



# OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA



11) Número de publicación: 2 745 718

51 Int. Cl.:

G01N 33/574 (2006.01) C12Q 1/6886 (2008.01)

(12)

## TRADUCCIÓN DE PATENTE EUROPEA

**T3** 

Fecha de presentación y número de la solicitud europea: 25.09.2015 E 15186777 (7)
 Fecha y número de publicación de la concesión europea: 04.09.2019 EP 3009842

(54) Título: Método para apoyar el diagnóstico de riesgo de recurrencia de un cáncer colorrectal, programa y sistema de computadora

(30) Prioridad:

26.09.2014 JP 2014197018 14.04.2015 JP 2015082645 16.09.2015 JP 2015183024

Fecha de publicación y mención en BOPI de la traducción de la patente: **03.03.2020** 

73) Titular/es:

SYSMEX CORPORATION (100.0%) 5-1, Wakinohama-Kaigandori 1-chome, Chuo-ku Kobe-shi, Hyogo 651-0073, JP

(72) Inventor/es:

GOTO, KENGO; YOSHIDA, YUICHIRO y OTOMO, YASUHIRO

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

#### **Observaciones:**

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

## **DESCRIPCIÓN**

Método para apoyar el diagnóstico de riesgo de recurrencia de un cáncer colorrectal, programa y sistema de computadora

#### Campo técnico

5

10

20

35

La presente invención se refiere a un método para apoyar un diagnóstico de riesgo de recurrencia de un cáncer colorrectal. Particularmente, la presente invención se refiere a un método que comprende: la obtención de datos de los niveles de expresión de los genes que pertenecen a grupos genéticos predeterminados que se relaciona con los ácidos nucleicos extraídos de tejidos de un paciente con un cáncer colorrectal; y el apoyo al diagnóstico de un riesgo de recurrencia del cáncer colorrectal en el paciente basándose en los datos obtenidos de los niveles de expresión, un programa y un sistema de computadora.

#### 15 Antecedentes

El cáncer colorrectal es un término genérico para tumores del ciego, intestino grueso y recto. De manera similar que, para muchos tipos de cáncer, la detección precoz es importante para tratar el cáncer colorrectal. En el tratamiento del cáncer, se pueden utilizar fármacos anticáncer con potentes efectos secundarios. En este caso, los pacientes son forzados a soportar cargas pesadas. Con el fin de reducir las cargas del paciente, es importante para los médicos seleccionar el tratamiento óptimo para los pacientes. Con este fin, los médicos necesitan conocer con precisión el estadio de progresión del cáncer, la malignidad, y las afecciones de los pacientes.

Una predicción precisa del pronóstico del paciente es importante para mejorar la Calidad de vida (QOL) del pronóstico. La clasificación de Dukes (es decir, el procedimiento histopatológico) se conoce como un método de predicción del pronóstico del cáncer colorrectal. La clasificación de Dukes se utiliza internacionalmente de manera amplia e incluye un proceso de clasificación en cualquiera de Dukes A, B, C, y D dependiendo del grado de invasión del cáncer. Aunque la clasificación de Dukes se lleva a cabo por los métodos a simple vista, algunos médicos son susceptibles de errar. También existe un problema ya que se produce fácilmente una diferencia de diagnóstico debido a que los tejidos de cáncer colorrectal se obtienen de diferentes centros médicos.

En los últimos años, se han llevado a cabo estudios para la predicción del pronóstico del cáncer utilizando marcadores genéticos que se enfocan en las tasas de aumento-disminución de los niveles de expresión de ciertos genes. Por ejemplo, el documento US 2008/058432 A desvela un análisis molecular para predecir la recurrencia del cáncer colorrectal en pacientes diagnosticados de cáncer colorrectal o pacientes que reciben tratamiento para el cáncer colorrectal. La técnica que se desvela en el documento US 2008/058432 A es un método para la predicción del pronóstico de recurrencia de ciertos cánceres colorrectales. En consecuencia, el método no se puede utilizar para todos los tipos de cáncer colorrectal.

- 40 El documento WO 2007/100859 desvela un método de identificación de pacientes que es probable que sean tratados satisfactoriamente con un régimen terapéutico basado en la expresión de uno o más genes seleccionados de entre un grupo que consiste en 14 genes diferentes, incluyendo ANGPTL2, en una muestra de tejido tumoral de un paciente.
- 45 El documento WO 2013/079309 desvela un método para la predicción del resultado, tal como la recurrencia, de un paciente que padece un cáncer colorrectal basándose en el nivel de expresión de más de 100 genes incluyendo el ANGPTL2, donde un aumento del nivel de expresión de dichos genes es indicativo de un aumento de la probabilidad de un resultado negativo del paciente.
- 50 Sillars-Hardebol et al. (2012. Cellular Oncology 35(4):293-300) desvela que la ganancia en 20g se correlaciona con un mal pronóstico en los pacientes de cáncer colorrectal. Sheffer et al. (2009. Proceedings of the National Academy of Sciences 06 (17):7131-7136) desvela que el CCDC68, que se localiza en el cromosoma 18, está regulado negativamente en tumores de cáncer de colon primarios en comparación con el tejido normal, pero no menciona que el CCDC68 se pueda utilizar para predecir la recurrencia del cáncer colorrectal. Letellier et al. (2014. British Journal 55 of Cancer 111(4):726-735) desvela que los niveles del SOCS6 están reducidos en el cáncer colorrectal y pueden servir como biomarcador diagnóstico para los pacientes de cáncer colorrectal, pero menciona que el SOCS6 no se asocia con el resultado de la enfermedad. Dunne et al. (2014. Clinical Cancer Research 20(1):164-175) desvela que los altos niveles de expresión de AXL se asocian con una pobre supervivencia total en los tejidos de cáncer colorrectal es estadio temprano. De Angelis et al. (2001. International Journal of Colorectal Disease 16(1):38-45) 60 desvela que los pacientes que padecen carcinoma colorrectal con pérdida de brazos cromosómicos 1p, 4q, 8p, 14q o 18q o ganancia de brazos cromosómicos 20q tienen tiempos de supervivencia más cortos que los que no tienen estas aberraciones.

Un objetivo de la presente invención es proporcionar un método para apoyar el diagnóstico del riesgo de recurrencia del cáncer colorrectal que es fiable en distintos casos de cáncer colorrectal, un programa y un sistema de computadora.

Los presentes inventores se han dedicado a hacer estudios repetitivos y han descubierto que el cáncer colorrectal se puede clasificar en tres tipos por análisis de agrupamiento. Después, han descubierto que los tres tipos se asocian con el pronóstico del cáncer colorrectal y el resultado obtenido es suficientemente estable, y completan la presente invención.

#### Sumario de la invención

15

20

25

40

45

De acuerdo con la presente invención, se proporciona un método para apoyar un diagnóstico de riesgo de 10 recurrencia de un cáncer colorrectal. El método comprende las etapas de:

llevar a cabo una primera medición para medir los niveles de expresión de C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407 en una muestra biológica recolectada de un paciente con un cáncer colorrectal, una segunda medición para medir los niveles de expresión de ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y una tercera medición para medir los niveles de expresión de ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y determinar el riesgo de recurrencia del cáncer colorrectal del paciente basándose en los resultados de la primera medición, la segunda medición y la tercera medición.

En la etapa de determinación del método,

el riesgo de recurrencia se puede determinar que es alto cuando la media de los niveles de expresión de los genes del tercer grupo de genes es mayor o igual a un valor de referencia del tercer grupo de genes independientemente de los niveles de expresión de los genes del primer y segundo grupos genéticos;

el riesgo de recurrencia se puede determinar que es medio cuando la media de los niveles de expresión de los genes del tercer grupo de genes es menor que un valor de referencia del tercer grupo de genes y la media de los niveles de expresión de los genes del primer grupo genético es menor que un valor de referencia para el primer grupo genético independientemente de los niveles de expresión de los genes del segundo grupo genético;

30 el riesgo de recurrencia se puede determinar que es medio cuando la media de los niveles de expresión de los genes del tercer grupo de genes es menor que un valor de referencia del tercer grupo de genes y la media de los niveles de expresión de los genes del primer grupo genético es mayor o igual a un valor de referencia para el primer grupo genético y la media de los niveles de expresión de los genes del segundo grupo genético es mayor o igual a un valor de referencia para el segundo grupo genético; y

el riesgo de recurrencia se puede determinar que es bajo cuando la media de los niveles de expresión de los genes del tercer grupo de genes es menor que un valor de referencia del tercer grupo de genes y la media de los niveles de expresión de los genes del primer grupo genético es mayor o igual a un valor de referencia para el primer grupo genético y la media de los niveles de expresión de los genes del segundo grupo genético es menor que un valor de referencia para el segundo grupo genético.

Los niveles de expresión de los genes se pueden medir con una micromatriz en la etapa de medición.

El riesgo de recurrencia se puede determinar que es alto cuando un grupo determinado que tiene un riesgo medio de recurrencia tiene una mutación del gen KRAS, y el riesgo de recurrencia puede determinarse que es bajo cuando el grupo determinado que tienen un riesgo medio de recurrencia no tiene la mutación del gen KRAS.

De acuerdo con la presente invención, también se proporciona un programa de computadora ejecutado por una computadora. El programa de computadora hace que la computadora ejecute un procesamiento que comprende:

la recepción de una muestra biológica recolectada de un paciente con un cáncer colorrectal, los niveles de expresión de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles de expresión de un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y

determinar un riesgo de recurrencia de cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos

De acuerdo con la presente invención, también se proporciona un sistema de computadora para determinar el riesgo de recurrencia de un cáncer colorrectal. El sistema de computadora comprende una computadora que comprende un procesador y una memoria. La memoria almacena un programa de computadora que hace que la computadora ejecute un procesamiento que comprende:

la recepción de una muestra biológica recolectada de un paciente con un cáncer colorrectal, los niveles de

3

60

expresión de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles de expresión de un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y

la determinación del riesgo de recurrencia del cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos.

10

5

De acuerdo con la presente invención, se proporciona un método, un programa de computadora y un sistema de computadora para apoyar un diagnóstico de riesgo de recurrencia de un cáncer colorrectal que sea fiable.

#### Breve descripción de los dibujos

15

30

60

- La Fig. 1 es una vista esquemática que muestra un ejemplo de un aparato de apoyo al diagnóstico que se utiliza para un método de apoyo al diagnóstico de la presente invención.
- La Fig. 2 es un diagrama de bloques que muestra una configuración funcional del software del aparato de apoyo al diagnóstico.
- 20 La Fig. 3 es un diagrama de bloques que muestra una configuración del hardware del aparato de apoyo al diagnóstico.
  - La Fig. 4 es un diagrama de flujo que muestra la operación del aparato de apoyo al diagnóstico.
  - La Fig. 5 es una vista que muestra el resultado de la clasificación de los casos en un conjunto de entrenamiento en los grupos de riesgo de recurrencia.
- La Fig. 6 ilustra curvas de Kaplan-Meier que muestran el riesgo de recurrencia en los respectivos grupos de riesgo.
  - La Fig. 7 es una vista que muestra el resultado de la clasificación de los casos en un conjunto de entrenamiento un conjunto 1 de validación en los grupos de riesgo de recurrencia.
  - La Fig. 8 es una vista que muestra el resultado de la clasificación de los casos en un conjunto 2 de validación en los grupos de riesgo de recurrencia.
  - La Fig. 9 es una vista que muestra el resultado de la clasificación de los casos en un conjunto 3 de validación en los grupos de riesgo de recurrencia.
  - La Fig. 10 ilustra curvas de Kaplan-Meier de acuerdo con la clasificación de Dukes de casos del conjunto 3 de validación.
- La Fig. 11 es una vista que muestra el resultado de la clasificación de los casos en un conjunto 4 de validación en los grupos de riesgo de recurrencia.
  - La Fig. 12 ilustra curvas de Kaplan-Meier de acuerdo con la clasificación de Dukes de casos del conjunto 4 de validación.
- La Fig. 13 ilustra curvas de Kaplan-Meier de acuerdo con la clasificación de Dukes de casos en un conjunto de entrenamiento.
  - La Fig. 14 ilustra curvas de Kaplan-Meier que muestran el resultado de la clasificación de riesgo de recurrencia de especímenes que se habían determinado como un grupo de riesgo medio en el Ejemplo 4, de acuerdo con la presencia de mutaciones del gen KRAS en los especímenes.
- La Fig. 15 ilustra curvas de Kaplan-Meier que muestran el resultado de la clasificación de riesgo de recurrencia de especímenes que se habían determinado como un grupo de riesgo medio en el Ejemplo 5, de acuerdo con la presencia de mutaciones del gen KRAS en los especímenes.
  - La Fig. 16 es una vista que muestra el resultado de la clasificación del riesgo de recurrencia en muestras de tejido FFPE 18 del Ejemplo 8.
- La Fig. 17 ilustra las curvas de Kaplan-Meier que muestran el resultado de la clasificación de riesgo de recurrencia estratificado del grupo de riesgo medio en muestras de tejido FFPE 18 del Ejemplo 8 de acuerdo con la presencia de mutaciones del gen KRAS.
  - La Fig. 18 ilustra curvas de Kaplan-Meier obtenidas en el Ejemplo 10.
  - La Fig. 19 es un diagrama de flujo que muestra una operación del aparato de apoyo al diagnóstico.

## 55 Descripción detallada de las realizaciones preferidas

En el método de apoyo al diagnóstico de un riesgo de recurrencia de cáncer colorrectal descrito en el presente documento (al que se hace referencia de aquí en adelante como "método de apoyo al diagnóstico"), las etapas incluyen llevar a cabo una primera medición para medir los niveles de expresión de una pluralidad de genes seleccionados de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407 en una muestra biológica recolectada de un paciente con un cáncer colorrectal, una segunda medición para medir los niveles de expresión de una pluralidad de genes seleccionados de entre un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1,

DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y una tercera medición para medir los niveles de expresión de una pluralidad de genes seleccionados de entre un tercer

grupo genético que consiste en ANGPTL2, AXL, C1R, CIS, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM.

No existe una limitación particular en cuanto a la "muestra biológica" a condición de que sea una muestra biológica que contenga ácidos nucleicos derivados de células tumorales de un paciente con un cáncer colorrectal (por ejemplo, de ARNm). Por ejemplo, se puede utilizar un espécimen clínico. Ejemplos específicos de un espécimen clínico incluyen sangre, suero o tejidos recolectados mediante cirugía o biopsia. Adicionalmente, se puede utilizar una muestra fijada en formalina embebida en parafina (FFPE) de un tejido recolectado de un sujeto como muestra biológica.

10

15

El método de la realización puede incluir una etapa de extracción de ADN de la muestra biológica antes de la etapa de medición. La extracción de ADN de la muestra biológica se puede llevar a cabo mediante cualquier método conocido en la técnica. Por ejemplo, se puede obtener un extracto de ADN mediante un procedimiento que incluye la centrifugación de la muestra biológica para precipitar las células que contienen el ADN, destruir física o químicamente las células y retirar los desechos celulares. La operación también se puede llevar a cabo utilizando un kit de extracción de ADN disponible en el mercado o similar.

