcc cctcggggtc gcgcgccac gatgctgcag ggccctggct 1260
cgctgctgct gctcttctc gcctcgact gctgcctggg 1300
    
```

<210> 36  
 <211> 1300  
 <212> ADN  
 <213> *Homo Sapiens*

10

<400> 36



```

ttttctattt aaatatatat tatatattta aaaaagtgtc ctcccagagc taataaccgtt 60
gctagcagct cttcctgccg ccacaccggg caaagtccac ccaactgccc agtggtgagg 120
gccaccatgg gcggcccccac ctggagaggt gctgctcaca gcaaacagct ccaactgcgc 180
cttcgcctcg ccttccaggg agcccagcca ggcccactgg gtatttaca gcagacctcc 240
ctcgcctcag ccttccgaa ccctgttag ttgggaaacc acctgtctgc accgcagcta 300
gagaaccgag gagaggagcc gctagtctaa agggctgttg gttgaaatta ggaagcagtg 360
taaagaaaaa gaaaaaaaaa gtttgggagg ccaaggcagg agcatcacct gaggtcagga 420
gttcgagacc agcctggcta acatggtgaa accccgtctc tactaaaaat acaaaaaatt 480
agcgggggca tggtagcacg cggtgtaat cccagctact cgggaggctg aggcaggaga 540
atcgcctgaa cccgggaggc ggaggttgca gtcagcggag atagcgccat tgactccag 600
cttagcaac aagagcgaaa ctgtctcaa aaaaaaagt cttcataatt tcatgggtt 660
gcaagtatga tccaggctcc ccgcttctct gcaagccaat gcgagttaat tacagcgtcc 720
gccctggtct ctctccacc caccgctga tccattcccc ttcttttct cccctgtct 780
ctttcctcct cccccttta tttatgtatt tttggtttg ttttttaagg ggtgttgagc 840
cgcgtctggt tctagtaaac cgaaccgct cgcgaggag gcgattggt cccgcgccg 900
tgacggacgt ggtaacgagt gcggtcgc ccgcccggag ctgattggt gcgcggggcg 960
gctccgaggg ctcgccgta ggagccccgc gactccagc cctgcagcct ccggagtcag 1020
tgccgcgcgc ccgcccacc gcgccttct gctcgcgca cctccggag ccggggcgca 1080
cccagcccgc agcgcgcct ccccgccgc gccgcctcc accgcaggcc gagggccgc 1140
actgcccggg gggaccgggc agcagcttgc ggccgaggag ccgggcaacg ctggggactg 1200
cgccttttgt ccccgagggt ccctggaagt ttgcggcagg acgcgcgcgg ggaggcggcg 1260
gaggcagccc cgacgtcgcg gagaacaggg cgcagagccg 1300

```

<210> 37  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de PCR

<400> 37  
 ttcgggtttt ttgtttta attc 24

10

<210> 38  
 <211> 25  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <223> Cebador de PCR

<400> 38  
 gaaaaccata acgacgtact aacgt 25

20

<210> 39  
 <211> 25  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

25

<400> 39  
 ttttgggttt tttgtttt aattt 25

<210> 40

<211> 28  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de PCR

<400> 40  
 acaaaaacca taacaacata ctaacatc 28

10

<210> 41  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

15

<400> 41  
 gtaatttaga ttctggaggg c 21

<210> 42  
 <211> 19  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> Cebador de PCR

<400> 42  
 cgaccacaaa aaacgaaa 19

25

<210> 43  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

30

<400> 43  
 ttgtaattt agatttga gggc 24

<210> 44  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

35

<220>  
 <223> Cebador de PCR

<400> 44  
 ccaaccacaaa aaaacacaaa ca 22

40

<210> 45  
 <211> 21  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> Cebador de PCR  
 <400> 45  
 gtatttttt cgcgagaaat c 21  
 <210> 46  
 <211> 19  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Cebador de PCR  
 <400> 46  
 aatcgtaacc gctacaacc 19  
 <210> 47  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Cebador de PCR  
 <400> 47  
 aaagtatttt tttgtgaga aatt 24  
 <210> 48  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Cebador de PCR  
 <400> 48  
 cccaatcata accactacaa cc 22  
 <210> 49  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Cebador de PCR  
 <400> 49  
 gttattttta attc gatcgg tc 22  
 <210> 50  
 <211> 18  
 <212> ADN  
 <213> Secuencia artificial  
 <220>

<223> Cebador de PCR

<400> 50  
aaaccgctcg aaaaaaaa 18

5 <210> 51  
<211> 25  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> Cebador de PCR

10 <400> 51  
ttggtattt ttaattgat tggtt 25

<210> 52  
<211> 21  
<212> ADN  
15 <213> Secuencia artificial

<220>  
<223> Cebador de PCR

<400> 52  
ccaaaaccac tcaaaaaaa a 21

20 <210> 53  
<211> 21  
<212> ADN  
<213> Secuencia artificial

<220>  
25 <223> Cebador de PCR

<400> 53  
tcgtttttg gtatagtggt c 21

<210> 54  
<211> 18  
30 <212> ADN  
<213> Secuencia artificial

<220>  
<223> Cebador de PCR

<400> 54  
35 caaatccgcg caactaaa 18

<210> 55  
<211> 24  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> Cebador de PCR

