

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 863 531**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.11.2017** **PCT/IB2017/056875**

87 Fecha y número de publicación internacional: **11.05.2018** **WO18083646**

96 Fecha de presentación y número de la solicitud europea: **03.11.2017** **E 17804665 (2)**

97 Fecha y número de publicación de la concesión europea: **23.12.2020** **EP 3374527**

54 Título: **Método para determinar un carcinoma de células escamosas de cabeza y cuello**

30 Prioridad:

04.11.2016 IT 201600111174

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

11.10.2021

73 Titular/es:

**ALMA MATER STUDIORUM - UNIVERSITÀ DI
BOLOGNA (100.0%)**

**Via Zamboni, 33
40126 Bologn, IT**

72 Inventor/es:

**MORANDI, LUCA;
GISSI, DAVIDE BARTOLOMEO y
TARSITANO, ACHILLE**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 863 531 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

Método para determinar un carcinoma de células escamosas de cabeza y cuello

- 5 La presente invención se refiere a un método realizado *in vitro* para determinar un carcinoma de células escamosas de cabeza y cuello, preferiblemente en una etapa temprana, con base en la medición del nivel o niveles de hipermetilación y/o hipometilación de islas de CpG del conjunto de genes.

Antecedentes

- 10 El cáncer oral y faríngeo, agrupados juntos, es el sexto cáncer más común en el mundo. La incidencia anual estimada es de aproximadamente 600.000 por año, dos tercios de estos casos ocurren en países en desarrollo. Las tasas de mortalidad de estos tumores se han mantenido sin cambios (50% dentro de los 5 años posteriores al diagnóstico) y se relacionan principalmente con el tabaquismo y la ingesta de alcohol. Los pacientes con cáncer oral suelen ser diagnosticados en un estadio avanzado, lo que se asocia a peor pronóstico y mayor morbilidad por radio y quimioterapia. Además, la calidad de vida del paciente está extremadamente comprometida, ya que la terapia quirúrgica generalmente implica mutilaciones, lo que a menudo tiene efectos significativos sobre la deglución, el habla y la apariencia física.

- 20 En particular, el carcinoma oral de células escamosas (OSCC) es el carcinoma de células escamosas de cabeza y cuello más frecuente (HNSCC, las enfermedades neoplásicas que afectan a las regiones de cabeza y cuello) y suele estar precedido por lesiones premalignas orales (OPML). Las características clínicas e histológicas del OPML no pueden proporcionar suficiente información para identificar las lesiones, lo que conduce a un alto riesgo de sufrir una transformación maligna y desarrollar un OSCC durante el seguimiento. Además, los pacientes afectados por OSCC pueden desarrollar un segundo OSCC primario, con una frecuencia que oscila entre el 17% y el 30%.

- 30 Aunque la cavidad bucal es fácilmente accesible para su examen, varios factores limitan la identificación y el tratamiento temprano de las OPML. Para este propósito, el estándar de oro actual para la detección y el cribado es la palpación visual y táctil durante un examen extraoral e intraoral por parte del profesional de la salud en un examen dental o físico de rutina.

- 35 Sin embargo, esta enfermedad no es fácil de identificar en sus primeras etapas y con frecuencia ha eludido a los profesionales médicos y dentales porque puede estar "oculta" o escondida de la observación a simple vista. De hecho, el tejido de apariencia normal a menudo puede ocultar la verdad dentro de las células debajo de la superficie de la mucosa.

- 40 Si la enfermedad se identifica en etapas tempranas (etapa I o etapa II), el momento ideal para la identificación es antes de que las células displásicas hayan podido atravesar la membrana basal, la tasa de supervivencia general a cinco años es superior al 80 por ciento.

- 45 Desafortunadamente, con demasiada frecuencia las manifestaciones de esta enfermedad invasiva y devastadora se detectan en las etapas tardías (etapas III-IV), cuando las lesiones han avanzado típicamente tan profundamente que es imposible de tratar sin una intervención quirúrgica radical y una pérdida significativa de la calidad de vida del paciente.

- En particular, el OSCC o lesión intraepitelial escamosa de alto grado (HG-SIL) se diagnostica habitualmente basándose en una biopsia incisional. Sin embargo, la biopsia incisional requiere un abordaje quirúrgico mínimamente invasivo que puede generar molestias y ser rechazado por el paciente.

- 50 Parece evidente a partir de los comentarios informados anteriormente que se necesitan herramientas mejoradas de prevención del cáncer oral, detección temprana, diagnóstico y manejo clínico para identificar a los pacientes de alto riesgo, tales como aquellos expuestos al consumo de tabaco y alcohol, pacientes sin acceso adecuado a cuidado de la salud, y con lesiones de alto riesgo tales como leucoplasia, que pueden progresar hasta lesiones de carcinoma. En particular, teniendo en cuenta la extensión del HNPCC, existe una gran necesidad de desarrollar un método no invasivo para detectar precozmente este tipo de tumores, y en particular las lesiones premalignas orales (OPML) que pueden desarrollar lesiones neoplásicas, con el fin de reducir la carga de OSCC.

- 60 Durante la última década, se ha propuesto la PCR cuantitativa específica de metilación (qMSP) precedida por el tratamiento con bisulfito como un método para evaluar biomarcadores útiles en la detección de OSCC. En este sentido, se han estudiado previamente varios genes para el estado de metilación del promotor en tejidos de OSCC. Guerrero - Preston et al., 2014 y la solicitud de patente relacionada WO2015066170A1 se refieren a un análisis genómico y epigenómico integrado completo en HNSCC, que se centra en la identificación de genes que tienen metilación del promotor concurrente, mutación y subregulación negativa de la expresión. En particular, estos documentos evaluaron el patrón de metilación de *PAX1*, *PAX5*, *ZIC4* y *PLCB1*. Además, Morandi et al., desarrollaron un método no invasivo para la detección temprana de OSCC y HG-SIL mediante el análisis de modificaciones epigenéticas en la mucosa oral evaluando el estado de metilación de *GP1BB* y *ZAP70* (Morandi, L. 2015).

Sin embargo, el estudio mencionado anteriormente está limitado por el número de casos y por la baja sensibilidad y especificidad del método. Por lo tanto, todavía se siente la necesidad de identificar métodos que permitan la detección temprana y precisa de HNSCC, en particular HG-SIL/OSCC. Esto no es fácil de obtener también porque identificar los genes clave y localizar los sitios de metilación informativos y eficientes en sus secuencias no son actividades triviales y obvias. De hecho, es bien sabido, por ejemplo, que la red de genes implicada en la carcinogénesis es realmente compleja y amplia y que la región promotora de los genes abarca más de mil pares de bases y contiene aproximadamente cien sitios potenciales de metilación. El documento WO201112880 se refiere a biomarcadores de hipermetilación para la detección de cáncer de células escamosas de cabeza y cuello. De acuerdo con Demokan et al., Int. J. Cancer: 127, 2351-2359 2010, KIFA y EDNRB están metilados diferencialmente en HNSCC primario y enjuagues salivales. La presente invención resuelve las necesidades descritas anteriormente con un método que implica la medición del nivel o niveles de metilación de CpG/ islas de CpG informativas en regiones genómicas específicas de un número seleccionado de genes.

La capacidad de correlacionar los niveles de metilación de varias CpG/ islas de CpG de un número seleccionado de genes otorga al método una precisión superior a cualquier otro método actualmente disponible. Además, el método descrito permite la detección temprana de HNSCC con alta especificidad y sensibilidad a partir de una muestra biológica, tal como un cepillado oral.

Breve divulgación de los dibujos

Las Figuras 1A-B muestran el diagrama de flujo de un ejemplo de aplicación del método de la invención a un conjunto específico de genes.

Las Figuras 2A-D muestran el análisis ROC que discrimina a los OSCC frente a los donantes sanos normales que muestran las tres CpG de mejor rendimiento de cada gen.

Las Figuras 3A-D muestran el gráfico del perfil de metilación de los genes evaluados. Para cada grupo de muestras, las líneas representan la media de metilación para cada posición. Los asteriscos indican una significancia estadística calculada por la prueba de Kruskal-Wallis.

Las Figuras 4 y 5 muestran el análisis ROC de las combinaciones específicas de genes ejemplificados.

La Figura 6 muestra los diagramas de caja obtenidos utilizando las puntuaciones calculadas a partir del algoritmo que mostraron una diferencia significativa entre grupos (prueba de Kruskal-Wallis $T = 78,8587$, $p < 0,05$).

Definiciones

En el contexto de la presente invención, "carcinoma de células escamosas de cabeza y cuello (HNSCC)" son tumores que se desarrollan en las membranas mucosas de la boca, nariz o garganta y, en general, se clasifican por su ubicación. De hecho, este tipo de carcinoma puede ocurrir en la boca (cavidad oral), en la parte media de la garganta cerca de la boca (orofaringe), en el espacio detrás de la nariz (cavidad nasal y senos paranasales), en la parte superior de la garganta cerca de la cavidad nasal (nasofaringe), en la laringe o en la parte inferior de la garganta cerca de la laringe (hipofaringe).