Como se utiliza en el presente documento, la expresión "el primer grupo genético" es un término general para genes presentes en una región desde 18q21 a 18q23 en el brazo largo del cromosoma 18. Específicamente, el primer grupo consiste en genes representados por los siguientes símbolos genéticos: C18orf22 (fase de lectura abierta 22 del cromosoma 18), C18orf55 (fase de lectura abierta 55 del cromosoma 18), CCDC68 (dominio superenrrollado que contiene 68), CNDP2 (CNDP dipeptidasa 2 (familia M20 de metalopeptidasas)), CYB5A (citocromo b5 tipo A (microsómico)), LOC400657 (LOC400657 hipotético), LOC440498 (proteína tipo 1 de unión al factor de choque térmico), MBD2 (proteína 2 del dominio de unión a metil-CpG), MBP (proteína mielínica básica), MYO5B (miosina VB), NARS (asparginil-tARN sintetasa), PQLC1 (bucle PQ repetido que contiene 1), RTTN (Rotatin),SEC11C (homólogo C de SEC11 (S. cerevisiae)), SOCS6 (supresor de la señalización 6 de citocina), TNFRSF11A (miembro 11a de la superfamilia del receptores del factor de necrosis tumoral, activador NFKB), TXNL1 (tiorredoxina tipo 1), TXNL 4A (tiorredoxina tipo 4A), BPS4B (homólogo B de la proteína vacuolar de clasificación 4 (S. cerevisiae)), y ZNF407 (proteína 407 en dedos de zinc).

30

35

40

Como se utiliza en el presente documento, la expresión "el segundo grupo genético" es un término general para genes presentes en una región desde 20q11 a 20q13 en el brazo largo del cromosoma 20. Específicamente, el segundo grupo genético consiste en genes representados por los siguientes símbolos genéticos: ASXL1 (peines sexuales tipo 1 adicionales (Drosophila)), C20orf112 (fase de lectura abierta 112 del cromosoma 20), C20orf177 (fase de lectura abierta 177 del cromosoma 20), CHMP4B (proteína 4B modificadora de cromatina), COMMD7 (dominio COMM que contiene 7), CPNE1 (copina I), DIDO1 (obliterador 1 inductor de muerte), DNAJC5 (homólogo del DnaJ (Hsp40) miembro 5 de la subfamilia C), KIF3B (miembro 3B de la familia de cinesinas), NCOA6 (coactivador 6 del receptor nuclear), PHF20 (proteína 20 en dedo PHD), PIGU (biosíntesis del ancla de fosfatidilinositol glicano, clase U), PLAGL2 (gen tipo 2 del adenoma pleomórfico), POFUT1 (proteína Ofucosiltransferasa 1), PPP1R3D (proteína fosfatasa 1, subunidad reguladora (inhibidora) 3D), PTPN1 (proteína tirosina fosfatasa, no-receptor tipo 1), RBM39 (proteína 39 del motivo de unión al ARN), TAF4 (TAF4 ARN polimerasa II, factor asociado a la proteína de unión de la caja TATA (TBP), 135 kDa), y TCFL5 (factor de transcripción tipo 5 (hélice-bucle-hélice básico)).

Como se utiliza en el presente documento la expresión "el tercer grupo genético" es un término general para genes a los que se hace referencia biológicamente como "genes relacionados con el estroma" o "genes relacionados con EMT". Específicamente, el tercer grupo genético consiste en genes representados por los siguientes símbolos genéticos: ANGPTL2 (angiopoyetina tipo 2), AXL (receptor AXL de tirosina cinasa), C1R (componente 1 del complemento, subcomponente s), CALHM2 (modulador 2 de la homeostasis del calcio), CTSK (catepsina K), DCN (Decorina), EMP3 (proteína 3 de la membrana epitelial), GREM1 (gremlina 1, superfamilia del nudo de cisteína, homólogo (Xenopus laevis)), ITGAV (integrina, alfa V (receptor de vitronectina, polipéptido alfa, antígeno CD51)), KLHL5 (kelch tipo 5 (Drosophila)), MMP2 (metalopeptidasa 2 de la matriz (gelatinasa A, gelatinasa de 72 kDa, colagenasa tipo IV de 72 kDa)), RAB34 (RAB34, miembro de la familia RAS de oncogenes), SELM (selenoproteína M), SRGAP2P1 (SLIT-ROBO Rho pseudogén 1 de la proteína 2 activadora de la GTPasa), y VIM (Vimentina). Como se utiliza en el presente documento, se puede hacer referencia al tercer grupo genético como "grupo genético relacionado con el estroma".

El método de apoyo al diagnóstico descrito en el presente documento determina el riesgo de recurrencia de cáncer colorrectal utilizando los tres grupos genéticos.

60

Como se utiliza en el presente documento, la expresión "productos de la transcripción de genes" significa los productos obtenidos por la transcripción de los genes, incluyen un ácido ribonucleico (ARN), específicamente, el ARN mensajero (ARNm).

65 Como se utiliza en el presente documento, la expresión "el nivel de expresión de los productos de la transcripción de genes" significa la cantidad de productos de transcripción genética en la muestra biológica o la cantidad de

sustancias que reflejan la cantidad de los productos de transcripción genética en la muestra biológica. Por lo tanto, el método de apoyo al diagnóstico descrito en el presente documento puede medir la cantidad de productos de la transcripción genética (ARNm) o la cantidad de ácidos desoxirribonucleicos complementarios (ADNc) o ácidos ribonucleicos complementarios (ARNc) derivados de los ARNm. En general, la cantidad de ARNm en una muestra biológica es muy pequeña. Por lo tanto, se mide preferentemente la cantidad de ADNc o ARNc derivada del mismo por transcripción inversa o transcripción *in vitro* (IVT).

Los productos de la transcripción genética se pueden extraer de la muestra biológica mediante un método de extracción de ARN conocido en la técnica. Por ejemplo, se puede obtener un extracto de ARN mediante un procedimiento que incluye la centrifugación de la muestra biológica para precipitar las células que contienen el ARN, destruir física o enzimáticamente las células y retirar los desechos celulares. La extracción de ARN también se puede llevar a cabo utilizando un kit de extracción de ARN disponible en el mercado o similar.

También se puede llevar a cabo un tratamiento para la retirada de un contaminante del extracto de producto de la transcripción genética como se ha descrito anteriormente. Dicho contaminante, es el ARNm de globina cuando la muestra biológica es sangre, se deriva de la muestra biológica y preferentemente estará ausente de la medición de los niveles de expresión de los productos de la transcripción genética.

El extracto de producto de la transcripción genética resultante se mide en cuanto a los niveles de expresión de una pluralidad (dos o más) de genes seleccionados de entre el primer al tercer grupo genético. Se pueden medir al menos 5 o más de los genes de cada uno de los grupos; o al menos 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 o si fuera apropiado se pueden medir al menos 17, 18, 19, 20, así como todos los genes. El número exacto de la pluralidad de genes medidos no tiene que ser el mismo para cada grupo genético. Para cada grupo, se pueden medir los niveles de expresión de cinco o más genes, por lo que se pueden reducir las variaciones biológicas o los errores de medición que se producen cuando la expresión de los genes predeterminados es inesperadamente alta o baja. En consecuencia, el diagnóstico de un riesgo de recurrencia puede apoyarse con una fiabilidad más alta.

Los niveles de expresión de los genes se pueden medir por cualquier método conocido en sí mismo. En el método de apoyo al diagnóstico de un riesgo de recurrencia de cáncer colorrectal descrito en el presente documento, la medición se lleva a cabo preferentemente mediante un ensayo en chip de ácido nucleico, es decir, un ensayo denominado de micromatriz.

30

35

55

65

Cuando los niveles de expresión de los productos de transcripción genética se miden utilizando una micromatriz, el procedimiento puede incluir: poner en contacto los ADNc o ARNc, que se preparan a partir del extracto de producto de transcripción genética o los productos de la transcripción genética con sondas de ácido nucleico de aproximadamente 20 a 25 meros fijadas en un sustrato; y medir el cambio de fluorescencia, coloración, corriente, o cualquier otro índice para determinar la presencia o ausencia de hibridación, de manera que se puedan determinar los niveles de expresión de los productos diana de la transcripción genética.

- 40 Se puede utilizar al menos una sonda de ácido nucleico para un producto de la transcripción genética, y se puede utilizar una pluralidad de sondas dependiendo de la longitud del producto de la transcripción genética. La secuencia de la sonda puede determinarse apropiadamente por un experto en la técnica de acuerdo con la secuencia del producto de la transcripción genética que se va a medir.
- Por ejemplo, se puede utilizar el sistema GeneChip System disponible en Affymetrix, Inc., en el método de medición de niveles de expresión de los productos de la transcripción genética utilizando un chip de ácido nucleico.

Cuando se utiliza un chip de ácido nucleico, los productos de la transcripción genética o los ADNc o ARNc de los mismos se pueden fragmentar de manera que se pueda facilitar la hibridación con las sondas de ácido nucleico. La fragmentación se puede llevar a cabo por cualquier método conocido en la técnica, tal como un método que utilice una nucleasa tal como una ribonucleasa o una desoxirribonucleasa.

En la etapa de medición, la medición de los niveles de expresión de la pluralidad de genes se puede llevar a cabo respectivamente para cada gen o se puede llevar a cabo simultáneamente para algunos o todos los genes. Por ejemplo, cuando se utiliza un chip de ácido nucleico, la primera medición, la segunda medición y la tercera medición se pueden llevar a cabo simultáneamente en un único chip de ácido nucleico.

Las cantidades de producto de la transcripción genética o los ADNc o ARNc de los mismos que se ponen en contacto con las sondas de ácido nucleico en el chip de ácido nucleico pueden ser en general desde aproximadamente 5 µg a 20 µg. Las condiciones de contacto son generalmente de 45 °C durante aproximadamente 16 horas.

Si los productos de la transcripción genética o los ADNc o ARNc de los mismos se hibridan o no y cuánto con las sondas de ácido nucleico se puede detectar utilizando una sustancia fluorescente o un colorante o basándose en un cambio inducido por la hibridación en la cantidad de flujo de corriente en el chip de ácido nucleico.

Cuando se mide la hibridación mediante la detección de una sustancia fluorescente o un colorante, los productos de la transcripción genética o ADNc o ARN de los mismos se marcan preferentemente con una sustancia marcadora para la detección de la sustancia fluorescente o el colorante. Los ejemplos de la sustancia marcadora pueden incluir sustancias marcadoras que se utilizan en general en la técnica. En general, se puede mezclar un nucleótido biotinilado o un ribonucleótido biotinilado como sustrato de nucleótido o ribonucleótido en la síntesis de los ADNc o ARNc de manera que se puedan obtener ADNc o ARNc marcados con biotina. Los ADNc o ARNc marcados con biotina se pueden acoplar con avidina o estreptavidina, que es una pareja de unión con la biotina, en el chip de ácido nucleico. La unión de avidina o estreptavidina a una sustancia fluorescente o colorante hace posible detectar la hibridación. Ejemplos de una sustancia fluorescente incluyen el isotiocianato de fluoresceína (FITC), la proteína fluorescente verde (GFP), luciferina, y ficoeritrina. En general, está disponible en el mercado un conjugado ficoeritrina-estreptavidina y por lo tanto se utiliza convenientemente.

De manera alternativa, se pone en contacto un anticuerpo marcado contra la avidina o estreptavidina, con la avidina o estreptavidina de manera que se pueda detectar la sustancia fluorescente o el colorante del anticuerpo marcado.

No existe una limitación particular para los niveles de expresión de los productos de la transcripción genética obtenidos en esta etapa a condición de que puedan ser valores que indiquen relativamente la cantidad de cada producto de la transcripción genética en la muestra biológica. Cuando se lleva a cabo la medición utilizando el chip de ácido nucleico, los niveles de expresión pueden ser señales obtenidas del chip de ácido nucleico, que se basen en la intensidad de fluorescencia, la intensidad de coloración, la cantidad de corriente, y similares.

Dichas señales se pueden medir utilizando un analizador de chips de ácido nucleico.

10

15

20

Los datos de los niveles de expresión de los genes obtenidos de las muestras biológicas recolectadas de una pluralidad de pacientes con cáncer colorrectal se pueden agrupar mediante un agrupamiento. El procedimiento de agrupamiento se puede llevar a cabo, por ejemplo, mediante el método de la distancia de la aproximación más estrecha. El método de la distancia de la aproximación más estrecha es un método de agrupamiento jerárquico que incluye las etapas de: calcular una distancia entre los elementos y la unión secuencial de los elementos cercanos entre ellos con el fin de llevar a cabo un agrupamiento paso a paso. Como se utiliza en el presente documento, una "muestra biológica" se corresponde con un "elemento". Una "diferencia entre los niveles de expresión de genes" entre los "elementos" se corresponde con una "distancia".

En la etapa de determinación, se pueden emplear distintos métodos de análisis.

- Por ejemplo, la determinación se lleva a cabo basándose en la comparación del nivel de expresión de los genes seleccionados del primer grupo genético en la muestra biológica con respecto a un primer valor de referencia, la comparación del nivel de expresión de los genes seleccionados de entre el segundo grupo genético en la muestra biológica con respecto a un segundo valor de referencia, y la comparación del nivel de expresión de los genes seleccionados de entre el tercer grupo genético en la muestra biológica respecto a un tercer valor de referencia de los genes. En la presente divulgación, un valor de referencia es un valor de un umbral o en otras palabras un valor de corte. El nivel de expresión se puede calcular a partir de los niveles de expresión de genes. El nivel de expresión puede ser una media, un valor de la mediana o un valor modal de los niveles de expresión de los genes.
- Posteriormente, en la etapa de determinación de los métodos proporcionados en el presente documento, se determina el riesgo de recurrencia de un cáncer colorrectal basándose en los datos del nivel de expresión de los genes obtenidos en la etapa de medición. En particular, el riesgo de recurrencia de un cáncer colorrectal se determina basándose en los datos de los niveles de expresión solo de la pluralidad de genes de los grupos genéticos identificados anteriormente.
- 50 Por ejemplo, el riesgo de recurrencia se determina que es alto cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es mayor o igual que un valor de referencia independientemente del nivel de expresión de los genes seleccionados de entre el primer y el segundo grupo genético en la etapa de determinación.
- En otras palabras, el riesgo de recurrencia se determina que es alto cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es mayor o igual que un valor de referencia, incluso aunque el nivel de expresión de los genes seleccionados de entre el primer y el segundo grupo genético sea mayor o igual que un valor de referencia, o sea menor que un valor de referencia.
- Por ejemplo, el riesgo de recurrencia se determina que es medio cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es menor que un valor de referencia y el nivel de expresión de los genes seleccionados de entre el primer grupo genético es menor que un valor de referencia independientemente del nivel de expresión de los genes seleccionados de entre el segundo grupo genético en la etapa de determinación.
- Por ejemplo, el riesgo de recurrencia se determina que es medio cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es menor que un valor de referencia y el nivel de expresión de los genes seleccionados de entre el primer grupo genético es menor que un valor de referencia, incluso aunque el nivel de

expresión de los genes seleccionados de entre el segundo grupo genético sea mayor o igual que un valor de referencia, o sea menor que un valor de referencia.

Por ejemplo, el riesgo de recurrencia se determina que es medio cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es menor que un valor de referencia, el nivel de expresión de los genes seleccionados de entre el primer grupo genético es mayor o igual que un valor de referencia de los genes, y el nivel de expresión de los genes seleccionados de entre el segundo grupo genético es mayor o igual que un valor de referencia en la etapa de determinación.

Se proporciona en el presente documento un método para apoyar un diagnóstico del riesgo de recurrencia de cáncer colorrectal, donde el riesgo de recurrencia se determina que es bajo cuando el nivel de expresión de los genes seleccionados del tercer grupo genético es menor que los valores de referencia, el nivel de expresión de los genes seleccionados de entre el primer grupo genético es mayor o igual que los valores de referencia de los genes, y el nivel de expresión de los genes seleccionados de entre el segundo grupo genético es menor que los valores de referencia en la etapa de determinación.

