

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 714 582**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.07.2012 E 16184484 (0)**

97 Fecha y número de publicación de la concesión europea: **12.12.2018 EP 3150720**

54 Título: **Procedimiento de predicción de la respuesta a la quimioterapia en un paciente que padece o está en riesgo de desarrollar un cáncer de mama recurrente**

30 Prioridad:

28.07.2011 EP 11175852

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

29.05.2019

73 Titular/es:

**MYRIAD INTERNATIONAL GMBH (100.0%)
Building S19, Nattermannallee 1
50829 Cologne, DE**

72 Inventor/es:

**GEHRMANN, MATHIAS;
WEBER, KARSTEN;
KRONENWETT, RALF;
PETRY, CHRISTOPH y
BRASE, JAN**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 714 582 T3

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

DESCRIPCIÓN

Procedimiento de predicción de la respuesta a la quimioterapia en un paciente que padece o está en riesgo de desarrollar un cáncer de mama recurrente

Campo de la técnica

- 5 La presente invención se refiere a procedimientos para la predicción de la respuesta de un tumoral a la quimioterapia. Más específicamente, la presente invención se refiere a la predicción de la respuesta a agentes quimioterápicos en particular, pero sin limitarse a un conjunto neoadyuvante, basándose en las mediciones de los niveles de expresión genética en muestras tumorales de pacientes con cáncer de mama.

Antecedentes de la invención

- 10 El cáncer de mama es el tipo de tumor más común y una de las causas principales de muertes relacionadas con el cáncer en las mujeres (Jemal y col., CA Cancer J Clin., 2011). Se estima que cada diez mujeres una desarrollará un cáncer de mama en su vida. Aunque la incidencia ha aumentado a lo largo de los años, la mortalidad ha disminuido constantemente debido a los avances en la detección precoz y el desarrollo de nuevas estrategias eficaces de tratamiento.

- 15 Los pacientes de cáncer de mama se tratan frecuentemente con radioterapia, terapia hormonal o quimioterapia citotóxica después de la cirugía (tratamiento adyuvante) para controlar las células tumorales residuales y reducir el riesgo de recurrencia. La quimioterapia incluye el uso combinado de varios agentes citotóxicos, aunque la antraciclina y las estrategias de tratamiento basadas en taxanos han demostrado ser superiores en comparación con otras terapias de combinación convencionales (Misset y col., J Clin Oncol., 1996, Henderson y col., J Clin Oncol., 2003).

- 20 Se aplica comúnmente una quimioterapia sistémica para reducir la probabilidad de recurrencia en los positivos a HER2/neu y en tumores que carecen la expresión del receptor estrogénico y el receptor de HER2/neu (triple negativo, básico). La decisión de tratamiento más desafiante concierne a los tumores luminales (positivos al receptor estrogénicos y negativos a HER2/neu) para los que los factores clínicos clásicos como gradación, tamaño tumoral o implicación de ganglios linfáticos no proporcionan una clara respuesta a la cuestión de si utilizar quimioterapia o no.

- 25 Para reducir el número de pacientes que padecen graves efectos secundarios sin un claro beneficio de la terapia sistémica, existe una gran necesidad de nuevos biomarcadores moleculares para predecir la sensibilidad a la quimioterapia y permitir de esta manera una estrategia de tratamiento más a medida.

- 30 La quimioterapia también se puede aplicar en el conjunto neoadyuvante (preoperatorio) en el que los pacientes de cáncer de mama reciben terapia sistémica antes de que se hayan eliminado las células tumorales restantes por cirugía. La quimioterapia neoadyuvante del cáncer de mama precoz da lugar a tasas de respuesta clínica altas del 70-90 %. Sin embargo, en la mayoría de los que responden clínicamente, la evaluación patológica del resto del tumor revela la presencia de focos celulares tumorales residuales. Una erradicación completa de las células cancerosas en la mama y ganglios linfáticos después del tratamiento neoadyuvante se llama una respuesta patológica completa (pCR) y se observa solo en un 10-25 % de todos los pacientes. La pCR es un marcador sustituto apropiado de la supervivencia libre de enfermedad y un fuerte indicador del beneficio de la quimioterapia.

- 35 La estrategia de tratamiento preoperatorio proporciona la oportunidad de evaluar directamente la respuesta de un tumor particular a la terapia aplicada: la reducción de la masa tumoral en respuesta a la terapia se puede monitorizar directamente. Para los pacientes con una baja probabilidad de respuesta, se deberían considerar otras estrategias terapéuticas. Se puede analizar biomarcadores para las biopsias centrales preterapéuticas para identificar los marcadores predictivos más valiosos. Una estrategia común es aislar el ARN de las biopsias centrales para el análisis de la expresión genética antes de la terapia neoadyuvante. A continuación, el éxito terapéutico se puede evaluar por reducción del tumor y correlacionarse con los datos de expresión genética.

- 40 Se ha demostrado que los ensayos multigenéticos predictivos tales como DLDA30 (Hess y col., J Clin Oncol., 2006) proporcionan información más allá de los parámetros clínicos como la gradación del tumor y el estado de receptores hormonales en los pacientes de cáncer de mama tratados con terapia neoadyuvante. Sin embargo, el ensayo multigenético predictivo DLDA30 se estableció sin considerar el estado del receptor estrogénico. Por lo tanto, el ensayo puede reflejar diferencias fenotípicas entre el los que responden completamente y los que no responden, siendo los que responden predominantemente negativos a ER y positivos a HER2/neu (Tabchy y col., Clin Can Res, 2010).

- 45 Adicionalmente, los ensayos multigenéticos establecidos para el pronóstico se analizaron en el conjunto neoadyuvante para evaluar si los ensayos pronósticos podían predecir también la quimiosensibilidad. Un ejemplo, es el Índice de Grado Genómico (GGI), un ensayo multigenético para definir el grado histológico basándose en los perfiles de expresión genética (Sotiriou y col., JNCI, 2006). Se demostró por Liedtke y sus colaboradores que un GGI alto se asocia con un aumento de quimiosensibilidad en pacientes con cáncer de mama tratados con terapia neoadyuvante (Liedtke, J Clin Oncol, 2009).

Aunque se ha demostrado que las firmas genéticas predicen la respuesta a la terapia, los estudios de validación a gran escala que incluyen los datos de seguimiento clínico se han omitido y hasta ahora ninguno de ellos se utiliza comúnmente para guiar las decisiones de tratamiento en la rutina clínica.

5 El documento WO 2010/076322 A1 desvela un procedimiento para la predicción de una respuesta Y/o el beneficio de la quimioterapia en un paciente que padece un cáncer que comprende las etapas de (i) clasificar un tumor en al menos dos clases, (ii) determinar en una muestra tumoral la expresión de al menos un marcador genético indicativo de una respuesta a la quimioterapia para un tumor de cada clase respectiva, (iii) dependiendo de dicha expresión genética, predecir dicha respuesta y/o beneficio; en el que dicho al menos un marcador genético comprende un gen seleccionado de entre el grupo que consiste en TMSL8, ABCC1, EGFR, MVP, ACOX2, HER2/NEU, MYH11, TOB1, AKR1C1, ERBB4, NFKB1A, TOP2A, AKR1C3, ESR1, OLFM1, TOP2B, ALCAM, FRAP1, PGR, TP53, BCL2, GADD45A, PRKAB1, TUBA1A, C16orf45, HIF1A, PTPRC, TUBB, CA12, IGKC, RACGAP1, UBE2C, CD14, 1KBKB, S100A7, VEGFA, CD247, KRT5, SEPT8, YBX1, CD3D, MAPK3, SLC2A1, CDKN1A, MAPT, SLC7A8, CHPT1, MLPH, SPON1, CXCL13, MMP1, STAT1, CXCL9, MMP7, STC2, DCN, MUC1, STMN1 y combinaciones de los mismos.

15 Maia Chanrion y col., exponen en Clin Cancer Res 2008; 14(6) March 15, 2008, p. 1744-1752 una firma de expresión genética que puede predecir la recurrencia de un cáncer de mama primario tratado con tamoxifeno. El estudio desvelado identifica una firma molecular que especifica un subgrupo de pacientes que no obtienen un beneficio con el tratamiento con tamoxifeno. Estos pacientes pueden elegirse por lo tanto para terapias endocrinas alternativas y/o quimioterapia.

20 El documento WO 2009/158143A1 desvela procedimientos para la clasificación y evaluación del pronóstico de un sujeto que tiene cáncer de mama. Los procedimientos incluyen la predicción de un subtipo de cáncer de mama utilizando un algoritmo supervisado, preparado para estratificar sujetos basándose en el subtipo intrínseco de cáncer de mama. El modelo de predicción se basa en el perfil de expresión genética de los genes intrínsecos que se enumeran en la Tabla 1. Este modelo de predicción se puede utilizar para predecir con precisión el subtipo intrínseco de un sujeto diagnosticado o sospechoso de tener un cáncer de mama. Se proporcionan adicionalmente composiciones y procedimientos para la predicción del resultado o respuesta a la terapia en una sujeto diagnosticado o sospechoso de tener un cáncer de mama. Estos procedimientos son útiles para la guía o determinación de opciones de tratamiento para un sujeto que padece cáncer de mama. Los procedimientos de la invención incluyen adicionalmente medios para la evaluación de los perfiles de expresión genética, incluyendo micromatrices y ensayos de reacción en cadena de la polimerasa, así como kits que comprenden los reactivos para la práctica de los procedimientos de la invención.

El documento WO 2006/119593 desvela procedimientos y sistemas para la determinación pronóstica en muestras tumorales, midiendo la expresión genética en una muestra tumoral y aplicando un índice del grado de expresión genética (GGI) o una puntuación de recaída (RS) para dar lugar a una puntuación numérica del riesgo.

35 Karen J Taylor y col., exponen en Breast Cancer Research 2010, 12: R39 cambios dinámicos en la expresión genética *in vivo* para predecir el pronóstico de pacientes con cáncer de mama tratados con tamoxifeno.

40 El documento WO 2008/006517A2 desvela procedimientos y kits para la predicción de un resultado probable de la quimioterapia en un paciente con cáncer. Más específicamente, la invención se refiere a la predicción de la respuesta tumoral a la quimioterapia basándose en mediciones de los niveles de expresión de un pequeño conjunto de marcadores genéticos. El conjunto de marcadores genéticos es útil para la identificación de subtipos de cáncer de mama que responden a la quimioterapia basada en taxanos, tales como, por ejemplo, una quimioterapia basada en taxano-antraciclina-ciclofosfamida (por ejemplo, basado en Taxotere (docetaxel)-Adriamicina (doxorubicina)-ciclofosfamida, es decir, (TAC)).

45 El documento WO 2009/114836 A1 desvela conjuntos de genes que son útiles para evaluar el pronóstico y/o predecir la respuesta del cáncer, por ejemplo, el cáncer colorrectal a la quimioterapia. También se desvela un ensayo del cáncer validado clínicamente, por ejemplo, un ensayo colorrectal, para la evaluación del pronóstico y/o predicción de la respuesta del paciente a la quimioterapia, utilizando un análisis de la expresión. El uso de material de biopsia archivado embebido en parafina para el ensayo de todos los marcadores en los conjuntos de genes relevantes se acomoda para, y por lo tanto es compatible con el tipo más ampliamente disponible de material de biopsia.

55 El documento WO 2011/120984 A1 desvela procedimientos, kits y sistemas para el pronóstico del resultado de una enfermedad para el cáncer de mama, comprendiendo dicho procedimiento: (a) la determinación en una muestra tumoral de dicho paciente los niveles de expresión de ARN de al menos 2 de los siguientes 9 genes: UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP; (b) la combinación matemática de los valores de niveles de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para dar lugar a una puntuación combinada, en la que dicha puntuación combinada es indicativa de un pronóstico de dicho paciente; y kits y sistemas para llevar a cabo dicho procedimiento.

Definiciones

A menos de que se defina otra cosa, los términos técnicos y científicos que se utilizan en el presente documento tienen el mismo significado que entiende comúnmente un experto habituado en la técnica a la que pertenece la presente invención.

5 “Predicción de la respuesta a la quimioterapia”, en el significado de la invención, se entenderá que es el acto de la determinación de un resultado probable de quimioterapia citotóxica en un paciente afectado por un cáncer. La predicción de una respuesta se hace preferentemente en referencia a valores de probabilidad para conseguir un resultado deseado o no deseado de la quimioterapia. Los procedimientos predictivos de la presente invención se pueden utilizar clínicamente para tomar decisiones de tratamiento eligiendo las modalidades de tratamiento más apropiadas para un paciente en particular.

10 La “respuesta de un tumor a la quimioterapia”, en el significado de la invención, se refiere a cualquier respuesta del tumor a la quimioterapia citotóxica, preferentemente a un cambio de la masa y/o volumen del tumor después del inicio de la quimioterapia neoadyuvante y/o prolongación de tiempo de las metástasis distantes o el tiempo de muerte a continuación de la quimioterapia neoadyuvante o adyuvante. La respuesta del tumor se puede evaluar en una situación neoadyuvante en la que el tamaño de un tumor después de la intervención sistémica se puede
15 comparar con el tamaño y dimensiones iniciales según se mide por CT, PET, mamografía, ultrasonidos o palpación, habitualmente registrado como la “respuesta clínica” de un paciente. La respuesta también se puede evaluar por medición con un calibre o el examen patológico del tumor después de la biopsia o la resección quirúrgica. La respuesta se puede registrar de una manera cuantitativa como el porcentaje de cambio en el volumen tumoral o de una manera cualitativa como “sin cambios” (NC), “remisión parcial” (PR), “remisión completa” (CR) u otros criterios
20 cualitativos. La evaluación de la respuesta tumoral se puede hacer pronto después del inicio de la terapia neoadyuvante, por ejemplo, después de pocas horas, días, semanas o preferentemente después de pocos meses. Un punto final típico para la evaluación de la respuesta es al terminar la quimioterapia neoadyuvante o a la retirada quirúrgica de las células tumorales residuales y/o el lecho tumoral. Esto es normalmente tres meses después del inicio de la terapia neoadyuvante. La respuesta puede evaluarse también comparando el tiempo hasta las
25 metástasis distantes o la muerte de un paciente después de la quimioterapia neoadyuvante o adyuvante con el tiempo hasta las metástasis distantes o la muerte de un paciente no tratado con quimioterapia.

El término “tumor” como se utiliza en el presente documento, se refiere a todo crecimiento celular neoplásico, sea maligno o benigno, y todos los tejidos y células cancerosos y precancerosos.

30 El término “cáncer” se refiere o describe la afección fisiológica en mamíferos que se caracteriza normalmente por un crecimiento celular sin regulación. El término “cáncer” como se utiliza en el presente documento incluye carcinomas (por ejemplo, carcinomas *in situ*, carcinoma invasivo, carcinoma metastático) y afecciones premalignas, cambios neomórficos independientemente de su origen histológicos. El término “cáncer” no se limita a un estadio, grado, característica histomorfológica, invasividad, agresividad o malignidad de un tejido o agregación celular afectados. En particular, el cáncer en estadio 0, el cáncer en estadio I, el cáncer en estadio II, el cáncer en estadio III, el cáncer en
35 estadio IV, el cáncer de grado I, el cáncer de grado II, el cáncer de grado III, el cáncer maligno y los carcinomas primarios están incluidos.

La expresión “quimioterapia citotóxica” se refiere a distintas modalidades de tratamiento que afectan a la proliferación y/o la supervivencia celular. El tratamiento puede incluir la administración de agentes alquilantes, antimetabolitos, antraciclinas, alcaloides vegetales, inhibidores de la topoisomerasa y otros agentes antitumorales,
40 incluyendo anticuerpos monoclonales inhibidores de la cinasa. En particular, el tratamiento citotóxico se puede referir a un tratamiento con taxano. Los taxanos son alcaloides vegetales que bloquean la división celular evitando la función de microtúbulos. El taxano prototípico es el producto natural paclitaxel, conocido originalmente como taxol y derivado por primera vez de la corteza del tejo del pacífico. El docetaxel es un análogo semisintético del paclitaxel. Los taxanos aumentan la estabilidad de los microtúbulos, evitando la separación de cromosomas durante la anafase.

45 El término “terapia” se refiere a una administración secuencial o simultánea en el tiempo de un antitumoral, y/o antivasculador, y/o anti-estroma, y/o estimulante o supresor inmunitario, y/o agentes proliferativos de células de la sangre, y/o radioterapia, y/o hipertermia, y/o hipotermia para la terapia del cáncer. La administración de estos se puede llevar a cabo de un modo adyuvante o neoadyuvante. La composición de dicho “protocolo” puede variar en la dosis de cada uno de los agentes individuales, marco de tiempo de aplicación y frecuencia de administración en una
50 ventana de terapia definida. Actualmente, está en investigación distintas combinaciones de distintos fármacos y/o procedimientos físicos, y distintos calendarios. Una “quimioterapia que contiene taxano/antraciclina” es una modalidad de terapia que comprende la administración de taxano y/o antraciclina y derivados terapéuticamente eficaces de los mismos.

La expresión “quimioterapia neoadyuvante” se refiere a un régimen de terapia preoperatoria que consiste en un
55 panel de agentes hormonales, quimioterápicos y/o anticuerpos, que tienen el objetivo de reducir el tumor primario, haciendo de esta manera que la terapia local (cirugía o radioterapia) menos destructiva o más eficaz, haciendo posible una cirugía conservadora del pecho y la evaluación de la respuesta de la sensibilidad del tumor ha agentes específicos *in vivo*.

La expresión “implicación de ganglios linfáticos” significa un paciente que se ha diagnosticado previamente de una

- metástasis a los ganglios linfáticos. Se englobará la metástasis tanto los ganglios linfáticos de drenaje, los ganglios linfáticos cercanos y los ganglios linfáticos lejanos. Este diagnóstico previo por sí mismo no forma parte del procedimiento inventivo. Más bien es una precondition para seleccionar pacientes cuyas muestras se pueden utilizar para una realización de la presente invención. Se puede llegar a este diagnóstico previo por cualquier procedimiento adecuado conocido en la técnica, incluyendo, pero sin limitarse a retirada del ganglio linfático y análisis patológico, análisis de biopsia, análisis *in vitro* de biomarcadores indicativos de la metástasis, procedimientos de creación de imágenes (por ejemplo, tomografía computarizada, rayos X, creación de imágenes por resonancia magnética, ultrasonidos), y hallazgos intraoperatorios.
- 5
- La expresión “respuesta patológica completa” (pCR), como se utiliza en el presente documento, se refiere a una desaparición completa o ausencia de células tumorales invasivas en la mama y/o ganglios linfáticos según se evalúa mediante examen histopatológico de un espécimen quirúrgico después de la quimioterapia neoadyuvante.
- 10
- El término “marcador” o “biomarcador” se refiere a una molécula biológica, por ejemplo, un ácido nucleico, péptido, proteína, hormona, etc., cuya presencia o concentración se puede detectar y correlacionar con una afección conocida tal como un estado de enfermedad.
- 15
- La expresión “marcador predictivo” se refiere a un marcador que se puede utilizar para predecir la respuesta clínica de un paciente hacia un tratamiento determinado.
- El término “pronóstico”, como se utiliza en el presente documento, se refiere a una evaluación individual de la malignidad de un tumor, o la respuesta esperada si no existe una terapia farmacológica. Por el contrario, el término “predicción” se refiere a una evaluación individual de la malignidad de un tumor, o la respuesta esperada si la terapia contiene un fármaco en comparación con la malignidad o respuesta sin este fármaco.
- 20
- El término “inmunohistoquímica” o IHC se refiere al procedimiento de localización de proteínas en células de una sección de tejido que aprovecha el principio de que los anticuerpos se unen específicamente a antígenos de los tejidos biológicos. La tinción inmunohistoquímica se utiliza ampliamente en el diagnóstico y tratamiento del cáncer. Los marcadores moleculares específicos son característicos de tipos de cáncer particulares. La IHC también se utiliza ampliamente en la investigación básica para entender la distribución y localización de biomarcadores en diferentes partes de un tejido.
- 25
- El término “muestra”, como se utiliza en el presente documento, se refiere a una muestra obtenida de un paciente. La muestra puede ser cualquier tejido o fluido biológico. Dichas muestras incluyen, pero no se limitan a, esputo, sangre, suero, plasma, células sanguíneas (por ejemplo, leucocitos), tejido, muestras de biopsia centrales o por aguja fina, fluidos corporales que contienen células, ácidos nucleicos flotantes libres, orina, líquido peritoneal, líquido pleural, o células de estos. Las muestras biológicas también pueden incluir secciones de tejidos tales como secciones congeladas o fijadas que se toman con fines histológicos o células microdisecadas o partes extracelulares de las mismas. Una muestra biológica que se va a analizar es un material tisular de una lesión neoplásica que se toma por aspiración o punción, escisión o por cualquiera otro procedimiento quirúrgico que da lugar a una biopsia o material celular resecado. Dicha muestra biológica puede comprender células obtenidas de un paciente. Las células se pueden encontrar en un “frotis” celular, por ejemplo, por aspiración del pezón, lavado ductal, biopsia con aguja fina o de una descarga espontánea del pezón. En otra realización, la muestra es un fluido corporal. Dichos fluidos incluyen, por ejemplo, fluidos sanguíneos, suero, plasma, linfa, fluidos ascíticos, fluidos ginecológicos, u orina, pero no se limita a estos fluidos.
- 30
- Una “muestra tumoral” es una muestra que contiene material tumoral por ejemplo material tisular de una lesión neoplásica tomada por aspiración o punción, escisión o por cualquier otro procedimiento quirúrgico que da lugar a la biopsia o material celular resecado, incluyendo el material conservado tal como el material congelado reciente, material fijado con formalina, material embebido en parafina, y similares. Dicha muestra biológica puede comprender celular obtenidas de un paciente. Las células se pueden encontrar en un “frotis” celular recolectado mediante una aspiración en el pezón, lavado ductal, biopsia de aguja fina o a partir de una descarga del pezón provocada o espontánea. En otra realización, la muestra en un fluido corporal. Dichos fluidos incluyen, por ejemplo, diluidos sanguíneos, suero, plasma, linfa, fluidos ascíticos, fluidos ginecológicos, u orina, pero no se limita a estos fluidos.
- 35
- La expresión “niveles de expresión combinados matemáticamente”, en el significado de la invención se entenderá como la derivación de un valor numérico a partir de un nivel de expresión determinado de un gen y la aplicación de un algoritmo a uno o más de dichos valores numéricos para obtener un valor numérico combinado o puntuación combinada.
- 50
- Una “puntuación” en el significado de la invención se entenderá como un valor numérico, que se relaciona con el resultado de una enfermedad del paciente y/o respuesta de un tumor a la quimioterapia. El valor numérico se deriva combinando los niveles de expresión de marcadores genéticos utilizando coeficientes pre-especificados en un algoritmo matemático. Los niveles de expresión se pueden emplear como valores CT o CT delta obtenidos por una RT-PCR cinética, como los valores de fluorescencia relativa o absoluta obtenidos mediante micromatrices o mediante cualquier procedimiento útil para cuantificar niveles absolutos o relativos de ARN. La combinación de estos niveles de expresión se puede conseguir, por ejemplo, multiplicando cada nivel de expresión con un coeficiente
- 55

definido o especificado y sumando dichos productos para dar lugar a una puntuación. La puntuación se puede derivar también de los niveles de expresión junto con otra información, por ejemplo, de datos clínicos tales como el tamaño tumoral, estado de ganglios linfáticos o gradación del tumor ya que dichas variables se pueden codificar también como números en una ecuación. La puntuación se puede utilizar en una escala continua para predecir la respuesta de un tumor a la quimioterapia y/o el resultado de la enfermedad de un paciente. Se pueden aplicar valores de corte para distinguir subgrupos clínicos relevantes. Los valores de corte de dichas puntuaciones se pueden determinar de la misma manera que los valores de corte para los marcadores diagnósticos convencionales y se conocen bien por los expertos en la técnica. Una manera útil para la determinación de dicho valor de corte es construir una curva de eficacia diagnóstica (curva ROC) basándose en todos los valores de corte concebibles, determinar el punto único en la curva ROC con la proximidad más estrecha con la esquina izquierda superior (0/1) en el gráfico ROC. Obviamente, la mayoría de las veces, los valores de corte se determinarán por procedimientos menos formalizados escogiendo la combinación de sensibilidad y especificidad determinada por dicho valor de corte que proporciona la información médica más beneficiosa para el problema investigado.

La expresión “un procedimiento basado en PCR” como se utiliza en el presente documento se refiere a procedimientos que comprenden una reacción en cadena de polimerasa (PCR). Esta es una estrategia para amplificar exponencialmente ácidos nucleicos, como el ADN y el ARN, mediante replicación enzimática, sin utilizar un organismo vivo. Como la PCR es una técnica *in vitro*, se puede llevar a cabo sin restricciones en la forma del ADN, y se puede modificar extensamente para llevar a cabo una amplia matriz de modificaciones genéticas. Cuando se hace para determinar los niveles de expresión, un procedimiento basado en PCR puede utilizarse, por ejemplo, para detectar la presencia de un ARNm determinado mediante (1) transcripción inversa del agrupamiento del ARNm completo (el denominado transcriptoma) en ADNc que la ayuda de la enzima transcriptasa inversa, y (2) detectar la presencia de un ADNc determinado con ayuda de los cebadores respectivos. Esta estrategia es conocida comúnmente como PCR de transcriptasa inversa (rtPCR). Además, los procedimientos basados en PCR comprenden, por ejemplo, una PCR en tiempo real, y, particularmente adecuada para los análisis de los niveles de expresión una PCR cinética o cualitativa (qPCR).