En el contexto de la presente invención, el "carcinoma oral de células escamosas (OSCC)" representa el 95% de todas las formas de HNSCC y durante la última década, su incidencia ha aumentado en un 50%. OSCC, surge de la acumulación de cambios genéticos y anomalías epigenéticas en las vías de señalización que están asociadas con el cáncer, lo que resulta en fenotipos que facilitan el desarrollo de OSCC. Se deriva del epitelio escamoso estratificado de la mucosa oral. La mayoría de los OSCC se diagnostican en una fase tardía, en las etapas III o IV, lo que disminuye notablemente las posibilidades de supervivencia y conduce a un deterioro significativo de la calidad de vida del paciente.

La "Lesión intraepitelial escamosa de alto grado (HG-SIL)" ha sido introducida por Gale et al., 2014. La clasificación proporciona dos grados: 1) lesiones de grado bajo y 2) de grado alto y, específicamente para la laringe, un carcinoma *in situ* de grado adicional (CIS) que debe separarse de los SIL laríngeos de alto grado. Este estudio proporcionó criterios morfológicos claros con los que definir los grupos de pronóstico. El estudio de seguimiento retrospectivo demostró una diferencia muy significativa en el riesgo de progresión maligna entre las lesiones de bajo y alto grado, lo que mejora el manejo del paciente y la toma de decisiones clínicas.

En el contexto de la presente invención, la "Lesión Premaligna Oral (OPML)" es un tejido alterado morfológicamente en el que es más probable que ocurra cáncer oral que en su contraparte aparentemente normal. Recientemente, la Organización Mundial de la Salud (OMS) identificó las siguientes OPML: leucoplasia, eritroplasia, lesión palatina por tabaquismo inverso, liquen plano oral, fibrosis submucosa oral (SMF) y lupus eritematoso discoide.

En el contexto de la presente invención, "metilación del ADN" significa modificaciones epigenéticas de ácidos nucleicos que alteran su accesibilidad y estructura, preferiblemente la estructura de la cromatina, regulando así los patrones de expresión génica. Pueden ser modificados por influencias exógenas y, como tales, pueden contribuir o ser el resultado de alteraciones ambientales del fenotipo o patofenotipo.

En el contexto de la presente invención, las "islas de CpG" son regiones del genoma humano con un contenido elevado de G + C. Estas regiones tienen generalmente una longitud de aproximadamente 1 kb y, por lo general, se solapan con la región promotora del 60-70% de todos los genes humanos. También están presentes en repeticiones como elementos transponibles y funcionan para reprimir la transcripción. Dentro de estas regiones, la mayoría de los pares de CpG se modifican químicamente por la unión covalente de un grupo metilo a la posición C5 del anillo de citosina. La metilación aberrante del ADN de las CpG en la proximidad de los sitios de inicio de la transcripción a menudo conduce a alteraciones en la función genética y la desregulación de la vía en el cáncer humano.

En el contexto de la presente invención, "región de ADN genómico" significa regiones de ADN genómico tales como el promotor del gen, el 5'UTR, el cuerpo del gen, el 3'UTR y las orillas (se refiere a la secuencia de ADN que se produce hasta a aproximadamente 2 kb de distancia de una isla de CpG de densidad de CpG comparativamente baja que se encuentra cerca de islas de CpG tradicionales).

Descripción detallada

Un primer aspecto de la presente invención se refiere a un método para detectar (para determinar la presencia de) un carcinoma de células escamosas de cabeza y cuello (HNSCC) que comprende las etapas de:

(i) Purificar un ácido nucleico, preferiblemente el ADN, más preferiblemente el ADN genómico, de una muestra aislada obtenida de un individuo;

(ii) Tratar el ácido nucleico purificado, preferiblemente el ADN, con bisulfito, preferiblemente bisulfito de sodio;

(iii) Amplificar el ácido nucleico después de la etapa (ii) utilizando al menos un par de cebadores, permitiendo dicho par de cebadores la amplificación de al menos una secuencia de la región genómica de al menos un gen de cada uno de los genes comprendidos en el siguiente conjunto de genes (en otras palabras, al menos una parte de la región genómica/cromosómica de un gen, que es una parte de su sitio de mapeo genómico que comprende las regiones codificadoras y reguladoras del gen):

LRRTM1, PARP15, ZAP70, miR193, miR296, hTERT y, opcionalmente, de al menos uno de los siguientes genes: FLI1, LINC00599, GP1BB, EPHX3, ITGA4, NTM, KIF1A; dicha secuencia comprende al menos un CpG, preferiblemente al menos una isla de CpG;

(iv) Medir el nivel de metilación de al menos una CpG/isla de CpG de la secuencia amplificada de la región genómica y perteneciente al siguiente conjunto de genes:

LRRTM1, PARP15, ZAP70, miR193, miR296, hTERT y, opcionalmente, de al menos uno de los siguientes genes: FLI1, LINC00599, GP1BB, EPHX3, ITGA4, NTM, KIF1A;

En los que la hipermetilación de al menos una CpG/isla de CpG que cae en la secuencia de al menos uno de los siguientes genes de los conjuntos: *EPHX3*, *KIF1A*, *LRRTM1*, *FLI1*, *ITGA4*, *LINC00599*, *NTM*, *PARP15*, *ZAP70*, *miR193* y/o la hipometilación de al menos una CpG/isla de CpG que cae en la secuencia de al menos uno de los siguientes genes de los conjuntos: *miR296*, *GP1BB*, *hTERT* son indicativos de una muestra positiva para HNSCC en la que la hipermetilación y/o hipometilación es/se evalúan con respecto al nivel de metilación medido para las mismas CpG/islas de CpG de las mismas regiones de los mismos genes en una muestra aislada de un individuo sano (control negativo), preferiblemente en una población de individuos sanos.

La Figura 1 - A y B resumen en un diagrama de flujo las etapas clave del método de la invención. En particular, el diagrama de flujo se refiere a un ejemplo no limitativo de una aplicación del método de la invención a un conjunto específico de genes.

La Figura 1-A resume una realización específica de la invención que implica 1) recolectar la muestra biológica y purificar el ADN; 2) tratar el ADN con bisulfito; y 3) amplificar las regiones genómicas de interés.

La Figura 2-B resume una realización específica que ejemplifica la etapa de medición del nivel de metilación, en otras palabras, la última fase del método que permite la determinación de la eventual positividad de la muestra para un HNSCC.

El método de la invención permite determinar si la muestra biológica pertenece a un individuo afectado por un HNSCC (positivo para HNSCC). Preferiblemente, el HNSCC se detecta (determina) en una etapa temprana. Por lo tanto, la detección es una detección en etapa temprana. Por lo tanto, el método también se puede utilizar como herramienta de pronóstico y/o para el seguimiento de un tratamiento terapéutico de un HNSCC. Finalmente, el método de la invención también es útil en la prevención de este tipo de carcinomas.

Por lo tanto, el método de la invención también es para pronosticar un carcinoma de células escamosas de cabeza y cuello y/o para el seguimiento de un tratamiento terapéutico de un HNSCC y/o para la prevención de un HNSCC.

En este contexto, *EPHX3* es el acrónimo de epóxido hidrolasa o los siguientes sinónimos: *FLJ22408*, *ABHD9*, *EH3*; *KIF1A* es el acrónimo del miembro 1 de la familia de las quinesinas o los siguientes sinónimos: *UNC104*, *C2orf20*, *ATSV*, *HSN2C*, *SPG30*, *MRD9*; *LRRTM1* es el acrónimo de la repetición neuronal 1 transmembrana rica en leucina o los siguientes sinónimos: *FLJ32082*; *FLI1* es el acrónimo del protooncogén Fli-1, factor de transcripción de *ETS* o los siguientes sinónimos: *EWSR2*, *SIC-1*; *ITGA4* es el acrónimo de la subunidad alfa 4 de la integrina o los siguientes

sinónimos: *IA4*, *CD49D*; *LINC00599* es el acrónimo del ARN 599 que codifica proteína no intergénica larga o los siguientes sinónimos: *MIRN124A1*, *MIRN124-1*, *MIR124A1*, *MIR124A*, *Rncr3*, *mir-124-1*; *NTM* es el acrónimo de neurotrimina o los siguientes sinónimos: *NT*, *NTRI*, *IGLON2*, *HNT*; *PARP15* es el acrónimo del miembro 15 de la familia de la poli(ADP-ribosa) polimerasa o los siguientes sinónimos: *FLJ40597*, *pART7*, *ARTD7*, *BAL3*; *ZAP70* es el acrónimo de la hebra zeta de la proteína quinasa 70 asociada al receptor de células T o los siguientes sinónimos: *IMD48*, *TZK*, *STD*, *ZAP-70*, *SRK*, *STCD*, *ADMIO2*; *miR193* es el acrónimo de microARN 193a o los siguientes sinónimos: *MIRN193A*, *mir-193a*, *hsa-mir-193a*, *hsa-mir-193*, *MIRN193*; *miR296* es el acrónimo de microARN 296 o los siguientes sinónimos: *hsa-mir-296*, *mir-296*, *MIRN296*, *miRNA296*; *GP1BB* es el acrónimo de la subunidad beta plaquetaria de la glicoproteína Ib o los siguientes sinónimos: *GPIBB*, *GPIbbeta*, *CD42C*, *BS*, *BDPLT1*, *CD42c*; *hTERT* es el acrónimo de transcriptasa inversa de telomerasa o los siguientes sinónimos: *TRT*, *TCS1*, *DKCA2*, *EST2*, *PFBMFT1*, *CMM9*, *hEST2*, *DKCB4*, *hTRT*, *TP2*.