Como se utiliza en el presente documento, el "valor de referencia" que se fija para cada grupo genético es un valor que puede determinar si los genes del grupo genético están sobre expresados. El "valor de referencia" para el primer grupo genético, por ejemplo, se puede obtener de la siguiente manera. Primero, en un grupo de pacientes particular, se calcula el valor medio de los niveles de expresión de los genes. Concretamente, midiendo los niveles de expresión de C18orf22 en cada paciente en el grupo de pacientes y luego dividiendo la suma de los niveles de expresión por el número de pacientes, se puede obtener el valor medio de los niveles de expresión de C18orf22 en el grupo de pacientes. De la misma manera, se pueden obtener los valores medios de los otros genes en el primer grupo genético. Dividiendo la suma de los valores medios de los niveles de expresión de los genes por el número de genes, se puede obtener el valor medio del primer grupo genético en el grupo de pacientes particular. El valor medio resultante se puede utilizar como el "valor de referencia". De la misma manera, también se pueden obtener los "valores de referencia" para el segundo grupo genético y el tercer grupo genético.

20

25

30

40

45

50

55

60

65

Aunque el valor de referencia ejemplificado anteriormente es el "valor medio", se puede utilizar a su vez otro valor distinto del valor medio, tal como el valor de la mediana o el valor modal.

El valor de referencia se obtiene preferentemente antes de la etapa de medición y la etapa de determinación.

Preferentemente, la suma de los niveles de expresión de los genes en una muestra biológica se divide por el número de los genes para dar el valor medio de los niveles de expresión de los genes en la muestra biológica, y el valor medio resultante de los niveles de expresión de los genes en la muestra biológica se compara con el valor de referencia descrito anteriormente.

Por lo tanto, se entenderá que el "valor de referencia" puede fijarse apropiadamente para cada grupo genético a partir de los datos acumulados de los niveles de expresión de los genes seleccionados de entre el primer al tercer grupo genético. Por ejemplo, los valores de referencia son valores medios de los datos de los niveles de expresión de los genes en una pluralidad de pacientes. Según aumenta el número de pacientes diana (n), los valores medios de los datos de los niveles de expresión convergen. Por lo tanto, es posible reducir las variaciones de los valores medios

El riesgo de recurrencia se puede determinar también basándose en un patrón de expresión de genes en una muestra biológica y patrones de expresión de genes en grupos de pacientes. En este método, en vez de establecer una pluralidad de "grupos genéticos", se establece una pluralidad de grupos de pacientes de acuerdo con el riesgo de recurrencia. Específicamente, un grupo de pacientes se clasifica en tres grupos, a saber, un grupo de paciente que se determina que tiene un alto riesgo de recurrencia (también llamado, de aquí en adelante, "grupo de alto riesgo"), un grupo de pacientes que se determina que tiene un riesgo medio de recurrencia (también llamado, de aquí en adelante, "grupo de recurrencia (también llamado, de aquí en adelante, "grupo de bajo riesgo").

Los niveles de expresión de los genes que se analizan en las muestras de cada grupo de pacientes se obtienen, y luego se calculan los valores medios. Cuando el grupo de alto riesgo tiene 100 pacientes y con el fin de calcular el valor medio de los niveles de expresión de C18orf22, por ejemplo, la suma de los niveles de expresión de C18orf22 en los 100 pacientes se divide por 100 y el valor resultante es el valor medio de los niveles de expresión de C18orf22 en el grupo de alto riesgo. De la misma manera, se calculan los valores medios de los niveles de expresión de C18orf22 en el grupo de riesgo medio y el grupo de bajo riesgo. Se lleva a cabo el análisis en una pluralidad de genes y por lo tanto se calcula el valor medio de los niveles de expresión de cada uno de la pluralidad de genes. No se calcula el valor medio de los valores medios. Un conjunto de datos está compuesto de los valores medios de los niveles de expresión obtenidos de esta manera para el grupo de alto riesgo se refiere a un patrón de expresión obtenidos de esta manera para el grupo de riesgo medio, y un conjunto de datos que está compuesto por los valores medios de los niveles de expresión obtenidos de medio, y un conjunto de datos que está compuesto por los valores medios de los niveles de expresión obtenidos de

esta manera para el grupo de bajo riesgo es refiere a un patrón de expresión del grupo de bajo riesgo. Si se analizan 55 genes, el patón de expresión de cada grupo de riesgo incluye 55 valores.

Los patrones de expresión para los grupos de riesgo se obtienen antes de la medición de la expresión de los genes 5 en la muestra biológica y la determinación del riesgo de recurrencia.

Entonces se miden los niveles de expresión de los genes en la muestra biológica. Un conjunto de datos que está compuesto de los niveles de expresión de los genes medidos en la etapa de medición hace referencia al patrón biológico de la muestra biológica. Si se analizan 55 genes, el patón de expresión de la muestra biológica incluye 55 valores.

Se analiza la correlación entre el patrón de expresión para los genes en la muestra biológica y el patrón de expresión de cada grupo de riesgo. Se identifica un grupo de riesgo que presenta la correlación más alta con el patrón de expresión de la muestra biológica. El riesgo de recurrencia de la muestra biológica se determina que es el riesgo de recurrencia correspondiente con el grupo de riesgo identificado de esta manera. En el caso donde los niveles de expresión de los genes en la muestra biológica presenten la correlación más alta con el grupo de alto riesgo, por ejemplo, la muestra biológica se determina como que tiene un alto riesgo de recurrencia.

En el análisis de correlación anterior, se pueden emplear distintos métodos. La determinación del riesgo de recurrencia se puede hacer, por ejemplo, calculando un coeficiente de correlación entre el patrón de expresión de la muestra biológica y el patrón de expresión del grupo de alto riesgo, un coeficiente de correlación entre el patrón de expresión de la muestra biológica y el patrón de expresión del grupo de riesgo medio, y un coeficiente de correlación entre el patrón de expresión de la muestra biológica y el patrón de expresión del grupo de bajo riesgo, comparando los coeficientes de correlación resultantes entre ellos y clasificando la muestra biológica en el grupo de riesgo que presente el mayor coeficiente de correlación. En el caso donde el coeficiente de correlación con el patrón de expresión para el grupo de alto riesgo, por ejemplo, la muestra biológica se clasifica en el grupo de alto riesgo y se determina que tienen un alto riesgo de recurrencia.

El cálculo de los coeficientes de correlación se puede llevar a cabo por un método conocido. Por ejemplo, se pueden calcular los coeficientes de correlación de acuerdo con la correlación de Spearman, correlación del momento-producto de Pearson, la correlación de intervalos de Kendall, o similares.

Como método de análisis de la correlación con los grupos de riesgo, se puede emplear también el análisis de agrupamiento. El análisis se puede llevar, por ejemplo, de la siguiente manera.

Los valores de los niveles de expresión de genes en cada uno de una pluralidad de pacientes se obtienen con antelación (en este punto, aún no se ha llevado a cabo la clasificación en grupo de alto riesgo, grupo de riesgo medio, y grupo de bajo riesgo). En la etapa de medición, se miden los niveles de expresión de los genes en una muestra biológica. El análisis de agrupamiento se lleva a cabo para clasificar los pacientes y la muestra biológica en un grupo de alto riesgo, un grupo de riesgo medio, y un grupo de bajo riesgo basándose en los niveles de expresión de los genes en los pacientes y los niveles de expresión de los genes en la muestra biológica. Basándose en el grupo de riesgo en el que se clasifica la muestra biológica, se puede determinar el riesgo de recurrencia para la muestra biológica.

En vez de las técnicas de análisis descritas anteriormente, se pueden utilizar un análisis de discriminación lineal, máquina vectorial de soporte, o similares.

En cualquiera de los distintos métodos de análisis descritos anteriormente, la determinación se hará de la siguiente manera.

Cuando el nivel de expresión de los genes que pertenecen al tercer grupo genético es alto, se determina que el riesgo de recurrencia es alto.

Cuando el nivel de expresión de los genes que pertenecen al primer grupo genético es bajo y los niveles de expresión de los genes que pertenecen al tercer grupo genético son bajos, se determina que el riesgo de recurrencia es medio.

Cuando el nivel de expresión de los genes que pertenecen al primer grupo genético es alto, el nivel de expresión de los genes que pertenecen al segundo grupo es alto, y el nivel de expresión de los genes que pertenecen al tercer grupo genético es bajo, se determina que el riesgo de recurrencia es medio.

Cuando el nivel de expresión de los genes que pertenecen al primer grupo genético es alto, el nivel de expresión de los genes que pertenecen al segundo grupo es bajo, y el nivel de expresión de los genes que pertenecen al tercer grupo genético es bajo, se determina que el riesgo de recurrencia es bajo.

En otras palabras,

65

60

10

15

35

40

la expresión regulada positivamente de los genes que pertenecen al tercer grupo genético indica alto riesgo, la expresión regulada negativamente de los genes que pertenecen al primer grupo genético y la expresión regulada negativamente de los genes que pertenecen al tercer grupo genético indica un riesgo medio, y la expresión regulada positivamente de los genes que pertenecen al primer grupo genético y la expresión regulada negativamente de los genes que pertenecen al tercer grupo genético indica un riesgo medio, y la expresión regulada positivamente de los genes que pertenecen al primer grupo genético, la expresión regulada negativamente de los genes que pertenecen al segundo grupo genético, y la expresión regulada negativamente de los genes que pertenecen al tercer grupo genético indica un riesgo bajo.

- 10 En una realización más preferida de la presente invención, el riesgo de recurrencia se determina que es alto cuando el grupo determinado que tiene un riesgo medio de recurrencia tiene una mutación del gen KRAS, y el riesgo de recurrencia se determina que es bajo cuando el grupo no tiene la mutación del gen KRAS.
- El gen KRAS es un tipo de oncogén ras, que está presente en una posición de 25,36 Mb a 25,4 Mb en el cromosoma 12. El gen transmite la señal del receptor del factor de crecimiento epidérmico (EGFR) al núcleo y tiene la función de facilitar la proliferación de las células. La secuencia de bases del ADNc de KRAS se representa como SEQ ID NO: 56. Esta secuencia de bases se conoce como el número de registro: AF493917 en la base de datos del genoma humano GenBank.
- 20 La mutación del gen KRAS indica preferentemente una mutación que se produce en la secuencia de bases GGTGGC que se corresponde con el 12º y 13º codón localizado en el exón 2 del gen (bases 34 a 39) o una mutación que se produce en la secuencia de bases CAA que se corresponde con el 61º codón localizado en el exón 3 (bases 182 a 184).
- No existe ninguna limitación particular como método para la medición de la presencia de la mutación KRAS y el método se puede llevar a cabo por cualquier método conocido por el experto en la técnica. La medición de la presencia de la mutación KRAS se puede llevar a cabo utilizando un análisis de secuencia.
- No existe una limitación particular en cuento a la mutación del gen KRAS. Cuando se mutan cualquiera de las bases de los codones anteriores, se puede determinar la presencia de la mutación. Es preferible que el objetivo sea una mutación de la secuencia de bases que dé como resultado la mutación de la secuencia de aminoácidos de la proteína KRAS (es decir, mutaciones distintas de mutaciones silentes, por ejemplo, una mutación en mal sentido, mutación sin sentido, y mutación de cambio de fase). El tipo de mutación se considera que es una sustitución, deficiencia, eliminación o adición de nucleótidos. En la presente realización, se prefiere la sustitución. Ejemplos específicos de dicha sustitución incluyen la Sustitución de G34 por A, la Sustitución de G35 por A, C o T, la sustitución de G38 por A, la sustitución de C182 por A, y la sustitución de A184 por C o T.
- Como se ha descrito anteriormente, la presencia de la mutación KRAS se añade a los criterios de determinación de manera que el grupo de riesgo medio se puede clasificar en un grupo de alto riesgo y un grupo de bajo riesgo y todo se puede clasificar en los dos grupos de alto y bajo riesgo. En consecuencia, el grupo de riesgo medio se puede clasificar en el grupo alto y el grupo bajo, y se puede proporcionar una información útil para muchos casos.
- También se desvela en el presente documento un programa de computadora para hacer que una computadora ejecute el procesamiento de determinación de un riesgo de recurrencia de cáncer colorrectal en un paciente.

  Ejemplos del producto del programa de computadora incluyen programas descargables mediante internet y medios que almacenen los programas.
  - Por ejemplo, se ejemplifican los programas para hacer que la computadora ejecute el siguiente procesamiento.
- El procesamiento de la recepción, en una muestra biológica recolectada de un paciente con un cáncer colorrectal, de los niveles de expresión de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles de expresión de un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo
- PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y el procesamiento de la determinación del riesgo de recurrencia de cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos.
- De aquí en adelante, se describirá un aparato ejemplar para ejecutar el método como se describe en el presente documento con referencia a los dibujos. La Fig. 1 es una vista esquemática que muestra un ejemplo de un aparato de apoyo al diagnóstico que se utiliza para determinar un riesgo de recurrencia de cáncer colorrectal en un paciente. El aparato de apoyo al diagnóstico 1 que se muestra en la Fig. 1 incluye un dispositivo de medición 2 y un sistema de computadora 3 conectado al dispositivo de medición 2.
  - En el ejemplo, el dispositivo de medición 2 es un dispositivo de medición para el chip de ácido nucleico. El

dispositivo de medición 2 obtiene los niveles de expresión propia de los genes y la información de los niveles de expresión de los genes, tal como la tonalidad de la fluorescencia coloreada del chip de ácido nucleico o la intensidad de fluorescencia. Cuando se fija la muestra biológica recolectada de un paciente con cáncer colorrectal en el dispositivo de medición 2, el dispositivo de medición 2 obtiene la información de los niveles de expresión de los genes de la muestra biológica y transmite la información obtenida al sistema de computadora 3.

Cuando se determina que el espécimen tiene un riesgo medio, se determina si el riesgo de recurrencia de cáncer colorrectal es alto o bajo, el aparato de apoyo al diagnóstico 1 incluye adicionalmente un dispositivo de medición de mutación 4, además del sistema de computadora 3, conectado al dispositivo de medición 2 y el dispositivo de medición 2.

10

15

20

25

30

35

40

45

50

55

60

65

En el ejemplo, el dispositivo de medición de la mutación 4 obtiene la presencia de la mutación del gen KRAS en la muestra biológica. Cuando se fija la muestra biológica recolectada de un paciente con cáncer colorrectal en el dispositivo de medición 4, el dispositivo de medición 4 obtiene la información de la presencia de la mutación del gen KRAS en la muestra biológica y transmite la información obtenida al sistema de computadora 3.

El sistema de computadora 3 incluye un cuerpo principal 3a de la computadora, una unidad de entrada 3b que incluye un teclado y un ratón, y una unidad de pantalla 3c que incluye un LCD o CRT para presentar la información del espécimen, el resultado de la determinación o similar en ella. El sistema de computadora 3 recibe la información de los niveles de expresión de los genes y la información de la presencia de la mutación del gen KRAS (si fuera necesario) a partir del dispositivo de medición 2 y el dispositivo de medición de mutación 4. Entonces, el sistema de computadora 3 ejecuta un programa para determinar un riesgo de recurrencia de cáncer colorrectal en un sujeto basándose en la información. Los datos de "clasificación en dos grupos es necesaria" descrita posteriormente puede introducirse a partir de la unidad de entrada 3b.

La Fig. 2 es un diagrama de bloques que muestra un bloque funcional del software del cuerpo principal 3a de la computadora del aparato de ayuda al diagnóstico 1. Como se muestra en la Fig. 2, la computadora incluye una unidad de recepción 301, una unidad de almacenamiento 302, una unidad de cálculo 303, una unidad de determinación 304, y una unidad de salida 305. La unidad de recepción 301 está conectada comunicándose con el dispositivo de medición 2 y el dispositivo de medición de mutación 4 (si fuera necesario) mediante la red. La información necesaria para ejecutar la determinación del riesgo de recurrencia de un cáncer colorrectal, específicamente la información de si la medición de la presencia o no de la mutación del gen KRAS en el espécimen que se ha determinado que tiene un riesgo medio (clasificación en dos grupos), se puede introducir en la unidad de determinación 304 a través de la unidad de entrada 3b.