<400> 55

gttttgttt ttgtatagt gggt 24

<210> 56  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de PCR

<400> 56  
 caaatccaca caactaaaaa c 21

10

<210> 57  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <223> Cebador de PCR

<400> 57  
 tggaacgttt tggtgttac 20

20

<210> 58  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

25

<400> 58  
 tacctcgaaa actcctacg 20

<210> 59  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <223> Cebador de PCR

<400> 59  
 gtggaatggt ttggtgta t 21

35

<210> 60  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

40

<400> 60  
 ttacctcaaa aactccctac a 21

<210> 61

<211> 20  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de PCR

<400> 61  
 cgtgtttgt cgggatattc 20

10

<210> 62  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

15

<400> 62  
 cccgtgtct cctataatca 20

<210> 63  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> Cebador de PCR

<400> 63  
 gtgtgtttg ttgggatatt t 21

25

<210> 64  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de PCR

30

<400> 64  
 cccatcatc tcctataatc a 21

<210> 65  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

35

<220>  
 <223> Cebador de PCR

<400> 65  
 cgggttag cgtagattc 20

40

<210> 66  
 <211> 18  
 <212> ADN

<213> Secuencia artificial

<220>  
<223> Cebador de PCR

5 <400> 66  
cgtcgaataa ccactccc 18

<210> 67  
<211> 23  
<212> ADN  
<213> Secuencia artificial

10 <220>  
<223> Cebador de PCR

<400> 67  
gtgtgggtg tagttagat ttt 23

15 <210> 68  
<211> 21  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> Cebador de PCR

20 <400> 68  
tcatcaaata accactccca a 21

<210> 69  
<211> 20  
<212> ADN  
25 <213> Secuencia artificial

<220>  
<223> Cebador de PCR

<400> 69  
agttaattgg tcgtcggtc 20

30 <210> 70  
<211> 20  
<212> ADN  
<213> Secuencia artificial

<220>  
35 <223> Cebador de PCR

<400> 70  
cgaataacga ctacaccgaa 20

<210> 71  
<211> 23  
<212> ADN  
40 <213> Secuencia artificial

<220>

<223> Cebador de PCR  
  
 <400> 71  
 attagttaat tgggtgttg ttt      23  
  
 5      <210> 72  
       <211> 23  
       <212> ADN  
       <213> Secuencia artificial  
  
 <220>  
 <223> Cebador de PCR  
  
 10     <400> 72  
       acacaaataa caactacacc aaa      23  
  
 <210> 73  
 <211> 20  
 <212> ADN  
 15     <213> Secuencia artificial  
  
 <220>  
 <223> Cebador de PCR  
  
 <400> 73  
 aggagttagc ttttagcgc      20  
  
 20     <210> 74  
       <211> 20  
       <212> ADN  
       <213> Secuencia artificial  
  
 <220>  
 25     <223> Cebador de PCR  
  
 <400> 74  
 acgactcaa cggaactac      20  
  
 <210> 75  
 <211> 23  
 30     <212> ADN  
       <213> Secuencia artificial  
  
 <220>  
 <223> Cebador de PCR  
  
 <400> 75  
 35     agtaggagtt ttgttttag tgt      23  
  
 <210> 76  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
  
 40     <220>  
       <223> Cebador de PCR  
  
       <400> 76



acaactcaa cacaaactac aaa 23

<210> 77  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de secuenciación

<400> 77  
 gggtttttt gttttaatt 20

10

<210> 78  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de secuenciación

15

<400> 78  
 accaaaaacc actcacaac tc 22

<210> 79  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> Cebador de secuenciación

<400> 79  
 ggagggtaga tgatttgag aa 22

25

<210> 80  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de secuenciación

30

<400> 80  
 ctccccttc ccctaactac ta 22

<210> 81  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

35

<220>  
 <223> Cebador de secuenciación

<400> 81  
 agggggtgtt atttttttt tttt 24

40

<210> 82

<211> 19  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Cebador de secuenciación

<400> 82  
 cccaatccct atccctacc 19

10

<210> 83  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de secuenciación

15

<400> 83  
 tttttggayg agtttggtta tt 22

<210> 84  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> Cebador de secuenciación

<400> 84  
 ccacctaatt caaacatata acc 23

25

<210> 85  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de secuenciación

30