Una “micromatriz”, a la que se hace referencia también en el presente documento como “biochip” o “chip biológico” es una matriz de regiones que tienen una densidad de regiones separadas de al menos aproximadamente 100/cm², y preferentemente al menos aproximadamente 1000/cm². Las regiones en una micromatriz tienen dimensiones típicas, por ejemplo, diámetros en el intervalo de entre aproximadamente 10-250 nm, y se separan de otras regiones de la matriz por aproximadamente la misma distancia.

La expresión “procedimiento basado en hibridación”, como se utiliza en el presente documento, se refiere a procedimientos que imparten un procedimiento de combinación complementaria de ácidos nucleicos de cadena sencilla o análogos de nucleótidos en una única molécula de doble cadena. Los nucleótidos o análogos de nucleótidos se unirán a su complementario en condiciones normales, de manera que dos cadenas perfectamente complementarias se unirán entre ellas fácilmente. En bioanalítica, hay sondas de cadena sencilla, muy a menudo marcadas, con el fin de encontrar secuencias diana complementarias. Si dichas secuencias existente en la muestra, las sondas se hibridarán con dichas secuencias que se pueden detectar debido al marcador. Otros procedimientos basados en hibridación comprenden procedimientos de micromatriz y/o biochip. Aquí las sondas están inmovilizadas en una fase sólida que entonces se exponen a una muestra. Si existen ácidos nucleicos complementarios en la muestra, se hibridarán con las sondas y de esta manera se detectarán. Estas estrategias se conocen también como “procedimientos basados en matriz”. Otro procedimiento basado en hibridación más es la PCR, que se ha descrito anteriormente. Cuando es para la determinación de niveles de expresión. Los procedimientos basados en hibridación pueden utilizarse, por ejemplo, para determinar la cantidad de ARNm para un gen determinado.

La expresión “marcador genético” como se utiliza en el presente documento, se refiere a un gen expresado diferencialmente cuyo patrón de expresión se puede utilizar como parte de un procedimiento predictivo, pronóstico o diagnóstico en la evaluación de la neoplasia maligna o cáncer, o que, de manera alternativa se puede utilizar en procedimientos para la identificación de compuestos útiles para el tratamiento o prevención de neoplasias malignas y cáncer de cabeza y cuello, colon o mama en particular. Un marcador genético puede tener también las características de un gen diana.

Un “algoritmo” es un procedimiento que lleva a cabo una secuencia de operaciones para producir una información.

La expresión “medición a nivel proteico”, como se utiliza en el presente documento, se refiere a procedimientos que permiten la determinación cuantitativa y/o cualitativa de una o más proteínas en una muestra. Estos procedimientos incluyen, entre otros, la purificación de proteínas, incluyendo la ultracentrifugación, precipitación y cromatografía, así como el análisis proteico y la determinación, incluyendo la inmunohistoquímica, inmunofluorescencia, ELISA (inmunoensayo ligado a enzimas), RIA (radioinmunoensayo) o el uso de micromatrices proteicas, exploración de dos híbridos, procedimientos de transferencia que incluyen la transferencia de Western, electroforesis en gel de una o dos dimensiones, enfoque isoeléctrico así como procedimientos que se basan en espectrometría de masas como MALDI-TOF y similares.

La expresión “PCR cinética” o “PCR cuantitativa” (qPCR) se refiere a cualquier tipo del procedimiento PCR que permita la cuantificación de la matriz en una muestra. La PCR cuantitativa en tiempo real comprende diferentes

técnicas de actuación o detección del producto como, por ejemplo, la técnica TaqMan o la técnica LightCycler. La técnica TaqMan, por ejemplo, utiliza una sonda fluorogénica marcada dual. La PCR TaqMan en tiempo real mide la acumulación de un producto mediante el fluoróforo durante los estadios exponenciales de la PCR, más que en el punto final como en la PCR convencional. El aumento exponencial del producto se utiliza para determinar el umbral de ciclos, CT, es decir, el número de ciclos de PCR en el que se detecta un aumento exponencial significativo de la fluorescencia, y que se correlaciona directamente con el número de copias de la matriz de ADN presente en la reacción. La construcción de la reacción es muy similar a la PCR convencional, pero se lleva a cabo en un ciclador térmico en tiempo real que permite la medición de moléculas fluorescentes en los tubos de PCR. A diferencia de la PCR regular, en la PCR TaqMan en tiempo real se añade una sonda a la reacción, es decir, un oligonucleótido de cadena sencilla complementario de un segmento de 20-60 nucleótidos de la matriz de ADN y localizado entre los dos cebadores. Se une covalentemente un indicador fluorescente o fluoróforo (por ejemplo, 6-carboxifluoresceína, con acrónimo: FAM, o tetraclorofluoresceína de acrónimo TET) y un interruptor (por ejemplo, tetrametilrodamina, con el acrónimo TAMRA, o dihidrociclopírrrolindol tripéptido "aglutinante de agrupamiento menor", con acrónimo MGB) al extremo 5' y 3' de la sonda, respectivamente [2]. La estrecha proximidad entre el fluoróforo y el interruptor unido a la sonda inhibe la fluorescencia del fluoróforo. Durante la PCR, según comienza la síntesis de ADN, la actividad de la exonucleasa de 5' a 3' de la polimerasa Taq degrada la proporción de la sonda que se ha hibridado a la matriz (de ahí su nombre: Taq polimerasa + PacMan). La degradación de la sonda libera el fluoróforo de esta y rompe la estrecha proximidad al interruptor, aminorando de esta manera el efecto de interrupción y permite la fluorescencia de fluoróforo. Por lo tanto, la fluorescencia detectada en el ciclador término de PCR en tiempo real es directamente proporcional al fluoróforo liberado y la cantidad de matriz de ADN presente en la PCR.

"Cebador" y "sondas", en el significado de la invención, tendrá el significado ordinario de este término que es bien conocido por el experto en la técnica de la biología molecular. En una realización preferida de la invención "cebador" y "sondas" se entenderá que son moléculas de polinucleótido que tienen una secuencia idéntica, complementaria, homóloga u homóloga a la complementaria de regiones de un polinucleótido diana que se va a detectar o cuantificar. En otra realización más, también están comprendidos los análogos de nucleótido y/o morfolinos para su uso como cebadores y/o sondas. Las "sondas marcadas individualmente", en el significado de la invención se entenderán como que son sondas moleculares que comprenden un polinucleótido, oligonucleótido o análogo de nucleótido y un marcador, útil en la detección o cuantificación de la sonda. Los marcadores preferidos son moléculas fluorescentes, moléculas luminiscentes, moléculas radioactivas, moléculas enzimáticas y/o moléculas interruptoras.

Objetivo de la invención

Un objetivo de la presente invención es proporcionar un procedimiento mejorado para la predicción de una respuesta de un tumor en un paciente que sufre o tienen el riesgo de desarrollar un cáncer de mama recurrente a la quimioterapia que contiene taxano-antraciclina.

Es otro objetivo de la presente invención evitar una quimioterapia citotóxica adyuvante o neoadyuvante innecesaria en pacientes que sufren especialmente un cáncer de mama.

Es otra divulgación ofrecer un sistema de ensayo diagnóstico más robusto y específico que la inmunohistoquímica convencional para las muestras de tejido fijado de rutina clínica que ayuden mejor al médico a seleccionar modalidades de tratamiento individualizado.

En una realización más preferida, el procedimiento desvelado se puede utilizar para seleccionar una terapia adecuada para una enfermedad neoplásica, particularmente cánceres de mama.

Es otra divulgación detectar nuevas dianas para fármacos dirigidos disponibles últimamente, o para determinar fármacos que ya se han desarrollado.

Sumario de la invención

Antes de describir la invención en detalle, se tiene que entender que la presente invención no se limita a las partes componentes particulares de los dispositivos descritos o etapas de procesamiento de los procedimientos descritos ya que dichos dispositivos y procedimientos pueden variar. también se tiene que entender que la terminología que se utilizan en el presente documento es con fines solamente de describir realizaciones particulares, y no se pretende que sean limitantes. Se debe señalar que, como se utiliza en la memoria descriptiva y las reivindicaciones adjuntas, las formas singulares de "un", "una" y "el" incluye el singular y/o los referentes en plural a menos de que el contexto dicte claramente otra cosa. Se tiene que entender además que, en el caso de que se den intervalos de parámetros que estén delimitados por valores numéricos, se considera que los intervalos incluyen estos valores de limitación.

Los problemas anteriores se resuelven con los procedimientos proporcionados por la invención.

El estado del receptor estrogénico se determina en general utilizando inmunohistoquímica. El estatus de HER2/neu (ERBB2) se determina generalmente utilizando inmunohistoquímica e hibridación fluorescente *in situ*. Sin embargo, el estado del receptor estrogénico y el estatus de HER2/neu (ERBB2) pueden, para los fines de la invención, determinarse por cualquier procedimiento adecuado, por ejemplo, inmunohistoquímica, hibridación fluorescente *in situ* (FISH), o análisis de expresión genética.

La presente invención se refiere a un procedimiento para predecir una respuesta y/o un beneficio a la quimioterapia que contiene taxano/antraciclina incluyendo la quimioterapia neoadyuvante como se define en la reivindicación 1.

5 El documento WO 2011/120984 A1 utiliza, sin embargo, los nueve genes para la predicción del resultado del cáncer de mama en un tumor positivo al receptor estrogénico y negativo a HER2 de un paciente de cáncer de mama, que no se relaciona con el procedimiento de la presente invención que predice una respuesta o el beneficio de la quimioterapia. Los genes de la presente invención se utilizan para un objetivo diferente.

En una divulgación, el procedimiento comprende:

10 (a) la determinación en una muestra tumoral de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP, indicativos de una respuesta a la quimioterapia para un tumor.

(b) la combinación matemática de los valores del nivel de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para dar lugar a una puntuación combinada, en el que dicha puntuación combinada es predictiva de dicha respuesta y/o beneficio de la quimioterapia.

En una divulgación adicional el procedimiento de la invención comprende:

15 (a) determinar en una muestra de dicho paciente los niveles de expresión de ARN de los siguientes 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP; indicativos de una respuesta a la quimioterapia para un tumor

20 **mientras que** BIRC5 puede sustituirse por UBE2C o TOP2A o RACGAP1 o AURKA o NEK2 o E2F8 o PCNA o CYBRD1 o DCN o ADRA2A o SQLE o CXCL12 o EPHX2 o ASPH o PRSS16 o EGFR o CCND1 o TRIM29 o DHCR7 o PIP o TFAP2B o WNT5A o APOD o PTPRT a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

mientras que UBE2C puede sustituirse por BIRC5 o RACGAP1 o TOP2A o AURKA o NEK2 o E2F8 o PCNA o CYBRD1 o ADRA2A o DCN o SQLE o CCND1 o ASPH o CXCL12 o PIP o PRSS16 o EGFR o DHCR7 o EPHX2 o TRIM29 a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

25 **mientras que** DHCR7 puede sustituirse por AURKA, BIRC5, UBE2C o por cualquier otro gen que pueda sustituir a BIRC5 o UBE2C a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

mientras que STC2 puede sustituirse por INPP4B o IL6ST o SEC14L2 o MAPT o CHPT1 o ABAT o SCUBE2 o ESR1 o RBBP8 o PGR o PTPRT o HSPA2 o PTGER3 a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

30 **mientras que** AZGP1 puede sustituirse por PIP o EPHX2 o PLAT o SEC14L2 o SCUBE2 o PGR a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

mientras que RBBP8 puede sustituirse por CELSR2 o PGR o STC2 o ABAT o IL6ST a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

35 **mientras que** IL6ST puede sustituirse por INPP4B o STC2 o MAPT o SCUBE2 o ABAT o PGR o SEC14L2 o ESR1 o GJA1 o MGP o EPHX2 o RBBP8 o PTPRT o PLAT a condición de que después de la sustitución se seleccionen 8 genes diferentes; y

mientras que MGP puede sustituirse por APOD o IL6ST o EGFR a condición de que después de la sustitución se seleccionen 8 genes diferentes;

40 (b) la combinación matemática de los valores del nivel de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para dar lugar a una puntuación combinada, en la que dicha puntuación combinada es predictiva de dicha respuesta y/o beneficio de la quimioterapia.

Los procedimientos de la invención son adecuados para predecir una respuesta a la quimioterapia que contiene taxano/antraciclina, preferentemente en tumores negativos a HER2/neu, positivos al receptor estrogénico (luminal), preferentemente en el modo neoadyuvante.

45 De acuerdo con un aspecto de la divulgación se proporciona un procedimiento como se ha descrito anteriormente, en el que dicho nivel de expresión se determina como un nivel de ARNm. De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicho nivel de expresión se determina como un nivel de expresión genética.

50 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicho nivel de expresión se determinar mediante al menos uno de

un procedimiento basado en PCR

un procedimiento basado en micromatrices

un procedimiento basado en hibridación, y

una estrategia de secuenciación y/o secuenciación de siguiente generación.

55 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en

el que dicha determinación de los niveles de expresión es una muestra tumoral fijada en formalina embebida en parafina o en una muestra tumoral fresca congelada.

5 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que el nivel de expresión de dicho al menos un gen marcador se determina como un patrón de expresión respecto a al menos un gen de referencia o un valor de expresión medio computado.

De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicha etapa de combinación matemática comprende una etapa de aplicación de un algoritmo a valores representativo de un nivel de expresión de un gen determinado.

10 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicho algoritmo es una combinación lineal de dichos valores representativos de un nivel de expresión de un gen determinado.

De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que un valor representativo de un nivel de expresión de un gen determinado se multiplica por un coeficiente.

15 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que se determina uno, dos o más umbrales para dicha puntuación combinada y se discrimina entre alto y bajo riesgo, alto, medio y bajo riesgo, o más grupos de riesgo aplicando el umbral de la puntuación combinada.

20 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que una puntuación combinada alta es indicativa del beneficio de una terapia más agresiva, por ejemplo, una quimioterapia citotóxica. El experto entiende que una "puntuación alta" a este respecto se refiere a un valor de referencia o valor de corte. El experto entiende adicionalmente que dependiendo del algoritmo particular que se utilice para obtener la puntuación combinada, también una puntuación "baja" por debajo de un valor de corte o de referencia puede ser indicativo del beneficio de una terapia más agresiva, por ejemplo, una quimioterapia citotóxica.

25 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que la información con respecto al estado ganglionar del paciente se procesa en la etapa de la combinación matemática de los valores de expresión de los genes para generar una puntuación combinada.

30 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicha información con respecto al estado ganglionar es un valor numérico ≤ 0 si dicho estado ganglionar es negativo y dicha información es un valor numérico > 0 si dicho estado ganglionar es positivo o desconocido. En realizaciones ejemplares de la invención se asigna al estado ganglionar negativo el valor 0, se asigna a un estado ganglionar desconocido el valor 0,5 y se asigna a un estado ganglionar positivo el valor 1. Se pueden escoger otros valores que reflejen una ponderación diferente del estado ganglionar en un algoritmo.

De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicha información con respecto al tamaño del tumor del paciente se procesa en la etapa de combinación matemática de los valores del nivel de expresión de los genes para generar una puntuación combinada.

35 De acuerdo con un aspecto de la invención se proporciona un procedimiento como se ha descrito anteriormente, en el que dicha información con respecto al estado ganglionar y el tamaño del tumor del paciente se procesa en la etapa de combinación matemática de los valores de nivel de expresión de los genes para generar una puntuación combinada.

40 La divulgación se refiere adicionalmente a un kit para llevar a cabo un procedimiento como se ha descrito anteriormente, comprendiendo dicho kit un conjunto de oligonucleótidos capaces de unirse específicamente a secuencias o fragmentos de secuencia de genes en una combinación de genes, en la que

(i) dicha combinación comprende al menos los 8 genes UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP; o

45 (ii) dicha combinación comprende al menos los 8 genes UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, y MGP.

La divulgación se refiere adicionalmente a un producto que es un programa de computadora capaz de procesar los valores representativos de un nivel de expresión de una combinación de genes combinando matemáticamente dichos valores para generar una puntuación combinada, en la que dicha puntuación combinada es predictiva de dicha respuesta y/o beneficio de la quimioterapia en dicho paciente.

50 Dicho producto que es un programa de computadora se almacena en un vehículo de datos o se ejecuta en una computadora, el personal operativo puede introducir los valores de expresión obtenidos del nivel de expresión de los genes respectivos. El producto que es un programa de computadora puede entonces aplicar un algoritmo para producir una puntuación combinada indicativa del beneficio de la quimioterapia citotóxica en un determinado paciente.

Si el producto que es un programa de computadora se almacena en un vehículo de datos o se ejecuta en una computadora, el personal operativo puede introducir los valores de expresión obtenidos del nivel de expresión de los genes respectivos. El producto que es un programa de computadora puede entonces aplicar un algoritmo para producir una puntuación combinada indicativa del beneficio de la quimioterapia citotóxica en un determinado paciente.

Los procedimientos de la presente invención tienen la ventaja de proporcionar una predicción fiable de la respuesta y/o el beneficio de la quimioterapia basada en el uso de solo un pequeño número de genes. Se ha descubierto procedimientos de la presente invención son especialmente adecuados para analizar la respuesta y/o beneficio de la quimioterapia de pacientes con tumores clasificados como positivos a ESR1 y negativos a ERBB2.

10 **Descripción detallada de la invención**

Se desvelan detalles, distintivos, características y ventajas adicionales del objeto de la invención en sub-reivindicaciones, y la siguiente descripción de las respectivas figuras y ejemplos, en que, de una manera ejemplar, presentan realizaciones preferidas de la presente invención. Sin embargo, estos dibujos no se deberían entender como que limiten el ámbito de la invención.

15 Se recuperaron cuatro conjuntos de datos de expresión genética disponibles públicamente (Affymetrix HG-U133A) del depósito de datos omnibus de expresión genética (GEO). Todos los pacientes de cáncer de mama analizados se trataron con quimioterapia neoadyuvante de antraciclina o taxano/antraciclina. Los archivos celulares de micromatrices eran MAS5 normalizado con un procedimiento a escala global y una intensidad de direccionamiento de 500. Se utilizó la respuesta patológica completa (pCR) como punto final primario para la evaluación de la respuesta al tratamiento. El análisis se llevó a cabo en todos los pacientes con cáncer de mama negativo a HER2/neu y en el subconjunto de pacientes de cáncer de mama positivos a ER, negativos a HER2 de acuerdo con niveles de corte pre-especificados (conjunto de sondas ERBB2 216836 < 6000 = negativo a HER2/neu, conjunto de sondas ERBB2 21836 < 6000 y conjunto de sondas ESR1 > 1000 = positivos a ER/negativos a HER2/neu).

25 Se examinó la puntuación T5 en 374 pacientes de cáncer de mama negativos a HER2 tratados con la terapia neoadyuvante (Figura 1). Entre los 374 pacientes, 63 tumores (un 16,8 %) se clasificaron con un bajo riesgo T5, mientras que 311 tumores (un 83,2 %) eran de alto riesgo T5. Solamente uno de los tumores de bajo riesgo T5 consiguió una pCR después de la terapia neoadyuvante, mientras que 84 de los 85 eventos de pCR se clasificaron como de alto riesgo T5. La sensibilidad de la puntuación T5 era del 99 % y el valor predictivo negativo del 98 % con un área bajo la curva de eficacia diagnóstica de 0,69 (Figura 1).

30 La Figura 1 muestra:

(a) Distribución de la puntuación T5 en 374 pacientes de cáncer de mama negativos a HER2/neu (85 eventos pCR frente a 289 muestras con enfermedad residual); Ensayo de Mann-Whitney de dos lados.

35 (b) Utilizando la puntuación 5 como el corte pre-especificado de T5, la sensibilidad era del 99 %, la especificidad del 21 %, el valor predictivo negativo del 98 %, y el valor predictivo positivos del 27 % con un área bajo la curva de eficacia diagnóstica de 0,69.

40 Se examinó la puntuación T5 en 221 pacientes de cáncer de mama negativos a HER2 tratados con terapia neoadyuvante (Figura 2). Entre los 221 pacientes, 61 tumores (un 27,6 %) se clasificaron como de bajo riesgo T5, mientras que 160 tumores (un 72,4 %) eran de alto riesgo T5. Solamente uno de los tumores de bajo riesgo T5 consiguieron una pCR después de la terapia neoadyuvante, mientras que 24 de los 25 eventos de pCR se clasificaron como de alto riesgo T5. La sensibilidad de la puntuación T5 era del 96 % y el valor predictivo negativo del 98 % con un área bajo la curva de eficacia diagnóstica de 0,73 (Figura 2).

La Figura 2 muestra:

45 (c) distribución de la puntuación T5 en 221 pacientes de cáncer de mama positivos al receptor estrogénico y negativos a HER2/neu (25 eventos de pCR vs. 196 muestras con enfermedad residual); Ensayo de Mann-Whitney de dos lados

(d) utilizando la puntuación 5 como el corte pre-especificado de T5, la sensibilidad era del 96 %, la especificidad del 30 %, el valor predictivo negativo del 98 %, y el valor predictivo positivo del 15 % con un área bajo la curva de eficacia diagnóstica de 0,73.

50 Se desvelan en el presente documento combinaciones únicas de marcadores genéticos que se pueden combinar en un algoritmo para el nuevo ensayo predictivo presentado aquí. Técnicamente, el procedimiento de la invención se puede practicar utilizando dos tecnologías: 1) El aislamiento del ARN total de tejido tumoral reciente o fijado y 2) una RT-PCR cuantitativa de los ácidos nucleicos aislados. De manera alternativa, se contempla medir los niveles de expresión utilizando tecnologías alternativas, por ejemplo, por micromatrices, en particular matrices Affymetrix U-133 o por medición de un nivel proteico.

- Los procedimientos de la invención se basan en la determinación cuantitativa de especies de ARN aisladas del tumor con el fin de obtener los valores de expresión y el posterior análisis bioinformático de dichos valores de expresión determinados. Las especies de ARN se pueden aislar de cualquier tipo de muestra tumoral, por ejemplo, muestras de biopsia, muestras de frotis, material tumoral resecado, tejido tumoral reciente congelado o de tejido tumoral fijado en formalina y embebido en parafina. Primero, se determinan los niveles de ARN de los genes que codifican combinaciones específicas de los genes UBE2C, BIRC5, DHCR7, RACGAP1, AURKA, PVALB, NMU, STC2, AZGP1, RBBP8, IL6ST, MGP, PTGER3, CXCL12, ABAT, CDH1, y PIP o combinaciones específicas de los mismos, como se indica. Basándose en estos valores de expresión se calcula una puntuación predictiva mediante una combinación matemática, por ejemplo, de acuerdo con las fórmulas T5, T1, T4 o T5b (véase posteriormente).
- 5 Una puntuación alta indica un aumento de la probabilidad de una respuesta patológica completa después del tratamiento con quimioterapia neoadyuvante, un valor de puntuación bajo indica una disminución en la probabilidad de desarrollar una respuesta patológica completa después del tratamiento neoadyuvante. En consecuencia, una alta puntuación también indica que el paciente es un paciente de alto riesgo que se beneficiará de una terapia más agresiva, por ejemplo, una quimioterapia citotóxica.
- 10 La Tabla 1, a continuación, muestra las combinaciones de genes utilizadas para cada algoritmo.

Tabla 1: Combinación de genes para los algoritmos respectivos

Gen	Algo_T1	Algo_T4	Algo_T5	Algo_T5b
UBE2C			X	
BIRC5	X	X	X	
DHCR7		X	X	X
RACGAP1		X		X
AURKA	X			
PVALB	X	X		
NMU	X			X
STC2	X	X	X	
AZGP1			X	X
RBBP8	X		X	X
IL6ST		X	X	X
MGP			X	X
PTGER3	X	X		
CXCL12	X	X		
ABAT		X		
CDH1	X			
PIP	X			

La Tabla 2, a continuación, muestra el ID del conjunto de sondas Affy y el mapeo del ID de diseño de TaqMan de los genes marcadores de la presente invención.

- 20 Tabla 2: Símbolo del gen, ID del conjunto de sondas Affy y mapeo del ID de diseño de TaqMan

Gen	ID del diseño	ID del conjunto de sondas
UBE2C	R65	202954_at
BIRC5	SC089	202095_s_at
DHCR7	CAGMC334	201791_s_at
RACGAP1	R125-2	222077_s_at
AURKA	CAGMC336	204092_s_at
PVALB	CAGMC339	205336_at
NMU	CAGMC331	206023_at
STC2	R52	203438_at
AZGP1	CAGMC372	209309_at
RBBP8	CAGMC347	203344_s_at
IL6ST	CAGMC312	212196_at
MGP	CAGMC383	202291_s_at
PTGER3	CAGMC315	213933_at
CXCL12	CAGMC342	209687_at
ABAT	CAGMC338	209460_at
CDH1	CAGMC335	201131_s_at

La Tabla 3, a continuación, presenta los nombres completos, el ID genético Entrez, el número de acceso de GenBank y la localización cromosómica de los genes marcadores de la presente invención.