La unidad de recepción 301 recibe la información transmitida desde el dispositivo de medición 2 y el dispositivo de medición de mutación 4. La unidad de almacenaje 302 almacena los valores de referencia necesarios para la determinación, fórmulas para el cálculo de los niveles de expresión de los genes, y el procesamiento de los programas. La unidad de cálculo 303 calcula los niveles de expresión de los genes de acuerdo con las fórmulas almacenadas utilizando la información obtenida por la unidad de recepción 301. La unidad de determinación 304 determina si los niveles de expresión de los genes que se han obtenido por la unidad de recepción 301 o calculados por la unidad de cálculo 303 y mayores o iguales que los valores de referencia almacenados en la unidad de almacenamiento 302. La unidad de salida 305 transmite el resultado de determinación de la unidad de determinación 304 (como el resultado de la determinación del riesgo de recurrencia de cáncer colorrectal en el sujeto) a la unidad de pantalla 3c.

Cuando se determina que el espécimen tiene un riesgo medio, se determina si el riesgo de recurrencia de cáncer colorrectal es alto o bajo, la unidad de recepción 301 obtiene la información transmitida desde el dispositivo de medición 4, además de la información transmitida desde el dispositivo de medición 2. La unidad de almacenamiento 302 almacena secuencias no mutantes en el gen KRAS, además de los valores de referencia necesarios para la determinación, y las fórmulas para calcular los niveles de expresión de los genes. La unidad de cálculo 303 calcula los niveles de expresión de los genes de acuerdo con las fórmulas almacenadas utilizando la información obtenida por la unidad de recepción 301. La unidad de determinación 304 determina si los niveles de expresión de los genes que se han obtenido por la unidad de recepción 301 o calculados por la unidad de cálculo 303 y mayores o iguales que los valores de referencia almacenados en la unidad de almacenamiento 302. Adicionalmente, la unidad de determinación 304 determina la presencia de la mutación del gen KRAS basándose en si las secuencias del gen KRAS obtenidas por la unidad de recepción 301 coinciden o no con las secuencias no mutantes del gen KRAS almacenado en la unidad de almacenamiento 302. La unidad de salida 305 transmite el resultado de determinación de la unidad de determinación 304 (como el resultado de la determinación del riesgo de recurrencia de cáncer colorrectal en el sujeto) a la unidad de pantalla 3c.

La Fig. 3 es un diagrama de bloques que muestra la configuración del hardware del cuerpo principal 3a de la computadora que se muestra en la Fig. 2. Como se muestra en la Fig. 3, el cuerpo principal 3a de la computadora incluye una CPU (Unidad central de procesamiento) 30, una ROM (Memoria solo de lectura) 31, una RAM 32, un disco duro 33, una interfaz entrada/salida 34, un dispositivo de lectura 35, una interfaz de comunicaciones 36 y una interfaz de salida de imágenes 37. La CPU 30, la ROM 31, la RAM (memoria de acceso aleatorio) 32, el disco duro

33, la interfaz de entrada/salida 34, el dispositivo de lectura 35, la interfaz de comunicaciones 36 y la interfaz de salida de imágenes 37 se conectan entre ellos mediante un bus 38 para permitir la comunicación de datos.

La CPU 30 puede ejecutar el programa de computadora almacenado en la ROM 31 y el programa de computadora cargado en la RAM 32. La CPU 30 ejecuta los programas de computadora, de esta manera se ejecutan cada una de las funciones que se muestran en la Fig. 2. En consecuencia, el sistema de computadora 3 funciona como un aparato de apoyo al diagnóstico para determinar un riesgo de recurrencia del cáncer colorrectal en un sujeto.

La ROM 31 se configura para que incluya una ROM oculta, PROM, EPROM, EEPROM o similares. La ROM 31 almacena el programa de computadora que se va a ejecutar por la CPU 30, como se ha descrito anteriormente y los datos utilizados para la ejecución.

La RAM 32 se configura para que incluye una SRAM, DRAM o similar. La RAM 32 se utiliza para la lectura de los programas de computadora almacenados en la ROM 31 y el disco duro 33. Cuando se ejecutan estos programas de computadora, la RAM 32 se utiliza como el área de trabajo de la CPU 30.

Los programas de computadora que se van a ejecutar por la CPU 30, tal como un sistema operativo y un programa de una aplicación (programa de computadora para la determinación del riesgo de recurrencia de un cáncer colorrectal en un sujeto), y los datos que se van a utilizar para la ejecución del programa de computadora se instalan en el disco duro 33.

El dispositivo de lectura 35 se configura para que incluya una unidad de disco flexible, una unidad de CD-ROM, una unidad de DVD-ROM, y similares. El dispositivo de lectura 35 puede leer el programa de computadora o los datos almacenados en un medio de grabación 40 portátil.

Por ejemplo, la interfaz de entrada/salida 34 se configura para que incluya una interfaz en serie tal como un USB, IEEE 1394 o RS-232C, una interfaz paralela tal como una SCSI, IDE o IEEE 1284, y una interfaz análoga que incluya un convertidor D/A, un convertidor A/D o similares. La unidad de entrada 3b que incluye un teclado y un ratón se conecta en la interfaz de entrada/salida 34. Un operador puede introducir distintas instrucciones en el cuerpo principal de la computadora 3a utilizando la unidad de entrada 3b.

La interfaz de comunicaciones 36 es, por ejemplo, una interfaz de Ethernet (marca registrada) o similar. El cuerpo principal de la computadora 3a puede enviar datos de impresión a una impresora mediante la interfaz de comunicaciones 36.

La interfaz de salida de imágenes 37 se conecta a una unidad de pantalla 3c que incluye una LCD, CRT o similar. De esta manera, se puede enviar una señal de imagen que se corresponde con los datos de imagen desde la CPU 30 a la unidad de pantalla 3c. La unidad de pantalla 3c presenta una imagen (en la pantalla) de acuerdo con la entrada de la señal de imagen.

Posteriormente, se describirá un procedimiento para la determinación del riesgo de recurrencia de un cáncer colorrectal en un sujeto mediante el aparato de apoyo al diagnóstico 1. La Fig. 4 es un diagrama de flujo de la determinación del riesgo de recurrencia de cáncer colorrectal. Un ejemplo que se describe en el presente documento es el caso donde se calcula la intensidad de fluorescencia a partir de la información de la fluorescencia coloreada obtenida utilizando una muestra biológica de un sujeto, se calculan los niveles de expresión de los genes a partir de la intensidad de fluorescencia obtenida, y se determina si los niveles de expresión obtenidos son mayores o iguales que los valores de referencia.

Primero, la unidad de recepción 301 del aparato de apoyo al diagnóstico 1 obtiene la información de la fluorescencia coloreada con respecto a los niveles de expresión de los genes seleccionados del tercer grupo genético a partir del dispositivo de medición 2 (etapa S1-1). A continuación, la unidad de cálculo 303 calcula la intensidad de fluorescencia de la información obtenida y la transmite a la unidad de almacenamiento 302 (etapa S1-2). La unidad de cálculo 303 calcula los niveles de expresión de los genes de acuerdo con las fórmulas almacenadas utilizando la intensidad de fluorescencia almacenada (etapa S1-3).

A continuación, la unidad de determinación 304 determina si los niveles de expresión calculados en la etapa S1-3 son mayores o iguales que los valores de referencia almacenados en la unidad de almacenamiento 302 (etapa S1-4). Cuando los niveles de expresión son mayores o iguales que los valores de referencia, la rutina continúa a la etapa S1-5. La unidad de determinación 304 transmite el resultado de la determinación indicando que el riesgo de recurrencia del cáncer colorrectal en el sujeto es alto (alto riesgo) a la unidad de salida 305. Por otra parte, cuando los niveles de expresión son menores que los valores de referencia, la rutina continúa a la etapa S1-6.

La unidad de recepción 301 del aparato de apoyo al diagnóstico 1 obtiene la información de la fluorescencia coloreada con respecto a los niveles de expresión de los genes seleccionados del primer grupo genético a partir del dispositivo de medición 2 (etapa S1-6). La unidad de cálculo 303 calcula la intensidad de fluorescencia de la información obtenida y la transmite a la unidad de almacenamiento 302 (etapa S1-7). La unidad de cálculo 303

12

40

10

15

20

25

30

35

50

45

55

60

calcula los niveles de expresión de los genes de acuerdo con las fórmulas almacenadas utilizando la intensidad de fluorescencia almacenada (etapa S1-8).

A continuación, la unidad de determinación 304 determina si los niveles de expresión calculados en la unidad de cálculo 303 son mayores o iguales que los valores de referencia almacenados en la unidad de almacenamiento 302 (etapa S1-9). Cuando los niveles de expresión son mayores o iguales que los valores de referencia, la rutina continúa a la etapa S1-11. Cuando los niveles de expresión son menores que los valores de referencia, la rutina continúa a la etapa S1-10. La unidad de determinación 304 determina que el riesgo de recurrencia del cáncer colorrectal en el sujeto es medio (riesgo medio), y entonces la rutina continúa a la etapa S1-17.

10

15

La unidad de recepción 301 del aparato de apoyo al diagnóstico 1 obtiene la información de la fluorescencia coloreada con respecto a los niveles de expresión de los genes seleccionados del segundo grupo genético a partir del dispositivo de medición 2 (etapa S1-11). La unidad de cálculo 303 calcula la intensidad de fluorescencia de la información obtenida y la transmite a la unidad de almacenamiento 302 (etapa S1-12). La unidad de cálculo 303 calcula los niveles de expresión de los genes de acuerdo con las fórmulas almacenadas utilizando la intensidad de fluorescencia almacenada (etapa S1-13).

A continuación, la unidad de determinación 304 determina si los niveles de expresión calculados en la unidad de cálculo 303 son mayores o iguales que los valores de referencia almacenados en la unidad de almacenamiento 302 (etapa S1-14). Cuando los niveles de expresión son mayores o iguales que los valores de referencia, la rutina continúa a la etapa S1-15. La unidad de determinación 304 determina que el riesgo de recurrencia del cáncer colorrectal en el sujeto es medio (riesgo medio), y entonces la rutina continúa a la etapa S1-17. Por otra parte, cuando los niveles de expresión son menores que los valores de referencia de la etapa S1-14, la rutina continúa a la etapa S1-16. La unidad de determinación 304 transmite el resultado de la determinación indicando que el riesgo de recurrencia del cáncer colorrectal en el sujeto es bajo (bajo riesgo) a la unidad de salida 305.

Cuando se ha determinado que el espécimen tiene un riesgo medio de recurrencia del cáncer colorrectal mediante la etapa S1-10 o la etapa S1-15, los datos de "es necesaria la clasificación en dos grupos" se introducen desde la unidad de entrada 3b (etapa S1-17). En este caso, se determina en estos especímenes si el riesgo de recurrencia del cáncer colorrectal es alto o bajo basándose en la medición de la mutación del gen KRAS.

Cuando los datos de "es necesaria la clasificación en dos grupos" no se introduce, la rutina continúa a la etapa S1-18. El resultado de la determinación indicando que el riesgo de recurrencia del cáncer colorrectal en el sujeto es medio (riesgo medio) se transmite a la unidad de salida 305.

35

30

Por otra parte, cuando es necesaria la clasificación en dos grupos, la rutina continúa a la etapa S1-19. En el espécimen que se ha determinado que tiene un riesgo medio se determina si el riesgo de recurrencia de cáncer colorrectal es alto o bajo basándose en la presencia de la mutación del gen KRAS (etapa S1-19). Se utiliza el dispositivo de medición de la mutación 4 para llevar a cabo este procesamiento.

40

45

50

55

65

La unidad de recepción 301 obtiene la información de la secuencia del gen KRAS en el sujeto que se ha determinado que tiene un riesgo medio (etapa S1-19). Después, la unidad de determinación 304 compara la secuencia obtenida del gen KRAS con las secuencias no mutantes del gen KRAS almacenadas en la unidad de almacenamiento 302 y determina si está presente cualquier mutación en el gen KRAS en la muestra biológica del sujeto (etapa S1-20). Cuando el gen KRAS tiene una mutación, la rutina continúa a la etapa S1-21. La unidad de determinación 304 transmite el resultado de la determinación indicando que el riesgo de recurrencia del cáncer colorrectal en el sujeto es alto (alto riesgo) a la unidad de salida 305. Por otra parte, cuando el gen KRAS no tiene una mutación, la rutina continúa a la etapa S1-22. La unidad de determinación 304 transmite el resultado de la determinación indicando que el riesgo de recurrencia del cáncer colorrectal en el sujeto es bajo (bajo riesgo) a la unidad de salida 305.

La unidad de salida 305 transmite el resultado de la determinación del riesgo de recurrencia de cáncer colorrectal en el sujeto y permite a la unidad de pantalla 3c presentar el resultado de la determinación (etapa S1-23). En consecuencia, el aparato de apoyo al diagnóstico 1 puede proporcionar información para apoyar la determinación de si el riesgo de recurrencia del cáncer colorrectal en el sujeto es alto, medio o bajo a un médico o similar.

También se desvela en el presente documento sistema adecuado para la determinación de un riesgo de recurrencia de cáncer colorrectal en un sujeto.

60 La unidad de almacenamiento 302 almacena un programa de computadora para hacer que el sistema de computadora 3 ejecute el siguiente procesamiento:

la recepción, en una muestra biológica recolectada de un paciente con un cáncer colorrectal, de los niveles de expresión de una pluralidad de genes seleccionados de entre un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles expresión de

una pluralidad de genes seleccionados de entre un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de una pluralidad de genes seleccionados de entre un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y la determinación del riesgo de recurrencia del cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos.

El riesgo de recurrencia del cáncer colorrectal del sujeto se determina basándose en el resultado del análisis obtenido en la etapa de análisis anterior. Por ejemplo, es posible proporcionar un resultado de la determinación, tal como una posibilidad alta, media o baja del riesgo de recurrencia del cáncer colorrectal en el sujeto. La provisión del resultado de la determinación anterior ayuda a que el médico o similar diagnostique la posibilidad de un cáncer colorrectal.

#### 15 Ejemplos

5

20

25

30

35

45

50

55

#### Ejemplo 1:

Examen de la clasificación con respecto al pronóstico de los pacientes con cáncer colorrectal

<Método>

Entre los conjuntos de datos GSE14333 de la matriz U133 plus 2.0 del GeneChip del genoma humano (Affymetrix, Inc.) (obtenido en el Gene Expression Omnibus del NCBI (URL; http://www.ncbi.nlm.nih.gov/geo/),se utilizaron 72 casos de pacientes con cáncer colorrectal (cáncer de colon) como conjunto de entrenamiento. Se utilizaron el softwares de análisis de los datos de la matriz (Expression Console v1.1 (fabricado por Affymetrix, Inc.)), softwares de hoja de cálculo (Office Excel 2002, 2007 (fabricado por Microsoft)), software de análisis de agrupamiento (Cluster3.0, Java (marca comercial registrada)) Treeview (disponible en; http://bonsai. hgc.jp/~mdehoon/software/cluster/software.htm)), y software del análisis estadístico (MedCalc (fabricado por MedCalc)) para llevar a cabo distintos análisis.

Se utilizó MAS5 para normalizar los datos. Entre todas las sondas del GeneChip, las sondas cuyo símbolo genético no estaba claro y las sondas que tenían un valor medio de señal de la expresión menor de 300 se excluyeron del análisis. En las sondas en las que lo genes correspondientes se solapaban entre ellos, la sonda que tenía el máximo valor medio de señal de expresión se utilizó como representativa y el resto de las sondas se excluyeron. Después de la transformación Z de los valores de señal, se llevó a cabo un agrupamiento jerárquico no supervisado de acuerdo con el método de la distancia de aproximación más estrecha. La medida de similitud se definió como el coeficiente de correlación de Pearson.