<400> 85  
 ttttaygtag ttggygagg 20

35

<210> 86  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador de secuenciación

<400> 86  
 ctccctaaac tataaccccc ct 22

40

<210> 87  
 <211> 22  
 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> Cebador de secuenciación  
 <400> 87  
 atttagtttg agtagtygg tg 22  
 <210> 88  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> Cebador de secuenciación  
 <400> 88  
 ctccatccta acaatccata aa 22  
 <210> 89  
 <211> 17  
 <212> ADN  
 <213> Secuencia artificial  
 15 <220>  
 <223> Cebador de secuenciación  
 20 <400> 89  
 ggggggtgtt tagaggg 17  
 <210> 90  
 <211> 23  
 <212> ADN  
 25 <213> Secuencia artificial  
 <220>  
 <223> Cebador de secuenciación  
 <400> 90  
 cctcctacca aaaactacaa acc 23  
 30 <210> 91  
 <211> 19  
 <212> ADN  
 <213> Secuencia artificial  
 35 <220>  
 <223> Cebador de secuenciación  
 <400> 91  
 agaaggggtt taagagagg 19  
 <210> 92  
 <211> 19  
 <212> ADN  
 40 <213> Secuencia artificial  
 <220>

<223> Cebador de secuenciación

<400> 92  
actatcccca aaaaaaaccc 19

5 <210> 93  
<211> 22  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> Cebador de secuenciación

10 <400> 93  
atttggggg ttttagattg ga 22

<210> 94  
<211> 22  
<212> ADN  
15 <213> Secuencia artificial

<220>  
<223> Cebador de secuenciación

<400> 94  
ccaaaaaaccc actatattcc ca 22

20 <210> 95  
<211> 22  
<212> ADN  
<213> Secuencia artificial

<220>  
25 <223> Cebador de secuenciación

<400> 95  
gatgtagatg gtttggty gg 22

<210> 96  
<211> 18  
<212> ADN  
30 <213> Secuencia artificial

<220>  
<223> Cebador de secuenciación

<400> 96  
35 caaacraaaa ccatcccc 18

<210> 97  
<211> 20  
<212> ADN  
<213> Secuencia artificial

40 <220>  
<223> Cebador de secuenciación

<400> 97

gggtttttt gttttaatt      20

<210> 98  
<211> 22  
<212> ADN  
5      <213> Secuencia artificial

<220>  
<223> Cebador de secuenciación

<400> 98  
accaaaaacc actcacaac tc      22

10      <210> 99  
<211> 25  
<212> ADN  
<213> Secuencia artificial

15      <220>  
<223> Cebador de secuenciación

<400> 99  
ggaaaaatga aggagattta aatt      25

20      <210> 100  
<211> 18  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> Cebador de secuenciación

25      <400> 100  
aataacctaa acrcccc      18

30

35

40

**REIVINDICACIONES**

1. Un para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer, o si un individuo está recayendo tras un tratamiento contra el cáncer, que comprende el siguiente paso:

5 a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora, primer exón o intrón, de al menos un gen de una muestra, obtenida de dicho individuo, donde dicho gen es:

10 CNRIP1, p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927, y donde dicha secuencia comprende una secuencia de ácido nucleico constituida por la SEQ ID NO:7.

2. El método de acuerdo con la reivindicación 1 para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar o está recayendo tras un tratamiento contra el cáncer en el sistema aéreo-digestivo, que comprende los siguientes pasos:

15 a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG en una secuencia de ácido nucleico en la región promotora, primer exón o intrón de CNRIP1 (p. ej., tal como lo identifica la identificación génica ensembl ENSG00000119865, identificación de entrada 25927, y donde dicha secuencia comprende una secuencia de ácido nucleico constituida por la SEQ ID NO.: 7, en una muestra obtenida a partir de dicho individuo,

20 b) comparar dicho nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG respecto a una referencia; e

25 c) identificar si es probable que el individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar dicho cáncer, o esté recayendo tras el tratamiento contra dicho cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG son más elevados que los de la referencia e identificar si es poco probable que un individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar dicho cáncer, o esté recayendo tras el tratamiento contra dicho cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG está por debajo de los de la referencia.