Símbolo oficial	Nombre completo oficial	ID del gen Entrez	Número de acceso	Localización
UBE2C	Enzima E2C conjugada con ubiquitina	11065	U73379	20q13. 12
BIRC5	IAP baculovírica 5 que contiene repeticiones	332	U75285	17q25
DHCR7	7-deshidrocolesterol reductasa	1717	AF034544	11q13. 4
STC2	estaniocalcina 2	8614	AB012664	5q35.2
RBBP8	Proteína 8 de unión del retinoblastoma	5932	AF043431	18q11. 2
IL6ST	Transductor de señal de interleucina 6	3572	M57230	5q11
MGP	Proteína Gla de la matriz	4256	M58549	12p12. 3
AZGP1	alfa-2-glicoproteína 1, de unión al zinc	563	BC005306	11q22. 1
RACGAP1	Proteína activadora 1 de la Rac GTPasa	29127	NM_013277	12q13
AURKA	aurora cinasa A	6790	BC001280	20q13
PVALB	parvalbúmina	5816	NM_002854	22q13. 1
NMU	neuromedina U	10874	X76029	4q12
PTGER3	Receptor 3 de prostaglandina E (subtipo EP3)	5733	X83863	1p31.2
CXCL12	Ligando 12 de quimiocina (motivo C-X-C) (factor 1 derivado del estroma celular)	6387	L36033	10q11. 1
ABAT	4-aminobutirato aminotransferasa	18	L32961	16p13. 2
CDH1	caderina 1, tipo 1, E-caderina (epitelial)	999	L08599	16q22. 1
PIP	Proteína inducida por prolactina	5304	NMM_002652	7q32-qter

Ejemplo de algoritmo T5:

- 5 El algoritmo T5 es un comité de cuatro miembros donde cada miembro es una combinación lineal de dos genes. Las fórmulas matemáticas para T5 se muestran a continuación, la notación es la misma que para T1. T5 se puede calcular a partir solo de los datos de expresión genética.

$$\text{Miembro de riesgo 1} = 0,434039 [0,301..0,567] * (0,939 * \text{BIRC5} -3,831) -0,491845 [-0,714..-0,270] * (0,707 * \text{RBBP8} -0,934)$$

10 $\text{Miembro de riesgo 2} = 0,488785 [0,302..0,675] * (0,794 * \text{UBE2C} -1,416) -0,374702 [-0,570..-0,179] * (0,814 * \text{IL6ST} -5,034)$

$$\text{Miembro de riesgo 3} = -0,39169 [-0,541..-0,242] * (0,674 * \text{AZGP1} -0,777) + 0,44229 [0,256..0,628] * (0,891 * \text{DHCR7} -4,378)$$

$$\text{Miembro de riesgo 4} = -0,377752 [-0,543..-0,212] * (0,485 * \text{MGP} +4,330)$$

15 $-0,177669 [-0,267..-0,088] * (0,826 * \text{STC2} -3,630)$

$$\text{riesgo} = \text{Miembro de riesgo 1} + \text{Miembro de riesgo 2} + \text{Miembro de riesgo 3} + \text{Miembro de riesgo 4}$$

Los coeficientes de la izquierda de cada línea se calcularon como coeficientes de regresión de riesgos proporcionales COX, los números entre corchetes denotan una unión de confianza del 95 % para estos coeficientes. En otras palabras, en vez de multiplicar el término $(0,939 * \text{BIRC5} -3,831)$ con 0,434039, se puede multiplicar con cualquier coeficiente entre 0,301 y 0,567 y sigue dando un resultado predictivo en las uniones de confianza del 95 %. Los términos entre paréntesis a la derecha de cada línea denotan una transferencia de plataforma de la PCR a Affymetrix: Las variables PVALB, CDH1, ... denotan las expresiones basadas en PCR normalizadas por los genes de referencia (valores Ct delta), el término completo entre paréntesis corresponde con el logaritmo (de base 2) de los valores de expresión de la micromatriz Affymetrix de los conjuntos de sonda correspondientes.

- 25 Ejemplo de algoritmo T5clin:

El algoritmo T5clin es una puntuación combinada que consiste en la puntuación T5 y los parámetros clínicos (estado ganglionar y tamaño tumoral).

ES 2 714 582 T3

$$T5_{clin} = 0,35 * t + 0,64 * n + 0,28*s$$

Donde t codifica el tamaño tumoral (1: < 1 cm, 2: > 1 cm a < 2 cm, 3: > 2 cm a < 5 cm, 4: > 5 cm), y n el estado ganglionar (1: negativo, 2: 1 a 3 ganglios positivos, 3: 4 a 10 ganglios positivos, 4: > 10 ganglios positivos).

En una realización preferida, la puntuación del umbral de T5_{clin} es de 3,3.

5 Ejemplo de algoritmo T1:

El algoritmo T1 es un comité de tres miembros donde cada miembro es una combinación lineal de hasta cuatro variables. En general las variables pueden ser expresiones genéticas o variables clínicas. En T1 la única variable no genética es el estado ganglionar codificado como 0 si el paciente es negativo en los ganglios linfáticos y 1, si el paciente es positivo en los ganglios linfáticos. Las fórmulas matemáticas para T1 se muestran a continuación.

$$10 \quad \text{Miembro de riesgo 1} = +0,193935 [0,108..0,280] * (0,792 * PVALB -2,189) -0,240252 [-0,400..-0,080] * (0,859 * CDH1 -2,900) -0,270069 [-0,385..-0,155] * (0,821 * STC2 -3,529) + 1,2053 [0,534..1,877] * \text{estado ganglionar}$$

$$\text{Miembro de riesgo 2} = -0,25051 [-0,437..-0,064] * (0,558 * CXCL12 +0,324) -0,421992 [-0,687..-0,157] * (0,715 * RBBP8 -1,063) +0,148497 [0,029..0,268] * (1,823 * NMU -12,563) +0,293563 [0,108..0,479] * (0,989 * BIRC5 -4,536)$$

$$15 \quad \text{Miembro de riesgo 3} = +0,308391 [0,074..0,543] * (0,812 * AURKA-2,656) -0,225358 [-0,395..-0,055] * (0,637 * PTGER3 + 0,492) -0,116312 [-0,202..-0,031] * (0,724 * PIP + 0,985)$$

$$\text{riesgo} = + \text{Miembro de riesgo 1} + \text{Miembro de riesgo 2} + \text{Miembro de riesgo 3}$$

20 Los coeficientes de la izquierda de cada línea se calcularon como coeficientes de regresión de riesgos proporcionales COX, los números entre corchetes denotan uniones con un 95 % de confianza para estos coeficientes. Los términos entre paréntesis a la derecha de cada línea denotan una transferencia de plataforma de la PCR a Affymetrix: Las variables PVALB, CDH1, ... denotan las expresiones basadas en PCR normalizada por los genes de referencias, el término completo entre paréntesis corresponde con el logaritmo (de base 2) de los valores de expresión de micromatriz Affymetrix de los conjuntos de sondas correspondientes.

Ejemplo de algoritmo T4:

25 El algoritmo T4 es una combinación de motivos. Los 10 genes superiores de varios análisis de los conjuntos de datos Affymetrix y los datos de la PCR se agruparon en motivos. Los genes que no pertenecen a un agrupamiento se utilizaron como motivos genéticos únicos. Se encontraron los coeficientes de regresión de riesgos proporcionales en un análisis multivariado.

30 En general, los motivos pueden ser expresiones genéticas únicas o la media de expresiones genéticas de genes correlativos. Las fórmulas matemáticas para T4 se muestran a continuación.

$$\text{prolif} = ((0,84 [0,697..0,977] * RACGAP1 -2,174) + (0,85 [0,713..0,988] * DHCR7 -3,808) + (0,94 [0,786..1,089] * BIRC5 -3,734)) / 3$$

$$\text{motiv2} = ((0,83 [0,693..0,96] * IL6ST -5,295) + (1,11 [0,930..1,288] * ABAT -7,019) + (0,84 [0,701..0,972] * STC2 -3,857)) / 3$$

$$35 \quad \text{ptger3} = (\text{PTGER3} * 0,57 [0,475..0,659] + 1,436)$$

$$\text{cxcl2} = (\text{CXCL12} * 0,53 [0,446..0,618] + 0,847)$$

$$\text{pvalb} = (\text{PVALB} * 0,67 [0,558..0,774] -0,466)$$

40 Los factores y compensación de cada gen denotan una transferencia de plataforma de PCR a Affymetrix: Las variables RACGAP1, DHCR7, ..., denotan las expresiones basadas en PCR normalizadas por CALM2 y PPIA, el término completo entre paréntesis se corresponde con el logaritmo (de base 2) de los valores de expresión de micromatriz Affymetrix de los conjuntos de sondas correspondientes.

Los números entre corchetes denotan uniones del 95 % de confianza para estos factores.

45 Como el algoritmo se lleva a cabo aún mejor en combinación con una variable clínica se añadió el estado ganglionar. En T4 el estado ganglionar se codifica como 0, si los ganglios linfáticos del paciente son negativos y 1, si el paciente tiene los ganglios linfáticos positivos. Con esto, el algoritmo T4 es:

$$\text{riesgo} = -0,32 [-0,510..-0,137] * \text{motiv2}$$

$$+ 0,65 [0,411..0,886] * \text{prolif}$$

- 0,24 [-0,398..-0,08] * ptger3
- 0,05 [-0,225..0,131] * cxcll2
- + 0,09 [0,019..0,154] * pvalb
- + estado ganglionar

5 Los coeficientes del riesgo se calcularon como coeficientes de regresión de riesgos proporcionales, los números entre corchetes denotan las uniones con un 95 % de confianza para estos coeficientes.

El algoritmo T5b es un comité de dos miembros donde cada miembro es una combinación lineal de cuatro genes. Las fórmulas matemáticas de T5b se muestran a continuación, la notación es la misma que para T1 y T5. En T5b una variable no genética es el estado ganglionar, codificada como 0, si el paciente tiene los ganglios linfáticos negativos y 1, si el paciente tiene los ganglios linfáticos positivos y 0,5 si el estado de los ganglios linfáticos es desconocido. T5b se define por:

$$\begin{aligned} \text{Miembro de riesgo1} = & 0,359536 [0,153..0,566] * (0,891 * \text{DHCR7} -4,378) \\ & -0,288119 [-0,463..-0,113] * (0,485 * \text{MGP} + 4,330) \\ & + 0,257341 [0,112..0,403] * (1,118 * \text{NMU} -5,128) \\ & -0,337663 [-0,499..-0,176] * (0,674 * \text{AZGP1} -0,777) \end{aligned}$$

$$\begin{aligned} \text{Miembro de riesgo2} = & -0,374940 [-0,611..-0,139] * (0,707 * \text{RBBP8} -0,934) \\ & -0,387371 [-0,597..-0,178] * (0,814 * \text{IL6ST} -5,034) \\ & + 0,800745 [0,551..1,051] * (0,860 * \text{RACGAP1} -2,518) \\ & +0,770650 [0,323..1,219] * \text{Estado ganglionar} \end{aligned}$$

$$20 \quad \text{riesgo} = \text{Miembro de riesgo1} + \text{Miembro de riesgo2}$$

El experto entiende que estos algoritmos representan ejemplos particulares y que se basan en la información con respecto a la asociación de la expresión genética con la predicción de la respuesta terapéutica.

Simplificación del algoritmo empleando subconjuntos de genes

25 El "ejemplo de algoritmo T5" es un comité predictor que consiste en 4 miembros con 2 genes de interés cada uno. Cada miembro es un predictor independiente auto-contenido de recurrencia distante y/o respuesta a la terapia, cada miembro adicional contribuye a la robustez y potencia predictiva del algoritmo. La ecuación posterior muestra el "Ejemplo de algoritmo T5"; para facilitar la lectura se han reducido el número de decimales a 2; el intervalo entre corchetes enumera el intervalo estimado de los coeficientes (media +/- 3 desviaciones típicas).

Algoritmo T5

$$\begin{aligned} 30 \quad & + 0,41 [0,21..0,61] * \text{BIRC5} -0,33 [-0,57..-0,09] * \text{RBBP8} \\ & +0,38 [0,15..0,61] * \text{UBE2C} -0,30 [-0,55..-0,06] * \text{IL6ST} \\ & -0,28 [-0,43..-0,12] * \text{AZGP1} +0,42 [0,16..0,68] * \text{DHCR7} \\ & -0,18 [-0,31..-0,06] * \text{MGP} -0,13 [-0,25..-0,02] * \text{STC2} \\ & \text{c-índices: conjunto preparado} = 0,724, \end{aligned}$$

35 Los nombres de genes en el algoritmo denotan la diferencia de la expresión de ARNm del gen en comparación con uno o más genes constitutivos como se describe posteriormente.

Analizando una cohorte diferente de la cohorte de búsqueda (234 muestras tumorales) era sorprendente aprender que algunas simplificaciones del "algoritmo T5 original" aun daban una actuación diagnóstica no significativamente inferior al algoritmo T5. La simplificación más directa era reducir el comité predictor a un solo miembro. Ejemplos de la actuación de los "comités de un solo miembro" se muestran a continuación:

solo el miembro 1:

$$+ 0,41 [0,21..0,61] * \text{BIRC5} -0,33 [-0,57..-0,09] * \text{RBBP8}$$

c-índices: conjunto preparado = 0,653, cohorte independiente = 0,681

solo el miembro 2:

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

c-índices: conjunto preparado = 0,664, cohorte independiente = 0,696

solo el miembro 3:

5 $-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$

c-índices: conjunto preparado = 0,666, cohorte independiente = 0,601

solo el miembro 4:

$$-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2$$

c-índices: conjunto preparado = 0,668, cohorte independiente = 0,593

10 La actuación de los comités de un miembro como se muestra en una cohorte independiente de 234 muestras está notablemente reducida en comparación con la actuación del algoritmo completo.

15 Combinando gradualmente más de uno, pero menos de cuatro miembros en un nuevo algoritmo de comité predictor pronóstico, frecuentemente daba lugar a un aumento pequeño pero significativo en la actuación diagnóstica en comparación con un comité de un miembro. Era sorprendente aprender que había mejoras importantes en algunas combinaciones de miembros del comité mientras que otras combinaciones casi o no daban mejoras. Inicialmente, la hipótesis era que una combinación de miembros que representan motivos biológicos similares como se refleja por los genes empleados daba una mejora más pequeña que la combinación de miembros que reflejaban motivos biológicos distintivamente diferentes. Pero este no era el caso. No se pudo identificar ninguna regla pare predecir la combinación de algunos genes que generaran un algoritmo que presentara más potencia pronóstica que otra combinación de genes. Las combinaciones prometedoras solo se podían seleccionar basándose en datos experimentales. Las combinaciones identificadas de miembros de comité combinados para dar lugar a algoritmos simplificados, aunque potentes se muestran a continuación.

solo miembros 1 y 2:

$$+ 0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

25 $+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$

c-índices: conjunto preparado = 0,675, cohorte independiente = 0,712

solo miembros 1 y 3:

$$+ 0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

30 c-índices: conjunto preparado = 0,697, cohorte independiente = 0,688

solo miembros 1 y 4:

$$+ 0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

$$-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2$$

c-índices: conjunto preparado = 0,705, cohorte independiente = 0,679

35 solo miembros 2 y 3:

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

c-índices: conjunto preparado = 0,698, cohorte independiente = 0,670

solo miembros 1, 2 y 3:

40 $+ 0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$

$$+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST$$

$$-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7$$

c-índices: conjunto preparado = 0,701, cohorte independiente = 0,715

5 También es posible no omitir los miembros del comité completo sino un único gen o genes de diferentes miembros del comité, pero se necesita volver a probar el algoritmo completo. Aun así, puede ser ventajoso llevarlo a cabo. La actuación de algoritmos simplificados generados por omisión de todos los miembros o genes individuales es prácticamente idéntica.

Variantes de algoritmo por sustitución genética

10 Los algoritmos descritos, tales como el "Ejemplo de algoritmo T5" anterior, se puede modificar también sustituyendo uno o más genes por uno o más de otros genes. El fin de dichas modificaciones es remplazar genes difíciles de medir en una plataforma específica por un gen más directo de ensayar en esta plataforma. Aunque dicha transferencia no necesariamente genere una actuación mejorada en comparación con el algoritmo de partida, puede dar lugar a pistas para implantar el algoritmo pronóstico para una plataforma diagnóstica en particular. En general la sustitución de un gen por otro gen, mientras se mantiene el poder diagnóstico del algoritmo predictivo se puede conseguir mejor sustituyendo un gen por un gen co-expresado con una alta correlación (demostrado, por ejemplo, por el coeficiente de correlación de Pearson). Aunque se tiene que tener en mente que la expresión de ARNm de dos genes altamente correlativos en una plataforma puede parecer bastante independiente entre ellos cuando se evalúan en otra plataforma. En consecuencia, dicha aparentemente fácil sustitución cuando se reduce a la práctica experimental puede generar resultados pobres decepcionantes, así como resultados sorprendentemente fuertes, siempre dependiendo de imponderables de la plataforma empleada. Repitiendo este procedimiento se pueden sustituir varios genes.

20 La eficacia de dicha estrategia se puede demostrar evaluando la actuación predictiva de la puntuación del algoritmo T5 y sus variantes en cohortes de validación. La siguiente tabla muestra el c-índice con respecto al punto final de recurrencia distante en dos cohortes de validación.

Variante	Estudio de validación A	Estudio de validación B
Algoritmo T5 original	c-índice = 0,718	c-índice = 0,686
Omisión de BIRC5 (configurando la expresión a alguna constante)	c-índice = 0,672	c-índice = 0,643
Sustitución de BIRC5 por UBE2C (sin ajuste del coeficiente)	c-índice = 0,707	c-índice = 0,678

25 Se puede ver que la omisión de uno de los genes de T5, que se muestra aquí para BIRC5, por ejemplo, se reduce normalmente la actuación predictiva. Sustituyendo este con otro gen da lugar aproximadamente a la misma actuación.

30 Un procedimiento mejor de sustitución de un gen es volver a probar el algoritmo. Como T5 consiste en un comité de cuatro miembros independientes se tiene que volver a probar solo el miembro que contiene el gen sustituido. Las siguientes ecuaciones demuestran las sustituciones de genes del algoritmo T5 mostrado anteriormente preparado en una cohorte de 234 pacientes de cáncer de mama. Solo se muestra un miembro posteriormente, para el cálculo del c-índice los miembros restantes se utilizando sin cambios del algoritmo T5 original. El intervalo entre corchetes enumera el intervalo estimado de los coeficientes; media +/- 3 desviaciones típicas.

Miembro 1 de T5

35 Miembro original 1:

$$+ 0,41 [0,21..0,61] * BIRC5 -0,33 [-0,57..-0,09] * RBBP8$$

c-índices: conjunto preparado = 0,724, cohorte independiente = 0,705 se sustituye BIRC5 por TOP2A en el miembro 1:

$$+ 0,47 [0,24..0,69] * TOP2A -0,34 [-0,58..-0,10] * RBBP8$$

40 c-índices: conjunto preparado = 0,734, cohorte independiente = 0,694 se sustituye BIRC5 por RACGAP1 en el miembro 1:

$$+0,69 [0,37..1,00] * RACGAP1 -0,33 [-0,57..-0,09] * RBBP8$$

c-índices: conjunto preparado = 0,736, cohorte independiente = 0,743 se sustituye RBBP8 por CELSR2 en el miembro 1:

45
$$+ 0,38 [0,19..0,57] * BIRC5 -0,18 [-0,41..0,05] * CELSR2$$

ES 2 714 582 T3

- c-índices: conjunto preparado = 0,726, cohorte independiente = 0,680 se sustituye RBBP8 por PGR en el miembro 1:
+ 0,35 [0,15..0,54] * BIRC5 -0,09 [-0,23..0,05] * PGR
- c-índices: conjunto preparado = 0,727, cohorte independiente = 0,731
- 5 Miembro 2 de T5:
Miembro original 2:
+0,38 [0,15..0,61] * UBE2C -0,30 [-0,55..-0,06] * IL6ST
- c-índices: conjunto preparado = 0,724, cohorte independiente = 0,725 se sustituye UBE2C por RACGAP1 en el miembro 2:
- 10 +0,65 [0,33..0,96] * RACGAP1 -0,38 [-0,62..-0,13] * IL6ST
- c-índices: conjunto preparado = 0,735, cohorte independiente = 0,718 se sustituye UBE2C por TOP2A en el miembro 2:
- +0,42 [0,20..0,65] * TOP2A -0,38 [-0,62..-0,13] * IL6ST
- c-índices: conjunto preparado = 0,734, cohorte independiente = 0,700 se sustituye IL6ST por INPP4B en el miembro 2:
- 15 + 0,40 [0,17..0,62] * UBE2C -0,25 [-0,55..0,05] * INPP4B
- c-índices: conjunto preparado = 0,725, cohorte independiente = 0,686 se sustituye IL6ST por MAPT en el miembro 2:
- + 0,45 [0,22..0,69] * UBE2C -0,14 [-0,28..0,01] * MAPT
- 20 c-índices: conjunto preparado = 0,727, cohorte independiente = 0,711
- Miembro 3 de T5:
Miembro original 3:
-0,28 [-0,43..-0,12] * AZGP1 +0,42 [0,16..0,68] * DHCR7
- c-índices: conjunto preparado = 0,724, cohorte independiente = 0,705 se sustituye AZGP1 por PIP en el miembro 3:
- 25 -0,10 [-0,18..-0,02] * PIP +0,43 [0,16..0,70] * DHCR7
- c-índices: conjunto preparado = 0,725, cohorte independiente = 0,692 se sustituye AZGP1 por EPHX2 en el miembro 3:
- 0,23 [-0,43..-0,02] * EPHX2 +0,37 [0,10..0,64] * DHCR7
- c-índices: conjunto preparado = 0,719, cohorte independiente = 0,698 se sustituye AZGP1 por PLAT en el miembro 3:
- 30 -0,23 [-0,40..-0,06] * PLAT +0,43 [0,18..0,68] * DHCR7
- c-índices: conjunto preparado = 0,712, cohorte independiente = 0,715 se sustituye DHCR7 por AURKA en el miembro 3:
- 35 -0,23 [-0,39..-0,06] * AZGP1 +0,34 [0,10..0,58] * AURKA
- c-índices: conjunto preparado = 0,716, cohorte independiente = 0,733
- Miembro 4 de T5:
Miembro original 4:
-0,18 [-0,31..-0,06] * MGP -0,13 [-0,25..-0,02] * STC2
- 40 c-índices: conjunto preparado = 0,724, cohorte independiente = 0,705 se sustituye MGP por APOD en el miembro 4:

ES 2 714 582 T3

-0,16 [-0,30..-0,03] * APOD -0,14 [-0,26..-0,03] * STC2

c-índices: conjunto preparado = 0,717, cohorte independiente = 0,679 se sustituye MGP por EGFR en el miembro 4:

-0,21 [-0,37..-0,05] * EGFR -0,14 [-0,26..-0,03] * STC2

5 c-índices: conjunto preparado = 0,715, cohorte independiente = 0,708 se sustituye STC2 por INPP4B en el miembro 4:

-0,18 [-0,30..-0,05] * MGP -0,22 [-0,53..0,08] * INPP4B

c-índices: conjunto preparado = 0,719, cohorte independiente = 0,693 se sustituye STC2 por SEC14L2 en el miembro 4:

10 -0,18 [-0,31..-0,06] * MGP -0,27 [-0,49..-0,06] * SEC14L2

c-índices: conjunto preparado = 0,718, cohorte independiente = 0,681

Se puede ver que las sustituciones de genes únicos identificados experimentalmente por cuantificación con una PCR cuantitativa afectan normalmente la actuación predictiva del algoritmo T5, solo insignificamente según se evalúa mediante el c-índice.

15 La siguiente tabla muestra los candidatos potenciales a sustitución genética para los genes del algoritmo T5. Cada gen candidato se muestra en una celda de la tabla: el nombre del gen continúa con el coeficiente de correlación absoluta de Pearson de la expresión del gen original en el algoritmo T5 y el candidato de sustitución y el ID de conjunto de sondas U133A.