A partir del resultado del análisis de agrupamiento, se consideró un agrupamiento genético el que satisfacía dos condiciones: (1) reflejo de funciones biológicas importantes; y (2) contribución a la generación de agrupamientos de casos característicos. El agrupamiento genético se definió como un módulo funcional y se extrajo. El método de clasificación de los grupos de riesgo de recurrencia se construyó llevando a cabo repetidamente el agrupamiento de acuerdo con una combinación de los módulos funcionales.

<Resultados>

La Fig. 5 muestra el resultado de la clasificación de los casos en un conjunto de entrenamiento en grupos de riesgo de recurrencia. Basándose en el valor medio de los niveles de expresión de los genes en todos los casos de pacientes, se determinarán a continuación las tasas de aumento-disminución de los niveles de expresión. Por ejemplo, cuando los niveles de expresión de ciertos genes son mayores o iguales que el valor medio anterior, se determina que los niveles de expresión han aumentado relativamente. Cuando los niveles de expresión de ciertos genes son menores que el valor medio anterior, se determina que los niveles de expresión han disminuido relativamente. En referencia a la Fig. 5, los casos que presentaban una disminución relativa del nivel de expresión del grupo genético en el brazo largo del cromosoma 18 (a los que se hace referencia de aquí en adelante como "primer grupo genético" o "módulo de pérdida 18q") o los casos que presentan un aumento relativo del nivel de expresión del grupo genético en el brazo largo del cromosoma 20 (a los que se hace referencia de aquí en adelante como "segundo grupo genético" o "'módulo Amp 20q") se extrajeron del conjunto de entrenamiento. Los casos se definieron como un tipo B. Casos que tenían patrones de expresión del módulo de pérdida 18q y los del módulo Amp 20q opuestos a los del tipo B se extrajeron del conjunto de entrenamiento. Los casos se definieron como un tipo A. En el conjunto de entrenamiento, los casos caracterizados por la fuerte expresión en el grupo genético relacionado con el estroma parecía que no tenían relación con los niveles de expresión de los genes de los tipos A y B. En consecuencia, estos casos se definieron como un tipo C independiente. La Tabla 1 muestra los genes que forman los tres módulos funcionales utilizados.

65

[Tabla 1- parte 1]

	10/ 1 1 1	[Tabla 1- parte 1]	lin	100010110
Módulo	Símbolo del gen	ID del conjunto de sondas		SEQ ID NO:
	C18orf22	219419_at	NM_024805	SEQ ID NO: 1
	C18orf55	223180_s_at	BC000892	SEQ ID NO: 2
	CCDC68	220180_at	NM_025214	SEQ ID NO: 3
	CNDP2	217752_s_at	NM_018235	SEQ ID NO: 4
	CYB5A	209366_x_at	M22865	SEQ ID NO: 5
	LOC400657	226924_at	AI016355	SEQ ID NO: 6
	LOC440498	232594_at	AK001829	SEQ ID NO: 7
	MBD2	202484_s_at	AF072242	SEQ ID NO: 8
	MBP	225407_at	N37023	SEQ ID NO: 9
Pérdida de	MY05B	225301_s_at	Al991160	SEQ ID NO: 10
18q	NARS	200027_at	NM_004539	SEQ ID NO: 11
	PQLC1	218208_at	NM_025078	SEQ ID NO: 12
	RTTN	227072_at	BG167480	SEQ ID NO: 13
	SEC11C	223299_ at	AF212233	SEQ ID NO: 14
	SOCS6	227542_at	AU157543	SEQ ID NO: 15
	TNFRSF11A	238846_at	AW026379	SEQ ID NO: 16
	TXNL1	201588 at	NM 004786	SEQ ID NO: 17
	TXNL4A	202836_s_at	NM_006701	SEQ ID NO: 18
	VPS4B	218171 at	AF195514	SEQ ID NO: 19
	ZNF407	227768 at	AB051490	SEQ ID NO: 20
	ASXL1	212237 at	N64780	SEQ ID NO: 21
	C20orf112	225224 at	AL034550	SEQ ID NO: 22
	C20orf177	225313 at	Al627538	SEQ ID NO: 23
	CHMP4B	225498 at	AV713673	SEQ ID NO: 24
	COMMD7	224815 at	AA148301	SEQ ID NO: 25
	CPNE1	206918 s at	NM 003915	SEQ ID NO: 26
	DID01	218325 s at	NM 022105	SEQ ID NO: 27
	DNAJC5	224611 s at	AL118506	SEQ ID NO: 28
	KIF3B	203943 at	NM 004798	SEQ ID NO: 29
Amp 20q	NCOA6	208979 at	AF128458	SEQ ID NO: 30
7 411P 20q	PHF20	209422 at	AL109965	SEQ ID NO: 31
	PIGU	225903 at	AL118520	SEQ ID NO: 32
	PLAGL2	202925 s at	NM 002657	SEQ ID NO: 33
	POFUT1	212349 at	AL045513	SEQ ID NO: 34
	PPP1R3D	204554 at	AL109928	SEQ ID NO: 35
	PTPN1	202716 at	NM 002827	SEQ ID NO: 36
	RBM39	207941 s at	NM 004902	SEQ ID NO: 37
	TAF4	213090 s at	AI744029	SEQ ID NO: 38
	TCFL5	204849 at	NM 006602	SEQ ID NO: 39
	ANGPTL2	213001 at	AF007150	SEQ ID NO: 40
	AXL	202686 s at	NM 021913	SEQ ID NO: 40
				SEQ ID NO: 42
	C1R C1S	212067_s_at	AL573058	
		208747_s_at	M18767	SEQ ID NO: 43
	CALHM2	57715_at	W72694	SEQ ID NO: 44
Grupo	CTSK	202450_s_at	NM_000396	SEQ ID NO: 45
genético	DCN	201893_x_at	AF138300	SEQ ID NO: 46
relacionado	EMP3	203729_at	NM_001425	SEQ ID NO: 47
con el	GREM1	218469_at	NM_013372	SEQ ID NO: 48
estroma	ITGAV	202351_at	AI093579	SEQ ID NO: 49
	KLHL5	226001_at	AK002174	SEQ ID NO: 50
	MMP2	201069_at	NM_004530	SEQ ID NO: 51
	RAB34	1555630_a_at	AF327350	SEQ ID NO: 52
	SELM	226051_at	BF973568	SEQ ID NO: 53
	SRGAP2P1	229067_at	BF977829	SEQ ID NO: 54
	VIM	201426 s at	Al922599	SEQ ID NO: 55

La Tabla 2 muestra el número de casos clasificados en cada uno de los tipos como se ha descrito anteriormente (relación de abundancia) así como la tasa de recurrencia del cáncer colorrectal.

5

	[Tabla 2]				
Relación de abundancia Tasa de recurr					
Tipo A	30,6 % (22)	4,5 % (1)			
Tipo B	33,3 % (24)	12,5 % (3)			
Tipo C	36,1 % (26)	23,1 % (6)			
Total	100 % (72)	13,9 % (10)			

En referencia a la Tabla 2, entre todos los 72 casos, 22 casos se clasificaron en el tipo A, 24 casos se clasificaron en el tipo B, y 26 casos se clasificaron en el tipo C. Las tasas de recurrencia del cáncer colorrectal en el tipo A, el tipo B y el tipo C fueron del 4,5 %, 12,5 % y 23,1 %, respectivamente.

10

La Fig. 6 ilustra las curvas de Kaplan-Meier creadas para cada uno de los tipos clasificados. En referencia a la Fig. 6, se ha descubierto que las tasas de supervivencia libre de recurrencia postoperatoria en los tipos respectivos se diferencian mucho de unos a otros.

Los resultados que se muestran en la Tabla 2 y la Fig. 6 indican que el tipo A se puede definir con el grupo de bajo riesgo de recurrencia, el tipo B se puede definir como el grupo de riesgo de recurrencia medio, y el tipo C se puede definir como el grupo de alto riesgo de recurrencia. De aquí en adelante, se puede hacer referencia al grupo de bajo riesgo, el grupo de riesgo medio, y el grupo de alto riesgo colectivamente como "grupo de riesgo de recurrencia".

20

#### Ejemplo 2:

1. Verificación de la fiabilidad de la clasificación en grupos de riesgo de recurrencia

## 25 < Método>

Entre los conjuntos de datos GSE14333 de la matriz U133 plus 2.0 del GeneChip del genoma humano (Affymetrix, Inc.) (obtenido en el Gene Expression Omnibus del NCBI (URL; <a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a>), se utilizaron 74 casos de pacientes, que no se habían utilizado en el conjunto de entrenamiento, como un conjunto de validación 1. Los casos del conjunto de entrenamiento y los casos del conjunto de validación 1 se seleccionaron de manera que los especímenes de estos casos fueran de diferentes centros médicos.

El total de los 72 casos del conjunto de entrenamiento y los 74 casos del conjunto de validación (146 casos) se agruparon utilizando los genes de la Tabla 1 de la misma manera que en el Ejemplo 1.

35

30

## <Resultados>

La Fig. 7 muestra el resultado de la clasificación de los casos en el conjunto de entrenamiento y el conjunto de validación 1 en los grupos de riesgo de recurrencia. En referencia a la Fig. 7 se descubre que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 1 se obtuvieron de diferentes centros, sin embargo, no se formó un agrupamiento originado a partir de los diferentes centros, y todos los casos se clasificaron en cualquiera de los tres grupos de riesgo de recurrencia.

#### Ejemplo 3:

45

2. Verificación de la fiabilidad de la clasificación en grupos de riesgo de recurrencia

#### <Método>

Entre los conjuntos de datos GSE14333 de la matriz U133 plus 2.0 del GeneChip del genoma humano (Affymetrix, Inc.) (obtenido en el Gene Expression Omnibus del NCBI (URL; <a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a>), se utilizaron 53 casos de pacientes con cáncer colorrectal (cáncer de colon) como conjunto de validación 2. Los 53 casos se agruparon utilizando los genes de la Tabla 1 de la misma manera que en el Ejemplo 1.

#### 55 < Resultados >

La Fig. 8 muestra el resultado de la clasificación de los casos del conjunto de validación 2 en grupos de riesgo de

recurrencia. En referencia a la Fig. 7, se descubre que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 2 se obtuvieron de diferentes centros y las muestras biológicas se midieron con los GenChip en diferentes centros. Sin embargo, todos los casos del conjunto de validación 2 se clasificaron en cualquiera de los tres grupos de riesgo de recurrencia.

La Tabla 3 muestra el número de casos clasificados en cada uno de los tipos como se ha descrito anteriormente (relación de abundancia) así como la tasa de recurrencia del cáncer colorrectal.

г		LI.	- 21	
ш	ıaı	Olla	4 .SI	

	Relación de abundancia	Tasa de recurrencia
Grupo de bajo riesgo	43,4 % (23)	8,7 % (2)
Grupo de riesgo medio	47,2 % (25)	28,0 % (7)
Grupo de alto riesgo	9,4 % (5)	80,0 % (4)
Total	100 % (53)	24,5 % (13)

10

15

5

En referencia a la Tabla 3, entre el total de 53 casos, 23 casos se clasificaron en el grupo de bajo riesgo, 25 casos se clasificaron en el grupo de riesgo medio, y 5 casos se clasificaron en el grupo de alto riesgo. Las tasas de recurrencia del cáncer colorrectal en el grupo de bajo riesgo, el grupo de riesgo medio y el grupo de alto riesgo fueron del 8,7 %, 28,0 % y 80,0 %, respectivamente. El resultado de la Tabla 3 indica que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 2 se obtuvieron en diferentes centros y las muestras biológicas se midieron con el GenChip en diferentes centros, sin embargo, cada uno de los grupos de riesgo de recurrencia presentaban el mismo resultado que en el Ejemplo 1, sin estar influenciados por la diferencia de centro.

#### Ejemplo 4:

20

3. Verificación de la fiabilidad de la clasificación en grupos de riesgo de recurrencia

#### <Método>

Entre los conjuntos de datos GSE39582 de la matriz U133 plus 2.0 del GeneChip del genoma humano (Affymetrix, Inc.) (obtenido en el Gene Expression Omnibus del NCBI (URL; <a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a>), se utilizaron 258 casos de pacientes con cáncer colorrectal (cáncer de colon) como conjunto de validación 3. Los 258 casos se agruparon utilizando los genes de la Tabla 1 de la misma manera que en el Ejemplo 1.

#### 30 <Resultados>

La Fig. 9 muestra el resultado de la clasificación de los casos del conjunto de validación 3 en grupos de riesgo de recurrencia. En referencia a la Fig. 9, se descubre que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 3 se obtuvieron de diferentes centros y las muestras biológicas se midieron con los GenChip en diferentes centros. Sin embargo, todos los casos del conjunto de validación 3 se clasificaron en cualquiera de los tres grupos de riesgo de recurrencia.

La Tabla 4 muestra el número de casos clasificados en cada uno de los tipos como se ha descrito anteriormente (relación de abundancia) así como la tasa de recurrencia del cáncer colorrectal.

40

35

## [Tabla 4]

[1454 1]				
	Relación de abundancia	Tasa de recurrencia		
Grupo de bajo riesgo	28,7 % (74)	12,2 % (9)		
Grupo de riesgo medio	47,7 % (123)	23,6 % (29)		
Grupo de alto riesgo	23,6 (%)	39,3 % (24)		
Total	100 % (258)	24,0 % (62)		

En referencia a la Tabla 4, entre el total de 258 casos, 74 casos se clasificaron en el grupo de bajo riesgo, 123 casos se clasificaron en el grupo de riesgo medio, y 61 casos se clasificaron en el grupo de alto riesgo. Las tasas de recurrencia del cáncer colorrectal en el grupo de bajo riesgo, el grupo de riesgo medio y el grupo de alto riesgo fueron del 12,2 %, 23,6 % y 39,3 %, respectivamente.

La Fig. 10 ilustra las curvas de Kaplan-Meier creadas para cada uno de los tipos clasificados. En referencia a la Fig. 10, se ha descubierto que las tasas de supervivencia libre de recurrencia postoperatoria en los tipos respectivos se diferencian mucho de unos a otros. Los resultados de la Tabla 4 y la Fig. 10 indican que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 3 se obtuvieron en diferentes centros y las muestras biológicas se midieron con el GenChip en diferentes centros, sin embargo, cada uno de los grupos de riesgo de recurrencia presentaban el mismo resultado que en el Ejemplo 1, sin estar influenciados por la diferencia de centro.

#### Ejemplo 5:

10

15

4. Verificación de la fiabilidad de la clasificación en grupos de riesgo de recurrencia

Los datos de 85 casos de pacientes con cáncer colorrectal (cáncer de colon) se obtuvo midiendo los especímenes obtenidos del National Defense Medical College con la matriz U133 plus 2.0 del GeneChip del Genoma Humano (Affymetrix, Inc.). Los datos se utilizaron como un conjunto de validación 4. Los 85 casos se agruparon utilizando los genes de la Tabla 1 de la misma manera que en el Ejemplo 1.

#### <Resultados>

20 La Fig. 11 muestra el resultado de la clasificación de los casos del conjunto de validación 4 en grupos de riesgo de recurrencia. En referencia a la Fig. 11, se descubre que las muestras biológicas del conjunto de entrenamiento y el conjunto de validación 4 se obtuvieron de diferentes centros y las muestras biológicas se midieron con los GenChip en diferentes centros. Sin embargo, todos los casos del conjunto de validación 4 se clasificaron en cualquiera de los tres grupos de riesgo de recurrencia.
25

La Tabla 5 muestra el número de casos clasificados en cada uno de los tipos como se ha descrito anteriormente (relación de abundancia) así como la tasa de recurrencia del cáncer colorrectal.