30 3. Un método de acuerdo con la reivindicación 1-2, donde el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG se determina mediante secuenciación con bisulfito, reacción en cadena de la polimerasa específica de metilación (MSP) cualitativa y/o cuantitativa, pirosecuenciación, transferencia de Southern, barrido del genoma de los puntos de referencia de restricción (RLGS), extensión del cebador en un único nucleótido, micromatriz de islas de CpG, SNUPE, COBRA, espectrometría de masas, mediante la utilización de enzimas de restricción específicas de metilación, mediante la determinación del nivel de expresión de dichos genes o una combinación de estos.

35 4. Un método de acuerdo con la reivindicación 3, donde dicha PCR específica de metilación comprende la utilización de cebadores de ácidos nucleicos los cuales son capaces de hibridarse con una secuencia de ácido nucleico que comprende 2 sitios CpG y un residuo de citosina el cual no está en un sitio CpG.

40 5. Un método de acuerdo con cualquiera de la reivindicaciones precedentes, donde el cáncer del sistema aéreo-digestivo se selecciona a partir del grupo constituido por: tumores colorrectales, tumores pulmonares, carcinoma microcítico de pulmón, carcinomas pulmonares no microcíticos, tumores esofágicos, tumores gástricos, tumores pancreáticos, tumores hepáticos, tumores de la vesícula biliar y/o del conducto biliar, tumores del intestino delgado y tumores del intestino grueso.

45 6. Un método de acuerdo con cualquiera de las reivindicaciones precedentes, donde la muestra se obtiene a partir de sangre, excrementos, orina, líquido pleural, bilis, líquido bronquial, enjuagues bucales, piezas histológicas, ascitis, pus, líquido cefalorraquídeo, muestras obtenidas por aspiración, líquido folicular, tejido o moco.

7. Un método de acuerdo con cualquiera de las reivindicaciones precedentes, donde el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG se combina con al menos un marcador adicional.

50 8. Un método para determinar si un individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar o está recayendo tras un tratamiento contra el cáncer en el sistema aéreo-digestivo, que comprende los siguientes pasos:

a) determinar el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios

CpG en una secuencia de ácido nucleico que comprende una secuencia seleccionada a partir del grupo constituido por:

- 5
- 1) Una secuencia de ácido nucleico tal como la define la SEQ ID NO:7,
  - 2) Una secuencia de ácido nucleico la cual es complementaria a una secuencia tal como se define en 1);
  - 3) Una subsecuencia de una secuencia de ácido nucleico tal como se define en 1) o 2);
  - 4) Una secuencia de ácido nucleico la cual es idéntica en al menos un 75% a una secuencia tal como se define en 1), 2) o 3)

en una muestra obtenida de dicho individuo,

- 10
- b) comparar dicho nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG respecto a una referencia; e

- 15
- c) identificar si es probable que dicho individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar dicho cáncer, o esté recayendo tras el tratamiento contra dicho cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG son más elevados que los de la referencia e identificar si es poco probable que un individuo desarrolle, esté desarrollando o esté predispuesto a desarrollar dicho cáncer, o esté recayendo tras el tratamiento contra dicho cáncer, si el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG está por debajo del de la referencia.

- 20
- 9.** La utilización de un CNRIP1 en un ensayo diagnóstico donde se evalúa el nivel de metilación, el número de sitios CpG metilados o el estado de metilación de los sitios CpG como un indicador de si el individuo ha desarrollado, está desarrollando o está predispuesto a desarrollar cáncer en el sistema aéreo-digestivo, o si un individuo está recayendo tras el tratamiento contra el cáncer en el sistema aéreo-digestivo.