En el contexto de la presente invención, hipermetilación significa que el nivel de metilación medido de al menos una CpG/isla de CpG aumenta en una muestra analizada (aislada de un individuo sospechoso de estar afectado por un HNSCC) en comparación con una muestra negativa (una muestra biológica aislada de un individuo sano no afectado por un HNSCC), en la que la muestra negativa es preferiblemente una muestra de ADN aislada de la mucosa normal de individuos sanos no afectados por un HNSCC.

En el contexto de la presente invención, hipometilación significa que el nivel de metilación medido de al menos una CpG/ isla de CpG se reduce en una muestra analizada (aislada de un individuo sospechoso de estar afectado por un HNSCC) en comparación con una muestra negativa (una muestra biológica aislada de un individuo sano no afectado por un HNSCC), en la que la muestra negativa es preferiblemente una muestra de ADN aislada de la mucosa normal de individuos sanos no afectados por un HNSCC.

La muestra biológica de la etapa (i) se aísla de un individuo que puede estar afectado o se sospecha que está afectado por un HNSCC, preferiblemente ha sido tratado quirúrgicamente para un HNSCC, preferiblemente un OSCC. De acuerdo con una realización preferida, dicho HNSCC es cualquier tumor que se desarrolle en la cavidad oral, nariz y garganta. Preferiblemente, dicho HNSCC es cualquier carcinoma oral de células escamosas o su precursor, preferiblemente una lesión intraepitelial escamosa de alto grado (HG-SIL) o una lesión oral premaligna o un campo de cancerización. El campo de cancerización significa un área del epitelio de la mucosa formada por células genéticamente alteradas que eventualmente se relacionan clonalmente con el carcinoma y, por lo tanto, la posible causa de recurrencia local o tumores secundarios.

De acuerdo con una realización preferida, dicha lesión premaligna oral (OPML) se selecciona preferiblemente de: leucoplasia, eritroplasia, lesión palatina por fumar cigarrillos al revés, liquen plano oral, fibrosis submucosa oral (SMF) y lupus eritematoso discóide en el que es más probable que ocurra un cáncer oral que en su contraparte aparentemente normal y puede detectarse mediante la presente invención si la lesión es un precursor real de HNSCC.

Por lo tanto, el método es *ex vivo*. La muestra biológica es cualquier muestra biológica, preferiblemente dicha muestra se selecciona entre: cepillado de mucosa oral, plasma, saliva y cualquier muestra que contenga células epiteliales, siendo preferiblemente dichas células de la cavidad oral. De acuerdo con una realización preferida, la muestra se aísla de la cavidad bucal, preferiblemente de la mucosa oral, más preferiblemente mediante cepillado, preferiblemente utilizando un citocepillo o un hisopo bucal de ADN. Este procedimiento permite obtener una muestra que comprende células exfoliadas de la cavidad bucal, preferiblemente de la mucosa oral.

De acuerdo con una realización preferida, la recogida de la muestra está precedida por enjuagues bucales, al menos 1-3 veces, preferiblemente con clorhexidina al 0,12% o cualquier enjuague bucal germicida.

De acuerdo con otra realización preferida, la recogida de muestra se realiza cepillando suavemente una superficie de la región corporal de interés, preferiblemente dicha región presenta una lesión clínica, preferiblemente 1-10 veces, más preferiblemente aproximadamente cinco veces. Es aconsejable en el procedimiento de recolección evitar 1) recolectar sangre, y 2) no utilizar ningún anestésico local.

La muestra recogida se coloca preferiblemente en un recipiente que contiene una solución conservante (es decir, ADN /ARN Shield^{MR}), para permitir la conservación del ácido nucleico. Más preferiblemente, la muestra se almacena a baja temperatura, preferiblemente alrededor de + 4 °C.

De acuerdo con otra realización preferida, la muestra se somete a la etapa (i) que implica aislar (extraer-purificar) un ácido nucleico de la muestra biológica. Preferiblemente, el ácido nucleico aislado es ADN, preferiblemente el ADN genómico. Se puede utilizar cualquier técnica usada en un laboratorio para extraer ácidos nucleicos, en particular para extraer ADN de alta pureza, por ejemplo, el kit de purificación de ADN MasterPure^{MR}.

Después de la purificación, el ácido nucleico, preferiblemente el ADN genómico purificado, se trata con bisulfito, preferiblemente bisulfito de sodio (etapa ii).

El bisulfito convierte una citosina (C) no metilada en uracilo (U), de modo que el uracilo (U) se lee timina (T) cuando se secuenciar el ácido nucleico tratado. Esta conversión no afecta a una C metilada, que permanece C en la secuencia. Cualquier método que permita abordar este propósito debe considerarse una alternativa al tratamiento con bisulfito y, por lo tanto, parte de la divulgación.

El tratamiento con bisulfito del ácido nucleico se realiza de acuerdo con cualquier protocolo de laboratorio común o usando kits disponibles comercialmente, tales como el kit EZ Methylation-Lightning^{MR}. Preferiblemente, el tratamiento con bisulfito se realiza usando una cantidad de ácido nucleico, preferiblemente de ADN, que varía de 50 ng a 10 µg, preferiblemente de 250 ng a 1 µg.

Preferiblemente, el tratamiento con bisulfito implica una primera etapa de desnaturalización del ácido nucleico, preferiblemente a 98 °C durante 10-20 minutos, más preferiblemente aproximadamente 8 minutos y/o la incubación del ácido nucleico desnaturalizado con bisulfito de sodio. Preferiblemente, el bisulfito se usa en una concentración de aproximadamente 2,3-3,6 M, más preferiblemente en presencia de hidroquinona, preferiblemente en una concentración de aproximadamente 10 mM. La reacción se realiza preferiblemente a una temperatura que varía preferiblemente de 45 a 60 °C, más preferiblemente de 50 a 55 °C, aún más preferiblemente de aproximadamente 54 °C durante al menos una hora.

Al final del tratamiento, el ácido nucleico se purifica preferiblemente usando una tecnología de columna y más preferiblemente con un primer lavado a base de etanol, una etapa de desulfonación de 20 minutos a temperatura ambiente, dos etapas de lavado con etanol tamponado y la etapa de elución siguiendo las instrucciones del proveedor.

Después del tratamiento con bisulfito, el ácido nucleico, preferiblemente el ADN, se amplifica, preferiblemente usando un protocolo de PCR (reacción en cadena de la polimerasa). La amplificación se realiza preferiblemente utilizando al menos un par de cebadores, dicho par de cebadores permite la amplificación de al menos una secuencia, preferiblemente una porción/secuencia de la región genómica de al menos un gen de interés, comprendiendo dicha secuencia al menos una CpG/isla de CpG. Preferiblemente, el par de cebadores permite la amplificación de una secuencia de la región de ADN genómico de al menos un gen seleccionado entre: *EPHX3*, *KIF1A*, *LRRTM1*, *FLI1*, *ITGA4*, *LINC00599*, *NTM*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *GP1BB*, *hTERT* y cualquier combinación de los mismos. Preferiblemente, se amplifica al menos una de las siguientes combinaciones de genes, lo que significa que se amplifica al menos una secuencia de la región genómica (ADN) de al menos una de las siguientes combinaciones de genes, comprendiendo dicha región al menos una CpG/isla de CpG en el método de la invención:
b) *LRRTM1*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *hTERT*; y/o opcionalmente, de al menos uno de los siguientes genes: *FLI1*, *LINC00599*, *GP1BB*, *EPHX3*, *ITGA4*, *NTM*, *KIF1A*.

Sin embargo, cualquier combinación adicional de genes puede amplificarse de acuerdo con el método de la invención.

Preferiblemente, la región genómica de un gen significa cualquier región reguladora del gen, más preferiblemente la región promotora (generalmente secuencia arriba del sitio de inicio ATG de un gen) del gen, las UTR 5' y/o 3' y/o cualquier región codificante (cuerpo del gen), preferiblemente exones y/o intrones y/u orillas.