BIRC5	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
UBE2C (0,775), 202954_at	CELSR2 (0,548), 204029_at	BIRC5 (0,775), 202095_s_at	INPP4B (0,477), 205376_at	PIP (0,530), 206509_at	AURKA (0,345), 204092_s_at	APOD (0,368), 201525_at	INPP4B (0,500), 205376_at
TOP2A (0,757), 201292_at	PGR(0,392), 208305_at	RACGAP1 (0,756), TOP2A (0,753), 201292_at	STC2 (0,450), 203438_at	EPHX2 (0,369), 209368_at	BIRC5 (0,323), 202095_s_at	IL6ST (0,327), 212196_at	IL6ST (0,450), 212196_at
RACGAP1 (0,704), AURKA (0,681), 204092_s_at	STC2 (0,361), 203438_at	TOP2A (0,753), 201292_at	MAPT (0,440), 206401_s_at	PLAT (0,366), 201860_s_at	UBE2C (0,315), 202954_at	EGFR (0,308), 201983_s_at	SEC14L2 (0,417), 204541_at
NEK2 (0,680), 204026_s_at	ABAT (0,317), 209459_s_at	AURKA (0,694), 204092_s_at	SCUBE2 (0,418), 219197_s_at	SEC14L2 (0,351), 204541_at	202954_at	MAPT (0,414), 206401_s_at	MAPT (0,414), 206401_s_at
E2F8 (0,640), 219990_at	IL6ST (0,311), 212196_at	NEK2 (0,684), 204026_s_at	ABAT (0,389), 209459_s_at	SCUBE2 (0,331), 219197_s_at	202954_at	CHPT1 (0,410), 221675_s_at	CHPT1 (0,410), 221675_s_at
PCNA (0,544), 201202_at	ABAT (0,317), 209459_s_at	E2F8 (0,652), 219990_at	PGR (0,377), 208305_at	PGR (0,302), 208305_at	202954_at	ABAT (0,409), 209459_s_at	ABAT (0,409), 209459_s_at
CYBRD1 (0,462), 217889_s_at	PCNA (0,589), 201202_at	PCNA (0,589), 201202_at	SEC14L2 (0,356), 204541_at	208305_at	202954_at	SCUBE2 (0,406), 219197_s_at	SCUBE2 (0,406), 219197_s_at
DCN (0,439), 209335_at	CYBRD1 (0,486), 217889_s_at	CYBRD1 (0,486), 217889_s_at	ESR1 (0,353), 205225_at	ESR1 (0,353), 205225_at	202954_at	ESR1 (0,394), 205225_at	ESR1 (0,394), 205225_at
ADRA2A (0,416), 209869_at	ADRA2A (0,391), 209869_at	ADRA2A (0,391), 209869_at	GJA1 (0,335), 201667_at	GJA1 (0,335), 201667_at	202954_at	RBBP8 (0,361), 203344_s_at	RBBP8 (0,361), 203344_s_at
SQLI (0,415), 209218_at	DCN (0,384), 209335_at	DCN (0,384), 209335_at	MGP (0,327), 202291_s_at	MGP (0,327), 202291_s_at	202954_at	PGR (0,347), 208305_at	PGR (0,347), 208305_at
CXCL12 (0,388), 209687_at	SQLI (0,369), 209218_at	SQLI (0,369), 209218_at	EPHX2 (0,313), 209368_at	EPHX2 (0,313), 209368_at	202954_at	PTPRT (0,343), 205948_at	PTPRT (0,343), 205948_at
EPHX2 (0,362), 209368_at	CCND1 (0,347), 208712_at	CCND1 (0,347), 208712_at	RBBP8 (0,311), 203344_s_at	RBBP8 (0,311), 203344_s_at	202954_at	HSPA2 (0,317), 211538_s_at	HSPA2 (0,317), 211538_s_at
ASPH (0,352), 210896_s_at	ASPH (0,344), 210896_s_at	ASPH (0,344), 210896_s_at	PLAT (0,301), 201860_s_at	PLAT (0,301), 201860_s_at	202954_at	PTGER3 (0,314), 210832_x_at	PTGER3 (0,314), 210832_x_at
PRSS16 (0,352), 208165_s_at	CXCL12 (0,342), 209687_at	CXCL12 (0,342), 209687_at	PIP (0,328), 206509_at	PIP (0,328), 206509_at	202954_at		

(continuación)		AZGP1	DHCR7	MGP	STC2
BIRC5	RBBP8	IL6ST	UBE2C		
EGFR (0,346), 201983_s_at			PRSS16 (0,326), 208165_s_at		
CCND1 (0,331), 208712_at			EGFR (0,320), 201983_s_at		
TRIM29 (0,325), 202504_at			DHCR7 (0,315), 201791_s_at		
DHCR7 (0,323), 201791_s_at			EPHX2 (0,315), 209368_at		
PIP (0,308), 206509_at			TRIM29 (0,311), 202504_at		
TFAP2B (0,306), 214451_at					
WNT5A (0,303), 205990_s_at					
APOD (0,301), 201525_at					
PTPRT (0,301), 205948_at					

Las secuencias de los cebadores y sondas eran las siguientes:

Tabla 1 Secuencias de cebadores y sondas para los genes respectivos:

gen	sonda	Seq ID	Cebador directo	Seq ID	Cebador inverso	Seq ID
ABAT	TGCCCTAAGAGGCTCT- TCCTC	1	GGCAACTTGAGGTCT- GACTTTTG	2	GGTCAGCTCACAAAGTGGT- GTGA	3
ADRA2A	TTGTCCTTTCCCCCTC- CGTGC	4	CCCCAAGAGCTGTTAGG- TATCAA	5	TCAATGACATGATCT- CAACCAGAA	6
APOD	CATCAGCTCTCAACTCCT- GGTTTAAACA	7	ACTCACTAATGAAAAACG- GAAAGATC	8	TCACCTTCGATTTGAT- TCACAGTT	9
ASPH	TGGGAGGAAGGCAAGGT- GCTCATC	10	TGTGCCAACGAGACCAC- GAC	11	TCGTGCTCAAAGGAGT- CATCA	12
AURKA	CCGTACGCCGTGTGCTAG- GCAT	13	AATCTGGAGGCAAGGTT- GA	14	TCTGGATTTGCCCTCCTGT- GAA	15
BTRC5	AGCCAGATGACGAG- CCCATAGAGGAACA	16	CCCAGTGTTCCTCTGCT- TCAAAG	17	CAACCGGAGGAAT- GCTTTTT	18
CELSR2	ACTGACTTTTCCTTCT- GGAGCAGGTGGC	19	TCCAAGCATGTATTCCA- GACTTGT	20	TGCCACAGCCTC TTTT-TCT	21
CHPT1	CCACGGCCACCGAAGAG- GCAC	22	CGCTCGTGTCTCATCTC- CTACT	23	CCAGTGCACATAAAAAG- TATGTC	24
CXCL12	CCACAGCAGGGTTTCAG- GTTCC	25	GCCACTACCCCTCCTC- GAA	26	TCACCTTGCCAACAGT- TCTGAT	27
CYBRD1	AGGCATCGCCAT- CATCGTC	28	GTCACCGGCTTCGTCTTCA	29	CAGGTCCACGGCAGTCT- GT	30
DCN	TCTTTTCAGCAACCCG- GTCCA	31	AAGGCTTCTTATTCGGGT- GTGA	32	TGGATGGCTGTATCTC- CCAGTA	33
DHCR7	TGAGCGCCACCCCTCTC- GA	34	GGGCTCTGCTTCCCGATT	35	AGTCATAGGGCAAGCA- GAAAATTC	36
E2F8	CAGGATACCTAATC- CCTCTCAGCGAG	37	AAATGTCTCCCGCAACCTT- GTTTC	38	CTGCCCCACGGGATGAG	39
EPHX2	TGAAGCGGGAG- GACTTTTTGTAAA	40	CGATGAGAGTGTTTTATC- CATGCA	41	GCTGAGGCTGGGCTCT- TCT	42
ESR1	ATGCCCTTTTGCCGAT GCA	43	GCCAAAATTGTGTTTGTAT- GGATTA	44	GACAAAACCGAGTCACAT- CAGTAATAG	45
GJA1	TGCACAGCCTTTTGTATTC- CCCGAT	46	CGGGAAGCAC- CATCTCTAACTC	47	TTCATGTCCAG- CAGCTAGTTTTT	48
HSPA2	CAAGTCAGCAAACACG- CAAAA	49	CATGCACGAACTAAT- CAAAAATGC	50	ACATTATTCGAGGTT- TCTCTTTAATGC	51

(continuación)

gen	sonda	Seq ID	Cebador directo	Seq ID	Cebador inverso	Seq ID
IL6ST	CAAGCTCCACCTTC- CAAAGGACCT	52	CCCTGAATCCATAAAG- GCATACC	53	CAGCTTCGTTTTTC- CCTACTTTTT	54
INPP4B	TCCGAGCGCTGGATT- GCATGAG	55	GCACCAGTTACACAAG- GACTTCCTT	56	TCCTCTATGCGGCATCCT- TCTC	57
MAPT	AGACTATTTGCACACT- GCCGCCT	58	GTGGCTCAAAGGA- TAATATCAAACAC	59	ACCTTGCTCAGGTCAACT- GGTT	60
MGP	CCTTCATATCCCCTCAG- CAGAGATGG	61	GAAATGCCAA CCTTCATTAAACAGGA-	62	ATTGAGCTCGTGGACAG- GCCTTA	63
NEK2	TCCTGAACAAAATGAATCG- CATGTCCTACAA	64	ATTTGTTGGCACACCTTAT- TACATGT	65	AAGCAGCCCAATGACCA- GATa	66
PCNA	AAATACTAAAATGCGCCG- GCAATGA	67	GGCGTGAACCTCAC- CAGTA	68	CTTCGGCCCTTAGTG- TAATGATATC	69
PGR	TTGATAGAAACGCTGT- GAGCTCGA	70	AGCTCATCAAGGCAATT- GGTTT	71	ACAAGATCATGCAAGT- TATCAAGAAGTT	72
PIP	TGCATGGTGGTTAAAAC- TACCTCA	73	TGCTTGCAGTTCAAACA- GAATTG	74	CACCTTGTAGAGGGAT- GCTGCTA	75
PLAT	CAGAAAGTGGCCAT- GCCACCCCTG	76	TGGGAAGACATGAAT- GCACACTA	77	GGAGGTTGGCTTTAGCT- GAA	78
PRSS16	CACTGCCGGTCAC- CCACACCA	79	CTGAGGAGCACAGAACCT- CAACT	80	CGAACTCGGTACATGTCT- GATACAA	81
PTGER3	TCGGTCTGCTGGTCTC- CGCTCC	82	CTGATTGAAGATCATTI- TCAACATCA	83	GACGGCCATTGAGCTTAT- GG	84
PTPRT	TTGGCTTCTGGACACCC- CACA	85	GAGTTGGCCCTCTAC- CATTGC	86	GAGCGGGAACCTTGGGA- TAG	87
RACGAP1	ACTGAGAATCTCCACCCG- GCGCA	88	TCGCCAACTGGATAAATT- GGA	89	GAATGTGCGGAATCT- GTTTGAG	90
RBBP8	ACCGATTCCGCTACATTC- CACCCAAAC	91	AGAAATTGGCTTCCTGCT- CAAG	92	AAAACCAACTTC- CCAAAAATTCTCT	93
SCUBE2	CTAGAGGGTTCCAGGTC- CCATACGTGACATA	94	TGTGGATTCAAGT- TCAAGTCCAATG	95	CCATCTCGAACTATGTCT- TCAAATGAGT	96
SEC14L2	TGGGAGGCAI- GCAACGCGTG	97	AGGTCTTACTAAGCAGTC- CCATCTCT	98	CGACCAGGACCTGAACTC	99
SQLE	TATGCGTCTCCCAAAA- GAAGAACACCTCG	100	GCAAGCTTCTCCTCCTCCT- TCA	101	CCTTTAGCAGTTTTCTC- CATAGTTTTATATC	102

(continuación)

gen	sonda	Seq ID	Cebador directo	Seq ID	Cebador inverso	Seq ID
TFAP2B	CAACACCACCACTAACAG- GCACACGTC	103	GGCATGGACAAGATGT- TCTTGA	104	CCTCCTTGTCGCCAGTTT- TACT	105
TOP2A	CAGATCAGGACCAAGAT- GGTTCCACAT	106	CATTGAAGACGCTTCGT- TATGG	107	CCAGTTGTGATGGA- TAAAAATTAATCAG	108
TRIM29	TGCTGTCTCACTACCG- GCCATTCTAGC	109	TGAAAATCTGGCAAGCA- GACT	110	CAATCCCCTTGCCTTT- GTTG	111
UBE2C	TGAACACACATGCT- GCCGAGCTCTG	112	CTTCTAGGAGAACCACAA- CATTGATAGT	113	GTTTCTTGCAGGTACTTCT- TAAAAGCT	114
WNT5A	TATTCACATCCCCTCAGTT- GCAGTGAATTG	115	CTGTGGCTCTTAATTTATT- GCATAATG	116	TTAGTGCTTTTTTGCTT- TCAAGATCTT	117
STC2	TCTCACCTTGACCCT- CAGCCAAG	118	ACATTTGACAAAATTTCCCT- TAGGATT	119	CCAGGACGCAGCTTTAC- CAA	120

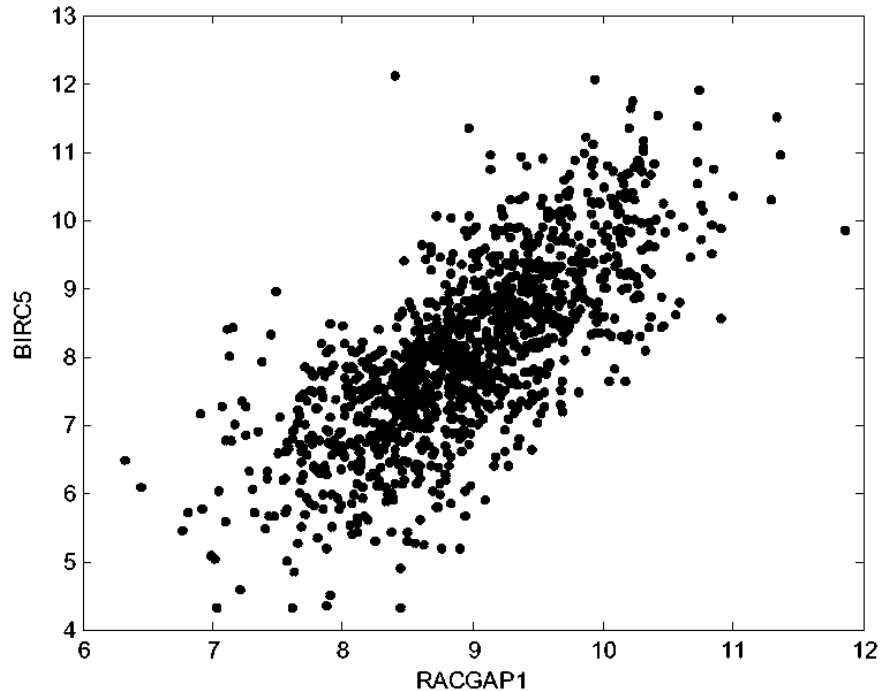
5 Una segunda alternativa para la selección no supervisada de posibles candidatos de sustitución genética se basa solo en los datos de Affymetrix. Esto tiene la ventaja de que se puede hacer solamente basándose en los datos ya publicados (por ejemplo, de www.nc-bi.nlm.nih.gov/geo/). Las siguientes Tablas enumeran los candidatos de sustitución del conjunto de sondas HG-U133a para los conjuntos de sondas utilizadas en los algoritmos T1-T5. Esto se basa en los datos de preparación de estos algoritmos. Las cabeceras de columna contienen el nombre del gen y la ID del conjunto de sondas en negrita. Entonces, se enumeran los 10 mejores conjuntos de sondas correlativos, donde cada celda de la tabla contiene el ID del conjunto de sondas, el coeficiente de correlación entre paréntesis y el nombre del gen.

UBE2C	BIRC5	DHCR7	RACGAP1	AURKA	PVALB	NMU	STC2
202954_at	202095_s_at	201791_s_at	222077_s_at	204092_s_at	205336_at	206023_at	203438_at
210052_s_at (0,82) TPX2	202954_at (0,82) UBE2C	201790_s_at (0,66) DHCR7	218039_at (0,79) NUSAP1	208079_s_at (0,89) STK6	208683_at (-0,33) CAPN2	205347_s_at (0,45) TMSL8	203439_s_at (0,88) STC2
202095_s_at (0,82) BIRC5	218039_at (0,81) NUSAP1	202218_s_at (0,48) FADS2	214710_s_at (0,78) CCNB1	202954_at (0,80) UBE2C	219682_s_at (0,30) TBX3	203764_at (0,45) DLG7	212496_s_at (0,52) JMJD2B
218009_s_at (0,82) PRC1	218009_s_at (0,79) PRC1	202580_x_at (0,47) FOXM1	203764_at (0,77) DLG7	210052_s_at (0,77) TPX2	218704_at (0,30) FLJ20315	203554_x_at (0,44) PTTG1	219440_at (0,52) RAI2
203554_x_at (0,82) PTTG1	202705_at (0,78) CCNB2	208944_at (-0,46) TGFB2	204026_s_at (0,77) ZWINT	202095_s_at (0,77) BIRC5		204962_s_at (0,44) CENPA	215867_x_at (0,51) CA12
208079_s_at (0,81) STK6	204962_s_at (0,78) CENPA	202954_at (0,46) UBE2C	218009_s_at (0,76) PRC1	203554_x_at (0,76) PTTG1		204825_at (0,43) MELK	214164_x_at (0,50) CA12
202705_at (0,81) CCNB2	203554_x_at (0,78) PTTG1	209541_at (-0,45) IGF1	204641_at (0,76) NEK2	218009_s_at (0,75) PRC1		209714_s_at (0,41) CDKN3	204541_at (0,50) SEC14L2
218039_at (0,81) NUSAP1	208079_s_at (0,78) STK6	201059_at (0,45) CTTN	204444_at (0,75) KIF11	201292_at (0,73) TOP2A		219918_s_at (0,41) ASPM	203963_at (0,50) CA12
202870_s_at (0,80) CDC20	210052_s_at (0,77) TPX2	200795_at (-0,45) SPARCL1	202705_at (0,75) CCNB2	214710_s_at (0,73) CCNB1		207828_s_at (0,41) CENPF	212495_at (0,50) JMJD2B
204092_s_at (0,80) STK6	202580_x_at (0,77) FOXM1	218009_s_at (0,45) PRC1	203362_s_at (0,75) MAD2L1	204962_s_at (0,73) CENPA		202705_at (0,41) CCNB2	208614_s_at (0,49) FLNB
209408_at (0,80) KIF2C	204092_s_at (0,77) STK6	218542_at (0,45) C10orf3	202954_at (0,75) UBE2C	218039_at (0,73) NUSAP1		219787_s_at (0,40) ECT2	213933_at (0,49) PTGER3

(continuación)

AZGP1	RBBP8	IL6ST	MGP	PTGER3	CXCL12	ABAT	CDH1
209309_at	203344_s_at	212196_at	202291_s_at	213933_at	209687_at	209460_at	201131_s_at
217014_s_at(0,92) AZGP1	36499_at(0,49) CELSR2	212195_at(0,85) IL6ST	201288_at(0,46) ARHGDI8	210375_at(0,74) PTGER3	204955_at (0,81)SRPX	209459_s_at (0,92)ABAT	201130_s_at(0,57) CDH1
206509_at(0,52) PIP	204029_at (0,45)CELSR2	204864_s_at(0,75) IL6ST	219768_at(0,42) VTCN1	210831_s_at (0,74)PTGER3	209335_at (0,81)DCN	206527_at(0,63) ABAT	221597_s_at(0,40) HSPC171
204541_at(0,46) SEC14L2	208305_at (0,45)PGR	211000_s_at(0,68) IL6ST	202849_x_at(- 0,41)GRK6	210374_x_at (0,73)PTGER3	211896_s_at (0,81)DCN	213392_at(0,54) MGC35048	203350_at(0,38) AP1G1
20670_at(0,45) XBP1	205380_at (0,43)PDZK1	214077_x_at(0,61) MEIS4	205382_s_at(0,40) DF	210832_x_at (0,73)PTGER3	201893_x_at (0,81)DCN	221666_s_at (0,49)PYCARD	209163_at(0,36) CYB561
209368_at(0,45) EPHX2	203303_at (0,41)TCTE1L	204863_s_at(0,58) IL6ST	200099_s_at(0,39) RPS3A	210834_s_at (0,55)PTGER3	203666_at (0,80)CXCL12	218016_s_at (0,48)POLR3E	210239_at(0,35) IRX5
218627_at(-0,43) FLJ11259	205280_at (0,38)GLRB	202089_s_at(0,57) SLC39A6	221591_s_at(- 0,37)FAM64A	210833_at(0,55) PTGER3	211813_x_at (0,80)DCN	214440_at(0,46) NAT1	200942_s_at(0,34) HSBP1
202286_s_at(0,43) TACSTD2	205279_s_at (0,38)GLRB	210735_s_at(0,56) CA12	214629_x_at(0,37) RTN4	203438_at(0,49) STC2	208747_s_at (0,79)C1S	204981_at(0,45) SLC22A18	209157_at(0,34) DNAJA2
213832_at(0,42)~ SORBS2	203685_at (0,38)BCL2	200648_s_at(0,52) GLUL	200748_s_at(0,37) FTH1	203439_s_at (0,46)STC2	203131_at (0,78)PDGFRA	212195_at(0,45) IL6ST	210715_s_at(0,33) SPINT2
204288_s_at(0,41) SERPINA3	203304_at(- 0,38)BAMBI	214552_s_at(0,52) RABEP1	209408_at(-0,37) KIF2C	212195_at(0,41) IL6ST	202994_s_at (0,78)FBLN1	204497_at(0,45) ADCY9	203219_s_at(0,33) APRT
202376_at(0,41) SERPINA3	205862_at (0,36)GREB1	219197_s_at(0,51) SCUBE2	218726_at(-0,36) DKFZp762E1312	217764_s_at (0,40)RAB31	208944_at (0,78)TGFB2	215867_x_at (0,45)CAT2	218074_at(0,33) FAM96B

Después de la selección de un gen o un conjunto de sondas se tiene que definir un mapeo matemático entre los valores de expresión del gen a sustituir y los del nuevo gen. Hay varias alternativas que se exponen en el presente documento basándose en el ejemplo “sustitución de los valores Ct delta de BIRC5 por RACGAP1”. En los datos preparados la distribución conjunta de las expresiones se parece a esta:



5

El coeficiente de correlación de Pearson es 0,73.

Una estrategia es crear una función de mapeo de RACGAP1 a BIRC5 por regresión. La regresión lineal es la primera elección y se produce en este ejemplo

$$\text{BIRC5} = 1,22 * \text{RACGAP1} - 2,85.$$

10 Utilizando esta ecuación se puede reemplazar fácilmente la variable BIRC5 en, por ejemplo, el algoritmo T5 por el lado derecho. En otros ejemplos pueden ser adecuadas una regresión robusta, regresión polinómica o pre-transformaciones no lineales no variadas.

15 El procedimiento de regresión asume la medición del ruido en BIRC5, pero sin ruido en RACGAP1. Por lo tanto, el mapeo no es simétrico con respecto al intercambio de las dos variables. Una estrategia de mapeo simétrica se basaría en dos transformaciones z no variadas.

$$z = (\text{BIRC5} - \text{media}(\text{BIRC5})) / \text{std}(\text{BIRC5})$$

y

$$z = (\text{RACGAP1} - \text{media}(\text{RACGAP1})) / \text{std}(\text{RACGAP1})$$

$$z = (\text{BIRC5} - 8,09) / 1,29 = (\text{RACGAP1} - 8,95) / 0,77$$

20
$$\text{BIRC5} = 1,67 * \text{RACGAP1} + -6,89$$

De nuevo, en otros ejemplos, pueden ser adecuadas otras transformaciones: normalización por mediana y/o mad, mapeos no lineales, u otros.

LISTADO DE SECUENCIAS

25 <110> Sividon Diagnostics GmbH

<120> Método para la predicción de la respuesta a la quimioterapia en un paciente que padece o está en riesgo de desarrollar un cáncer de mama recurrente

30 <130> 161573EP

<140> PCT/EP2012/064865
 <141> 30-07-2012

 5 <150> EP 11175852
 <151> 28-07-2011

 <160> 137

 10 <170> PatentIn versión 3.3

 <210> 1
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 1
 tcgccctaag aggctcttcc tc 22

 20 <210> 2
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

 25 <400> 2
 ggcaactga ggtctgactt ttg 23

 30 <210> 3
 <211> 22
 <212> ADN
 <213> *Homo sapiens*

 <400> 3
 ggfcagctca caagtgggtg ga 22
 35
 <210> 4
 <211> 22
 <212> ADN
 <213> *Homo sapiens*

 40 <400> 4
 ttgtctttc cccctccgt gc 22

 45 <210> 5
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

 50 <400> 5
 cccaagagc tgtaggtat caa 23

 55 <210> 6
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

 <400> 6
 tcaatgacat gatctcaacc agaa 24
 60
 <210> 7
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