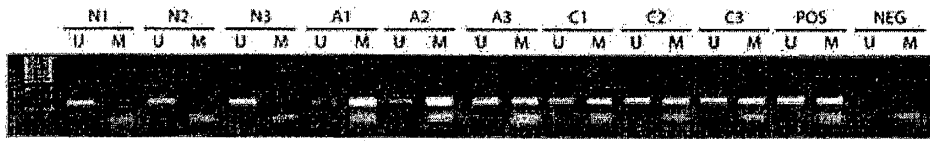


Fig. 1



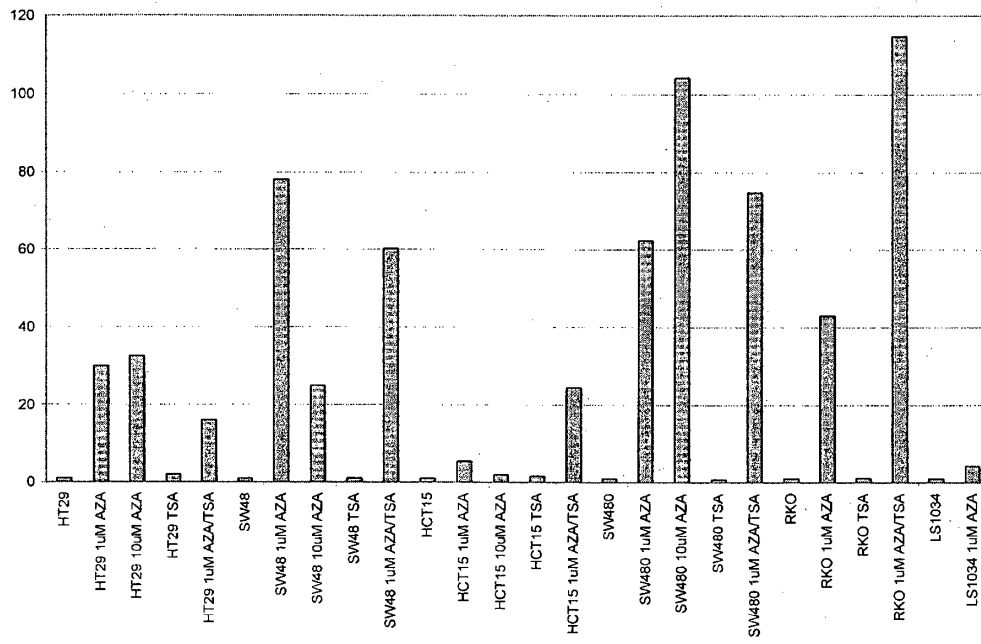


Fig. 2

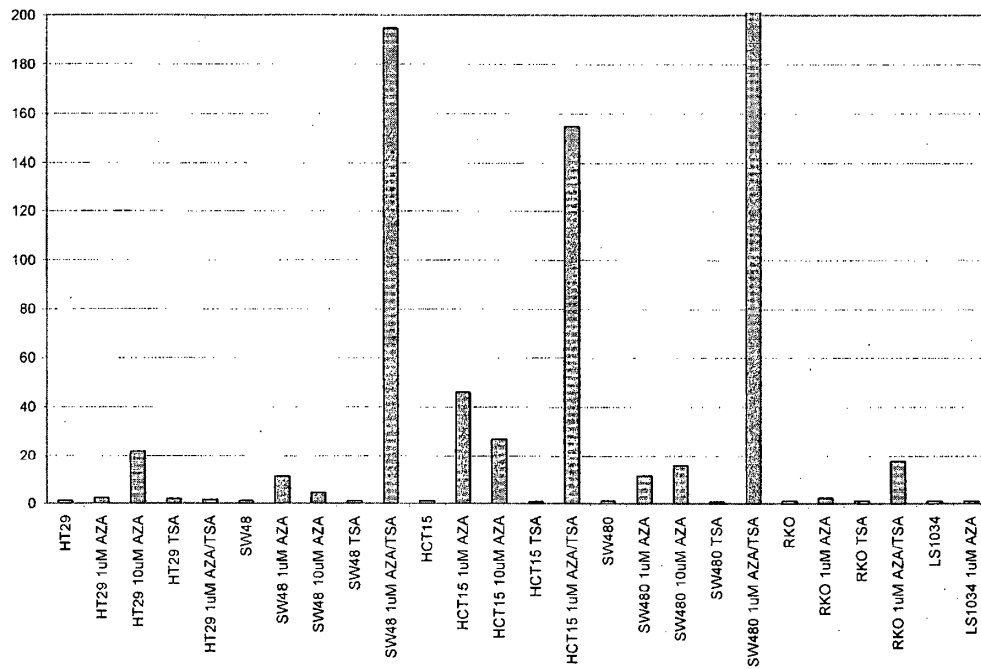