De acuerdo con una realización preferida de la invención, dicha región se selecciona preferiblemente de las SEQ ID NO: 53-65 en las que:

- SEQ ID NO: 53 es la región del promotor *PARP15* que abarca desde 122296564 hasta 122296725 de la hebra + del cromosoma 3 con respecto a las coordenadas de hg19 que muestra las siguientes CpG/islas de CpG que tienen las coordenadas: 122296564, 122296586, 122296613, 122296617, 122296623, 122296630, 122296630, 122296645, 122296649, 122296656, 122296663, 122296671, 122296677, 122296680, 122296692, 122296708, 122296710, 122296717, 122296723;
- SEQ ID NO: 54 es la región del promotor *ITGA4* que abarca desde 182322887 hasta 182323053 de la hebra + del cromosoma 2 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas: 182322890, 182322898, 182322902, 182322908, 182322915, 182322937, 182322956, 182322962, 182322980, 182322985, 182323004, 182323016, 182323024, 182323028;
- SEQ ID NO: 55 es la región del promotor *NTM* que abarca desde 131781042 hasta 131781185 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas: 131781046, 131781058, 131781060, 131781080, 131781082, 131781098, 131781098, 131781098, 131781110, 131781113, 131781122, 131781144, 131781151, 131781158, 131781167, 131781183;
- SEQ ID NO: 56 es la región del promotor *ZAP70* que abarca desde 98340750 hasta 98340885 de la hebra + del cromosoma 12 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas: 98340751, 98340762, 98340764, 98340766, 98340772, 98340775, 98340794, 98340801, 98340809, 98340813, 98340816, 98340823, 98340828, 98340834, 98340844, 98340854, 98340859, 98340869, 98340877, 98340881;
- SEQ ID NO: 57 es la región del promotor *MIR193A* que abarca desde 29886860 hasta 29887068 de la hebra + del cromosoma 17 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas:

29886868, 29886870, 29886875, 29886881, 29886885, 29886890, 29886903, 29886910, 29886913, 29886936, 29886939, 29886944, 29886949, 29886951, 29886964, 29886967, 29886971, 29886983, 29886992, 29886995, 29887001, 29887008, 29887015, 29887045, 29887049, 29887063;

– SEQ ID NO: 58 es la región del promotor *EPHX3* que abarca desde 15342826 hasta 15343049 de la hebra - del cromosoma 19 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas: 15342895, 15342889, 15342885, 15342875, 15342869, 15342866, 15342864, 15342855, 15342851, 15342848, 15342845, 15342840, 15342838, 15342831, 15343002, 15343000, 15342991, 15342988, 15342985, 15342983, 15342976, 15342962, 15342960, 15342934, 15342928, 15342918, 15342916, 15342913, 15342906;

SEQ ID NO: 59 es la región del promotor *KIF1A* que abarca desde 241759586 hasta 241759727 de la hebra - del cromosoma 2 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen las coordenadas: 241759726, 241759722, 241759719, 241759717, 241759715, 241759702, 241759697, 241759688, 241759686, 241759681, 241759676, 241759671, 241759665, 241759663, 241759656, 241759651, 241759649, 241759646, 241759643, 241759640, 241759632, 241759621, 241759618, 241759614, 241759602, 241759596, 241759588;

– SEQ ID NO: 60 es la región del promotor *LINC00599* que abarca desde 9760739 hasta 9760890 de la hebra - del cromosoma 8, con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 9760888, 9760881, 9760878, 9760870, 9760856, 9760852, 9760848, 9760839, 9760832, 9760822, 9760820, 9760815, 9760811, 9760809, 9760805, 9760803, 9760777, 9760764, 9760759, 9760752;

– SEQ ID NO: 61 es la región del promotor *FLI1* que abarca desde 128564020 hasta 128564160 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 128564022, 128564033, 128564051, 128564063, 128564065, 128564068, 128564078, 128564089, 128564106, 128564134, 128564158;

– SEQ ID NO: 62 es la región del promotor *GP1BB* que abarca desde 19710829 hasta 19710960 de la hebra + del cromosoma 22 con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 19710839, 19710842, 19710847, 19710855, 19710864, 19710880, 19710890, 19710902, 19710909, 19710916, 19710919, 19710937, 19710944, 19710946, 19710948, 19710952, 19710954, 19710956;

– SEQ ID NO: 63 es la región del promotor *MIR296* que abarca desde 57392355 hasta 57392545 de la hebra - del cromosoma 20, con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 57392543, 57392538, 57392536, 57392517, 57392501, 57392469, 57392452, 57392440, 57392437, 57392430, 57392419, 57392394, 57392386, 57392374, 57392363;

– SEQ ID NO: 64 es la región del promotor *LRRTM1* que abarca desde 80531676 hasta 80531807 de la hebra - del cromosoma 2, con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 80531806, 80531803, 80531799, 80531796, 80531789, 80531786, 80531783, 80531780, 80531776, 80531763, 80531759, 80531756, 80531752, 80531749, 80531745, 80531742, 80531738, 80531719, 80531702, 80531697, 80531695, 80531677;

– SEQ ID NO: 65 es la región del promotor de *TERT* humano (*hTERT*) que abarca desde 1279743 hasta 1279851 de la hebra - del cromosoma 21, con respecto a las coordenadas de hg19 que muestran las siguientes CpG/islas de CpG que tienen coordenadas: 1279847, 1279838, 1279829, 1279775, 1279758, 1279746.

En el contexto de la presente invención, se ha mapeado un locus de una CpG/isla de CpG de acuerdo con el genoma humano de referencia (hg19), utilizando blat (<https://genome.ucsc.edu/cgi-bin/hgBlat?command=start>), en el que se indica el número del cromosoma, la posición, la hebra positiva o negativa y el porcentaje de identidad. De acuerdo con una realización preferida al menos una secuencia de PARP15 amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón es el nombre del producto de amplificación) es la SEQ ID NO: 53 correspondiente a la región del ADN genómico (hebra +) mapeada en el cromosoma 3 y que tiene las coordenadas desde 122296564 a 122296723 con respecto a las coordenadas de hg19 y al menos una CpG/isla de CpG evaluada en de secuencia corresponde a una coordenada seleccionada de: 122296564, 122296586, 122296613, 122296617, 122296623, 122296630, 122296637, 122296645, 122296649, 122296656, 122296663, 122296671, 122296677, 122296680, 122296692, 122296708, 122296710, 122296717, 122296723 y combinación de las mismas, preferiblemente es 122296586.

De acuerdo con una realización preferida, al menos una secuencia de *ITGA4* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 54 correspondiente a la región del ADN genómico (hebra +) mapeada en el cromosoma 2 y que tiene coordenadas desde 182322887 a 182323053 con respecto a las coordenadas de hg19 y al menos una CpG/ isla de CpG evaluada en su secuencia corresponde a una coordenada seleccionada de: 182322890, 182322898, 182322902, 182322908, 182322915, 182322937, 182322956, 182322962,

182322980, 182322985, 182323004, 182323016, 182323024, 182323028; y combinación de los mismos, preferiblemente es 182322902.

De acuerdo con una realización preferida, la al menos una secuencia de *NTM* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 55 correspondiente a la región del ADN genómico (hebra +) mapeada en el cromosoma 11 y que tiene coordenadas desde 131781042 a 131781185 con respecto a las coordenadas de hg19 y el al menos una CpG/isla de CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 131781046, 131781058, 131781060, 131781080, 131781082, 131781084, 131781098, 131781113, 131781122, 131781144, 131781151, 131781158, 131781167, 131781183; y combinación de las mismas, preferiblemente es 131781082 o 131781167.

De acuerdo con una realización preferida, la al menos una secuencia de *ZAP70* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 56 correspondiente a la región del ADN genómico (hebra +) mapeada en el cromosoma 12 y que tiene coordenadas desde 98340750 a 98340885 con respecto a las coordenadas de hg19 y el al menos una CpG/isla de CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 98340751, 98340762, 98340764, 98340766, 98340772, 98340775, 9834080794, 9834080794, 98340751 98340813, 98340816, 98340823, 98340828, 98340834, 98340844, 98340854, 98340859, 98340869, 98340877, 98340881 y combinación de los mismos, preferiblemente 98340751 o 98340854.

De acuerdo con una realización preferida, la al menos una secuencia de *MIR193* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 57 correspondiente a la región del ADN genómico (hebra +) mapeada en cromosoma 17 y que las tiene coordenadas desde 29886860 a 29887068 con respecto a las coordenadas de hg19 y la al menos una isla CpG/CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 29886868, 29886870, 29886875, 29886881, 29886885, 29886890, 29886903, 29886910, 29886913, 29886936, 29886939, 29886944, 29886949, 29886951, 29886964, 29886967, 29886971, 29886983, 29886992, 29886995, 29887001, 29887008, 29887015, 29887045, 29887049, 29887063 y una combinación de los mismos, preferiblemente 29886964 y combinación de los mismos.

De acuerdo con una realización preferida, la al menos una secuencia de *EPHX3* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 58 correspondiente a la región de ADN genómico (hebra -) mapeada en cromosoma 19 y que tiene coordenadas desde 15342826 a 15343049 con respecto a las coordenadas de hg19 y la al menos una CpG/isla CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 15342895, 15342889, 15342885, 15342875, 15342869, 15342866, 15342864, 15342855, 15342851, 15342848, 15342845, 15342840, 15342838, 15342831, 15343002, 15343000, 15342991, 15342988, 15342985, 15342983, 15342976, 15342962, 15342960, 15342934, 15342928, 15342918, 15342916, 15342928, 15342918, 15342916 1534, preferiblemente 1534293413 y combinación de los mismos.

De acuerdo con una realización preferida, la al menos una secuencia de *KIF1A* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 59 correspondiente a la región de ADN genómico (hebra -) mapeada en el las cromosoma 2 y que tiene coordenadas desde 241759586 a 241759727 con respecto a las coordenadas de hg19 y la al menos una isla CpG/CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 241759726, 241759722, 241759719, 241759717, 241759715, 241759702, 241759697, 241759688, 241759686, 241759681, 241759676, 241759671, 241759665, 241759663, 241759656, 241759651, 241759649, 241759646, 241759643, 241759640, 241759632, 241759621, 241759618, 241759602 241759651 y 241759621 y combinaciones de los mismos;

De acuerdo con una realización preferida, la al menos una secuencia de *LINC00599* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 60 correspondiente a la región de ADN genómico (hebra -) mapeada en el cromosoma 8 y que tiene coordenadas desde 9760739 a 9760890 con respecto a las coordenadas de hg19 y el al menos una isla CpG/CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 9760888, 9760881, 9760878, 9760870, 9760856, 9760852, 9760848, 9760839, 9760832, 9760822, 9760820, 9760815, 9760811, 9760809, 9760805, 9760803, 9760777, 9760764, 9760759, 9760752 y combinación de las mismas, preferiblemente es 9760888.