[Tabla 5]				
	Relación de abundancia	Tasa de recurrencia		
Grupo de bajo riesgo	27,1 % (23)	0 % (0)		
Grupo de riesgo medio	30,6 % (26)	11,5 % (3)		
Grupo de alto riesgo	42,3 % (36)	22,2 % (8)		
Total	100 % (85)	12,9 % (11)		

30

En referencia a la Tabla 5, entre el total de 85 casos, 23 casos se clasificaron en el grupo de bajo riesgo, 26 casos se clasificaron en el grupo de riesgo medio, y 36 casos se clasificaron en el grupo de alto riesgo. Las tasas de recurrencia del cáncer colorrectal en el grupo de bajo riesgo, el grupo de riesgo medio y el grupo de alto riesgo fueron del 0 %, 11,5 % y 22,2 %, respectivamente.

35

La Fig. 12 ilustra las curvas de Kaplan-Meier creadas para cada uno de los tipos clasificados. En referencia a la Fig. 12, se ha descubierto que las tasas de supervivencia libre de recurrencia postoperatoria en los tipos respectivos se diferencian mucho de unos a otros.

40 El resultado de la Tabla 5 indica que el caso del conjunto de validación 4 presentaba el mismo resultado que el Ejemplo 1.

Como se ha descrito anteriormente, los casos de cáncer colorrectal se clasificaron en los tres grupos de riesgo de recurrencia mediante el análisis de los módulos funcionales. Cada uno de los grupos de riesgo de recurrencia tenían diferentes riesgos de recurrencia. Los resultados de los Ejemplos 1 a 5 indican que la clasificación en grupos de riesgo de recurrencia es un método de clasificación fiable que no está influenciado por las fuentes de obtención de los conjuntos de datos. Por lo tanto, se demuestra que el método para el apoyo de un diagnóstico del riesgo utilizando la clasificación en grupos de riesgo de recurrencia del cáncer colorrectal descrito en el presente documento daba resultados suficientemente estables y fiables.

50

## Ejemplo comparativo:

Predicción del pronóstico mediante el método convencional (Clasificación de Dukes)

55 Como controles comparativos para llevar a cabo la predicción del pronóstico, se analizó el tiempo de supervivencia

de 72 casos del conjunto de entrenamiento mediante la clasificación de Dukes. Los resultados se muestran en la Fig. 13. En la Fig. 13, Dukes A indica el cáncer confinado a la pared del intestino grueso, Dukes B indica el cáncer negativo en nodos que penetra en la pared del intestino grueso, y Dukes C indica el cáncer positivo en los nodos.

En referencia a la Fig. 2 y la Fig. 13, se determinó que 26 casos tenían un alto riesgo por el método de apoyo al diagnóstico descrito en el presente documento, mientras que se determinó que 15 casos tenían un alto riesgo por el método de determinación del Ejemplo comparativo (Dukes C). Existe una pequeña diferencia en la tasa de supervivencia libre de recurrencia entre los casos determinados como de Dukes A y los casos determinados como Dukes B basándose en el método de determinación del Ejemplo comparativo. Por otra parte, existen diferencias en la tasa de supervivencia libre de recurrencia entre los casos determinados como el grupo de bajo riesgo y los casos 10 determinados como el grupo de riesgo medio basándose en el método de apoyo al diagnóstico descrito en el presente documento. Este resultado sugiere que el método de apoyo al diagnóstico de un riesgo de recurrencia descrito en el presente documento hace posible que se determine el riesgo de recurrencia con alta precisión en comparación con la clasificación patológica convencional.

## Ejemplo 6:

15

20

25

30

1. Mejora de la realización de la predicción del pronóstico basándose en la estratificación del grupo de riesgo medio mediante la mutación del gen KRAS

Entre los especímenes determinados como el grupo de riesgo medio (como los resultados del análisis del Ejemplo 4), se determinó que los especímenes que tenían la mutación del gen KRAS eran de alto riesgo y se determinó que los especímenes que no tenían la mutación del gen KRAS eran de bajo riesgo. Todos los especímenes se dividieron en dos grupos (en referencia a la Fig. 14). Específicamente, la presencia de las mutaciones KRAS en el ADN de los especímenes se midió de la siguiente manera. Basándose en los resultados, todos los especímenes se dividieron en dos grupos.

Primero se preparó cada una de las soluciones Master Mix de PCR tenían la composición de la Tabla 6 a continuación.

[Tabla 6]			
	Volumen (para 5 tubos)		
10 x tampón Ex Taq	10 µl		
dNTP (2,5 mM)	8 µl		
TaKaRa EX Taq HS (1 unidad/ul)	2 μΙ		
Cebador directo (10 uM)	2 µl		
Cebador inverso (10 uM)	2 µl		
Volumen total	24 µl		

A continuación, se distribuyeron 10 ng de ADN en un tubo de PCR de 0,5 ml, y se añadió agua libre de nucleasas hasta hacer una cantidad total de 20 µl. Después, cada una de las soluciones Master Mix de PCR se añadió a este a 4,8 µl/tubo, que se mezcló. Los cebadores que se añadieron a cada solución Master Mix de PCR se muestran en la Tabla 7 a continuación. El par de cebadores de SEQ ID NO: 57 y 58 se utilizaron para amplificar la región que contiene el 12º y 13º codones en el exón 2 del gen KRAS. El par de cebadores de SEQ ID NO: 59 y 60 se utilizaron para amplificar la región que contiene el 61º codón localizado en el exón 3.

[Tabla 7]

Nombre del cebador	Secuencia de bases	SEQ ID NO:
Cebador directo de exón 2 de KRAS	CGATACACGTCTGCAGTCAAC	57
Cebador inverso de exón 2 de KRAS	ACCCTGACATACTCCCAAGG	58
Cebador directo de exón 3 de KRAS	AGGTGCTTAGTGGCCATTTG	59
Cebador inverso de exón 3 de KRAS	TGCATGGCATTAGCAAAGAC	60

Cada una de las soluciones Master Mix de PCR se fijó en un termociclador. Los programas siguientes se llevaron a cabo con el fin de amplificar las secuencias de los exones 2 y 3 del KRAS mediante PCR. Exón 2 de KRAS

95 °C: 10 min  $\rightarrow$  (94 °C: 1 min  $\rightarrow$  55 °C: 1 min  $\rightarrow$  72 °C: 1 min) x 38 ciclos  $\rightarrow$  72 °C: 10 min  $\rightarrow$  4 °C mantenidos

19

35

40

Exón 3 de KRAS

95 °C: 10 min  $\rightarrow$  (94 °C: 1 min  $\rightarrow$  63 °C: 1 min  $\rightarrow$  72 °C: 1 min) x 38 ciclos  $\rightarrow$  72 °C: 10 min  $\rightarrow$  4 °C mantenidos

Después de la amplificación se confirmó una única banda mediante electroforesis a través de gel de agarosa al 1 %. A continuación, 5 μl del producto de la PCR se distribuyeron en un tubo de PCR de 0,5 ml y se añadieron 2 μl de ExoSAP-IT, que se mezcló. La mezcla resultante se fijó en el termociclador y se llevó a cabo el siguiente programa: 37 °C 15 min → 80 °C 15 min → 4 °C mantenidos

Se añadieron 9,6 µl de cebador D o I (1 pmol/µl) y 9,4 µl de NFW a 2 µl del producto resultante, que se mezcló. El análisis de secuencia se confió a Operon Biotechnologies, Inc. Cuando las secuencias de bases que se iban a analizar se compararon con la secuencia de bases de SEQ ID NO: 1 y al menos se observaba una mutación, se determinó que tenía una mutación del gen KRAS. Como para los especímenes determinados como el grupo de riesgo medio en el Ejemplo 4, los especímenes que tenían la mutación del gen KRAS se clasificaron en el grupo de alto riesgo y los especímenes que no tenían la mutación del gen KRAS se clasificaron en el grupo de bajo riesgo.

Las curvas de Kaplan-Meier resultantes de los grupos se muestran en la Fig. 14. La comparación de las Fig. 14 a 10 muestra que los casos del grupo de riesgo medio se podían clasificar en dos grupos alto y bajo que tenían grandes diferencias en la tasa de supervivencia libre de recurrencia añadiendo los criterios de determinación de la presencia de la mutación del gen KRAS.

#### Ejemplo 7:

15

20

25

30

35

2. Mejora de la realización de la predicción del pronóstico basándose en la estratificación del grupo de riesgo medio mediante la mutación del gen KRAS

Entre los especímenes determinados como el grupo de riesgo medio (como los resultados del análisis del Ejemplo 5), se determinó que los especímenes que tenían la mutación del gen KRAS eran de alto riesgo y se determinó que los especímenes que no tenían la mutación del gen KRAS eran de bajo riesgo. Todos los especímenes se dividieron en dos grupos.

Las curvas de Kaplan-Meier resultantes de los grupos se muestran en la Fig. 15. La comparación de las Fig. 15 a 12 muestra que los casos del grupo de riesgo medio se podían clasificar en dos grupos alto y bajo que tenían grandes diferencias en la tasa de supervivencia libre de recurrencia añadiendo los criterios de determinación de la presencia de la mutación del gen KRAS.

## Ejemplo 8:

Verificación basada en tejido fijado con formalina embebido en parafina (FFPE)

Se prepararon los especímenes de tejido FFPE de 18 casos de los especímenes de tejido congelado (85 casos) utilizados en el Ejemplo 5. Los 18 especímenes se clasificaron en los grupos de riesgo de recurrencia. Más específicamente, se extrajeron los ARN totales de los especímenes de tejido FFPE utilizando el kit RNAeasy FFPE (QIAGEN). El pretratamiento del chip de ácido nucleico se llevó a cabo utilizando el kit Sensation Plus FFPE Amplification y 3'IVT Labeling (Affymetrix, Inc.). Los ARN totales obtenidos se midieron con el GeneChip. Los genes de la Tabla 1 se agruparon de la misma manera que en el Ejemplo 1.

La Fig. 16 muestra el resultado de la clasificación de los especímenes de tejido FFPE (18 especímenes) en grupos de riesgo de recurrencia. Como se muestra en la Fig. 16, se descubrió que cuando se utilizaban especímenes de tejido FFPE, los casos de cáncer colorrectal se podían clasificar en tres grupos de riesgo de recurrencia basándose en los niveles de expresión de los genes de la Tabla 1. La Tabla 8 muestra el número de especímenes de tejido FFPE (18 casos) clasificados en cada uno de los tipos como se ha descrito anteriormente (relación de abundancia) así como la tasa de recurrencia del cáncer colorrectal.

ГТ	᠘	h	ı	QΊ
	ιа	v	ıa	OI

	Relación de abundancia	Tasa de recurrencia
Grupo de bajo riesgo	22,2 % (4)	0 % (0)
Grupo de riesgo medio	33,3 % (6)	0 % (0)
Grupo de alto riesgo	44,4 % (8)	37,5 % (3)
Total	100 % (18)	16,7 % (3)

55

50

Entre el total de 18 casos, 4 casos se clasificaron en el grupo de bajo riesgo, 6 casos se clasificaron en el grupo de riesgo medio, y 8 casos se clasificaron en el grupo de alto riesgo. Las tasas de recurrencia del cáncer colorrectal en

el grupo de bajo riesgo, el grupo de riesgo medio y el grupo de alto riesgo fueron del 0 %, 0 % y 37,5 %, respectivamente. Estos resultados indican que cuando se utilizan los especímenes de tejido FFPE, los casos de cáncer colorrectal se pueden clasificar en los grupos de riesgo de recurrencia con alta precisión.

Como para los especímenes de tejido FFPE anteriores, 6 casos clasificados en el grupo de riesgo medio se sometieron a la medición de mutación del gen KRAS. En el total de los 6 casos, el resultado de la mutación del gen KRAS era negativo. Por lo tanto, en cada uno de los casos era posible clasificarlo en el grupo de bajo riesgo. La Fig. 17 ilustra las curvas de Kaplan-Meier de los grupos en los que se clasificaron los especímenes de tejido FFPE en grupos de riesgo de recurrencia en los Ejemplos 6 y 7. Como se muestra en la Fig. 17, se descubrió que cuando se utilizaban especímenes de tejido FFPE, se observaba una gran diferencia en las tasas de supervivencia libre de recurrencia postoperatoria en los respectivos tipos.

Los resultados de la Fig. 16 y la Tabla 8 muestran que cuando se utilizan los especímenes de tejido FFPE, los casos de cáncer colorrectal se pueden clasificar en grupos de riesgo de recurrencia de la misma manera que en el Ejemplo 1. Los resultados de la Fig. 17 indican que, al igual que en los Ejemplos 6 y 7 incluso si se utilizan especímenes de tejido FFPE, es posible mejorar la precisión de la clasificación en grupos de riesgo de recurrencia basándose en la presencia de la mutación del gen KRAS. La Tabla 9 a continuación muestra una tabla de correlación entre el resultado de la determinación en los especímenes de tejido congelados y el resultado de la determinación en los especímenes de tejido FFPE.

20

25

15

	[Tabla 9]				
FFPE					
Grupo de riesgo Grupo de bajo Grupo de a medio riesgo riesgo				Grupo de alto riesgo	
	Grupo de riesgo medio	5	0	1	
Congelado	Grupo de bajo riesgo	0	4	1	
	Grupo de alto riesgo	1	0	6	

La tasa de concordancia es alta del 83,3 %. Este resultado indica que cuando se utilizan especímenes de tejido FFPE, es posible determinar el riesgo de recurrencia, de manera similar al caso donde se utilizaron especímenes de tejido congelados.

Ejemplo 9: Método de determinación del riesgo de recurrencia utilizando el coeficiente de correlación

En el grupo de bajo riesgo, el grupo de riesgo medio, y el grupo de alto riesgo clasificados en el Ejemplo 1, se midieron los niveles de expresión de los 55 genes que se muestran en la Tabla 1. Basándose en los niveles de expresión medidos de esta manera, se obtuvieron el patrón de expresión del grupo de bajo riesgo, el patrón de expresión del grupo de riesgo medio, y el patrón de expresión del grupo de alto riesgo. Cada patrón de expresión incluye los valores medios de los genes.

35 Como especímenes se utilizaron los mismos especímenes del Ejemplo 4. Para cada espécimen, se midieron los niveles de expresión de los 55 genes que se muestran en la Tabla 1. Basándose en los niveles de expresión medidos de esta manera, se obtuvo el patrón de expresión de cada espécimen.

Se calculó un coeficiente de correlación entre el patrón de expresión de cada espécimen y los patrones de expresión para los grupos de riesgo de acuerdo con la correlación de intervalos de Spearman. Para cada espécimen, se identificó un grupo de riesgo que presentaba el mayor coeficiente de correlación.

La tasa de concordancia entre los resultados del Ejemplo 4 (clasificación del riesgo por análisis del agrupamiento) y los resultados del Ejemplo 9 se muestran en la Tabla 10.

		abla 10] Resultados del Ej	emplo 9	
		Grupo de bajo riesgo	Grupo de riesgo medio	Grupo de alto riesgo
Resultados del Ejemplo 4	Grupo de bajo riesgo	61	0	13
	Grupo de riesgo medio	6	116	1
	Grupo de alto riesgo	3	21	37

En referencia a la Tabla 10, la tasa de concordancia entre los resultados del Ejemplo 9 y los resultados del Ejemplo 4 era del 83 %. Este resultado indica que cuando se utiliza el coeficiente de correlación, es posible determinar el riesgo de recurrencia de un espécimen.

#### 5 **Ejemplo 10:** Determinación del riesgo de recurrencia utilizando la mutación del gen KRAS

Como para el grupo de pacientes determinado en el Ejemplo 9 que tenía un riesgo medio de recurrencia, se examinó la presencia o ausencia de una mutación del gen KRAS. En el grupo de riesgo medio, los especímenes que tienen la mutación del gen KRAS se clasificaron como que tenían un alto riesgo de recurrencia y los especímenes que no tenían la mutación del gen KRAS se clasificaron como que tenían un bajo riesgo de recurrencia.