Fig. 3

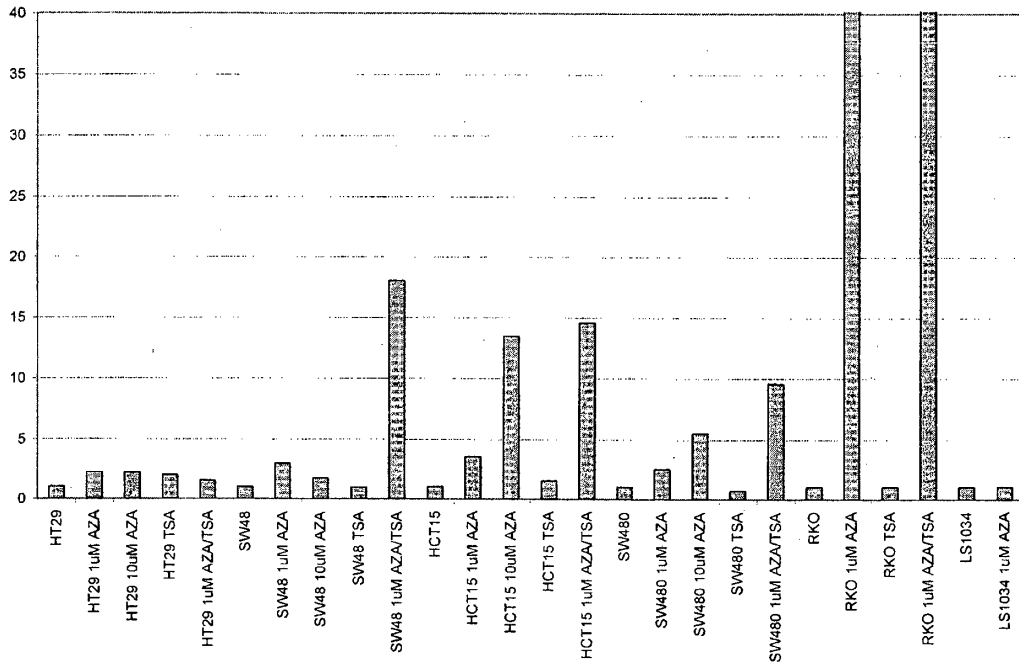


Fig. 4

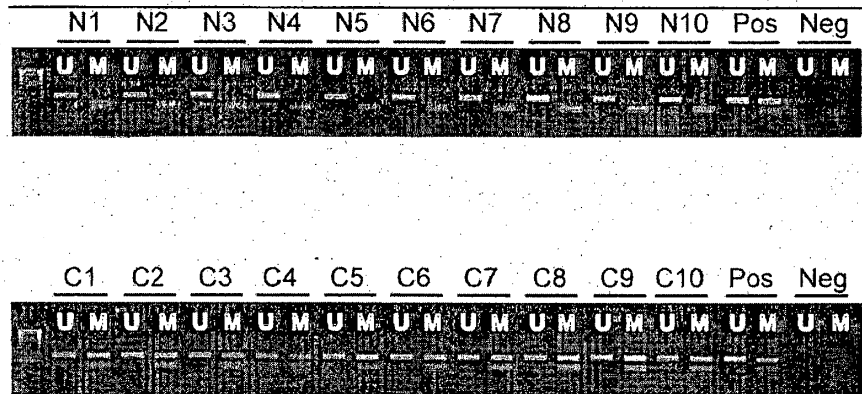


Fig. 5