De acuerdo con una realización preferida, la al menos una secuencia de *FLI1* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 61 correspondiente a la región del ADN genómico (hebra +) mapeada en cromosoma 11 y que tiene coordenadas desde 128564020 a 128564160 con respecto a las coordenadas de hg19 y el al menos una CpG/isla de CpG evaluado en esta secuencia corresponde a una coordenada seleccionada de: 128564022, 128564033, 128564051, 128564063, 128564065, 128564068, 128564078, 128564089, 128564106, 128564134, 128564158 y una combinación de los mismos, preferiblemente es 128564051 o 128564158.

De acuerdo con una realización preferida, la al menos una secuencia de *GP1BB* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 62 correspondiente a la región del ADN genómico (hebra +) mapeada en el cromosoma 22 y que tiene las coordenadas de 19710829 a 19710960 con respecto a las coordenadas de hg19 y la al menos una CpG/ isla de CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 19710839, 19710842, 19710847, 19710855, 19710864, 19710880, 19710890,

19710902, 19710909, 19710916, 19710919, 19710937, 19710944, 19710946, 19710948, 19710952, 19710954, 19710956, y combinación de las mismas, preferiblemente es 19710946 o 19710956.

De acuerdo con una realización preferida, la al menos una secuencia de *MIR296* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 63 correspondiente a la región del ADN genómico (hebra -) mapeada en el cromosoma 20 y que tiene coordenadas desde 57392355 a 57392545 con respecto a las coordenadas de hg19 y la al menos una CpG/ isla de CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 57392543, 57392538, 57392536, 57392517, 57392501, 57392469, 57392452, 57392440, 57392437, 57392430, 57392419, 57392394, 57392386, 57392374, 57392363, y combinación de las mismas, preferiblemente es 57392363 o 57392374. De acuerdo con una realización preferida, al menos una secuencia de *LRRTM1* amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 64 correspondiente a la región de ADN genómico (hebra -) mapeada en el cromosoma 2 y que tiene las coordenadas de 80531676 a 80531807 con respecto a las coordenadas de hg19 y la al menos una La isla CpG/CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 80531806, 80531803, 80531799, 80531796, 80531789, 80531786, 80531783, 80531780, 80531776, 80531763, 80531759, 80531756, 80531752, 80531749, 8053421753, 801753, 80531702, 80531697, 80531695, 80531677 y una combinación de los mismos, preferiblemente es 80531697 o 80531799.

De acuerdo con una realización preferida, la al menos una secuencia de *TERT* humana (*hTERT*) amplificada para el análisis del nivel de metilación de acuerdo con la etapa (iii) (amplicón) es la SEQ ID NO: 65 correspondiente a la región del ADN genómico (hebra -) mapeado en el cromosoma 21 y que tiene coordenadas desde 1279743 a 1279851 con respecto a las coordenadas de hg19 y la al menos una CpG/ isla de CpG evaluada en esta secuencia corresponde a una coordenada seleccionada de: 1279847, 1279838, 1279829, 1279775, 1279758, 1279746 y combinación del mismo, preferiblemente es 1279775, 1279747 o 1279758. De acuerdo con una realización preferida adicional, la fase (iii) de medición de las modificaciones epigenéticas (metilación) del conjunto (1) - *KIF1A* y *TERT* humana - implica la medición de 1) la hipermetilación de al menos una de la CpG/ isla de CpG de la SEQ ID NO: 59 informada anteriormente, más preferiblemente la CpG/ isla de CpG que tiene las coordenadas 241759686, 241759681, 241759651, 241759621 de SEQ ID NO: 59 y/o 2) la hipometilación de al menos uno de la CpG/ isla de CpG de SEQ ID NO: 65 indicada anteriormente, más preferiblemente la CpG/ isla de CpG que tiene las coordenadas 1279775 y/o 1279747 y/o 1279758 de SEQ ID NO: 65.

De acuerdo con una realización preferida, la fase (iii) de la medición de las modificaciones epigenéticas del conjunto (2) - *LRRTM1*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *hTERT*- implica medir 1) la hipermetilación de al menos uno de la CpG/ isla de CpG de la SEQ ID NO: 64 para *LRRTM1*, la SEQ ID NO: 53 para *PARP15*, la SEQ ID NO: 56 para *ZAP70*, SEQ ID NO: 57 para *miR193*, más preferiblemente la hipermetilación de la CpG/ isla de CpG que tiene las siguientes coordenadas: 80531799 y/o 80531697 para *LRRTM1*, 122296586 para *PARP*, 98340854 y/o 98340751 para *ZAP70*, 29886870 y/o 29886944 para *miR193* respectivamente y una combinación de los mismos; y/o 2) la hipometilación de al menos una de las islas CpG/CpG de la SEQ ID NO: 63 para *miR296*, la SEQ ID NO: 65 para *TERT*, preferiblemente la CpG/ isla de CpG que tiene las siguientes coordenadas: 57392374 y/o 57392363 para *miR296*, 1279758 y/o 1279775 y/o 1279747 para *TERT* respectivamente.

De acuerdo con otra realización preferida, la fase (iii) de medición de las modificaciones epigenéticas del conjunto (3) - *KIF1A*, *LRRTM1*, *ZAP70*, *miR193*, *miR296*, *hTERT*- implica la medición 1) la hipermetilación de al menos una de la CpG/ isla de CpG de la SEQ ID NO: 59 para *KIF1A*, la SEQ ID NO: 64 para *LRRTM1*, la SEQ ID NO: 56 para *ZAP70*, SEQ ID NO: 57 para *miR193*, más preferiblemente la hipermetilación de la CpG/ isla de CpG que tiene las siguientes coordenadas: 241759686, 241759681, 241759651, 241759621 para *KIF1A*, 80531799 y/o 80531697 para *LRRTM1*, 98340854 y/o 98340751 para *ZAP70*, 29886944 y/o 29886870 para *miR193* respectivamente y cualquier combinación de las mismas; y/o 2) la hipometilación de al menos una de las islas CpG/CpG de SEQ ID NO: 63 para *miR296*, SEQ ID NO: 65 para *TERT*, más preferiblemente la hipometilación de la CpG/ isla de CpG que tiene las siguientes coordenadas: 57392374 y/o 57392363 para *miR296*, 1279758 y/o 1279775 y/o 1279747 para *TERT* respectivamente.

De acuerdo con otra realización preferida, la fase (iii) de medición de las modificaciones epigenéticas del conjunto (4) - *KIF1A*, *LRRTM1*, *FLI1*, *LINC00599*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *GP1BB*, *hTERT*- implica la medición 1) la hipermetilación de al menos una de las islas CpG/CpG de SEQ ID NO: 59 para *KIF1A*, SEQ ID NO: 64 para *LRRTM1*, la SEQ ID NO: 61 para *FLI1*, la SEQ ID NO: 60 para *LINC00599*, la SEQ ID NO: 53 para *PARP15*, la SEQ ID NO: 56 para *ZAP70*, la SEQ ID NO: 57 para *miR193*, más preferiblemente la hipermetilación de la CpG/ isla de CpG que tiene las siguientes coordenadas: 241759686, 241759681, 241759651, 241759621 para *KIF1A*, 80531799 y/o 80531697 para *LRRTM1*, 128564158 y/o 128564051 para *FLI1*, 9760888 para *LINC0059*, 122296586 para *PARP15*, 98340854 y/o 98340751 para *ZAP70* y cualquier combinación de las mismas; y/o 2) la hipometilación de al menos una de las CpG/ isla de CpG de la SEQ ID NO: 63 para *miR296*, la SEQ ID NO: 62 para *GP1BB*, la SEQ ID NO: 65 para *TERT*, más preferiblemente la hipometilación del CpG/ isla CpG que tiene las siguientes coordenadas: 57392374 y/o 57392363 para *miR296*, 19710956 y/o 19710946 para *GP1BB*, 1279758 y/o 1279775 y/o 1279747 para *TERT*.

De acuerdo con una realización preferida adicional, la fase (iii) de medición de las modificaciones epigenéticas del conjunto (5) - *EPHX3*, *KIF1A*, *LRRTM1*, *FLI1*, *ITGA4*, *LINC00599*, *NTM*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *GP1BB*, *hTERT* - implica la medición de 1) la hipermetilación de al menos una de CpG/ isla CpG de *KIF1A*, preferiblemente 1-

4 islas CpG/CpG de KIF1A, más preferiblemente seleccionadas de SEQ ID NO: 59, SEQ ID NO: 58 para EPHX3, la SEQ ID NO: 64 para LRRTM1, SEQ ID NO: 61 para FLI1, SEQ ID NO: 54 para ITGA4, SEQ ID NO: 60 para LINC00599, SEQ ID NO: 55 para NTM, SEQ ID NO: 53 para PARP15, SEQ ID NO: 56 para ZAP70, SEQ ID NO: 57 para miR193, y/o 2) la hipometilación de al menos una de las islas CpG/CpG de SEQ ID NO: 63 para miR296, SEQ ID NO: 62 para GP1BB, SEQ ID NO: 65 para TERT humano.