 <400> 7

catcagctct caactcctgg ttaaca 27
 <210> 8
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 5
 <400> 8
 actcactaat ggaaaacgga aagatc 26
 10
 <210> 9
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 9
 tcacctcga ttgattcac agtt 24
 20
 <210> 10
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 25
 <400> 10
 tggaggaag gcaaggtgct catc 24
 30
 <210> 11
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 35
 <400> 11
 tgtccaacg agaccaagac 20
 40
 <210> 12
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 45
 <400> 12
 tcgtgctcaa aggagtcac a 21
 50
 <210> 13
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 55
 <400> 13
 ccgtcagcct gtgctaggca t 21
 60
 <210> 14
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 14
 aatctggagg caaggtcga 20
 <210> 15
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 15

tctggatttg cctcctgtga a 21
 <210> 16
 <211> 28
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 16
 10 agccagatga cgaccccata gaggaaca 28
 <210> 17
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 15 <400> 17
 cccagtggtt ctctgcttc aag 23
 <210> 18
 20 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 18
 25 caaccggacg aatgctttt 20
 <210> 19
 <211> 27
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 19
 actgacttfc ctctggagc aggtggc 27
 35 <210> 20
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 40 <400> 20
 tccaagcatg tattccagac ttgt 24
 <210> 21
 <211> 20
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 21
 50 tgcccacagc ctcttttct 20
 <210> 22
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 55 <400> 22
 ccacggccac cgaagaggca c 21
 <210> 23
 60 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 23

ogctcgtgct catctcctac t 21
 <210> 24
 <211> 24
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 24
 10 cccagtgcac ataaaaggta tgtc 24
 <210> 25
 <211> 22
 <212> ADN
 15 <213> *Homo sapiens*
 <400> 25
 ccaacagcagg gttcaggtt cc 22
 <210> 26
 20 <211> 19
 <212> ADN
 <213> *Homo sapiens*
 <400> 26
 25 gccactaccc cctcctgaa 19
 <210> 27
 <211> 22
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 27
 tcaccttgcc aacagttctg at 22
 <210> 28
 35 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 28
 40 agggcatcgc catcatcgtc 20
 <210> 29
 <211> 19
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 29
 50 gtcaccggct togtttca 19
 <210> 30
 <211> 19
 <212> ADN
 <213> *Homo sapiens*
 55 <400> 30
 caggtccacg gcagttctgt 19
 <210> 31
 60 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 31

tctttcagc aaccgggtcc a 21
 <210> 32
 <211> 22
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 32
 aaggctctt attcgggtgt ga 22
 10 <210> 33
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 15 <400> 33
 tggatggctg tatctccag ta 22
 <210> 34
 <211> 19
 20 <212> ADN
 <213> *Homo sapiens*
 <400> 34
 25 tgagcgccca ccctctga 19
 <210> 35
 <211> 18
 30 <212> ADN
 <213> *Homo sapiens*
 <400> 35
 gggctctgct tcccgatt 18
 35 <210> 36
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 40 <400> 36
 agtcataggg caagcagaaa attc 24
 <210> 37
 <211> 26
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 37
 caggatacct aatccctctc acgcag 26
 50 <210> 38
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 55 <400> 38
 aaatgtctcc gcaacctgt tc 22
 <210> 39
 <211> 17
 60 <212> ADN
 <213> *Homo sapiens*
 <400> 39

ctgccccag g gatgag 17
 <210> 40
 <211> 24
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 40
 tgaagcggga ggacttttg taaa 24
 10
 <210> 41
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 41
 cgatgagagt gtttatcca tgca 24
 20
 <210> 42
 <211> 19
 <212> ADN
 <213> *Homo sapiens*
 25
 <400> 42
 gctgaggctg ggctctct 19
 30
 <210> 43
 <211> 19
 <212> ADN
 <213> *Homo sapiens*
 35
 <400> 43
 atgcccttt gccgatgca 19
 40
 <210> 44
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 45
 <400> 44
 gccaaattgt gttgatgga ttaa 24
 50
 <210> 45
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 55
 <400> 45
 gacaaaaccg agtcacatca gtaatag 27
 60
 <210> 46
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 65
 <400> 46
 tgcacagcct tttgattcc ccgat 25
 70
 <210> 47
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 47

ogggaagcac catctctaac tc 22
 <210> 48
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 48
 ttcatgtcca gcagctagtt tttt 24
 <210> 49
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 49
 caagtcagca aacacgcaaa a 21
 <210> 50
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 50
 catgcacgaa ctaatcaaaa atgc 24
 <210> 51
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 51
 acattattcg aggtttctct ttaatgc 27
 <210> 52
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 52
 caagctccac cttccaaagg acct 24
 <210> 53
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 <400> 53
 ccctgaatcc ataaaggcat acc 23
 <210> 54
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 54
 cagcttcggt tttccctact tttt 24
 <210> 55
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 55

tccgagcgct ggattgcatg ag 22
 <210> 56
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 56
 gcaccagtta cacaaggact tcttt 25
 <210> 57
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 57
 tctctatgcg gcatccttct c 21
 <210> 58
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 <400> 58
 agactatttg cacactgccg cct 23
 <210> 59
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 59
 gtggctcaaa ggataatc aaacac 26
 <210> 60
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 60
 acctgtctca ggtcaactgg tt 22
 <210> 61
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 61
 ccttcatac ccctcagcag agatgg 26
 <210> 62
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 62
 ccttcattaa caggagaaat gcaa 24
 <210> 63
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 63

attgagctcg tggacaggct ta 22
 <210> 64
 <211> 30
 <212> ADN
 <213> *Homo sapiens*
 <400> 64
 tctgaacaa atgaatcgca tgcctacaa 30
 <210> 65
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 65
 attgtggc acacctatt acatgt 26
 <210> 66
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 66
 aagcagccca atgaccagat a 21
 <210> 67
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 67
 aaatactaaa atgcgccggc aatga 25
 <210> 68
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 68
 gggcgtgaac ctcaccagta 20
 <210> 69
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 69
 ctcggcct tagtgaatg atatc 25
 <210> 70
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 70
 ttgatagaaa cgctgtgagc tcga 24
 <210> 71
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 71

agctcatcaa ggcaatggt tt 22

 <210> 72
 <211> 28
 5 <212> ADN
 <213> *Homo sapiens*

 <400> 72
 acaagatcat gcaagttatc aagaagtt 28
 10
 <210> 73
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 73
 tgcatggggtg ttaaaacta cctca 25
 20
 <210> 74
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 25
 <400> 74
 tgcttgcaagt tcaaacagaa ttg 23
 30
 <210> 75
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 35
 <400> 75
 caccttgtag agggatgctg cta 23
 40
 <210> 76
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 45
 <400> 76
 cagaaagtgg ccatgccacc ctg 23
 50
 <210> 77
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 55
 <400> 77
 tgggaagaca tgaatgcaca cta 23
 60
 <210> 78
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 78
 ggaggtggg ctttagctga a 21
 <210> 79
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 79

cactgccggt cacccacacc a 21
 <210> 80
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 <400> 80
 ctgaggagca cagaacctca act 23
 <210> 81
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 81
 cgaactcggg acatgtctga taaa 25
 <210> 82
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 82
 tcggctcgt ggtctccgct cc 22
 <210> 83
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 83
 ctgattgaag atcatttca acatca 26
 <210> 84
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 84
 gacggccatt cagcttatgg 20
 <210> 85
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 85
 ttggctctg gacaccctca ca 22
 <210> 86
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 86
 gagttgtggc ctctaccatt gc 22
 <210> 87
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 87

gagcgggaac ctgggatag 20
 <210> 88
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 5
 <400> 88
 actgagaatc tccaccggc gca 23
 10
 <210> 89
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 89
 tcgccaactg gataaattgg a 21
 20
 <210> 90
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 25
 <400> 90
 gaatgtcgg aatctgttg ag 22
 30
 <210> 91
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 91
 accgattccg ctacattcca cccaac 26
 35
 <210> 92
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 40
 <400> 92
 agaaattggc ttctgctca ag 22
 45
 <210> 93
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 93
 aaaaccaact tcccaaaaat ttct 25
 50
 <210> 94
 <211> 31
 <212> ADN
 <213> *Homo sapiens*
 55
 <400> 94
 ctgaggggtt ccaggtcca tacgtgacat a 31
 60
 <210> 95
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 95

5 tgtagattca gttcaagtcc aatg 24
 <210> 96
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 96
 10 ccatctcgaa ctatgtcttc aatgagt 27
 <210> 97
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 15 <400> 97
 tgggaggcat gcaacgcgtg 20
 <210> 98
 20 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 98
 25 aggtcttact aagcagtccc atctct 26
 <210> 99
 <211> 18
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 99
 cgaccggcac ctgaactc 18
 35 <210> 100
 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 40 <400> 100
 tatgcgtctc ccaaagaag aacacctcg 29
 <210> 101
 45 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 101
 50 gcaagcttc ttctctctc a 21
 <210> 102
 <211> 31
 <212> ADN
 <213> *Homo sapiens*
 55 <400> 102
 ccttagcag tttctccat agttttatat c 31
 <210> 103
 60 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 103

caacaccacc actaacaggc acacgtc 27

5 <210> 104
<211> 22
<212> ADN
<213> *Homo sapiens*

10 <400> 104
ggcatggaca agatgttctt ga 22

15 <210> 105
<211> 22
<212> ADN
<213> *Homo sapiens*

20 <400> 105
cctccttgtc gccagtttta ct 22

25 <210> 106
<211> 28
<212> ADN
<213> *Homo sapiens*

30 <400> 106
cagatcagga ccaagatggt tcccacat 28

35 <210> 107
<211> 22
<212> ADN
<213> *Homo sapiens*

40 <400> 107
catgaagac gcttcgttat gg 22

45 <210> 108
<211> 27
<212> ADN
<213> *Homo sapiens*

50 <400> 108
ccagttgtga tggataaaaat taatcag 27

55 <210> 109
<211> 27
<212> ADN
<213> *Homo sapiens*

60 <400> 109
tgctgtctca ctaccggcca ttctaag 27

65 <210> 110
<211> 21
<212> ADN
<213> *Homo sapiens*

70 <400> 110
tggaaatctg gcaagcagac t 21

75 <210> 111
<211> 20
<212> ADN
<213> *Homo sapiens*

80 <400> 111

caatcccgtt gcctttgtg 20
 <210> 112
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 5
 <400> 112
 tgaacacaca tgctgccgag ctctg 25
 10
 <210> 113
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 15
 <400> 113
 cttctaggag aaccaacat tgatagt 27
 20
 <210> 114
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 25
 <400> 114
 gtttcttgca ggtacttctt aaaagct 27
 30
 <210> 115
 <211> 30
 <212> ADN
 <213> *Homo sapiens*
 <400> 115
 tattcacatc ccctcagttg cagtgaattg 30
 35
 <210> 116
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 40
 <400> 116
 ctgtggctct taatttattg cataatg 27
 45
 <210> 117
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 117
 ttagtgcttt ttgcttcaa gatctt 26
 50
 <210> 118
 <211> 23
 <212> ADN
 <213> *Homo sapiens*
 55
 <400> 118
 tctcacctg accctcagcc aag 23
 60
 <210> 119
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

ES 2 714 582 T3

<400> 119
acatttgaca aatttcctt aggatt 26

5
<210> 120
<211> 20
<212> ADN
<213> *Homo sapiens*

10
<400> 120
ccaggaogca gctttaccaa 20

15
<210> 121
<211> 783
<212> ADN
<213> *Homo sapiens*

<400> 121

ggcacgagcg agttcctgtc tctctgcaa cgccgcccgg atggcttccc aaaaccgoga 60

cccagccgcc actagcgtcg ccgcccgcg taaaggagct gagccgagcg ggggogccgc 120

ccggggtccg gtgggcaaaa ggctacagca ggagctgatg accctcatga tgtctggcga 180

taaagggatt tctgccttcc ctgaatcaga caaccttttc aaatgggtag ggaccatcca 240

tggagcagct ggaacagtat atgaagacct gaggtataag ctctcgctag agttccccag 300

tggctaccct tacaatgcgc ccacagtga gttcctcacg ccctgctatc accccaacgt 360

ggacacccag ggtaacatat gcctggacat cctgaaggaa aagtggctctg ccctgatga 420

tgtcaggacc attctgctct ccatccagag ccttctagga gaaccaaca ttgatagtcc 480

cttgaacaca catgctgccg agctctggaa aaaccccaca gcttttaaga agtacctgca 540

agaaaacctac tcaaagcagc tcaccagcca ggagccctga ccaggctgc ccagcctgtc 600

cttgtgtcgt ctttttaatt tttccttaga tggctctgtcc tttttgtgat ttctgtatag 660

gactctttat cttgagctgt ggtatTTTTG ttttgTTTT gtcttttaa ttaagcctcg 720

gttgagccct tgtatattaa ataaatgcat ttttgcctt ttttaaaaa aaaaaaaaaa 780

aaa 783

20
<210> 122
<211> 14796
<212> ADN
<213> *Homo sapiens*

25
<400> 122

ES 2 714 582 T3

tctagacatg cggatatatt caagctgggc acagcacagc agccccacco caggcagctt 60
gaaatcagag ctgggggtcca aagggaccac accccgaggg actgtgtggg ggtcggggca 120
cacaggccac tgcttcccc cgtctttctc agccattcct gaagtcagcc tcaactctgt 180
tctcagggat ttcaaagtgt cagagactct ggcacttttg tagaagcccc ttctggctct 240
aacttacacc tggatgctgt ggggctgcag ctgctgctcg ggctcgggag gatgctgggg 300
gcccgggtgcc catgagcttt tgaagctcct ggaactcggg ttgaggggtg ttcaggtcca 360
ggtggacacc tgggctgtcc ttgtccatgc atttgatgac attgtgtgca gaagtgaaaa 420
ggagttagcc cgggcatgct ggcttatgcc tgtaatocca gcactttggg aggctgaggc 480
gggtggatca cgaggtcagg agttcaatac cagcctggcc aagatggtga aaccocgtct 540
ctactaaaaa tacaaaaaaa ttagccgggc atggtggcgg gcgcatgtaa tcccagctac 600
tgggggggct gaggcagaga attgctggaa cccaggagat ggaggttgca gtgagccaag 660
attgtgccac tgcactgcac tccagcctgg cgacagagca agactctgtc tcaaaaaaaaa 720
aaaaaaaaag tgaaaaggag ttgttccttt cctccctcct gagggcaggc aactgctgcg 780
gttgccagtg gaggtggtgc gtcottggtc tgtgcoctggg ggccacccca gcagaggcca 840
tggtggtgcc agggcccggg tagcgagcca atcagcagga cccaggggcg acctgccaaa 900
gtcaactgga tttgataact gcagcgaagt taagtttctt gattttgatg attgtgttgt 960
ggttgtgtaa gagaatgaag tatttcgggg tagtatggta atgccttcaa cttacaaacg 1020
gttcaggtaa accaccata tacatacata tacatgcatg tgatatatac acatacaggg 1080
atgtgtgtgt gttcacatat atgaggggag agagactagg ggagagaaaag taggttgggg 1140
agagggagag agaaaaggaaa acaggagaca gagagagagc ggggagtaga gagagggaag 1200
gggtaagaga gggagaggag gagagaaagg gaggaagaag cagagagtga atgttaaagg 1260
aacaggcaa aacataaaca gaaaatctgg gtgaagggtg tatgagtatt ctttgtacta 1320
ttcttgcaat tatcttttat ttaaattgac atcgggcccgg gcgcagtggc tcacatctgt 1380
aatcccagca ctttgggagg ccgagggcagg cagatcactt gaggtcagga gtttgagacc 1440
agcctggcaa acatggtgaa accccatctc tactaaaaat acaaaaatta gcctggtgtg 1500
gtggtgcatg cctttaatct cagctactcg ggaggctgag gcaggagaat cgcttgaacc 1560
cgtggcgggg aggaggttgc agtgagctga gatcatgcca ctgcactcca gcctgggcca 1620
tagagcgaga ctcaagtttca aataaataaa taaacatcaa aataaaaagt tactgtatta 1680
aagaatgggg gcgggggtggg aggggtgggg agaggttgca aaaataaata aataaataaa 1740

ES 2 714 582 T3

taaaccccaa aatgaaaaag acagtggagg caccaggcct gcgtggggct ggaggggctaa 1800
 taaggccagg cctcttatct ctggccatag aaccagagaa gtgagtggat gtgatgccca 1860
 gctccagaag tgactccaga acaccctggt ccaaagcaga ggacacactg attttttttt 1920
 taataggctg caggacttac tgttggggg acgccctgct ttgcgaaggg aaaggaggag 1980
 tttgcocctga gcacaggccc ccaccctoca ctggggctttc cccagctccc ttgtottctt 2040
 atcacggtag tggcccagtc cctggcccct gactccagaa ggtggccctc ctggaaacct 2100
 aggtcgtgca gtcaacgatg tactcgccgg gacagcgatg tctgctgac tccatccctc 2160
 ccctgttcat ttgtccttca tgcccgtctg gagtagatgc tttttgcaga ggtggcacc 2220
 tgtaaagctc tcctgtctga cttttttttt ttttttagac tgagttttg tcttgttgcc 2280
 taggctggag tgcaatggca caatctcagc tcaactgcacc ctctgcctcc cgggttcaag 2340
 cgattctcct gcctcagcct cccgagtagt tgggattaca ggcatgcacc accacgccc 2400
 gctaattttt gtatttttag tagagacaag gtttcaccgt gatggccagg ctggtcttga 2460
 actccaggac tcaagtgatg ctctgccta ggcctctcaa agtgttggga ttacaggcgt 2520
 gagccactgc acccggcctg cacgcgttct ttgaaagcag tcgagggggc gctaggtgtg 2580
 ggcagggacg agctggcgcg gcgtcgctgg gtgcaccgcg accacgggca gagccaocg 2640
 gcgggaggac tacaactccc ggcacacccc gcgccgcccc gcctctactc ccagaaggcc 2700
 gcgggggggtg gaccgcctaa gagggcgtgc gctcccgaca tgccccggg cgcgccatta 2760
 accgccagat ttgaatcgcg ggacccttg gcagaggtgg cggcggcggc atgggtgccc 2820
 cgacgttgcc cctgcctgg cagcccttc tcaaggacca ccgcatctct acattcaaga 2880
 actggccctt cttggagggc tgcgcctgca ccccgagcg ggtgagactg cccggcctcc 2940
 tggggcccc cacgcccgc ttgcctgtc cctagcgagg cactgtgac tgggcctcgg 3000
 ggttacaagc cgcctcccc tcccgtcct gtcccagcg aggcactgt ggctgggccc 3060
 cttgggtcca ggcggcctc ccctccctgc tttgtoccca tcgaggcctt tgtggctggg 3120
 cctcgggggtt cgggctgcc acgtccactc acgagctgtg ctgtccctg cagatggccg 3180
 aggctggctt catccactgc cccactgaga acgagccaga cttggcccag tgtttcttct 3240
 gcttcaagga gctggaaggc tgggagccag atgacgacce catgtaagtc ttctctggcc 3300
 agcctcgatg ggctttggtt tgaactgagt tgtcaaaaga tttgagttgc aaagacactt 3360
 agtatgggag ggttgettcc caccctcatt gcttcttaa cagctgttgt gaacggatac 3420
 ctctctatat gctgggtgct tggatgatgct tacaacctaa ttaaactcct tttgaccaa 3480
 atgccttggg gtggacgtaa gatgcctgat gcctttcatg ttcaacagaa tacatcagca 3540
 gaccctgttg ttgtgaactc ccaggaatgt ccaagtgctt tttttgagat tttttaaaaa 3600

ES 2 714 582 T3

acagtttaat tgaatatataa cctacacagc acaaaaatta ccotttgaaa gtgtgcactt 3660
 cacactttcg gaggctgagg cgggcggatc acctgaggtc aggagttcaa gacctgcctg 3720
 gccaaacttg cgaaaccccg tctctactaa aaatacaaaa attagccggg catggtagcg 3780
 cacgcccgtg atcccagcta ctccggaggc taaggcagga gaatcgcttg aacctgggag 3840
 gcgagaggtg cagtgagccg agattgtgcc aatgcactcc agcctcggcg acagagcgag 3900
 actccgtcat aaaaataaaa aattgaaaaa aaaaaaagaa agaaagcata tacttcagtg 3960
 ttgttctgga tttttttctt caagatgcct agttaatgac aatgaaattc tgtactcgga 4020
 tggatctctg ctttccacac tgtaatgcca tattcttttc tcaccttttt ttctgtcgga 4080
 ttcagttgct tccacagctt taatTTTTTTT ccctggaga atcaccocag ttgtttttct 4140
 ttttgccag aagagagtag ctgttttttt tcttagtatg tttgctatgg tggttatact 4200
 gcatccccgt aatcactggg aaaagatcag tggatttctt cttgaaaatg aataagtgtt 4260
 atgatatttt cagattagag ttacaactgg ctgtcttttt ggactttgtg tggccatggt 4320
 ttcattgtaa tgcagttctg gtaacggtga tagtcagtta tacagggaga ctccoctagc 4380
 agaaaatgag agtgtgagct aggggggtccc ttggggaacc oggggcaata atgcccttct 4440
 ctgcccttaa tccttacagt gggccgggca cggtggctta cgctgtaat accagcactt 4500
 tgggaggccg aggcggggcg atcacgaggt caggagatcg agaccatctt ggctaatacg 4560
 gtgaaacccc gtctccacta aaaatacaaa aaattagccg ggcggtgggg tgggocctg 4620
 tagtcccagc tactcgggag gctgaggcag gagaatggcg tgaaccagag aggcggagct 4680
 tgcagtgagc cgagattgca cactgcact ccagcctggg cgacagaatg agactcogtc 4740
 tcaaaaaaaaa aaaaaaaaaa aaaaaatctt tacagtggat tacataacaa ttccagtga 4800
 atgaaattac ttcaaacagt tccttgagaa tggtggaggg atttgacatg taattccttt 4860
 ggacatatac catgtaacac tttccaact aattgctaag gaagtccaga taaatagat 4920
 acattagcca cacagatgtg gggggagatg tccacagga gagagaaggt gctaagaggt 4980
 gccatattgg aatgtggctt gggcaaagca ctgatgccat caacttcaga cttgacgtct 5040
 tactcctgag gcagagcagg gtgtgcctgt ggagggcgtg gggaggtggc ccgtggggag 5100
 tggactgccg ctttaatccc ttcagctgcc tttccgctgt tgttttgatt tttctagaga 5160
 ggaacataaa aagcattcgt ccggttgccg tttcctttct gtcaagaagc agtttgaaga 5220
 attaaccctt ggtgaatttt tgaaactgga cagagaaaga gccagaaca aaattgtatg 5280
 tattgggaat aagaactgct caaacctgt tcaatgtctt tagcactaaa ctacctagtc 5340
 cctcaaaggg actctgtggt ttccctcagga agcatttttt tttttttctt gagatagagt 5400
 ttcactcttg ttgccaggc tggagtgcaa tggtgcaatc ttggctcact gcaacctctg 5460
 cctctcgggt tcaagtgatt ctctgcctc agcctcccaa gtaactggga ttacagggaa 5520

ES 2 714 582 T3

gtgccaccac acccagctaa tttttgtatt tttagtagag atggggtttc accacattgc 5580
 ccaggctggt cttgaactcc tgacctcgtg attcgcaccac cttggcctcc caaagtgctg 5640
 ggattacagc cgtgaaccac cagcctcgtc tttttttttt ttgttctgag acacagtttc 5700
 actctgttac ccaggctgga gtaggggtggc ctgatctcgg atcactgcaa cctccgcctc 5760
 ctgggctcaa gtgatttgcc tgcttcagcc toccaagtag cagagattac aggcattgtc 5820
 caccacacc aggtaatttt tgtatttttg gtagagaaga ggtttcacca tgttgccag 5880
 gctggttttg aactcctgac ctcagggtgat ccaccgcct cagcctocca aagtgctgag 5940
 attataggtg tgagccacca cacctggcct caggaagtat ttttattttt aaatttattt 6000
 atttatttga gatggagtct tgctctgtcg ccaggctag agtgacagca cgggatctcg 6060
 gctcactgca agctccgccc ccagggttca agccattctc ctgcctcagc ctcccgagta 6120
 gctgggacta caggcgcccg ccaccacacc cgctaattt ttttgtattt ttagtagaga 6180
 cgggttttca cagtgttagc caggagggtc ttgatctcct gacctcgtga tctgcctgcc 6240
 tcggcctccc aaagtgctgg gattacaggt gtgagccacc acaccggct atttttattt 6300
 ttttgagaca gggactcact ctgtcacctg ggctgcagtg cagtggtaaa ccatagctca 6360
 ctgcagcctc gaactcctga gctcaagtga toctcccacc toatcctcac aagtaattgg 6420
 gactacaggt gcaccccacc atgccacct aatttattta tttatttatt tatttatttt 6480
 catagagatg agggttccct gtgttgctca ggctggtctt gaactcctga gctcacggga 6540
 tccttttgcc tgggcctccc aaagtgctga gattacaggc atgagccacc gtgccagct 6600
 aggaatcatt tttaaagccc ctaggatgct tgtgtgattt taaagctcct ggagtgtggc 6660
 cggataagt atataccggt ataagtaaat ccacatttt gtgtcagtat ttactagaaa 6720
 cttagtcatt tatctgaagt tgaaatgtaa ctgggcttta tttatttatt tatttattta 6780
 tttattttta atttttttt ttgagacgag tctcactttg tcaccagggc tggagtgcag 6840
 tggcacgata tcggctcact gcaacctctg cctccggggg tcaagcgatt ctctgcctt 6900
 agcctccga gtagctggga ctacaggcac gcaccaccat gcctggctaa tttttgtatt 6960
 tttagtagac ggggtttcac catgctggcc aagctggtct caaactcctg accttgtgat 7020
 ctgccgctt tagcctocca gagtgtcggg attacaggca tgagccacca tgcgtggtct 7080
 ttttaaaatt ttttgatttt ttttttttt gagacagagc cttgctctgt cgcaccagct 7140
 ggagtgcagt ggcacgatct cagctcacta caagctcgc ctcccggtt cagccattc 7200
 ttctgctca gcctcctgag tagctgggac tacaggtgcc caccaccacg cctggetaat 7260
 ttttttggc atttttatta gagacaaggt tcatcatgt tggccaggct ggtctcaaac 7320
 tcctgacctc aagtgatctg cctgcctcgg cctcccaag cgtgagatt acagggtgtg 7380

ES 2 714 582 T3

tctactgogc caggcctggg ogtcatatat tcttatttgc taagtctggc agccccacac 7440
agaataagta ctgggggatt ccatatcctt gtagcaaagc cctgggtgga gagtcaggag 7500
atgtttagt tctgtctctg ccacttgacg actttgagtt taagccagtc gtgctcatgc 7560
tttccttgct aaatagaggt tagacccct atcccatggt ttctcaggtt gcttttcagc 7620
ttgaaaattg tattcctttg tagagatcag cgtaaaataa ttctgtcctt atatgtggct 7680
ttattttaat ttgagacaga gtgtcactca gtcgccagc ctggagtgtg gtggtgcgat 7740
ctgggctcac tggacacctc acctcccagg ttcaagcgat tctcgtgcct caggctocca 7800
agtagctgag attataggtg tgtgccacca ggcccagcta acttttgtat ttttagtaga 7860
gacagggttt tgccatggtg gctaagctgg tctcgaactc ctggcctcaa gtgatctgcc 7920
cgccttgcca tcccaaagtg ctgggattac aggtgtgaac caccacacct ggcctcaata 7980
tagtggcttt taagtgctaa ggactgagat tgtgttttgt caggaagagg ccagttgtgg 8040
gtgaagcatg ctgtgagaga gcttgtcacc tggttgaggt tgtgggagct gcagcgtggg 8100
aactggaaag tgggctgggg atcatctttt tccaggtcag gggtcagcca gcttttctgc 8160
agcgtgccat agaccatctc ttagccctcg tgggtcagag tctctgttgc atattgtctt 8220
ttgttgtttt tcacaacctt ttagaaacat aaaaagcatt cttagccctg gggctggaca 8280
aaaaaaggcc atgacgggct gtatggattt ggcccagcag gcccttgctt gccaaagcct 8340
gttttagaca aggagcagct tgtgtgcctg gaaccatcat gggcacaggg gaggagcaga 8400
gtggatgtgg aggtgtgagc tggaaaccag gtcccagagc gctgagaaag acagagggtt 8460
tttgcccttg caagtagagc aactgaaatc tgacaccatc cagttccaga aagccctgaa 8520
gtgctggtgg acgctgcggg gtgctccgct ctagggttac agggatgaag atgcagtctg 8580
gtagggggag tccactcacc tgttgaaga tgtgattaag aaaagtagac tttcagggcc 8640
gggcatggtg gtcacgcct gtaatcccag cactttggga ggccgagggc ggtggatcac 8700
gaggtcagga gatcgagacc atcctggcta acatggtgaa accccgtctt tactaaaaat 8760
acaaaaaatt agctgggctg ggtggcgggc gcctgtagtc ccagctactc gggaggctga 8820
ggcaggagaa tggcgtgaac ctgggaggtg gagcttgctg tgagccgaga tcgcgccact 8880
gcactccagc ctgggcgaca gagcgagact ccgtctcaa aaaaaaaaaa aaagtaggct 8940
ttcatgatgt gtgagctgaa ggcgcagtag gcagaagtag aggcctcagt cctgcagga 9000
gaccctcgg tctctatctc ctgatagtc gaccagcca cactggaag aggggagaca 9060
ttacagcctg cgagaaaagt agggagattt aaaaactgct tggcttttat tttgaaactgt 9120
ttttttgtt tgtttgttt cccaattca gaatacagaa tacttttatg gatttgtttt 9180
tattacttta attttgaaac aatataatct ttttttgtt gttttttga gacagggtct 9240
tactctgtca occaggctga gtgcagtggg gtgatcttgg ctcacctcag cctcgacccc 9300