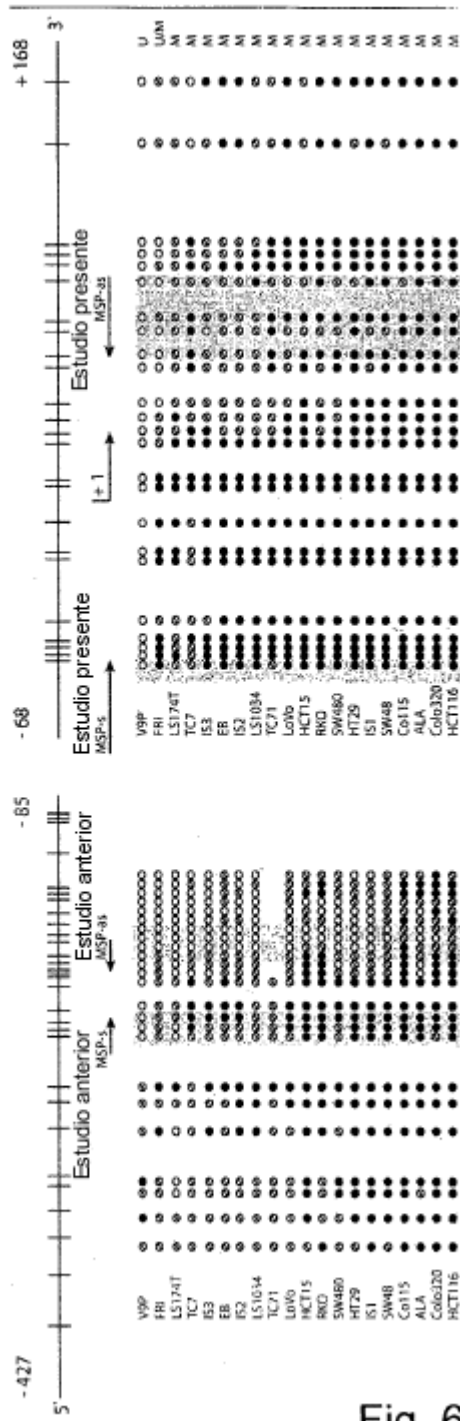


Fig. 6

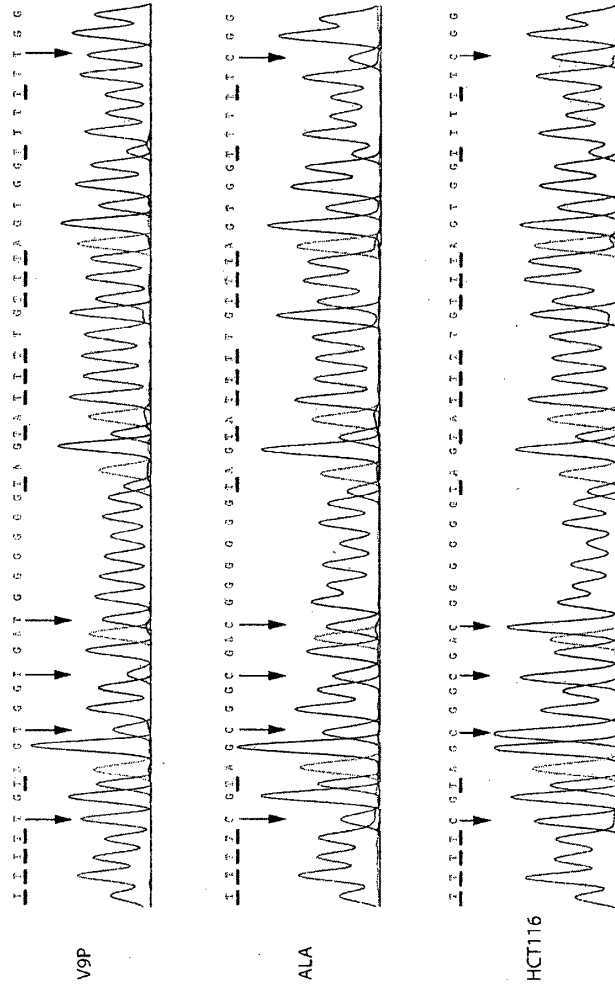


Fig. 7

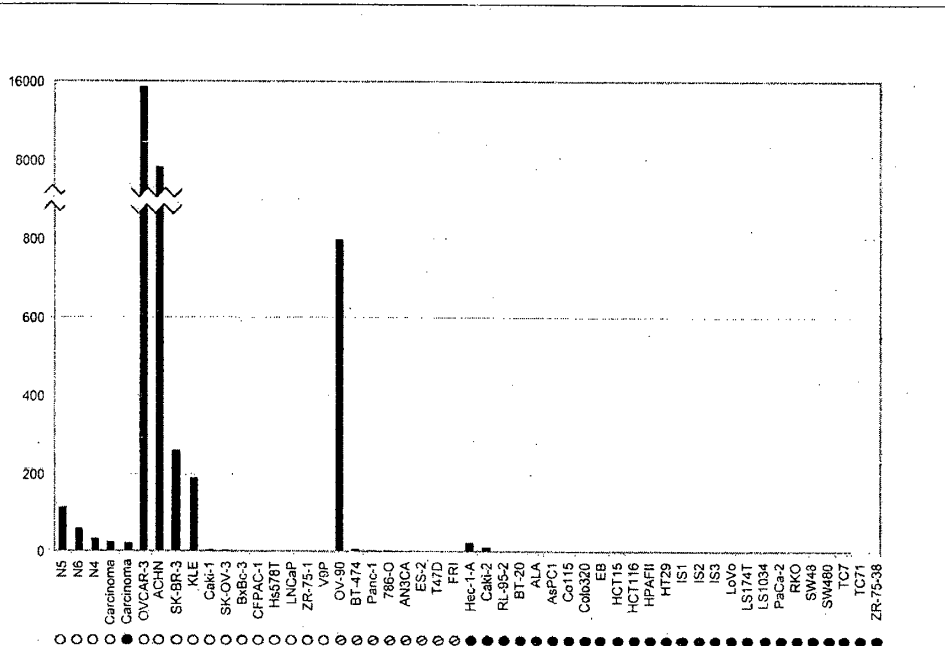