La tasa de concordancia entre los resultados del Ejemplo 6 y los resultados del Ejemplo 10 se muestran en la Tabla 11.

15

10

[Tabla 11]			
		Resultados del Ejemplo 10	
		Grupo de bajo riesgo	Grupo de alto riesgo
Resultados del Ejemplo 6	Grupo de bajo riesgo	146	13
	Grupo de alto riesgo	20	79

En referencia a la Tabla 11, la tasa de concordancia entre los resultados del Ejemplo 10 y los resultados del Ejemplo 6 era del 85 %. Este resultado indica que cuando se utiliza el coeficiente de correlación, es posible determinar el riesgo de recurrencia de un espécimen.

20

La Fig. 18 ilustra curvas de Kaplan-Meier creadas a partir de los resultados del Ejemplo 10. En referencia a la Fig. 18, añadiendo la presencia o ausencia de la mutación del gen KRAS a los criterios de determinación, era posible clasificar los especímenes en dos grupos, a saber, el grupo de alto riesgo y el grupo de bajo riesgo, para los que las tasas de supervivencia libre de recurrencia eran significativamente diferentes entre ellas.

25

Cuando se calcula un coeficiente de correlación para determinar el riesgo de recurrencia, también se puede utilizar el aparato de apoyo al diagnóstico que se muestra en la Fig. 1. El flujo del procesamiento para este caso se describe en referencia a la Fig. 19. La unidad de almacenamiento del aparato almacena, con anterioridad, el patrón de expresión del grupo de alto riesgo, el patrón de expresión del grupo de riesgo medio, y el patrón de expresión del grupo de bajo riesgo.

35

30

La unidad de recepción 301 del aparato de apoyo al diagnóstico 1 obtiene la información de la fluorescencia indicando los niveles de expresión de los genes de la muestra biológica a partir del dispositivo de medición 2 (etapa S2-1). A continuación, la unidad de cálculo 303 calcula la intensidad de fluorescencia de la información obtenida y la transmite a la unidad de almacenamiento 302 (etapa S2-2). La unidad de cálculo 303 calcula los niveles de expresión de los genes basándose en la intensidad de fluorescencia almacenada (y se obtiene el patrón de expresión de la muestra biológica) (etapa S2-3). A continuación, la unidad de determinación 304 lee el patrón de expresión para el grupo de alto riesgo, el patrón de expresión para el grupo de riesgo medio, y el patrón de expresión para el grupo de bajo riesgo almacenados en la unidad de almacenamiento 302, y entonces calcula, basándose en estos y el patrón de expresión para la muestra biológica obtenido en la etapa S2-3, un coeficiente de correlación entre el patrón de expresión para la muestra biológica y el patrón de expresión para el grupo de alto riesgo (de aquí en adelante, el coeficiente de correlación se llama también "coeficiente de correlación H"), un coeficiente de correlación entre el patrón de expresión para la muestra biológica y el patrón de expresión para el grupo de riesgo medio (de aquí en adelante, el coeficiente de correlación se llama también "coeficiente de correlación M"), y un coeficiente de correlación entre el patrón de expresión de la muestra biológica y el patrón de expresión del grupo de bajo riesgo (de aquí en adelante, el coeficiente de correlación se llama también "coeficiente de correlación L") (etapa S2-4).

45

50

Se lleva a cabo la determinación de si el coeficiente de correlación H es el mayor (etapa S2-5). En otras palabras, cuando el coeficiente de correlación H es mayor que el coeficiente de correlación M y el coeficiente de correlación H es mayor que el coeficiente de correlación L, se determina que el coeficiente de correlación H es el mayor. Cuando el coeficiente de correlación H es el mayor, la muestra biológica se clasifica en el grupo de alto riesgo y se determina que la muestra biológica tiene un alto riesgo de recurrencia (etapa S2-6).

55

Cuando se hace la determinación en la etapa S2-5 y el coeficiente de correlación H no es el coeficiente de correlación más alto, se lleva a cabo la determinación de si el coeficiente de correlación M es el mayor (etapa S2-7). En otras palabras, cuando el coeficiente de correlación M es mayor que el coeficiente de correlación H y el

- coeficiente de correlación M es mayor que el coeficiente de correlación L, se determina que el coeficiente de correlación M es el mayor. Cuando el coeficiente de correlación M es el mayor, la muestra biológica se clasifica en el grupo de riesgo medio y se determina que la muestra biológica tiene un riesgo medio de recurrencia (etapa S2-8).
- 5 Cuando se hace la determinación en la etapa S2-7 y el coeficiente de correlación M no es el coeficiente de correlación más alto, se determina que el coeficiente de correlación L es el mayor (etapa S2-9). Cuando el coeficiente de correlación L es el mayor, la muestra biológica se clasifica en el grupo de bajo riesgo y se determina que la muestra biológica tiene un bajo riesgo de recurrencia en la etapa S2-9.
- 10 La unidad de salida 305 transmite el resultado de la determinación del riesgo de recurrencia de cáncer colorrectal en el sujeto y permite a la unidad de pantalla 3c presentar el resultado de la determinación (etapa S2-10). En consecuencia, el aparato de apoyo al diagnóstico 1 puede proporcionar información para apoyar la determinación de si el riesgo de recurrencia del cáncer colorrectal en el sujeto es alto, medio o bajo a un médico o similar.
- El diagrama de flujo que se muestra en la Fig. 19 puede incluir, en vez de la etapa de determinación de si el coeficiente de correlación M es el mayor, una etapa de determinación de si el coeficiente de correlación L es el mayor. De manera alternativa, el diagrama de flujo que se muestra en la Fig. 19 puede incluir, en vez de la etapa de determinación de si el coeficiente de correlación H es el mayor, una etapa de determinación de si el coeficiente de correlación L es el mayor. En cualquier caso, se puede hacer la determinación de cuál de los coeficientes de correlación H M v L es el mayor. En cualquier caso, el orden en el que se llevan a cabo las etapas de determinación
- 20 correlación H, M y L es el mayor. En cualquier caso, el orden en el que se llevan a cabo las etapas de determinación no está limitado.

#### **REIVINDICACIONES**

1. Un método para apoyar un diagnóstico de un riesgo de recurrencia del cáncer colorrectal, que comprende las etapas de:

5

10

llevar a cabo una primera medición para medir los niveles de expresión de C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407 en una muestra biológica recolectada de un paciente con un cáncer colorrectal, una segunda medición para medir los niveles de expresión de ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y una tercera medición para medir los niveles de expresión de ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y determinar el riesgo de recurrencia del cáncer colorrectal del paciente basándose en los resultados de la primera medición, la segunda medición y la tercera medición.

15

- 2. El método de acuerdo con la reivindicación 1, donde
  - en la etapa de determinación,

los niveles de expresión medidos en la etapa de medición están correlacionados

- con los niveles de expresión de un grupo de alto riesgo medido por adelantado para las muestras biológicas de un grupo de pacientes en los que se ha determinado que tienen un alto riesgo de recurrencia, con los niveles de expresión de un grupo de riesgo medio medido por adelantado para las muestras biológicas de un grupo de pacientes en los que se ha determinado que tienen un riesgo medio de recurrencia, y con los niveles de expresión en un grupo de bajo riesgo medido por adelantado para las muestras biológicas de un grupo de pacientes en los que se ha determinado que tienen un bajo riesgo de recurrencia; y el riesgo de recurrencia, para la muestra biológica se clasifica en el grupo de riesgo que tenga la mayor.
  - el riesgo de recurrencia para la muestra biológica se clasifica en el grupo de riesgo que tenga la mayor correlación.
  - 3. El método de acuerdo con la reivindicación 2, donde
- 30 en la etapa de determinación,
  - se calcula un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los niveles de expresión del grupo de alto riesgo,
  - se calcula un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los niveles de expresión del grupo de riesgo medio.
- se calcula un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los niveles de expresión del grupo de bajo riesgo, y
  - el riesgo de recurrencia para la muestra biológica se clasifica en el grupo de riesgo que tenga el mayor coeficiente de correlación.
- 40 4. El método de acuerdo con la reivindicación 2, donde en la etapa de determinación,
  - se lleva a cabo un análisis de agrupamiento basándose en los niveles de expresión medidos en la etapa de medición, los niveles de expresión del grupo de alto riesgo, los niveles de expresión del grupo de riesgo medio, y los niveles de expresión del grupo de bajo riesgo, y entonces se clasifica el riesgo de recurrencia para la muestra biológica se clasifica en el grupo de riesgo que tenga la mayor correlación.
  - 5. El método de acuerdo con la reivindicación 1, donde en la etapa de determinación
  - se determina que el riesgo de recurrencia es alto cuando la media de los niveles de expresión de los genes del tercer grupo genético es mayor o igual a un valor de referencia para el tercer grupo genético independientemente de los niveles de expresión de los genes del primer y segundo grupos genéticos;
  - se determina que el riesgo de recurrencia es medio cuando la media de los niveles de expresión de los genes del tercer grupo genético es menor que los valores de referencia para el tercer grupo genético y la media de los niveles de expresión de los genes del primer grupo genético es menor que un valor de referencia para el primer grupo genético independientemente de los niveles de expresión de los genes del segundo grupo genético;
- se determina que el riesgo de recurrencia es medio cuando la media de los niveles de expresión de los genes del tercer grupo genético es menor que el valor de referencia para el tercer grupo genético, la media de los niveles de expresión de los genes del primer grupo genético es mayor o igual que el valor de referencia para el primer grupo genético y la media de los niveles de expresión de los genes del segundo grupo genético es mayor o igual a un valor de referencia para el segundo grupo genético; y
- se determina que el riesgo de recurrencia es bajo cuando la media de los niveles de expresión de los genes del tercer grupo genético es menor que el valor de referencia para el tercer grupo genético, la media de los niveles de expresión de los genes del primer grupo genético es mayor o igual que el valor de referencia para el primer grupo genético, y la media de los niveles de expresión de los genes del segundo grupo genético es menor que el valor de referencia para el segundo grupo genético.

65

45

50

6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, donde los niveles de expresión de los

genes se miden con una micromatriz en la etapa de medición.

10

15

30

45

60

65

- 7. El método de acuerdo con cualquiera de las reivindicaciones 2 a 6, donde se determina que el riesgo de recurrencia es alto cuando un grupo que se ha determinado que tiene un riesgo medio de recurrencia tiene una mutación del gen KRAS, y se determina que el riesgo de recurrencia es bajo cuando el grupo que se ha determinado que tiene un riesgo medio de recurrencia no tiene la mutación del gen KRAS.
- 8. Un programa de computadora ejecutado por una computadora, el programa de computadora hace que la computadora ejecute un procesamiento que comprende:

la recepción, en una muestra biológica recolectada de un paciente con un cáncer colorrectal, de los niveles de expresión de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles de expresión de un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y

determinar un riesgo de recurrencia de cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos.

- 9. Un sistema de computadora que se utiliza para determinar un riesgo de recurrencia del cáncer colorrectal, que comprende una computadora que comprende un procesador y una memoria, donde
- la memoria almacena un programa de computadora que hace que la computadora ejecute un procesamiento que 25 comprende:

la recepción, en una muestra biológica recolectada de un paciente con un cáncer colorrectal, de los niveles de expresión de un primer grupo genético que consiste en C18orf22, C18orf55, CCDC68, CNDP2, CYB5A, LOC400657, LOC440498, MBD2, MBP, MYO5B, NARS, PQLC1, RTTN, SEC11C, SOCS6, TNFRSF11A, TXNL1, TXNL4A, VPS4B, y ZNF407, los niveles de expresión de un segundo grupo genético que consiste en ASXL1, C20orf112, C20orf177, CHMP4B, COMMD7, CPNE1, DIDO1, DNAJC5, KIF3B, NCOA6, PHF20, PIGU, PLAGL2, POFUT1, PPP1R3D, PTPN1, RBM39, TAF4, y TCFL5, y los niveles de expresión de un tercer grupo genético que consiste en ANGPTL2, AXL, C1R, C1S, CALHM2, CTSK, DCN, EMP3, GREM1, ITGAV, KLHL5, MMP2, RAB34, SELM, SRGAP2P1, y VIM; y

la determinación del riesgo de recurrencia del cáncer colorrectal en el paciente basándose en los niveles de expresión recibidos.

10. El sistema de acuerdo con la reivindicación 9, donde

la memoria se puede operar para almacenar los niveles de expresión del grupo de alto riesgo, los niveles de expresión del grupo de riesgo medio, y los niveles de expresión del grupo de bajo riesgo, donde los niveles de expresión del grupo de alto riesgo se calculan a partir de los niveles de expresión de cada grupo genético para un grupo de pacientes que se ha determinado que tienen un alto riesgo de recurrencia,

los niveles de expresión del grupo de riesgo medio son valores que se calculan a partir de los niveles de expresión de cada grupo genético para un grupo de pacientes que se ha determinado que tienen un riesgo medio de recurrencia.

los niveles de expresión del grupo de bajo riesgo son valores que se calculan a partir de los niveles de expresión de cada grupo genético para un grupo de pacientes que se ha determinado que tienen un bajo riesgo de recurrencia, el procesador se puede operar para:

- recibir los niveles de expresión medidos en la etapa de medición,
  - la lectura, desde la memoria, de los niveles de expresión del grupo de alto riesgo, los niveles de expresión del grupo de riesgo medio, y los niveles de expresión del grupo de bajo riesgo,
  - comparar los niveles de expresión medidos en la etapa de medición con los niveles de expresión del grupo de alto riesgo,
- comparar los niveles de expresión medidos en la etapa de medición con los niveles de expresión del grupo de riesgo medio,
  - comparar los niveles de expresión medidos en la etapa de medición con los niveles de expresión del grupo de bajo riesgo, y
  - clasificar el riesgo de recurrencia para la muestra biológica en el grupo de riesgo que tenga la mayor correlación.