Todas las secuencias descritas en la presente solicitud de patente se enumeran en la Tabla I y las CpG/islas CpG para cada región están en negrita. Las secuencias que muestran una identidad del 80-99,9% con estas secuencias deben considerarse parte de la divulgación.

En general, para la etapa de amplificación, se genera preferiblemente una biblioteca de amplicones específica de locus, más preferiblemente una biblioteca de amplicones de bisulfito específica de locus. Para ello es aconsejable utilizar una plataforma seleccionada de: MiSEQ, NEXT500 o MiniSEQ - Illumina, IonTorrent, Pacific Biosciences y una plataforma 454 GSJunior.

En el contexto de la presente invención, IonTorrent es una de las máquinas de secuenciación de próxima generación que utiliza tecnologías de secuenciación masivamente paralelas para generar miles de megabases de información de secuencia mediante chips semiconductores. En el contexto de la presente invención, Illumina MiSEQ, NEXT500 o MiniSEQ son máquinas de secuenciación de próxima generación, que utilizan secuencia por química de síntesis para realizar la secuenciación paralela de miles de millones de fragmentos de ADN en paralelo con alta precisión, minimizando el sesgo de incorporación y reduciendo las tasas de error sin procesar en comparación con otras tecnologías, especialmente para homopolímeros. En cualquier caso, cualquier método para obtener el mismo propósito debe considerarse parte de la presente divulgación.

De acuerdo con una realización preferida, las etapas de amplificación implican una primera etapa de amplificación - una amplificación específica de la plantilla- y una segunda ronda de amplificación que permite el código de barras de los amplicones específicos de plantilla obtenidos de la primera etapa de amplificación.

La primera etapa de amplificación usa preferiblemente una estrategia múltiplex, preferiblemente usando pares de cebadores marcados y una ADN polimerasa, preferiblemente la ADN polimerasa Phusion^{MR} U (Figura 1-A).

Preferiblemente, las condiciones de amplificación para esta primera etapa de amplificación son las siguientes:

- 1) Al menos una etapa a aproximadamente 98 °C durante aproximadamente 2 min;
- 2) Al menos 1 ciclo a aproximadamente 98 °C durante aproximadamente 10 segundos, aproximadamente 62 °C durante aproximadamente 2 min, aproximadamente 72 °C durante aproximadamente 1 min;
- 3) Al menos 30-40, preferiblemente aproximadamente 35 ciclos, a aproximadamente 98 °C durante aproximadamente 10 segundos, aproximadamente 62 °C durante aproximadamente 40 segundos y aproximadamente 72 °C durante aproximadamente 20 segundos.

Preferiblemente, se añade una etapa de extensión final a aproximadamente 72 °C durante aproximadamente 5 min al final del último ciclo.

Mediante el uso de condiciones múltiplex, varias regiones de interés de los genes pueden amplificarse juntas con el fin de minimizar los costes y la mano de obra. Esto significa que al menos una región de interés para al menos un gen puede amplificarse (amplicón) al mismo tiempo usando condiciones múltiplex. Para los fines de la presente invención, el multiplex permite amplificar al menos una secuencia de al menos una región de ADN genómica que comprende al menos una CpG/isla de CpG de al menos un gen seleccionado entre: *PARP15*, *ITGA4*, *LRRTM1*, *TERT humano*, *NTM*, *MIR193a*, *KIF1A*, *LINC00599*, *ZAP70*, *EPHX3*, *MIR296*, *FLI1N*, *GP1BB* y combinaciones de los mismos, preferiblemente al menos una de las siguientes combinaciones de genes:

- (a) *PARP15*, *ITGA4*, *LRRTM1*, *TERT humano*; y/o
- (b) *NTM*, *MIR193a*, *KIF1A*; y/o
- (c) *LINC00599*, *ZAP70*, *EPHX3*; y/o
- (d) *MIR296*, *FLI1N*, *GP1BB*.

De acuerdo con una realización preferida, para la primera etapa de amplificación -la etapa de amplificación específica de la plantilla- el par de cebadores se selecciona de los siguientes pares de cebadores etiquetados:

- SEQ ID NO: 1 y 2 y/o SEQ ID NO: 27 y 28 para amplificar *PARP15* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 3 y 4 y/o SEQ ID NO: 29 y 30 para amplificar *ITGA4* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 5 y 6 y/o SEQ ID NO: 31 y 32 para amplificar *NMT* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o

- SEQ ID NO: 7 y 8 y/o SEQ ID NO: 33 y 34 para amplificar *ZAP70* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 9 y 10 y/o SEQ ID NO: 35 y 36 para amplificar *MIR193A* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- 5 – SEQ ID NO: 11 y 12 y/o SEQ ID NO: 37 y 38 para amplificar *EPHX3* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 13 y 14 y/o SEQ ID NO: 39 y 40 para amplificar *KIF1A* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- 10 – SEQ ID NO: 15 y 16 y/o SEQ ID NO: 41 y 42 para amplificar *LINC00599*, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 17 y 18 y/o SEQ ID NO: 43 y 44 para amplificar *FLI1*, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 19 y 20 y/o SEQ ID NO: 45 y 46 para amplificar *GP1BB*, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- 15 – SEQ ID NO: 21 y 22 y/o SEQ ID NO: 47 y 48 para amplificar *MIR296*, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 23 y 24 y/o SEQ ID NO: 49 y 50 para amplificar *LRRTM1*, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- 20 – SEQ ID NO: 25 y 26 y/o SEQ ID NO: 51 y 52 para amplificar *TERT* humano, preferiblemente con las plataformas IonTorrent e Illumina respectivamente.

Puede usarse cualquier combinación de pares de cebadores para cada plataforma dependiendo de las regiones de los genes a amplificar.

25 Cada par de cebadores etiquetados comprende dos cebadores (un cebador directo y otro inverso) para amplificar la región de interés. Cada cebador del par es un cebador de fusión que comprende 1) una porción específica de plantilla que hibrida la secuencia específica para amplificar (porción 3'); y 2) una secuencia universal (porción 5') que no tiene una región complementaria en el genoma (no hibrida con ninguna secuencia en el genoma). Por lo tanto, la secuencia universal se fusiona con una secuencia que se dirige a los extremos de la secuencia específicos de la plantilla- que

30 preferiblemente definen los límites del amplicón.

En el contexto de la presente invención, "secuencia universal" es una secuencia que no tiene una región complementaria en el genoma, es decir, no hibrida con ninguna secuencia en el genoma. En particular, esta secuencia es reconocida por una segunda combinación de cebadores que porta adaptadores y/o índices Illumina o IonTorrent.

35 Los dos cebadores directos o los dos cebadores inversos para cada región a amplificar para las plataformas IonTorrent e Illumina, por ejemplo, la SEQ ID NO: 1 y la SEQ ID NO: 27, tienen preferiblemente el mismo extremo 3' (ya que este extremo es el extremo específico de la plantilla) y un extremo 5' diferente dependiendo de la plataforma utilizada.

40 De acuerdo con una realización preferida, para mantener la direccionalidad de secuenciación, se deben diseñar diferentes secuencias universales (colas) para los dos cebadores de fusión de la primera etapa de amplificación (en la Figura 1-A se denominan extremos Univ-A y Univ- B de cada amplicón). Preferiblemente, después de la PCR, los productos de amplificación multiplex pueden combinarse, más preferiblemente los productos de amplificación del siguiente conjunto de genes se combinan (a) - (b) y (c) - (d) para purificación con perlas magnéticas.

45 Los primeros productos de amplificación (específicos de la plantilla), preferiblemente los productos de amplificación múltiplex denominados amplicones, y/o las combinaciones se purifican preferiblemente, después de la etapa de amplificación, preferiblemente usando SPRI-AMPure XT. Además, también se pueden cuantificar, después de la purificación, preferiblemente con el FluorometerQuantus^{MR} o Qubit^{MR} o similar.

50 Los amplicones, eventualmente después de la purificación o cuantificación, se usan como plantilla, preferiblemente en una cantidad de aproximadamente 100 ng, para una segunda ronda de amplificación. Esta segunda ronda es para codificar la muestra con códigos de barras e implica de 5 a 10, preferiblemente alrededor de 6 ciclos.

55 Esta segunda etapa de amplificación se realiza preferiblemente en presencia de un par de cebadores adicional (directo e inverso). Cada cebador del par comprende comenzar desde el extremo 5': 1) una secuencia adaptadora específica para la plataforma de secuenciación de próxima generación específica utilizada, tal como Illumina y/o IonTorrent, y/o 2) una secuencia de código de barras específica de la muestra, tal como identificadores múltiplex (MID) o índices, y/o 3) una etiqueta de tipo universal complementaria a la secuencia universal del par de cebadores utilizados en la primera

60 etapa de amplificación (Figura 1-A).