Fig. 8

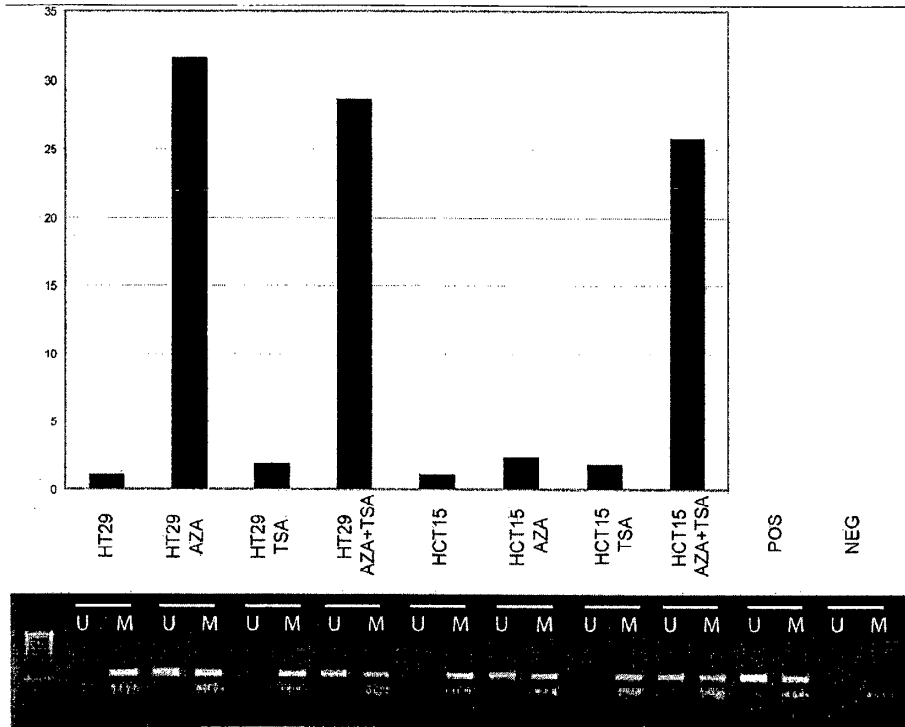


Fig. 9



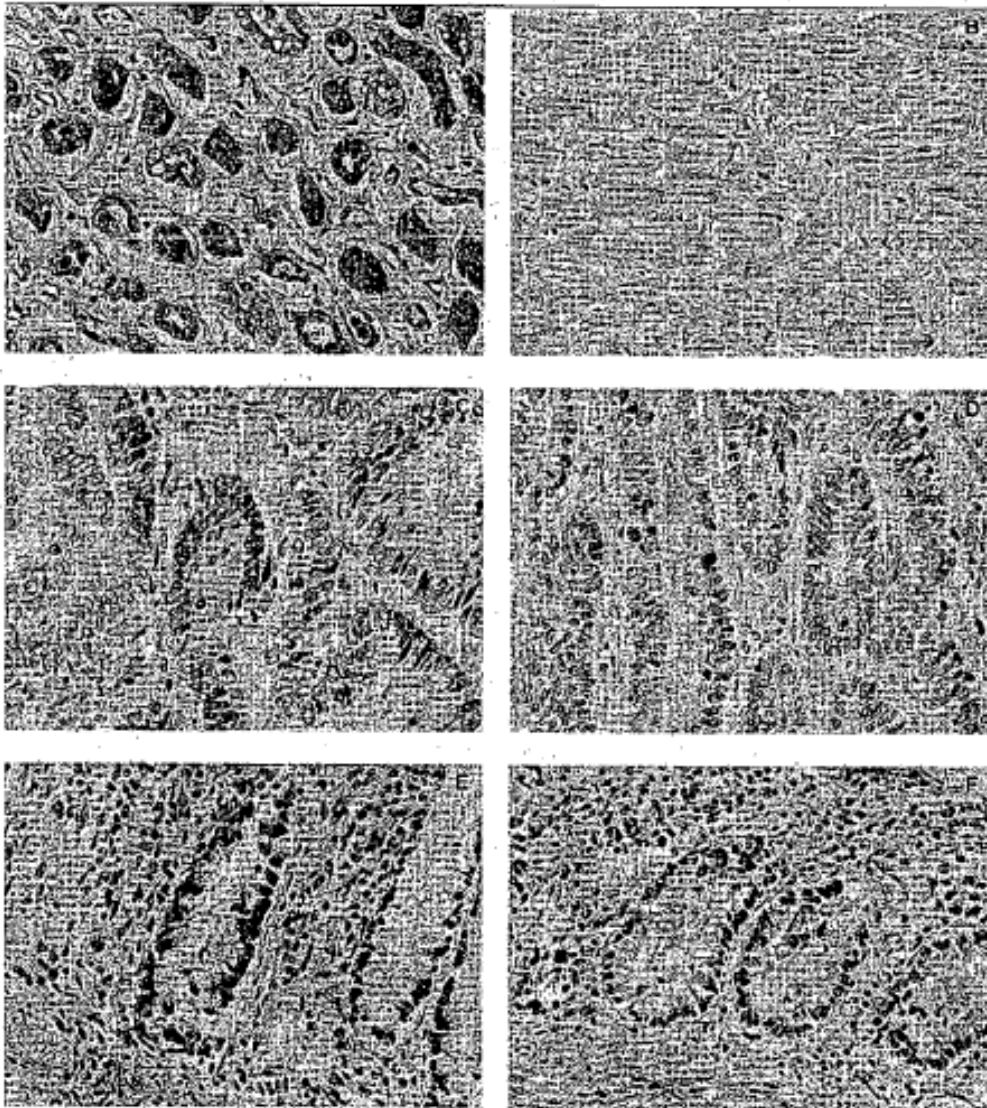


Figura 6

FIG. 10