ES 2 714 582 T3

ctgggctcaa atgattctcc cacctcagct toccaagtag ctgggaccac aggtgogtgt 9360
 gttgogctat acaaatcctg aagacaagga tgctgttgct ggtgatgctg gggattocca 9420
 agatcccaga tttgatggca ggatgcccct gtctgctgcc ttgccagggt gccaggaggg 9480
 cgctgctgtg gaagctgagg cccggccatc cagggcgatg cattgggccc tgattcttgt 9540
 tctgctgct gcctcgggtg ttagcttttg aaacaatgaa ataaattaga accagtgtga 9600
 aatcgatca ggggaataaat ttaatgtgga aataaactga acaacttagt tcttcataag 9660
 agtttacttg gtaaataact gtgatgagga caaacgaag cactagaagg agaggogagt 9720
 tgtagacctg ggtggcagga gtgttttggt tgttttcttt ggcagggtct tgctctgttg 9780
 ctcaggctgg agtacagtgg cacaatcaca gctcactata gcctcgacct cctggactca 9840
 agcaatcctc ctgcctcagc ctcccagtag ctgggactac aggcgcatgc caccatgcct 9900
 ggtaatttt aaatttttt tttctcttt tttgagatgg aatctcctc tgtcgcccag 9960
 gctggagtgc agtggogtga tctcgctga oggcaagctc cgcctcccag gttcactoca 10020
 ttcgcctgcc tcagcctccc aagtagctgg gactacaggc gctgggatta caaacccaaa 10080
 cccaaagtgc tgggattaca ggcgtgagcc actgcaccog gcctgttttg tctttcaata 10140
 gcaagagttg tgtttgette gccctacct ttagtggaaa aatgtataaa atggagatat 10200
 tgacctcac attgggggtg ttaaattata gcatgtatgc aaaggagctt cgctaattta 10260
 aggctttttt gaaagagaag aaactgaata atccatgtgt gtatatatat tttaaaagcc 10320
 atggtcatct ttccatatca gtaaagctga ggctccctgg gactgcagag ttgtccatca 10380
 cagtcatta taagtgcgct gctgggccag gtgcagtgge ttgtgctga atcccagcac 10440
 tttgggaggg caaggcagga ggatcattg agcccaggag ttttgaggog agcctgggca 10500
 atgtggccag acctcatctc ttcaaaaaat acacaaaaaa ttagccagge atggtggcac 10560
 gtgcctgtag tctcagctac tcaggaggct gaggtgggag gatcactttg agccttgag 10620
 gtcaaagctg cagtaagcca tgatcttgcc actgcattcc agcctggatg acagagcgag 10680
 accctgtctc taaaaaaaa aaaaaccaa cgggtgactg tttcttttt tcttatcaat 10740
 ttattatttt taaattaaat tttcttttaa taatttataa attataaatt tatattaaaa 10800
 aatgacaaat ttttattact tatacatgag gtaaaactta ggatatataa agtacatatt 10860
 gaaaagtaat tttttggctg gcacagtgge tcacacctgt aatcccagca ctttgggagg 10920
 ccgtggcggg cagatcacat gagatcatga gttcgagacc aacctgacca acatggagag 10980
 accccatctc tactaaaaat acaaaattag ccgggggtggt ggcgcatgcc tgtaatccca 11040
 gctactoggg aggctgaggc aggagaatct cttgaacctg ggaggcagag gttgoggtga 11100
 gccaaagatg tgcccttgca caccagccta ggcaacaaga gcgaaagtcc gtctcaaaaa 11160

ES 2 714 582 T3

aaaagtaatt tttttaagt taacctctgt cagcaaaaa atttaacca ataaaggtct 11220
 ttgttttta atgtagtaga ggagttaggg tttataaaaa atatgtagg gaagggggtc 11280
 cctggatttg ctaatgtgat tgtcatttgc cccttaggag agagctctgt tagcagaatg 11340
 aaaaaattgg aagccagatt cagggagggg ctggaagcaa aagaatttct gttcgaggaa 11400
 gagcctgatg tttgccaggg tctgtttaac tggacatgaa gaggaaggct ctggactttc 11460
 ctccaggagt ttcaggagaa aggtagggca gtggttaaga gcagagctct gcctagacta 11520
 gctgggggtgc ctgactagc tgggggtgcc agactagctg ggggtgctag actagctggg 11580
 tactttgagt ggctccttca gcctggacct cggtttctc acctgtatag tagagatatg 11640
 ggagcaocca gcgcaggatc actgtgaaca taaatcagtt aatggaggaa gcaggtagag 11700
 tgggtgctggg tgcataccaa gcactccgtc agtgtttcct gttattcgat gattaggagg 11760
 cagcttaaac tagagggagt tgagctgaat caggatgttt gtcccaggta gctgggaatc 11820
 tgcttagccc agtgcccagt ttatttaggt gctctctcag tgttccctga ttgttttttc 11880
 ctttgtcatc ttatctacag gatgtgactg ggaagctctg gtttcagtgt catgtgtcta 11940
 ttctttattt ccaggcaaag gaaaccaaca ataagaagaa agaatttgag gaaactgcga 12000
 agaaagtgcg ccgtgccatc gagcagctgg ctgccatgga ttgaggcctc tggccggagc 12060
 tgctgggtcc cagagtggct gcaccacttc cagggtttat tccctgggtc caccagcctt 12120
 cctgtgggce ccttagcaat gtcttaggaa aggagatcaa cttttcaaa ttagatgttt 12180
 caactgtgct cctgttttgt cttgaaagtg gcaccagagg tgcttctgcc tgtgcagcgg 12240
 gtgctgctgg taacagtggc tgcttctctc tctctctctc tttttgggg gctcattttt 12300
 gctgttttga ttcccgggct taccaggtga gaagtgaggg aggaagaagg cagtgtccct 12360
 tttgctagag ctgacagctt tgttcgcgtg ggcagagcct tccacagtga atgtgtctgg 12420
 acctcatggt gttgaggctg tcacagctct gagtgtggac ttggcaggtg cctgttgaat 12480
 ctgagctgca ggttccttat ctgtcacacc tgtgcctcct cagaggacag tttttttgtt 12540
 gttgtgtttt tttgtttttt ttttttgta gatgcatgac ttgtgtgtga tgagagaatg 12600
 gagacagagt ccctggctcc tctactgttt aacaacatgg ctttcttatt ttgtttgaat 12660
 tgtaattca cagaatagca caaactacaa ttaaaaactaa gcacaaagcc attctaagtc 12720
 attggggaaa cggggtgaac ttcaggtgga tgaggagaca gaatagagt ataggaagcg 12780
 tctggcagat actccttttg ccactgctgt gtgattagac aggccagtg agccgcgggg 12840
 cacatgctgg ccgctcctcc ctcagaaaaa ggcagtggcc taaatccttt ttaaatgact 12900
 tggtcogatg ctgtggggga ctggctgggc tgctgcaggc cgtgtgtctg tcagcccaac 12960
 cttcacatct gtcacgttct ccacacgggg gagagacgca gtccgccag gtccccgctt 13020
 tctttggagg cagcagctcc cgcagggctg aagtctggcg taagatgatg gatttgattc 13080

ES 2 714 582 T3

gocctcctcc ctgtcataga gctgcagggt ggattgttac agcttogctg gaaacctctg 13140
 gaggtcatct cggctgttcc tgagaaataa aaagcctgtc atttcaaaca ctgctgtgga 13200
 coctactggg tttttaaaat attgtcagtt tttcatogtc gtcoctagcc tgccaacagc 13260
 catctgcccc gacagccgca gtgaggatga gcgtcctggc agagacgcag ttgtctctgg 13320
 gcgcttgcca gagccacgaa ccccagacct gtttgatca tccgggctcc ttccgggcag 13380
 aaacaactga aatgcactt cagaccact tatttatgcc acatctgagt cggcctgaga 13440
 tagacttttc cctctaaact gggagaatat cacagtgggt tttgttagca gaaaatgcac 13500
 tocagcctct gtactcatct aagctgctta tttttgatat ttgtgtcagt ctgtaaattg 13560
 ataactcact ttaataactg ttgcttagta attggctttg tagagaagct ggaaaaaat 13620
 ggttttgtct tcaactcctt tgcatgccag gcggtgatgt ggatctggc ttctgtgagc 13680
 ctgtgctgtg ggcagggtg agctggagcc gccctctca gccgcctgc caggccttt 13740
 ccttaaaggc catccttaa accagacct catggctgcc agcaactgaa agcttctctg 13800
 acatctgtta ataaagcctt aggccttgt ctaagcgsaa ccgcctagac tttctttcag 13860
 ataatgtcc acatgtccat tttcaggtt ctctaagttg gagtggagtc tgggaagggt 13920
 tgtgaatgag gcttctgggc tatgggtgag gttccaatgg caggtagag cccctcgggc 13980
 caactgccat cctggaaagt agagacagca gtgcccgctg ccagaagag accagcaagc 14040
 caaactggag ccccatgac aggtgtcgc catgtggaaa gagtaactca caattgcaa 14100
 taaagtctca tgtggtttta tctactttt tttctttt ctttttttt gagacaagge 14160
 cttgccctcc caggctggag tgcagtggaa tgaccacagc tcaccgcaac ctcaaattct 14220
 tgcgttcaag tgaacctccc actttagcct cccaagtagc tgggactaca ggcgcagcc 14280
 atcacaccgc gctaattgaa aaatttttt tttgttttag atggaatctc actttgttgc 14340
 ccaggctggt ctcaaactcc tgggctcaag tgatcatcct gcttcagcgt ccgacttgtt 14400
 ggtattatag gcgtgagcca ctgggcctga cctagetacc attttttaat gcagaaatga 14460
 agactttagt aatgaaata acttgtccag gatagtcgaa taagtaactt ttagagctgg 14520
 gatttgaacc caggcaatct ggctccagag ctgggcctc actgctgaag gacactgtca 14580
 gottgggagg gtggctatgg toggctgtct gattctaggg agtgagggtc gtctttaag 14640
 caccocatto cattttcaga cagctttgtc agaaaggctg tcatatggag ctgacacctg 14700
 octccccaaag gcttccatag atcctctctg tacattgtaa ccttttattt tgaaatgaaa 14760
 attcacagga agttgtaagg ctagtacagg ggtacc 14796

<210> 123
 <211> 2597
 <212> ADN
 <213> *Homo sapiens*

ES 2 714 582 T3

<400> 123

gagcagcgcg cgcaagcagg ccaggggaag gtgggcgcag gtgaggggccc gaggtgtgcg 60
caggacttta gccggttgag aaggatcaag caggcatttg gagcacagggt gtctagaaac 120
ttttaagggg ccggttcaag aaggaaaagt tcccttctgc tgtgaaacta tttggcaaga 180
ggctggaggg cccaatggct gcaaaattgc aaccaacat tcccaaagcc aagagtctag 240
atggcgtcac caatgacaga accgcatctc aagggcagtg gggccgtgcc tgggaggtgg 300
actggttttc actggcgagc gtcactctcc tactgctggt cgcctcttc atcgtctact 360
acttcatcat ggcttgtgac cagtacagct gcgcctgac cggccctgtg gtggacatcg 420
tcaccggaca tgctcggctc tcggacatct gggccaagac tccacctata acgaggaaag 480
ccgcccagct ctataccttg tgggtcacct tccaggtgct tctgtacaag tctctccctg 540
acttctgcca taagtttcta cccggtctag taggaggcat ccaggagggg gccgtgactc 600
ctgcaggggt tgtgaacaag tatcagatca acggcctgca agcctggctc ctcacgcacc 660
tgctctgggt tgcaaacgct catctcctgt cctggttctc gcccaacctc atcttcgaca 720
actggatccc actgctgtgg tgcgccaaca tccttggtta tgccgtctcc accttcgcca 780
tggccaaggg ctacttcttc ccaccagcg ccagagactg caaattcaca ggcaatttct 840
tttacaacta catgatgggc atcgagtta accctcggat cgggaagtgg tttgacttca 900
agctgttctt caatggggcg cccgggatcg tcgcctggac cctcatcaac ctgtccttcg 960
cagcgaagca gccggagctc cacagccatg tgaccaatgc catggtcctg gtcaacgtcc 1020
tgcaggccat ctacgtgatt gacttcttct ggaacgaaac ctggtacctg aagaccattg 1080
acatctgoca tgaccacttc ggggtgtacc tgggctgggg cgactgtgtc tggctgcctt 1140
atctttacac gctgcagggt ctgtacttgg tgtaccacc cgtgcagctg tccaccccg 1200
acgcctgagg cgtcctgctg ctgggcctgg tgggctacta catcttcggg gtggccaacc 1260
accagaagga cctgttcgcg cgcacggatg ggcgctgcct catctggggc aggaagccca 1320
aggtcatcga gtgctcctac acatccgccc acgggcagag gcaccacagc aagctgctgg 1380
tgtcgggctt ctggggcgct gcccgccact tcaactacgt cggcgacctg atgggcagcc 1440
tggcctactg cctggcctgt ggcggtggcc acctgctgcc ctacttctac atcatctaca 1500
tggccatcct gctgaccac cgctgcctcc gggacgagca ccgctgcgcc agcaagtacg 1560
gccgggactg ggagcgtctac accgcccagc tgccttaccg cctgctgcct ggaatcttct 1620
aagggcacgc cctagggaga agccctgtgg ggctgtcaag agcgtgttct gccaggtcca 1680
tgggggctgg catcccagct ccaactcgag gagcctcagt ttctcatct gtaaacctgga 1740
gagagcccag cacttggcag gtgtccagta cctaatacag ctctgttctt tgcttttgcc 1800

ES 2 714 582 T3

```

ttcaagggaa ttccgagtgt ccagcactgc cgtattgcca gcacagacgg attttctcta 1860
atcagtgtcc ctgggcagga ggatgaccca gtcaccttta ctagtccttt ggagacaatt 1920
tacctgtatt aggagccag gccacgctac actctgocca cactggtgag caggaggtct 1980
tcccacgccc tgcattagg ctgcatttac tottgctaaa taaaagtggg agtggggcgt 2040
gcgcggtatc catgtattgc ctttcagctc tagatcccc tcccctgcct gctctgcagt 2100
cgtgggtggg gcccgctgcg cgtttctcct tggtagcgtg caoggtgttg aactgggaca 2160
ctggggagaa aggggctttc atgtcgtttc cttcctgctc ctgctgcaca gctgccagga 2220
gtgctctgcc tggagtctgc agacctcaga gaggtcccag cactggctgt ggctttcagg 2280
tgtaggcagg tgggctctgc ttcccgattc cctgtgagcg cccacctctc cgaagaatt 2340
ttctgtcttg cctgtgact gtgcagactc tggctogagc aaccgggga acttcacct 2400
caggggcctc tccacacctt ctccagcgag gaggtctcag tcccagcctc gggagggcac 2460
ctccttttct gtgctttctt ccctgaggea ttcttctca tccctagggt gttgtgtaga 2520
actcttttta aactctatgc tccgagtaga gttcatcttt atattaaact tcccctgttc 2580
aaaaaaaaa aaaaaaa 2597

```

- <210> 124
- <211> 1778
- 5 <212> ADN
- <213> *Homo sapiens*
- <400> 124

ES 2 714 582 T3

gaggagggaa aagggcgagca aaaaggaaga gtgggaggag gaggggagca caaaggatcc 60
 aggtctcccg acgggaggtt aataccaaga accatgtgtg ccgagcggct gggccagttc 120
 atgaccctgg ctttggtggt ggccacctt gaccggcgcc gggggaccga cgcaccaac 180
 ccaccogag gttcccaga caggagctcc cagcagaaag gccgcctgtc cctgcagaat 240
 acagcggaga tccagcactg tttggtcaac gctggcgatg tggggtgtgg cgtgtttgaa 300
 tgtttcgaga acaactcttg tgagattcgg ggcttacatg ggatttgcac gacttttctg 360
 cacaacgctg gaaaatttga tgcccagggc aagtcattca tcaaagacgc cttgaaatgt 420
 aaggcccacg ctctgcggca caggttcggc tgcataagcc ggaagtgcc gcccatcagg 480
 gaaatggtgt cccagttgca gggggaatgc taactcaagc acgacctgtg cggggctgcc 540
 caggagaaca cccgggtgat agtggagatg atccatttca aggacttgct gctgcacgaa 600
 ccctacgtgg acctcgtgaa cttgctgctg acctgtgggg aggaggtgaa ggaggccatc 660
 acccacagcg tgcaggttca gtgtgagcag aactggggaa gcctgtgctc catcttgagc 720
 ttctgcacct cggccatcca gaagcctccc acggcgcccc ccgagcgcca gcccaggtg 780
 gacagaacca agctctocag ggcccaccac ggggaagcag gacatcacct cccagagccc 840

 agcagtaggg agactggccg aggtgccaaag ggtgagcgag gtagcaagag ccacccaaac 900
 gcccatgccc gaggcagagt cgggggcctt ggggctcagg gaccttccgg aagcagcgag 960
 tgggaagacg aacagtctga gtattctgat atccggaggt gaaatgaaag gcctggccac 1020
 gaaatcttcc ctccacgccc tocattttct tatctatgga cattccaaa catttaacct 1080
 tagagagggg ggatgtcaca cgcaggattc tgtggggact gtggacttca tcgaggtgtg 1140
 tgttcgcgga acggacaggt gagatgagac ccctgggccc gtggggtctc aggggtgcct 1200
 ggtgaattct gcacttacac gtactcaagg gagcgcgccc gcgttatcct cgtacctttg 1260
 tcttcttcc atctgtgaag tcagtgggtg tcggccgctc tgttggtggg gaggtgaacc 1320
 agggaggggc agggcaaggc agggccccc gagctgggccc acacagtggg tgctgggccc 1380
 cgcocogaag cttctggtgc agcagcctct ggtgctgtct ccgcggaagt cagggcggct 1440
 ggattccagg acaggagtga atgtaaaaat aaatatcgtc tagaatgcag gagaaggggtg 1500
 gagaggaggc aggggcccag ggggtgcttg gtgccaaact gaaattcagt ttcttgtgtg 1560
 gggcttgcgg ttcagagctc ttggcgaggg tggagggagg agtgtcattt ctatgtgtaa 1620
 tttctgagcc attgtactgt ctgggctggg gggacactgt ccaagggagt ggcccctatg 1680
 agtttatatt ttaaccactg cttcaaactc cgatttcaact tttttattt atccagttat 1740
 atctacatat ctgtcatcta aataaatggc tttcaaac 1778

ES 2 714 582 T3

<211> 3246
 <212> ADN
 <213> *Homo sapiens*

5 <400> 125

```

    cgcaccatac cggcgcgggc acctggggag aaatggatgg agaagggacc tggctggaaa      60
    gctttgcccc gctgctctgc tccgccata agaggacccc tgaaatgtcc cgtgcagttt      120
    gttcaagtcc cctgtgtgat gaaatgtgcc tctcgcctta cccgtgtgag aatacctgtg      180
    gtgtggcagc gagtattttg gtatttgacc tgtccaaaga cgacttgata cctctataat      240
    gtaacagaaa aggtcagaaa atattaagca agtagaagtg tggagcatat taagcaagat      300
    gaacatctcg ggaagcagct gtggaagccc taactctgca gatacateta gtgactttaa      360
    ggacctttgg acaaaactaa aagaatgtca tgatagagaa gtacaagggt tacaagtaaa      420
    agtaaccaag ctaaaacagc aacgaatctt agatgcacaa agactagaag aattcttcac      480
    caaaaatcaa cagctgaggg aacagcagaa agtccttcat gaaaccatta aagttttaga      540
    agatcggtta agagcaggct tatgtgatcg ctgtgcagta actgaagaac atatgchgaa      600
    aaaacagcaa gagtttgaaa atatcggca gcagaatctt aaacttatta cagaacttat      660
    gaatgaaagg aatactctac aggaagaaaa taaaagctt tctgaacaac tccagcagaa      720
    
```


ES 2 714 582 T3

aattgagaat gatcaacagc atcaagcagc tgagcttgaa tgtgaggaag acgttattcc 780
agattcaccg ataacagcct tctcattttc tggcgtaaac cggctacgaa gaaaggagaa 840
ccccatgtc cgatacatag aacaaacaca tactaaattg gagcactctg tgtgtgcaaa 900
tgaaatgaga aaagtttcca agtcttcaac tcatccacaa cataatccta atgaaaatga 960
aattctagta gctgacactt atgaccaaag tcaatctcca atggcceaag cacatggaac 1020
aagcagctat accoctgata agtcatcttt taatttagct acagttgttg ctgaaacact 1080
tggacttggg gttcaagaag aatctgaaac tcaaggtccc atgagcccc ttggtgatga 1140
gctctaccac tgtctggaag gaaatcacia gaaacagcct tttgaggaat ctacaagaaa 1200
tactgaagat agtttaagat tttcagatto tacttcaaag actcctcctc aagaagaatt 1260
acctactcga gtgtcatctc ctgtatttgg agctacctct agtatcaaaa gtggtttaga 1320
tttgaataca agtttgtccc cttctctttt acagcctggg aaaaaaaaaac atctgaaaac 1380
actccctttt agcaacactt gtatatctag attagaaaaa actagatcaa aatctgaaga 1440
tagtgccott ttcacacatc acagtcttgg gtctgaagtg aacaagatca ttatccagtc 1500
atctaataaa cagatactta taaataaaaa tataagtga tccctaggtg aacagaatag 1560
gactgagtac ggtaaagatt ctaaacactga taaacatttg gagccctga aatcattggg 1620
aggccgaaca tccaaaagga agaaaactga ggaagaaagt gaacatgaag taagctgccc 1680
ccaagcttct tttgataaag aaaatgcttt cctttttcca atggataatc agttttccat 1740
gaatggagac tgtgtgatgg ataaacctct ggatctgtct gatcgatttt cagctattca 1800
gctcaagag aaaagccaag gaagtgagac ttctaaaaac aaatttaggc aagtgactct 1860
ttatgaggct ttgaagacca ttccaaaggg cttttcctca agccgtaagg cctcagatgg 1920
caactgcaog ttgcccgaag attccccagg ggagccctgt tcacaggaat gcatcatcct 1980
tcagcccttg aataaatgct ctccagacaa taaacctca ttacaataa aagaagaaaa 2040
tgetgtcttt aaaattctc tacgtccacg tgaaagtttg gagactgaga atgttttaga 2100
tgacataaag agtgctggtt ctcatgagcc aataaaaata caaaccaggt cagaccatgg 2160
aggatgtgaa cttgcatcag ttcttcagtt aaatccatgt agaactggta aaataaagtc 2220
tctacaaaac aaccaagatg tatcctttga aaatatccag tggagtatag atccgggagc 2280
agacctttct cagtataaaa tggatgttac tgtaatagat acaaaggatg gcagtcagtc 2340
aaaattagga ggagagacag tggacatgga ctgtacattg gttagtgaaa ccgttctctt 2400
aaaaatgaag aagcaagagc agaagggaga aaaaagttca aatgaagaaa gaaaaatgaa 2460
tgatagcttg gaagatatgt ttgatcggac aacacatgaa gagtatgaat cctgtttggc 2520
agacagtttc tcccaagcag cagatgaaga ggaggaattg tctactgcca caaagaaact 2580
acacactcat ggtgataaac aagacaaagt caagcagaaa gcgtttgtgg agccgtattt 2640

ES 2 714 582 T3

taaaggtgat gaaagagaga ctagcttgca aaattttcoct catattgagg tggttcggaa 2700
aaaagaggag agaagaaaac tgcttgggca cacgtgtaag gaatgtgaaa tttattatgc 2760
agatatgcca gcagaagaaa gagaaaagaa attggcttcc tgctcaagac accgattccg 2820
ctacattcca cccaacacac cagagaatth ttgggaagtt ggttttcoct ccaactcagac 2880
ttgtatggaa agaggttata ttaaggaaga tottgatcoct tgtcctcgtc caaaaagacy 2940
tcagcocttac aacgcaatat tttctccaaa aggcaaggag cagaagacat agacgttgaa 3000
acagaaacag aaggatgaag gacagttttt tccttcttag ttatttatag ttaaagttgg 3060
tactaaacat tgattttttt gatcttctgt aaatggattt ataatcagt tttctattga 3120
aaatgtttgt gatattttgc tttgacacct ttaaaacaat aaggcgcttt cattttgcac 3180
tctaacttaa gagtttttac tttatgtagt gatacctaata acaattttga aaatacaaaa 3240
aaaaaa 3246