11. El sistema de acuerdo con cualquiera de las reivindicaciones 9 o 10, donde el procesador se puede operar para:

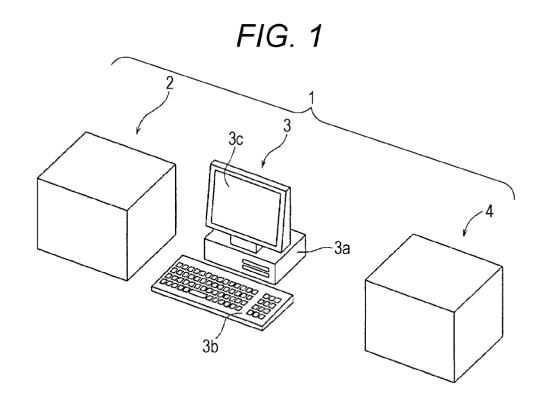
calcular un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los niveles de expresión del grupo de alto riesgo,

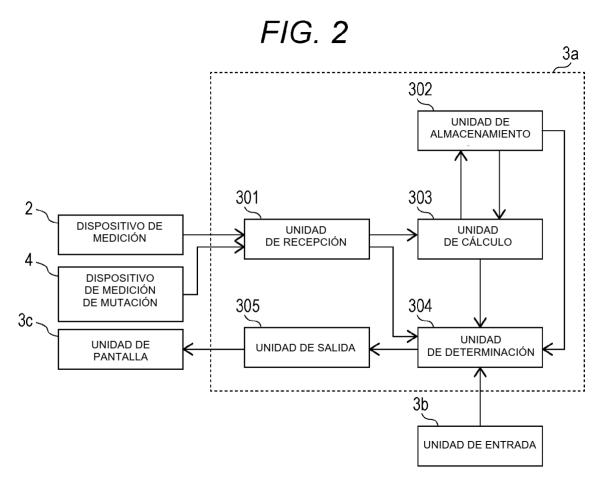
calcular un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los

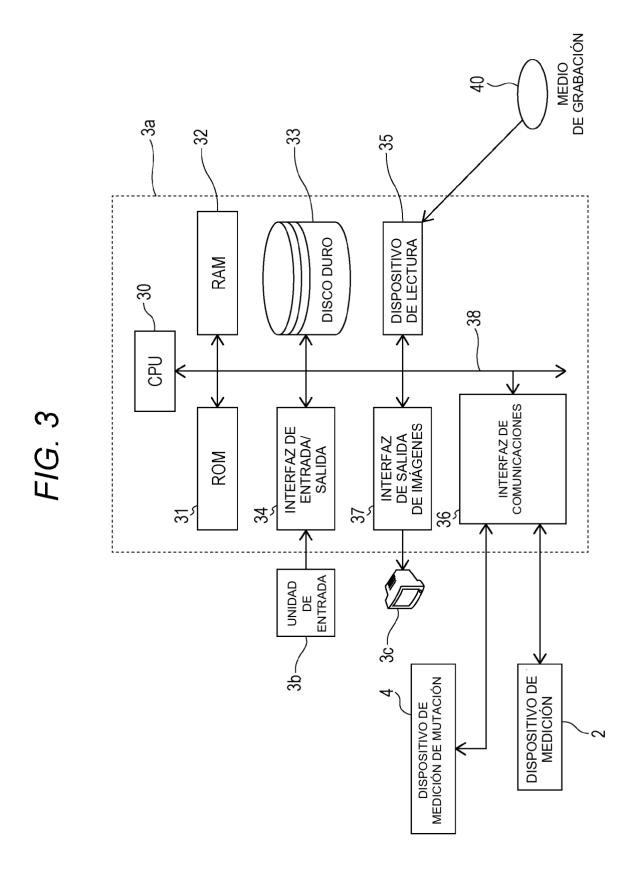
5

niveles de expresión del grupo de riesgo medio, calcular un coeficiente de correlación entre los niveles de expresión medidos en la etapa de medición y los niveles de expresión del grupo de bajo riesgo, y clasificar el riesgo de recurrencia para la muestra biológica en el grupo de riesgo que tenga el mayor coeficiente

de correlación.







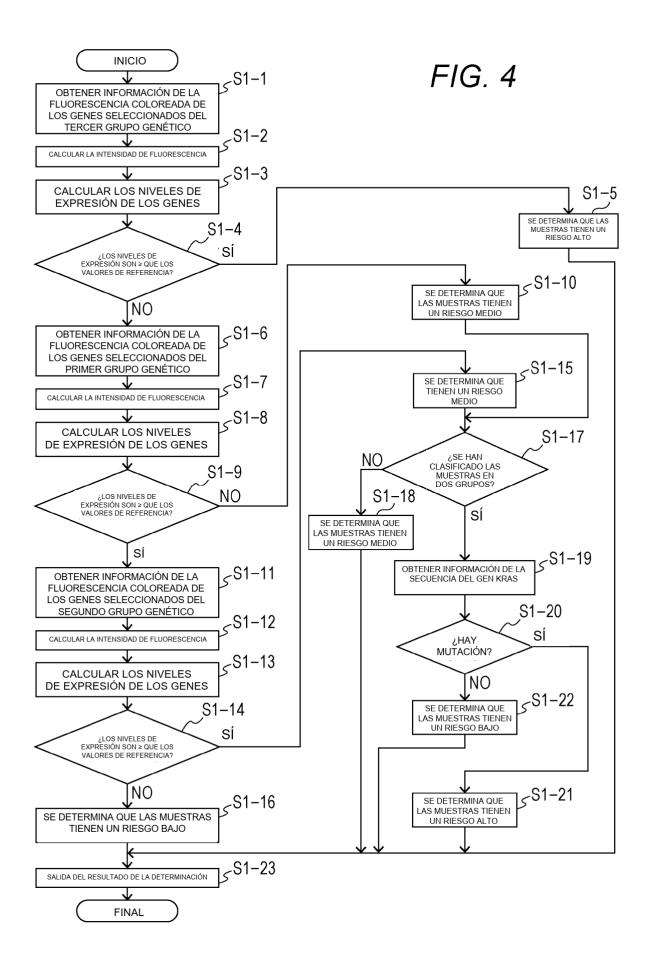


FIG. 5

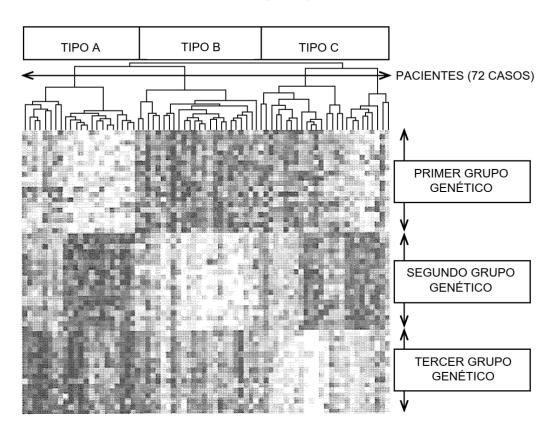


FIG. 6

100%

TIPO A (n=22)

TIPO B (n=24)

TIPO C (n=26)

TIPO C (n=26)

CURSO POSTOPERATORIO (AÑOS)

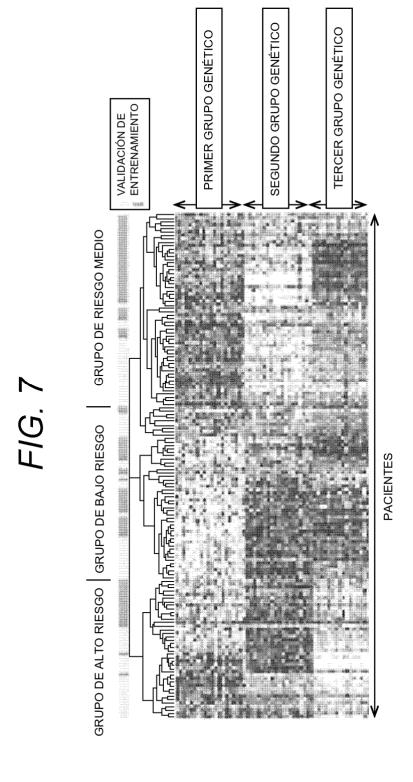
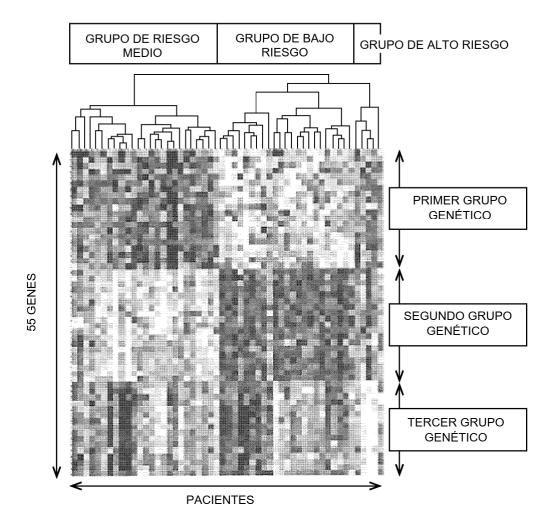


FIG. 8



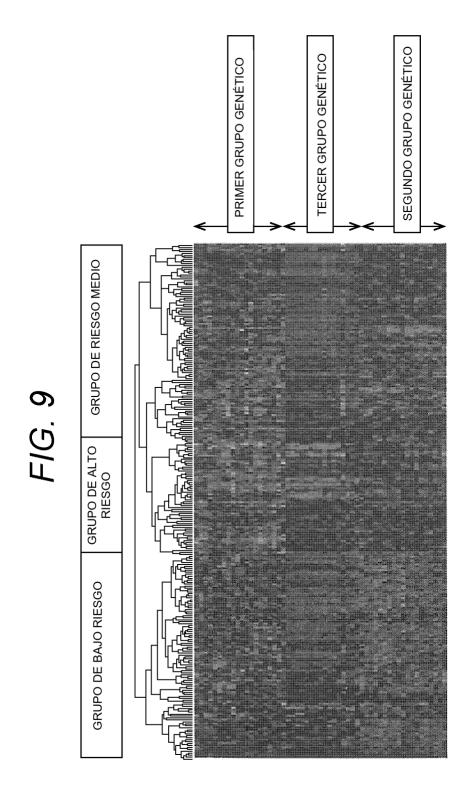
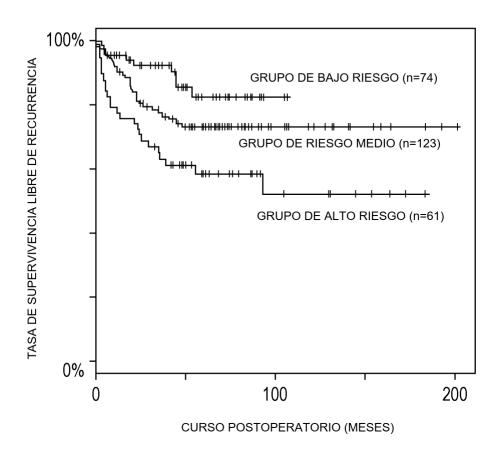


FIG. 10



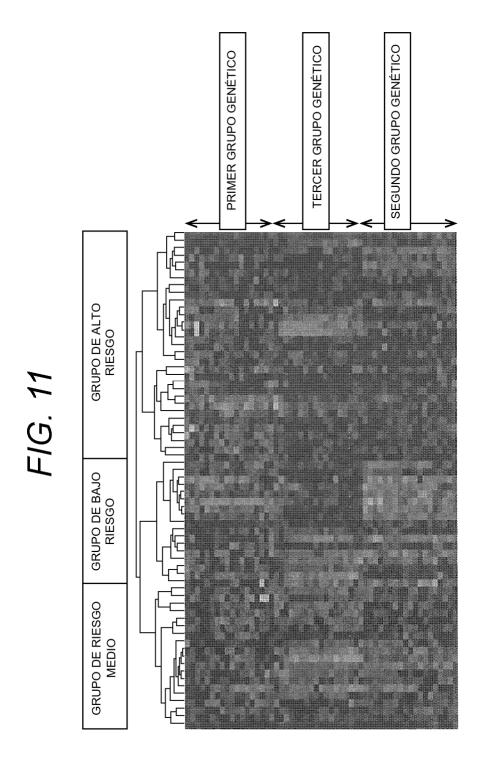


FIG. 12

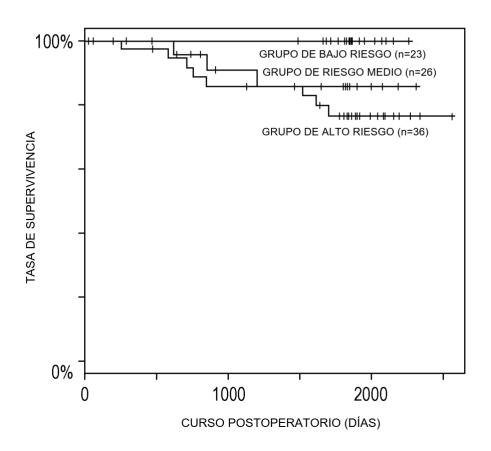


FIG. 13

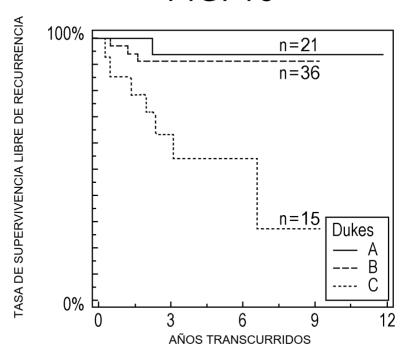
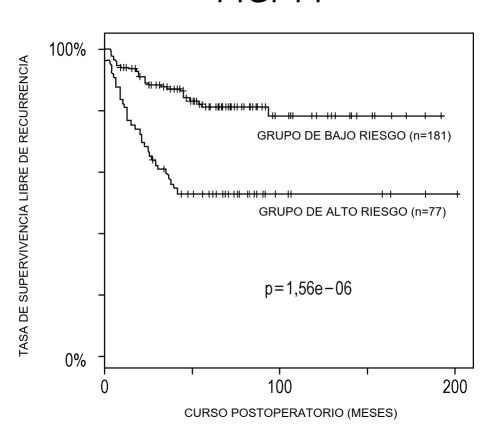
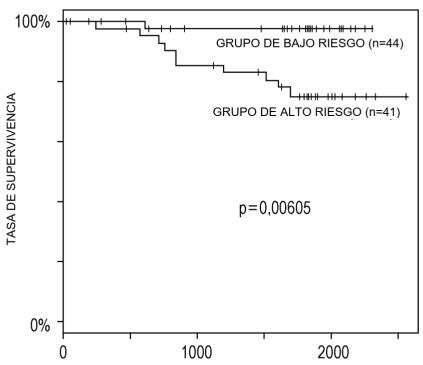


FIG. 14







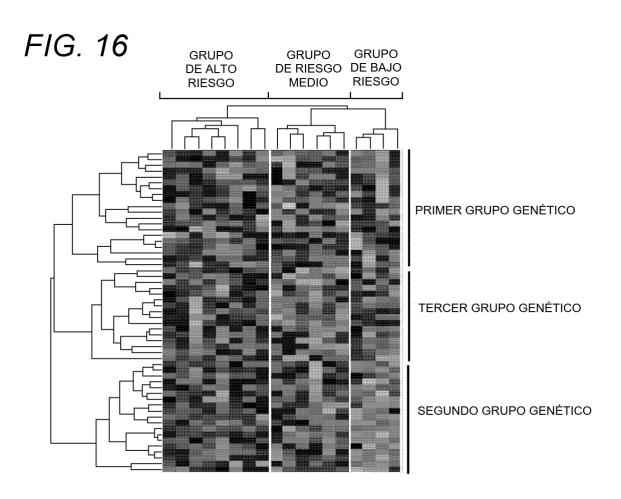


FIG. 17

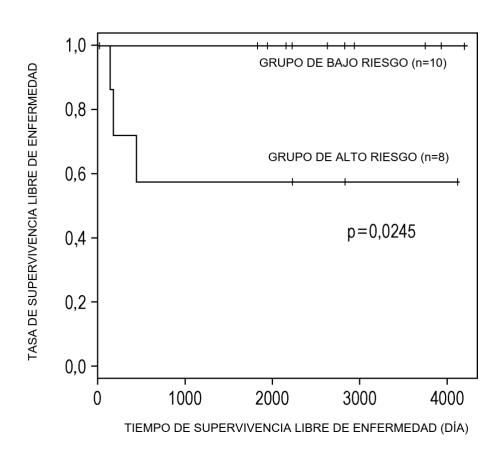


FIG. 18

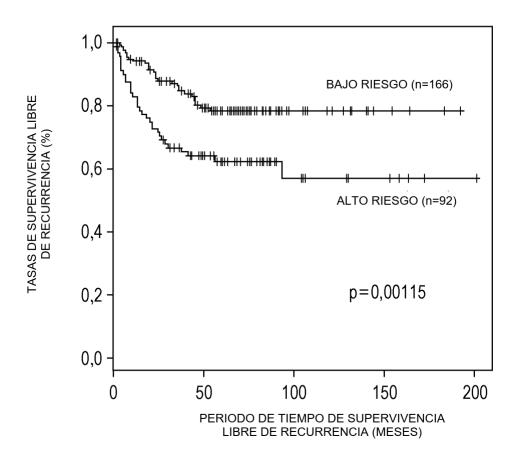


FIG. 19