En el contexto de la presente invención, "secuencias o índices de códigos de barras específicos de muestras" son secuencias cortas que se han introducido artificialmente en una secuencia diana específica que permite la identificación de cada muestra (en otras palabras, para identificar al individuo) en secuenciación paralela cuando las

a cada muestra para que se puedan distinguir y clasificar durante el análisis de datos. Este enfoque es rentable y rápido porque la combinación de muestras aumenta exponencialmente el número de muestras analizadas en un solo proceso.

- 5 En el contexto de la presente invención, los "identificadores o índices múltiplex" se insertan durante la segunda ronda de amplificación e incluyen también adaptadores específicos para la plataforma de secuenciación de próxima generación específica que se pretende usar, lo que permite la hibridación de secuencias diana con el chip o la celda de flujo.
- 10 En el contexto de la presente invención, "biblioteca de amplicones" significa una reunión de secuencias diana derivadas de varias muestras, que es necesario secuenciar en paralelo. Cualquier técnica de secuenciación debe considerarse parte de la presente divulgación.

Tabla I

SEQ ID NO: 1	PARP15-directo	GTAATACGACGGTCAGT- GTTGTTTTGAGTTTAGGGGTTT
SEQ ID NO: 2	PARP15-inverso	CAGGAAACAGCTATGAC- ACTCAACCCTACTATCCCCTTCC
SEQ ID NO: 3	ITGA4-directo	GTAATACGACGGTCAGT- GAGAGYGTATGGTTTGGGAAG
SEQ ID NO: 4	ITGA4-inverso	CAGGAAACAGCTATGAC- AAAAACRCATCAATCCAACCTCT
SEQ ID NO: 5	NMT-directo	GTAATACGACGGTCAGT- GTGTTTGAATTGTGTTGG
SEQ ID NO: 6	NMT-inverso	CAGGAAACAGCTATGAC- ATATTCCTCCCCCAAATCCA
SEQ ID NO: 7	ZAP70-directo	GTAATACGACGGTCAGT- TATTGTGGATYGGTAGAGTTTTG
SEQ ID NO: 8	ZAP70-inverso	CAGGAAACAGCTATGAC- CTCTCACCTCCAACCTCCA
SEQ ID NO: 9	MIR193A-directo	GTAATACGACGGTCAGT- GTGTAATTTTTGGAGGGTTGGGT
SEQ ID NO: 10	MIR193A-inverso	CAGGAAACAGCTATGAC- AAACTAAAACCTTTATAAACCAATTAATCC
SEQ ID NO: 11	EPHX3-directo	GTAATACGACGGTCAGT- GGYGATATGTGAGGTTGGT
SEQ ID NO: 12	EPHX3-inverso	CAGGAAACAGCTATGAC- CCAACRAAAAATCRCTCAAA
SEQ ID NO: 13	KIF1 A-directo	GTAATACGACGGTCAGT- ATAAATTAGTTGGYGATTGGAGTT
SEQ ID NO: 14	KIF1 A-inverso	CAGGAAACAGCTATGAC- CCCCRCATAAACACTAACATAAA
SEQ ID NO: 15	LINC00599-directo	GTAATACGACGGTCAGT- GTTAAGAATGGGGTTGGTTGAGTAT
SEQ ID NO: 16	LINC00599-inverso	CAGGAAACAGCTATGAC- CCCTATCTATCACAACTACACACCTC

(continuación)

SEQ ID NO: 17	FLI1-directo	GTAATACGACGGTCAGT- ATAGGGGAGTGAGGGTAGG
SEQ ID NO: 18	FLI1-inverso	CAGGAAACAGCTATGAC- CCCRCCRCTTACCTTAATA
SEQ ID NO: 19	GP1 BB-directo	GTAATACGACGGTCAGT- TTGTTTTGGTGATAGGAGAATAATG
SEQ ID NO: 20	GP1 BB-inverso	CAGGAAACAGCTATGAC- ACCTAACCAAAAATAACCCAAAAAC
SEQ ID NO: 21	MIR296-directo	GTAATACGACGGTCAGT- GGGTATTGGGGAATGGTGA
SEQ ID NO: 22	MIR296-inverso	CAGGAAACAGCTATGAC- CRAAAACCCCAACCTAACA
SEQ ID NO: 23	LRRTM1-directo	GTAATACGACGGTCAGT- GTTTGGTGGAGGGAAGGAGT
SEQ ID NO: 24	LRRTM1-inverso	CAGGAAACAGCTATGAC- TAACCRACRCCACCTTTTAC
SEQ ID NO: 25	hTERT-directo	GTAATACGACGGTCAGT- GGTTTTTGATTTTTATTTTTGTGGTT
SEQ ID NO: 26	hTERT-inverso	CAGGAAACAGCTATGAC- CCAACTTCCTCAAACCTATTTAA
SEQ ID NO: 27	PARP15-directo	TCGTCCGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GTTGTTTTGAGTTTAGGGGTTT
SEQ ID NO: 28	PARP15-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- ACTCAACCCTACTATCCCCTTCC
SEQ ID NO: 29	ITGA4-directo	TCGTCCGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GAGAGYGTATGGTTTGGGAAG
SEQ ID NO: 30	ITGA4-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-AAAAACRCATCAATCCAACCTCT
SEQ ID NO: 31	NMT-directo	TCGTCCGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GTGTTTGAATTGTGTTGG
SEQ ID NO: 32	NMT-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-ATATTCCTCCCCCAAATCCA
SEQ ID NO: 33	ZAP70-directo	TCGTCCGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-TATTGTGGATYGGTAGAGTTTTG
SEQ ID NO: 34	ZAP70-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-CTCTCACCTCCAACCTCCA

(continuación)

SEQ ID NO: 35	MIR193A-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GTGTAAATTTTTGGAGGGTTGGGT
SEQ ID NO: 36	MIR193A-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- AAACTAAACTTTATAAACCAATTAATCC
SEQ ID NO: 37	EPHX3-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GGYGATATGTYGGAGTTGGT
SEQ ID NO: 38	EPHX3-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-CCAACRAAAAATCRCTCAAA
SEQ ID NO: 39	KIF1 A-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN- ATAAATTAGTTGGYGATTGGAGTT
SEQ ID NO: 40	KIF1 A-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- CCCCRCATAAACACTAACATAAA
SEQ ID NO: 41	LINC00599-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN- GTTAAGAATGGGGTTGGTTGAGTAT
SEQ ID NO: 42	LINC00599-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- CCCTATCTATCACAACTACACACCTC
SEQ ID NO: 43	FLI1-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-ATAGGGGAGTGAGGGTAGG
SEQ ID NO: 44	FLI1-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-CCCRCCRCTTACCTTAATA
SEQ ID NO: 45	GP1 BB-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN- TTGTTTTGGTGATAGGAGAATAATG
SEQ ID NO: 46	GP1 BB-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- ACCTAACCAAAAATAACCCAAAAAC
SEQ ID NO: 47	MIR296-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GGGTATTGGGGAATGGTGA
SEQ ID NO: 48	MIR296-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-CRAAAACCCCAAACCTAACA
SEQ ID NO: 49	LRRTM1-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN-GTTTGGTGGAGGGAAGGAGT
SEQ ID NO: 50	LRRTM1-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN-TAACCRACRCCACCTTTTAC

(continuación)

SEQ ID NO: 51	hTERT-directo	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNN- GGTTTTTGTATTTTATTTTGTGGTT
SEQ ID NO: 52	hTERT-inverso	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGNNNN- CCAACTTCCTCAAACCCTATTTAA
SEQ ID NO: 53	amplicón PARP15 122296564- 122296725	CGACCCCCAGAGAACTTATGCACGGAGTTG CAGGTGTTACTTCCAGAGCCGGACGAGATC GGGAGGCGGGGAGCGTGCTGCCGGCCGG GAACCGTGGGGCGCGGAAGGCCTCCCGGC GCTCTTCTCCCGGAGTATGGTGAGGAGCG CGGGGGACGGGTGCG
SEQ ID NO: 54	amplicón ITGA4 182322887- 182323053	TCAC CGT CTGAG CGCACGT GCAC CGT GTCT CG CTTTAGGCTCCTAGTGGGTG CGCCCACTGC CAACTGGCT CGCCAACGCTT CAGTGATCAAT CC CGGGGCG ATTTACAGATGCAGGAT CGGA AAGAATCC CGGCCAGACGTGCGAACAGCTC CAGCTGGGTGAGTTGG
SEQ ID NO: 55	amplicón NTM 131781042- 131781185	GAAG CGGCTCCTGAGACGCGCCCA CACCTT TCACCTG CGCGCGCTTCCCCCTCCTCGGC CACCTTCCCG CGGAAGCAGCGAGGAGGG AGCCCCCTTTGG CGCTCCTCGTGGAACCG GTTTT CGAGGCTGGCAAAAGCCGA
SEQ ID NO: 56	amplicón ZAP70 98340750- 98340885	GCGAGTTCTACTCGCGCGACCCCGACGGGC TGCCCTGCAACCTG CGCAAGCCGTGCAACC GGCCGTGCGGCCTCGAGCCGCAGCCGGGG GTCTT CGACTGCCTGCGAGACGCCATGGTG CGTGA CTAC GTGCGCCA
SEQ ID NO: 57	amplicón MIR193A 29886860- 29887068	TTGGAGCC CGCGACCCGAGGT CGGG CGGG G CGGTGGA CTTTCC CGGGGAACGCCGCCT GAGGGACACCCAGAGCTT CGGCGGAGCGG AG CGCGGTGCACAGAGCCGGCGACCGGAC CCAGCCCC CGGGAAGCCCGTCGGGGACGCA CCCC GAACTCCGAGGATGGGAGCTGAGGG CTGGGTCTTTG CGGGCGAGATGAGGGTGTC GGATC
SEQ ID NO: 58	amplicón EPHX3 15342826- 15343049	G CCGCGTGGTCCCGGAGCGCGGCGACAT G CCGGAGCTGGTGGTGACCGCGCTGCTGG CGCCGTGCGGCCTGTGCTGAAGCTGCTGC GCGCCTTCATGTGGAGCCTGGTGTTCTCGG TGG CGCTGGTGGCCGCGGCGGTCTACGGC TGCATAGCGCTCAC GCACGTGCTGTGCCGG CCCC GGCGCGGCTGCTGCGGGCGCCGTG GAG CGCGTCCCCCGCCTG