<210> 126
<211> 3085
<212> ADN
<213> *Homo sapiens*

5

<400> 126

gagcagccaa aaggcccgcg gagtcgcgct gggccgcccc ggcgcagctg aaccgggggc 60
cgcgcctgcc aggccgacgg gtctggccca gcctggcgcc aaggggttgc tgcgctgtgg 120
agacgcggag ggtcgaggcg gcgcggcctg agtgaaccc aatggaaaa gcatgacatt 180
tagaagtaga agacttagct tcaaatcoct actccttcac ttactaattt tgtgatttgg 240
aaatatccgc gcaagatggt gacggtgcag acttgggtag tgcaagcctt gtttattttc 300
ctcaccactg aatctacagg tgaacttcta gatccatgtg gttatatcag tcctgaatct 360
ccagttgtac aacttcattc taatttcaact gcagtttgtg tgctaaagga aaaatgtatg 420
gattattttc atgtaaagtc taattacatt gtctggaaaa caaacattt tactattcct 480
aaggagcaat atactatcat aacagaaca gcatccagtg tcacctttac agatatagct 540
tcattaaata ttcagctcac ttgcaacatt cttacattcg gacagcttga acagaatggt 600
tatggaatca caataatttc aggcttgcct ccagaaaaac ctaaaaattt gagttgcatt 660
gtgaacgagg ggaagaaaat gaggtgtgag tgggatggtg gaagggaaac aacttggag 720
acaaacttca ctttaaaatc tgaatgggca acacacaagt ttgctgattg caagcaaaa 780
cgtgacaccc ccacctcatg cactgttgat tatttctactg tgtattttgt caacattgaa 840
gtctgggtag aagcagagaa tgcccttggg aaggttacat cagatcatat caattttgat 900
cctgtatata aagtgaagcc caatccgcca cataattht cagtgatcaa ctcagaggaa 960
ctgtctagta tottaaaatt gacatggacc aaccaagta ttaagagtgt tataatacta 1020

10

ES 2 714 582 T3

aatataaca ttcaatatag gaccaaagat gcctcaactt ggagccagat tcctcctgaa 1080
 gacacagcat ccacccgatc ttcattcact gtccaagacc ttaaaccottt tacagaatat 1140
 gtgttttagga ttcgctgtat gaaggaagat ggtaagggat actggagtga ctggagtgaa 1200
 gaagcaagtg ggatcaccta tgaagataga ccatctaaag caccaagttt ctggtataaa 1260
 atagatccat ccataactca aggctacaga actgtacaac tcgtgtggaa gacattgcct 1320
 ccttttgaag ccaatggaaa aatccttgat tatgaagtga ctctcacaag atggaaatca 1380
 catttcaaaa attacacagt taatgccaca aaactgacag taaatctcac aaatgatcgc 1440
 tatctagcaa ccoctaacagt aagaaatcct gttggcaaat cagatgcagc tgttttaact 1500
 atccctgcct gtgactttca agctactcac cctgtaatgg atcttaaagc attccccaaa 1560
 gataacatgc tttgggtgga atggactact ccaagggaat ctgtaaagaa atatatactt 1620
 gagtgggtgtg tgttatcaga taaagcacc tgtatcacag actggcaaca agaagatggt 1680
 accgtgcac gcacctattt aagagggaac ttagcagaga gcaaatgcta tttgataaca 1740
 gttactccag tatatgctga tggaccagga agccctgaat ccataaaggc ataccttaaa 1800
 caagctccac cttccaaagg acctactggt cggacaaaaa aagtagggaa aaacgaagct 1860
 gtcttagagt gggaccaact tcctgttgat gttcagaatg gatttatcag aaattatact 1920
 atattttata gaaccatcat tggaaatgaa actgctgtga atgtggattc ttcccacaca 1980
 gaatatacat tgtcctcttt gactagtgac acattgtaca tggtagaat ggcagcatac 2040
 acagatgaag gtgggaagga tggccagaa ttcactttta ctaccccaaa gtttgotcaa 2100
 ggagaaattg aagccatagt cgtgcctggt tgcttagcat tcctattgac aactcttctg 2160
 ggagtgtgt tctgctttaa taagcgagac ctaattaaaa aacacatctg gctaattggt 2220
 ccagatcctt caaagagtca tattgccag tggcaccctc aactcctcc aaggcacaat 2280
 ttaattcaa aagatcaaat gtattcagat ggcaatttca ctgatgtaag tgttgtggaa 2340
 atagaagcaa atgacaaaaa gccttttcca gaagatctga aatcattgga cctgttcaaa 2400
 aaggaaaaaa ttaatactga aggacacagc agtggatttg ggggtcttc atgcatgtca 2460
 tcttctaggc caagcatttc tagcagtgat gaaaatgaat cttcacaaaa cacttcgagc 2520
 actgtccagt attctaccgt ggtacacagt ggctacagac accaagttcc gtcagtccaa 2580
 gtettctcaa gatccgagtc taccagccc ttgttagatt cagaggagcg gccagaagat 2640
 ctacaattag tagatcatgt agatggcggg gatggtattt tgcccaggca acagtacttc 2700
 aacagaact gcagtcagca tgaatccagt ccagatattt cacattttga aaggtcaaag 2760
 caagtttcat cagtcaatga ggaagatfff gttagactta aacagcagat ttcagatcat 2820
 atttcacaat cctgtggatc tgggcaaatg aaaatgtttc aggaagtttc tgcagcagat 2880

ES 2 714 582 T3

gcttttggtc caggtactga gggacaagta gaaagatttg aaacagttgg catggaggct 2940
 gcgactgatg aaggcatgcc taaaagttac ttaccacaga ctgtacggca aggcggctac 3000
 atgcctcagt gaaggactag tagttcctgc tacaacttca gcagtaccta taaagtaaag 3060
 ctaaaatgat tttatctgtg aattc 3085

5 <210> 127
 <211> 585
 <212> ADN
 <213> *Homo sapiens*
 <400> 127

ctgagaactga cctgcaggac gaaacatga agagcctgat ccttcttgcc atcctggccg 60
 ccttagcggg agtaactttg tgttatgaat cacatgaaag catggaatct tatgaactta 120
 atcccttcat taacaggaga aatgcaaata ccttcatatc ccctcagcag agatggagag 180
 ctaaagtcca agagaggatc cgagaacgct ctaagcctgt ccacgagctc aataggggaag 240
 cctgtgatga ctacagactt tgcgaacgct acgccatggg ttatggatac aatgctgcct 300
 ataatcgcta cttcaggaag cgccgagga ccaaatgaga ctgaggggaag aaaaaaatc 360
 tctttttttc tggaggctgg cacctgattt tgtatcccc tgtagcagca ttactgaaat 420
 acataggctt atatacaatg cttctttcct gtatattctc ttgtctggct gcaccocctt 480
 ttcccccccc cagattgata agtaatgaaa gtgcactgca gtgaggggtca aaggagagtc 540
 10 aacatatgtg attgttccat aataaacttc tgggtgtgata ctttc 585

15 <210> 128
 <211> 1208
 <212> ADN
 <213> *Homo sapiens*
 <400> 128

ES 2 714 582 T3

aagcagacac aatggtaaga atggtgcctg tcctgctgtc totgctgctg cttctgggtc	60
ctgctgtccc ccaggagaac caagatggtc gttactctct gacctatgtc tacactgggc	120
tgtccaagca tgttgaagac gtccccgogt ttcaggocct tggotcaactc aatgacctoc	180
agttctttag atacaacagt aaagacagga agtctcagcc catgggactc tggagacagg	240
tggaaggaat ggaggattgg aagcaggaca gccaaactca gaaggccagg gaggacatct	300
ttatggagac cctgaaagac atcgtggagt attacaacga cagtaacggg tctcaogtat	360
tgcaggggaag gtttggttgt gagatcgaga ataacagaag cagcggagca ttctggaaat	420
attactatga tggaaaggac tacattgaat tcaacaaaga aatcccagcc tgggtccocct	480
tcgaccagc agcccagata accaagcaga agtgggaggc agaaccagtc taogtgcagc	540
ggccaaggc ttacctggag gaggagtgcc ctgcgactct gcggaatac ctgaaataca	600
gcaaaaatat cctggaccgg caagatcctc cctctgtggt ggtcaccagc caccaggccc	660
caggagaaaa gaagaaactg aagtgcctgg cctacgactt ctaccaggg aaaattgatg	720
tgcaactggac tcgggccggc gaggtgcagg agcctgagtt acggggagat gttcttcaca	780
atggaaatgg cacttaccag tcctgggtgg tgggtggcagt gccccgcag gacacagccc	840
cctactcctg ccacgtgcag cacagcagcc tggccagcc cctcgtggtg ccctgggagg	900
ccagctagga agcaagggtt ggaggcaatg tgggatctca gaccagtag ctgcccttc	960
tgcctgatgt gggagctgaa ccacagaaat cacagtcaat ggatccacaa ggocctgagga	1020
gcagtgtggg gggacagaca ggaggtggat ttggagaccg aagactggga tgccctgtctt	1080
gagtagactt ggacccaaaa aatcatctca ccttgagccc acccccaccc cattgtctaa	1140
tctgtagaag ctaataaata atcatcctc cttgcctagc aaaaaaaaaa aaaaaaaaaa	1200
aaaaaaaa	1208

<210> 129
 <211> 3315
 <212> ADN
 <213> *Homo sapiens*

 <400> 129

5

ES 2 714 582 T3

aataaatttc tctgtgattg gttggtgaag gttttcaaac cggagctgtg ggcgcggcgc	60
tgctctgccg ttgggtgagg cgcgagcga agtgaagggg ggcccagggt gggccaggct	120
gactgaatgt atctcctagc tatggactaa ataatacatg gggggaaata aacaagtatt	180
catgaggggtg aaaatgtgac ccagcaggaa aattacaact attttcaatt gacgttgaat	240
aggatgagtc atggaattta agtgatttac tgaagattat actactggta gatagaagag	300
ctaaagaaag atggatacta tgatgctgaa tgtgcggaat ctgtttgagc agcttgtgcg	360
ccgggtggag attctcagtg aaggaaatga agtccaattt atccagttgg cgaaggactt	420
tgaggatttc cgtaaaaagt ggcagaggac tgaccatgag ctggggaaat acaaggatct	480
tttgatgaaa gcagagactg agcgaagtgc tctggatggt aagctgaagc atgcacgtaa	540
tcaggtggat gtagagatca aacggagaca gagagctgag gctgactgcg aaaagctgga	600
acgacagatt cagctgattc gagagatgct catgtgtgac acatctggca gcattcaact	660
aagcgaggag caaaaatcag ctctggcttt tctcaacaga ggccaacat ccagcagcaa	720
tgctgggaac aaaagactat caaccattga tgaatctggt tccattttat cagatatcag	780
ctttgacaag actgatgaat cactggattg ggactcttct ttggtgaaga ctttcaaact	840
gaagaagaga gaaaagaggc gctctactag ccgacagttt gttgatggtc cccctggacc	900
tgtaaagaaa actcgttcca ttggctctgc agtagaccag gggaatgaat ccatagttgc	960
aaaaactaca gtgactgttc ccaatgatgg ogggccatc gaagctgtgt ccactattga	1020
gactgtgcca tattggacca ggagccgaag gaaaacaggt actttacaac cttggaacag	1080

ES 2 714 582 T3

tgactccacc ctgaacagca ggcagctgga gccagaact gagacagaca gtgtggggcac 1140
 gccacagagt aatggagggga tgcgcctgca tgactttggt tctaagacgg ttattaaacc 1200
 tgaatcctgt gttccatgtg gaaagcggat aaaatttggc aaattatctc tgaagtgtcg 1260
 agactgtcgt gtggtctctc atccagaatg tggggaccgc tgtccccttc cctgcattcc 1320
 taccctgata ggaacacctg tcaagattgg agagggaatg ctggcagact ttgtgtccca 1380
 gacttctcca atgatcccct ccattgttgt gcattgtgta aatgagattg agcaaagagg 1440
 tctgactgag acaggcctgt ataggatctc tggctgtgac cgcacagtaa aagagctgaa 1500
 agagaaattc ctcagagtga aaactgtacc cctcctcagc aaagtggatg atatccatgc 1560
 tatctgtagc cttctaaaag acttttctcg aaacctcaa gaacctcttc tgacctttog 1620
 ccttaacaga gcctttatgg aagcagcaga aatcacagat gaagacaaca gcatagctgc 1680
 catgtaccaa gctgttggtg aactgcccc aagccaacagg gacacattag ctttccctcat 1740
 gattcacttg cagagagtgg ctcagagtcc acatactaaa atggatgttg ccaatctggc 1800
 taaagtcttt ggccctacaa tagtggccca tgctgtgccc aatccagacc cagtgacaat 1860
 gttacaggac atcaagcgtc aacccaaggt ggttgagcgc ctgctttcct tgcctctgga 1920
 gtattggagt cagttcatga tgggtggagca agagaacatt gaccccctac atgtcattga 1980
 aaactcaaat gccttttcaa caccacagac accagatatt aaagtgagtt tactgggacc 2040
 tgtgaccact cctgaacatc agcttctcaa gactccttca tctagttccc tgtcacagag 2100
 agtccgttcc accctcacca agaacactcc tagatttggg agcaaaagca agtctgccac 2160
 taacctagga cgacaaggca acttttttgc ttctccaatg ctcaagtgaa gtcacatctg 2220
 cctgttactt cccagcattg actgactata agaaaggaca catctgtact ctgctctgca 2280
 gcctcctgta ctcaactacta cttttagcat tctccagget tttactcaag ttttaattgtg 2340
 catgagggtt ttattaaaac tatatatatc tccccttctc tctcctcaag tcacataata 2400
 tcagcacttt gtgctggtca ttgttgggag cttttagatg agacatcttt ccaggggtag 2460
 aagggttagt atggaattgg ttgtgattct ttttggggaa gggggttatt gttcctttgg 2520
 cttaaagcca aatgctgctc atagaatgat ctttctctag tttcatttag aactgatttc 2580
 cgtgagacaa tgacagaaac cctacctatc tgataagatt agcttgtctc aggggtgggaa 2640
 gtgggagggc agggcaaaga aaggattaga ccagaggatt taggatgcct ccttctaaga 2700
 accagaagtt ctcatcctcc attatgaact gagctataat atggagcttt cataaaaatg 2760
 ggatgcattg aggacagaac tagtgatggg agtatgcgta gctttgattt ggatgattag 2820
 gtctttaata gtgttgagtg gcacaacctt gtaaagtgtga aagtacaact cgtattttatc 2880
 tctgatgtgc tgctggctga actttggggt catttggggt caaagccagt ttttctttta 2940

ES 2 714 582 T3

```
aaattgaatt cattctgatg cttggccccc atacccccaa cttgtccag tggagoccaa 3000
cttctaaagg tcaatatatc atcctttggc atcccaacta acaataaaga gtaggctata 3060
agggagatt gtcaatattt tgtggtaaga aaagctacag tcattttttc tttgcacttt 3120
ggatgctgaa atttttccca tggaacatag ccacatctag atagatgtga gctttttctt 3180
ctgttaaaat tattcttaat gtctgtaaaa acgattttct tctgtagaat gtttgacttc 3240
gtattgacc ttatctgtaa aacacctatt tgggataata tttggaaaaa aagtaaatag 3300
ctttttcaaa atgaa 3315
```

<210> 130
<211> 2128
<212> ADN
<213> *Homo sapiens*

<400> 130

5

ES 2 714 582 T3

cttggaagac ttgggtcctt gggtcgcagg ctggagtgca atgggtgat ctcagctcac 60
 tgcaacctct gcttcctggg ttaagtgat tctcctgcct cagcctcccg agtagctggg 120
 attacaggca tcatggaccg atctaaagaa aactgcattt caggacctgt taaggctaca 180
 gctccagttg gaggtccaaa acgtgttctc gtgactcagc aatttccttg tcagaatcca 240
 ttacctgtaa atagtggcca ggctcagcgg gtcttgtgtc cttcaaattc ttcccagcgc 300
 gttcctttgc aagcacaaaa gcttgtctcc agtcacaagc cggttcagaa tcagaagcag 360
 aagcaattgc aggcaaccag tgtacctcat cctgtctcca ggccactgaa taacacccaa 420
 aagagcaagc agcccctgcc atcggcacct gaaaataatc ctgaggagga actggcatca 480
 aaacagaaaa atgaagaatc aaaaaagagg cagtgggctt tggagaactt tgaaattggt 540
 cgccctctgg gtaaaggaaa gtttgtaat gtttatttgg caagagaaaa gcaaagcaag 600
 tttattctgg ctcttaaagt gttatttaa gctcagctgg agaaagcccg agtggagcat 660
 cagctcagaa gagaagtaga aatacagtc caccttcggc atcctaatat tcttagactg 720
 tatggttatt tccatgatgc taccagagtc tacctaattc tggaatatgc accacttgga 780
 acagtttata gagaacttca gaaacttca aagtttgatg agcagagAAC tgctacttat 840
 ataacagaat tggcaaatgc cctgtcttac tgtcattcga agagagttat tcatagagac 900
 attaagccag agaacttact tcttgatca gctggagagc ttaaaattgc agattttggg 960
 tggtcagtac atgctccatc ttccaggagg accactctct gtggcacccct ggactacctg 1020
 ccccctgaaa tgattgaagg tcggatgcat gatgagaagg tggatctctg gagccttgga 1080
 gttcctttgct atgaattttt agttgggaag cctccttttg aggcaaacac ataccaagag 1140
 acctacaaaa gaatatcacg ggttgaattc acattccctg actttgtaac agagggagcc 1200
 agggacctca tttcaagact gttgaagcat aatcccagcc agaggccaat gctcagagaa 1260

ES 2 714 582 T3

gtacttgaac acccoctggat cacagcaaat tcatcaaaaac catcaaattg ccaaaaacaaa 1320
 gaatcagcta gcaaacagtc ttaggaatcg tgcaggggga gaaatcottg agccagggct 1380
 gccatataac ctgacaggaa catgctactg aagtttattt taccattgac tgctgcctc 1440
 aatctagaac gctacacaag aatatattgt tttactcagc aggtgtgctt taacctcctt 1500
 attcagaaag ctccacatca ataacatga cactctgaag tgaaagtagc cacgagaatt 1560
 gtgctactta tactggttca taatctggag gcaaggttcg actgcagccg ccccgtcagc 1620
 ctgtgctagg catggtgtct tcacaggagg caaatccaga gcctggctgt ggggaaagtg 1680
 accactctgc cctgaccccg atcagttaag gagctgtgca ataaccttcc tagtacctga 1740
 gtgagtgtgt aacttattgg gttggcgaag cctggtaaag ctggttggat gagtatgtga 1800
 ttctttttaa gtatgaaaat aaagatatat gtacagactt gtattttttc tctggtggca 1860
 ttcccttagg aatgctgtgt gtctgtcccg caccocggta ggctgattg ggtttctagt 1920
 cctccttaac cacttatctc ccatatgaga gtgtgaaaaa taggaacacg tgctctacct 1980
 ccatttaggg atttgcttgg gatacagaag aggccatgtg tctcagagct gtttaaggct 2040
 tattttttta aaacattgga gtcatagcat gtgtgtaaac tttaaatatg caaataaata 2100
 agtatctatg tcaaaaaaaaa aaaaaaaaa 2128

5 <210> 131
 <211> 586
 <212> ADN
 <213> *Homo sapiens*

<400> 131

ccagcctttc agtgcaggct ccagccctcc acccccaccc gagttgcagg atgtogatga 60
 cagacttgct gaacgctgag gacatcaaga aggcggtggg agcctttagc gctaccgact 120
 ccttcgacca caaaaagttc ttccaaatgg tcggcctgaa gaaaaagagt gcggatgatg 180
 tgaagaaggt gtttcacatg ctggacaagg acaaaagtgg cttcatcgag gaggatgagc 240
 tgggattcat cctaaaaggc ttctccccag atgccagaga cctgtctgct aaagaaacca 300
 agatgctgat ggctgctgga gacaaagatg gggacggcaa aattgggggt gacgaattct 360
 ccactctggt ggctgaaagc taagaagcac tgactgcccc tggctttcca cctctctgcc 420
 ctgaacaccc aatctcggcc cctctcgcca ccctcctgca tttctgttca gttcgtttat 480
 gttatttttt actccccat ccctgtggc cctctaataga caccattott ctggaaaatg 540
 ctggagaagc aataaaggtt gtaccagtca gaaaaaaaa aaaaaa 586

10
 15 <210> 132
 <211> 817
 <212> ADN
 <213> *Homo sapiens*

<400> 132

ES 2 714 582 T3

agtcoctgcgt ccgggccccg aggcgcagca gggcaccagg tggagcacca gctacgcgtg 60
 ggcagcgcga gcgtccctag caccgagcct ccgcagccg ccgagatgct gcgaacagag 120
 agctgcgcgc ccaggtcgcc cgccggacag gtggccgcgg cgtccccgct cctgctgctg 180
 ctgctgctgc tcgcctggtg cgcggggcgc tcgccgaggtg ctccaatatt acctcaagga 240
 ttacagcctg aacaacagct acagttgtgg aatgagatag atgatacttg ttcgtctttt 300
 ctgtccattg attctcagcc tcaggcatcc aacgcactgg aggagctttg ctttatgatt 360
 atgggaatgc taccaaagcc tcaggaacaa gatgaaaaag ataataactaa aaggttotta 420
 tttcattatt cgaagacaca gaagttgggc aagtcaaagt ttgtgtcgtc agttgtgcat 480
 cggttgctgc agctcgttcc tcacctgcat gagagaagaa tgaagagatt cagagtggac 540
 gaagaattcc aaagtccctt tgcaagtcaa agtcgaggat attttttatt caggccaagg 600
 aatggaagaa ggtcagcagg gttcatttaa aatggatgcc agctaatttt ccacagagca 660
 atgctatgga atacaaaatg tactgacatt ttgttttctt ctgaaaaaaaa tccttgctaa 720
 atgtaactctg ttgaaaatcc ctgtgttgtc aatgttctca gttgtaacaa tgttgtaaat 780
 gttcaatttg ttgaaaatta aaaaatctaa aaataaa 817

5 <210> 133
 <211> 1967
 <212> ADN
 <213> *Homo sapiens*

10 <400> 133

ES 2 714 582 T3

gaaggcgtgg ctccctcccg ggccagtgag cctggcgccg ccgcggccgc gtcccagcag 60
 cggagtaggg cggcggctgc gccccgcacc atgggggcag cccagcccca gccgcggtaa 120
 acgcccacct ccgcgcgcgc ccgcgcccgt ctgccccctc ccgctgcggc tctctggacg 180
 ccatcccctc ctcaoctcga agccaacatg aaggagaccg ggggctacgg aggggatgcc 240
 cccttctgca ccgcctcaa ccactoctac acaggcatgt gggcgcccga gcgttcggcc 300
 gaggcgcggg gcaacctcac ggcocotoca gggctctggc aggattgcgg atcgggtgcc 360
 gtggccttcc cgatcaccat gctgctcact ggtttcgtgg gcaacgcact ggccatgctg 420
 ctcgtgtcgc gcagctaccg gcgcggggag agcaagcgca agaagtcctt cctgctgtgc 480
 atcggctggc tggcgctcac cgacctggtc gggcagcttc tcaccacccc ggtcgtcacc 540
 gtcgtgtacc tgtccaagca gcgttgggag cacatcgacc cgtcggggcg gctctgcacc 600
 tttttcgggc tgaccatgac tgttttcggg ctctoctcgt tgttcatcgc cagcgcctatg 660
 gccgtogagc gggcgctggc catcagggcg ccgcactggt atgcgagcca catgaagacg 720
 cgtgccaccc gcgctgtgct gctcggcggtg tggtcggccg tgctcgcctt cgcctgctg 780
 ccggtgctgg gcgtgggcca gtacaccgtc cagtggcccg ggacgtggtg ctccatcagc 840

ES 2 714 582 T3

accgggagag ggggcaacgg gactagctct togcataact ggggcaacct tttcttcgcc 900
 tetgcctttg ccttcctggg gctcttggcg ctgacagtca ccttttctg caacctggcc 960
 accattaagg ccttgggtgc ccgctgccgg gccaaaggcca cggcatctca gtccagtgcc 1020
 cagtggggcc gcatcacgac cgagacggcc attcagotta tggggatcat gtgcgtgctg 1080
 toggctgct ggtctccgct cctgataatg atggtgaaa tgatcttcaa tcagacatca 1140
 gttgagcaact gcaagacaca cacggagaag cagaaagaat gcaacttctt cttaatagct 1200
 gttgcctgg cttcactgaa ccagatcttg gatccttggg tttacctgct gttaagaaag 1260
 atccttcttc gaaagttttg ccagatgaga aaaagaagac tcagagagca agctcctctt 1320
 cttcccacc ctactgtgat tgatccttca aggttctgtg ctcagccctt ccggttggtc 1380
 ttggatttgt cctttccgc catgtcttca tcacatcac aacttcact aacacttgcg 1440
 agcttcaaac ttcttagaga accctgcagt gtocagotaa gctgatgact tgaagataaa 1500
 tctgcctaac cctgggatga agtatctgtg aactatcttg acagcagatg aggaatcttg 1560
 gggaaattaa aacctgcctt tctgccagga tcacatcact ggaagctcca tgactctctt 1620
 tttgtaaaag aaaaaaaaaat cacagaaaca cccacctccc aaactattct cttttacttc 1680
 ttcccccaag cccacccccca aatataactg ttatccagaa gctggtatgt cctgtttcca 1740
 tacatgtttt tgtactttta ctatatctac atacatcaat taaacttatg tcctattggt 1800
 ttgtgaattt atatttgcgt atacattatc atatgtaaaa tttgcatttt tttattgaaa 1860
 attatgtttc ttgagattta tcacattga aacatggagc tctaaatcgt taattttaac 1920
 cgctatagag tattccataa tttgaataaa gcataatttg tttgtac 1967

<210> 134
 <211> 3524
 <212> ADN
 <213> *Homo sapiens*

5

<400> 134

ES 2 714 582 T3

tctccgtcag	ccgcattgcc	cgctcggcgt	ccggcccccg	accogtgetc	gtccgcccgc	60
ccgcccgcgc	gcccgggcca	tgaagccaa	ggtcgtggtc	gtgctgggoc	tcgtgctgac	120
cgcgctctgc	ctcagcgacg	ggaagcccgt	cagcctgagc	tacagatgcc	catgccgatt	180
cttcgaaagc	catgttgcc	gagccaacgt	caagcatctc	aaaattotca	acaactocaaa	240
ctgtgccctt	cagattgtag	cccggctgaa	gaacaacaac	agacaagtgt	gcattgaccc	300
gaagctaaag	tggattcagg	agtacctgga	gaaagcttta	aacaagaggt	tcaagatgtg	360
agagggtcag	acgcctgagg	aacccttaca	gtaggagccc	agctctgaaa	ccagtgttag	420
ggaagggcct	gccacagcct	ccoctgocag	ggcagggccc	caggcattgc	caagggcttt	480
gttttgaca	ctttgocata	ttttcaccat	ttgattatgt	agcaaaatac	atgacattta	540

ES 2 714 582 T3

tttttcattt agtttgatta ttcagtgca ctggcgacac gtagcagctt agactaaggc 600
 cattattgta cttgcottat tagagtgctt ttccacggag ccactcctct gactcagggc 660
 tcctggggtt tgtattctct gagctgtgca ggtggggaga ctgggctgag ggagcctggc 720
 cccatgggtca gccctagggt ggagagccac caagagggac gcctgggggt gccaggacca 780
 gtcaacctgg gcaaagccta gtgaaggctt ctctctgtgg gatgggatgg tggagggcca 840
 catgggaggc tcacccctt ctccatccac atgggagccg ggtctgcctc ttctgggagg 900
 gcagcagggc taccctgagc tgaggcagca gtgtgaggcc agggcagagt gagaccacgc 960
 cctcatcccg agcacctoca catcctccac gttctgtctca tcattctctg tctcatccat 1020
 catcatgtgt gtccaagact gtctccatgg ccccgcaaaa ggactctcag gaccaaaagct 1080
 ttcatgtaa ctgtgcacca agcaggaaat gaaaatgtct tgtgttacct gaaaacactg 1140
 tgcacatctg tgtcttgtgt ggaatattgt ccattgtcca atcctatgtt tttgttcaaa 1200
 gccagcgtcc tcctctgtga ccaatgtctt gatgcatgca ctgttcccc tgtgcagccg 1260
 ctgagcgagg agatgtcct tgggccctt gagtgcagtc ctgatcagag ccgtggctct 1320
 ttggggtgaa ctaccttgggt tccccactg atcacaaaaa catggtgggt ccatgggcag 1380
 agcccaaggg aattcgggtgt gcaccagggt tgaccocaga ggattgctgc cccatcagtg 1440
 ctccctcaca tgtcagtacc ttcaaactag ggccaagccc agcactgctt gaggaaaaca 1500
 agcattcaca acttgttttt ggtttttaa acccagtcca caaaataacc aatcctggac 1560
 atgaagattc tttcccaatt cacatctaac ctcatcttct tcaccatttg gcaatgccat 1620
 catctctctg cttcctctct ggccctctct gctctgcgtg tcacctgtgc ttcgggacct 1680
 tcccacagga catttctcta agagaacaat gtgctatgtg aagagtaagt caacctgcct 1740
 gacatttggg gtgttcccc cccactgagg gcagtcgata gagctgtatt aagccactta 1800
 aatgttcac ttttgacaaa ggcaagcact tgtgggtttt tgttttgtt ttcatcagt 1860
 cttacgaata cttttgcct ttgattaag actccagtta aaaaaattt taatgaagaa 1920
 agtgaaaaac aaggaagtca aagcaaggaa actatgtaac atgtaggaag taggaagtaa 1980
 attatagtga tgtaatcttg aattgtaact gttcgtgaat ttaataatct gtagggtaat 2040
 tagtaacatg tgttaagtat tttcataagt atttcaaatt ggagcttcat ggcagaaggc 2100
 aaacctatca acaaaaattg tccttaaac aaaaattaaa atcctcaatc cagctatggt 2160
 atattgaaaa aatagagcct gagggatctt tactagttat aaagatacag aactctttca 2220
 aaaccttttg aaattaacct ctactatac cagtataatt gagttttcag tggggcagtc 2280
 attatccagg taatccaaga tattttaaaa totgtcaogt agaacttggg tgtacctgcc 2340
 cccaatccat gaaccaagac cattgaattc ttggttgagg aaacaacat gaccctaaat 2400

ES 2 714 582 T3

cttgactaca gtcaggaaag gaatcatttc tatttctcct ccatgggaga aaatagataa 2460
 gagtagaaac tgcagggaaa attatttgca taacaattcc tctactaaca atcagctcct 2520
 tcctggagac tgcccagcta aagcaatatg catttaaata cagtcttcca tttgcaaggg 2580
 aaaagtctct tgtaatccga atctcttttt gctttcgaac tgctagtcaa gtgogtccac 2640
 gagctgttta ctagggatcc ctcatctgtc cctccgggac ctggtgctgc ctctacctga 2700
 cactcccttg ggctccctgt aacctcttca gaggcctcgc ctgccagctc tgtatcagga 2760
 cccagaggaa gggggccagag gctcgttgac tggctgtgtg ttgggattga gtctgtgcca 2820
 cgtgtatgtg ctgtggtgtg tccccctctg tccaggcact gagataccag cgaggaggct 2880
 ccagagggca ctctgcttgt tattagagat taacctctga gaaaaaagct tccgcttgga 2940
 gcagaggggc tgaatagcag aaggttgcac ctcccccaac cttagatggt ctaagtcttt 3000
 ccattggatc tcattggacc ctccatggt gtgatogtct gactggtgtt atcacogtgg 3060
 gctcctgac tgggagttga tcgcctttcc cagggtgctac acccttttcc agctggatga 3120
 gaatttgagt gctctgatcc ctctacagag cttcctgac tcattctgaa ggagcccat 3180
 tcctgggaaa tattccctag aaacttccaa atcccctaag cagaccactg ataaaaccat 3240
 gtagaaaatt tgttattttg caacctcgct ggactctcag tctctgagca gtgaatgatt 3300
 cagtgtaaa tgtgatgaat actgtatttt gtattgtttc aagtgcattc cccagataat 3360
 gtgaaaatgg tccaggagaa ggccaattcc tatacgcagc gtgctttaa aaataaataa 3420
 gaaacaactc tttgagaaac aacaatttct actttgaagt cataccaatg aaaaaatgta 3480
 tatgactta taattttcct aataaagttc tgtactcaaa tgta 3524

<210> 135
 <211> 1705
 <212> ADN
 <213> *Homo sapiens*
 <400> 135

5

ES 2 714 582 T3

ggctcactgc atotccggct cctggactca agcgattctc ctgcctcagg ctoccaaagg 60
 ggagagcagc aaaggggtgc cctgtccctc aaggggtcat ggccctocatg ttgctcgccc 120
 agcggctggc ctgcagcttc cagcacacgt accgcctgct ggtgcctgga tccagacaca 180
 ttagtcaagc tgcagccaaa gtgcagcttg aatttgatta tgatgggcct ctgatgaaga 240
 cggaagtccc agggcctaga tctcaggagt taatgaaaca gctgaatata attcagaatg 300
 cagaggctgt gcattttttc tgcaattacg aagagagccg aggcaattac ctgggtgatg 360
 tggacggcaa ccgaatgctg gatctttatt ccagatctc ctctgttccc ataggttaca 420
 ggcgcccggc cctcgtgaaa ctcatccaac agccacaaaa tgcgagcatg tttgtcaaca 480
 gaccgcacct cgaaatcctg cctccggaga actttgtgga gaagctccgg cagtcttgc 540

 tctcgggtgc tcccaaaggg atgtcccagc tcatcaccat ggccctgcggc tcctgctcca 600
 atgaaaacgc cttaaagacc atcttcatgt ggtaccggag caaggaaaga gggcagaggg 660
 gattctccaa agaggagctg gagacgtgca tgattaacca ggccccctgg tgccccgact 720
 acagcatcct ctcttcatg ggttccttcc atgggaggac catggggtgc ttagcgacca 780
 cgcactctaa agccattcac aagatcgata tcccttctt tgactggccc atcgcaccgt 840
 tcccacggct gaaataccct ctggaagagt ttgtgaaaga gaaccaacag gaagaggccg 900
 gctgtctgga agaggttgag gatctgattg tgaatatcg aaaaaagaag aagacggtgg 960
 ccgggatcat cgtggagccc atccagtccg aggggtggaga caaccatgca tccgatgact 1020
 tctttcggaa gctgagagac atcgccagga agcactgctg cgccttcttg gtggacgagg 1080
 tccagaccgg aggaggctgc acgggcaagt tctgggcca tgagcactgg ggccctggatg 1140
 acccagcaga cgtgatgacc ttcagcaaga agatgatgac tgggggcttc ttctcaagg 1200
 aggagttcag gcctaatgct ccctaccgga tcttcaacac gtggctgggg gaccctcca 1260
 agaacctgtt gctggctgag gtcacaaaca tcatcaagcg ggaggacctg ctaaataatg 1320
 cagcccatgc cgggaaggcc ctgctcacag gactgctgga cctccaggcc cggtacccc 1380
 agttcatcag cagggtgaga ggacgaggca ccttttgctc cttagatact cccgatgatt 1440
 ccatacggaa taagctcatt ttaattgcca gaaacaaagg tgtggtggtg ggtggctgtg 1500
 gtgacaaatc cattcgtttc cgtcccacgc tgggtgttcag ggatcaccac gctcacctgt 1560
 tcctcaatat tttcagtgac atcttagcag acttcaagta aagaagccat ttccactaca 1620
 gtgagaaagc ccggatccca acagttgtca aattgattag tttgcctaata tcatgttttc 1680
 acttaaaagt atcagaggtg gaatt 1705

<210> 136
 <211> 2808
 <212> ADN
 <213> *Homo sapiens*

5

<400> 136

ES 2 714 582 T3

ggaaagcacc tgtgagottg gcaagtcagt tcagagctcc agcccgctcc agcccggccc 60
gacccgaccg caccceggcgc ctgcctcget cgggctcccc ggccagccat gggcccttgg 120
agccgcagcc tctogggcct gctgctgctg ctgaggtctc ctcttggctc tcaggagcgg 180
agccctcctc cctgtttgac gcgagagcta cacgttcacg gtgccccggc gccacctgag 240
aagaggccgc gtctgggcag agtgaatfff gaagattgca ccggtcgaca aaggacagct 300
atfffctga caccgattcc gaaagtgggc acagatgggtg tgattacagt caaaaggcct 360
ctacggtttc ataaccaac agatccattt cttggtctac gctgggactc cacctacaga 420
aagttttcca ccaaagtcac gctgaatata gtggggcacc accaccgccc cccgccccat 480

ES 2 714 582 T3

caggcctcog tttctggaat ccaagcagaa ttgctcacat ttcccaactc ctctcctggc 540
 ctcagaagac agaagagaga ctgggttatt cctcccatca gctgcccaga aaatgaaaaa 600
 ggcccatttc ctaaaaacct ggttcagatc aaatccaaca aagacaaaaga aggcaaggtt 660
 ttctacagca tcaactggcca aggagctgac acaccccctg ttgggtgtctt tattattgaa 720
 agagaaacag gatggctgaa ggtgacagag cctctggata gagaacgcat tgccacatac 780
 actctcttct ctcacgctgt gtcacccaac gggaatgcag ttgaggatcc aatggagatt 840
 ttgatcacgg taaccgatca gaatgacaac aagcccgaat tcaccagga ggtctttaag 900
 gggctctgtca tggaagggtc tcttccagga acctctgtga tggaggtcac agccacagac 960
 gcggacgatg atgtgaacac ctacaatgcc gccatcgctt acaccatcct cagccaagat 1020
 cctgagctcc ctgacaaaaa tatgttcacc attaacagga acacaggagt catcagtggt 1080
 gtcaccactg ggctggaccg agagagtttc cctacgtata ccctgggtgt tcaagetgct 1140
 gaccttcaag gtgaggggtt aagcacaaca gcaacagctg tgatcacagt cactgacacc 1200
 aacgataatc ctccgatctt caatcccacc acgtacaagg gtcagggtcc tgagaacgag 1260
 gctaacgtcg taatcaccac actgaaagtg actgatgctg atgccccaa taccacagcg 1320
 tgggaggtcg tatacaccat attgaatgat gatggtggac aatttgtcgt caccacaaat 1380
 ccagtgaaca acgatggcat tttgaaaaca gcaaagggtt tggattttga ggccaagcag 1440
 cagtacattc tacacgtagc agtgacgaat gtggtacctt ttgaggtctc totcaccacc 1500
 tccacagcca ccgtcaccgt ggatgtgctg gatgtgaatg aaggccccat ctttgtgcct 1560
 cctgaaaaga gagtggaagt gtccgaggac tttggcgtgg gccaggaaat cacatcctac 1620
 actgcccagg agccagacac atttatggaa cagaaaataa catatcggat ttggagagac 1680
 actcgcaact ggctggagat taatccggac actggtgcc a tttccactcg ggctgagctg 1740
 gacagggagg attttgagca cgtgaagaac agcacgtaca cagcccta at catagctaca 1800
 gacaatggtt ctccagttgc tactggaaca gggacaactc tgctgatcct gtctgatgtg 1860
 aatgacaacg ccccatacc agaacctcga actatattct tctgtgagag gaatccaag 1920
 cctcaggtca taaacattca tgatgcagac cttoctcca atacatctcc cttcacagca 1980
 gaactaacac acggggcaggt gcccaactgg accattcagt acaacgaccc aacccaagaa 2040
 tctatcattt tgaagccaaa gatggcctta gaggtgggtg actacaaaat caatctcaag 2100
 ctcatggata accagaataa agaccaagtg accaccttag aggtcagcgt gtgtgactgt 2160
 gaaggggccg ccggcgtctg taggaaggca cagcctgtcg aagcaggatt gcaaattcct 2220
 gccattctgg ggattcttgg aggaattctt gctttgctaa ttctgattct gctgctcttg 2280
 ctgtttcttc ggaggagagc ggtggtcaaa gagcccttac tgccccaga ggatgacacc 2340
 cgggacaacg tttattacta tgatgaagaa ggagcggag aagaggacca ggactttgac 2400

ES 2 714 582 T3

ttgagccagc tgcacagggg octggagcgt cggcctgaag tgactcgtaa cgacgttgca 2460
 ccaaccctca tgagtgtccc ccggtatcct ccccgccctg ccaatccoga tgaaattgga 2520
 aattttattg atgaaaatct gaaagcggct gatactgacc ccacagcccc gccttatgat 2580
 tctctgctcg tgtttgacta tgaaggaagc ggttccgaag ctgctagtct gagctccctg 2640
 aactcctcag agtcagacaa agaccaggac tatgactact tgaacgaatg gggcaatccg 2700
 ttcaagaagc tggctgacat gtacggaggc ggcgaggacc actaggggac tcgagagagg 2760
 cggcccagac catgtgcaga aatgcagaaa tcagcgttct ggtgtttt 2808

<210> 137
 <211> 591
 <212> ADN
 <213> *Homo sapiens*
 <400> 137

5

cttctctggg acacattgcc ttctgttttc tccagcatgc gcttgctoca gtcctgttc 60
 agggccagcc ctgccaccct gtcctgggtt ctctgcctgc agttgggggc caacaaagct 120
 caggacaaca ctcggaagat cataataaag aattttgaca ttccaagtc agtacgtcca 180
 aatgacgaag tcaactgcagt gcttgcagtt caaacagaat tgaaagaatg catggtgggt 240
 aaaacttacc tcattagcag catccctcta caaggatgat ttaactataa gtatactgcc 300
 tgcctatgtg acgacaatcc aaaaaccttc tactgggact ttacaccaa cagaactgtg 360
 caaattgcag ccgtcgttga tgttattcgg gaattaggca tctgcctga tgatgctgct 420
 gtaatcccca tcaaaaacaa ccggttttat actattgaaa tcctaaaggt agaataatgg 480
 aagccctgtc tgtttgccac acccaggtga tttcctctaa agaaacttgg ctggaatttc 540
 tgctgtggtc tataaaataa acttcttaac atgcttaaaa aaaaaaaaaa a 591

10

REIVINDICACIONES

- 5 1. Un procedimiento de predicción de una respuesta y/o el beneficio de la quimioterapia que contiene taxano/antraciclina, incluyendo la quimioterapia neoadyuvante, en un paciente que padece o está en riesgo de desarrollar un cáncer de mama recurrente, comprendiendo dicho procedimiento las etapas de:
- (a) la determinación en una muestra de tumor de mama de dicho paciente de los niveles de expresión genética de al menos 3 de los siguientes genes: BIRC5, DHCR7, AZGP1, RBBP8, IL6ST, y MGP, indicativos de una respuesta a la quimioterapia de un tumor, y
- 10 (b) la combinación matemática de los valores del nivel de expresión para los genes de dicho conjunto cuyos valores se determinaron en la muestra tumoral para generar una puntuación combinada, en la que dicha puntuación combinada es predictiva de dicha respuesta y/o beneficio de la quimioterapia.
2. El procedimiento de la reivindicación 1 para la predicción de una respuesta a la quimioterapia citotóxica, preferentemente en tumores negativos a Her2/neu, positivos al receptor estrogénico (luminal), preferentemente en el modo neoadyuvante.
- 15 3. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicho nivel de expresión se determina como no proteica, tal como un nivel de expresión genética.
4. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicho nivel de expresión se determina mediante al menos uno de
- 20 un procedimiento basado en PCR,
un procedimiento basado en micromatrices, o
un procedimiento basado en hibridación, una estrategia de secuenciación y/o secuenciación de siguiente generación.
5. Un procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha determinación de los niveles de expresión es en una muestra tumoral fijada con formalina embebida en parafina o una muestra tumoral fresca congelada.
- 25 6. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que el nivel de expresión de dicho al menos un marcador genético se determina como un patrón de expresión con respecto a al menos un gen de referencia o a un valor de expresión medio computado.
7. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que dicha etapa de combinación matemática comprende una etapa de aplicación de un algoritmo a los valores representativos de un nivel de expresión de un gen determinado, en particular en el que dicho algoritmo es una combinación lineal de dichos valores representativos de un nivel de expresión de un gen determinado, o en el que un valor representativo del nivel de expresión de un gen determinado se multiplica por un coeficiente.
- 30 8. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que se determinan uno o más umbrales para dicha puntuación combinada y se discrimina entre alto y bajo riesgo, medio y bajo riesgo, o más grupos de riesgo aplicando el umbral en la puntuación combinada.
- 35 9. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que una puntuación combinada alta es indicativa del beneficio de una terapia más agresiva, por ejemplo, una quimioterapia citotóxica.
10. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que se procesa la información con respecto al estado ganglionar del paciente en la etapa de la combinación matemática de los valores del nivel de expresión de los genes para generar una puntuación combinada.
- 40 11. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que se procesa dicha información con respecto al tamaño tumoral del paciente en la etapa de combinación matemática de los valores del nivel de expresión de los genes para generar una puntuación combinada.
- 45 12. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que se procesa dicha información con respecto al estado ganglionar y el tamaño tumoral del paciente en la etapa de combinación matemática de los valores del nivel de expresión de los genes para generar una puntuación combinada.

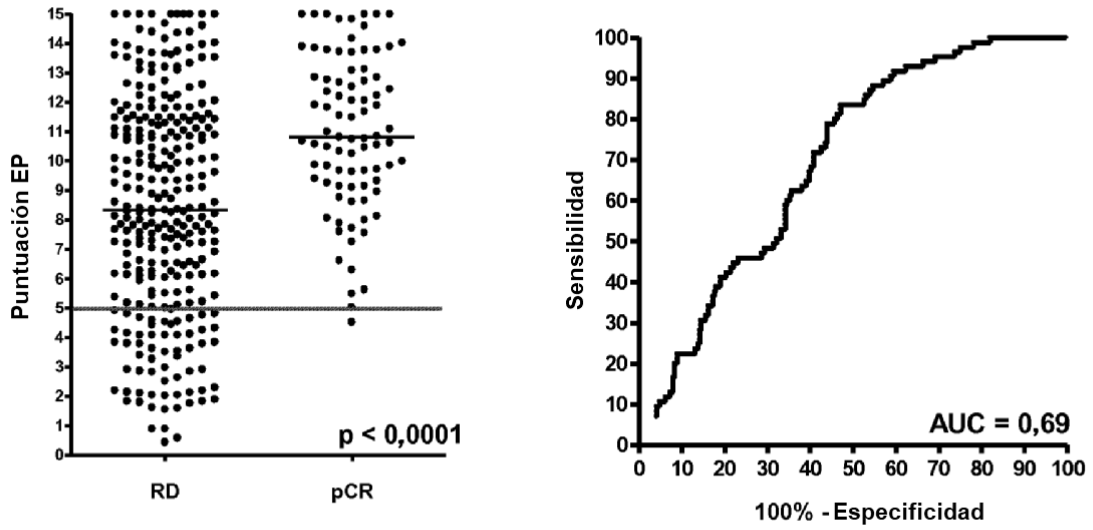


Figura 1

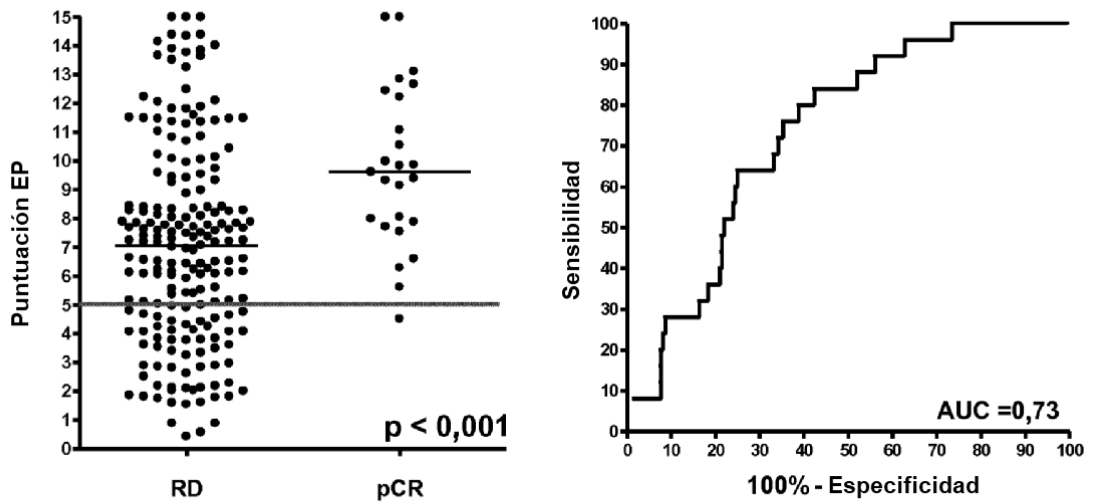


Figura 2

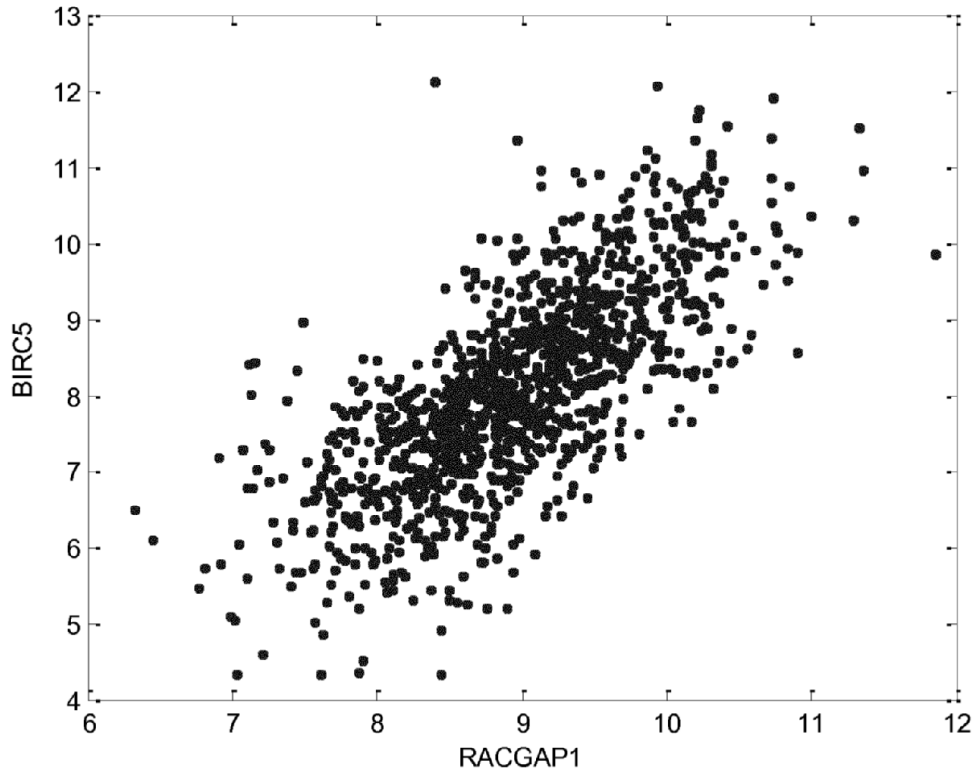


Figura 3