(continuación)

SEQ ID NO: 59	amplicón KIF1A 241459586- 241759727	T C G G A C G C C G C G C G C A G C C A A C C A G C G C C T C G G G A G G G G C G C G G G G C G G T C C G G G C C G G G G G C G C G C A G C C C G T C A C G C G G C G G C G C C G T C A C A G C G C A G A G C A G C C G G C G A G C G G C A G C A G C T C C G G G C T C G A G A G C C C G C
SEQ ID NO: 60	amplicón LINC00599	A C C G T G G G T C G G C G A G G G C C C G C C A A G G A A G G A G C G A C C G A C C G A G C C A G G C G C C C T C C G C A G A C C T C C G C G C A G C G G C C G C G G G C G
	9760739-9760890	C G A G G G A G G G G T C T G G A G C T C C T C C G G C T G C C T G T C C C G C A C C G A G C C C G T G G G G T G G G G A G
SEQ ID NO: 61	amplicón FLI1 128564020- 128564160	C T C G C A G G G G G C A C G C A G G G A G G G C C C A G G G C G C C A G G G A G G C C G C G C C G G G C T A A T C C G A A G G G G C T G C G A G G T C A G G C T G T A A C C G G G T C A A T G T G T G A A T A T T G G G G G G C T C G G C T G C A G A C T T G G C C A A T G G A C G
SEQ ID NO: 62	amplicón GP1BB 19710829- 19710960	T G T T G G T G A A C G T C G C A G C G G G T G T C C G A T G C T C C G T G T G C C C C T G A G A G C G G G T G G G A G C G G A A G C C T G A G C G G C C T G C G G C C T C C G G C G A T A G T G T G C T A T C T G C C G C T G C A G C G C G C G T C C G C G C G C G C
SEQ ID NO: 63	amplicón MIR296 57392355- 57392545	G C C G T G C C G C G C C T C T G C C A C A G T C T C C C G T G G G T C C C C A C T G C C G G G A G T T T A C C T A C C T C C C A C C C T C A G C C A C G A G C A C C T G C A G A T G C G G T G C C A T C C T C G C C G G T G C C C G C T C C C C T T C C G A G G C T C T C C A C C C T A C T G A C G C C T C C T C G G G A A G C C C C C G C T C C T T C C C C G A C C T C A C
SEQ ID NO: 64	amplicón LRRTM1 80531676- 80531807	G G C G C C G C T C G C C G C T C C C C G C G C C G C C G T C C G C A C C T C C C C A C C G C C C G C C G C C C G C C G C C C G C C G C C C G C A A A G C A T G A G T G A G C C C G C T C T G C A G C T G C C C G G G G C G C G A A T G G C A G G C T G T T T C C G
SEQ ID NO: 65	amplicón TERT humano 1279743- 1279851	G C T G C G G T G G C T G C G G T G A C C C C G T C A T C T G A G G A G A G T G T G G G G T G A G G T G G A C A G A G T G T G G C A T G A G G A T C C C G T G T G C A A C A C A C A T G C G G C C A G G A A C C C G T T
SEQ ID NO: 66	IonTorrent A	C C A T C T C A T C C C T G C G T G T C T C C G A C T C A G C T A A G G T A A C G A T - G T A A T A C G A C G G T C A G T
SEQ ID NO: 67	IonTorrent P	C C T C T C T A T G G G C A G T C G G T G A T - C A G G A A C A G C T A T G A C
SEQ ID NO: 68	P5 índice 503	A A T G A T A C G G C G A C C A C C G A G A T C T A C A C – T A T C C T C T - T C G T C G G C A G C G T C A G A T G T G T A A G A G A C A G

(continuación)

SEQ ID NO: 69	P7 índice 702	CAAGCAGAAGACGGCATACGAGAT - CTAGTACG- GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAG
SEQ ID NO: 70	Universal A directo	GTAATACGACGGTCAGT
SEQ ID NO: 71	Universal B inverso	CAGGAAACAGCTATGAC

La Tabla II resume varias informaciones sobre los genes analizados en el presente método, la información de mapeo y el tamaño de la región genómica amplificada, la referencia de la secuencia (referencia NM en el NCBI) y el número de CpG/ islas de CpG.

5

Tabla II

Gen	Mapa	RefSeq	Posición	Coordenadas de h19 de la UCSC	Longitud del amplicón	Posición respecto a TSS	Número de CpG de interrogación	Estado de la impresión
<i>ZAP70</i>	2q11.2	NM_001079	Exón 3	Hebra + del cromosoma 2: 98340750 - 98340885	180	+10705	21	No
<i>GP1BB</i>	22q11.21	NM_000407	Exón 1	Hebra + del cromosoma 22: 19710829 - 19710960	192	+315	18	No
<i>KIF1A</i>	2q37.3	NM_001244008	Exón 1	Hebra - del cromosoma 2: 241759586 - 241759727	189	-8	28	No
<i>PARP1</i> 5	3q21.1	NM_001113523	Exón 1	Hebra + del cromosoma 3: 122296564 - 122296725	206	+93	19	No
<i>ITGA4</i>	2q31.3	<u>NM 000885</u>	Exón 2	Hebra + del cromosoma 2: 182322887 - 182323053	214	+912	14	No
<i>NTM</i>	11q25	<u>NM 016522</u>	Exón 1	Hebra + del cromosoma 11: 131781042 - 131781185	190	+62	15	Sí, Maternal, sin SNP
<i>MIR19</i> 3A	17q11.2	<u>NR 029710</u>	Promotor	Hebra + del cromosoma 17: 29886860 - 29887068	256	-178	26	No
<i>MIR13</i> 7HG	1p21.3	<u>NR 046105</u>	Exón 1	Hebra - del cromosoma 1: 98511645 - 98511814	216	+136	18	No
<i>EPHX3</i>	19p13.12	<u>NM 024794</u>	Exón 1	Hebra - del cromosoma 19: 15342826 - 15343049	223	+215	29	No
<i>LINC00</i> 599	8p23.1	<u>NR 029668</u>	Exón 1	Hebra - del cromosoma 8: 9760739 - 9760890	199	+69	20	No
<i>FLI1</i>	11q24.3	<u>NM 001271010</u>	Exón 1	Hebra + del cromosoma 11: 128564020 - 128564160	186	+187	12	No

(continuación)

Gen	Mapa	RefSeq	Posición	Coordenadas de h19 de la UCSC	Longitud del amplicón	Posición respecto a TSS	Número de CpG de interrogación	Estado de la impresión
<i>MIR296</i>	20q13.32	NR 029844	Exón 1	Hebra - del cromosoma 20: 57392355 - 57392545	238	+180	15	Sí, paternal, sin SNP
<i>LRRTM1</i>	2p12	NM 178839	Promotor	Hebra - del cromosoma 2: 80531676 - 80531807	179	-431	24	Sí, paternal, sin SNP
<i>hTERT</i>	5p15.33	NM 198253	Promotor	Hebra - del cromosoma 5: 1279743 - 1279851	109	+14976	6	No, SNP en rs10069690

Después de determinar las secuencias de los amplicones, preferiblemente de acuerdo con la descripción anterior, la medición cuantitativa del nivel de metilación se realiza utilizando las herramientas comunes utilizadas para este propósito, preferiblemente seleccionadas entre: BISMAR, BS-Seeeker2, BISM, BSPAT y QUMA. La Figura 1-B resume una realización específica que ejemplifica la etapa de medición del nivel de metilación, en otras palabras, la última fase del método que permite la determinación de la eventual positividad de la muestra para HNSCC.

El análisis se basa en el hecho de que los tratamientos con bisulfito de las secuencias de ácidos nucleicos convierten una citosina (C) no metilada en uracilo (U), de manera que el uracilo (U) se lee como timina (T) cuando se secuencian el ácido nucleico tratado. Esta conversión no afecta a una C metilada que permanece como C en la secuencia. Por lo tanto, las lecturas de secuenciación que muestran C después de los tratamientos con bisulfito significan una citosina metilada, en lugar de una T en las lecturas de secuenciación después de los tratamientos con bisulfito significa una citosina no metilada.

De acuerdo con una realización preferida, tras la etapa de secuenciación, se filtra cualitativamente un archivo FASTQ que comprende todas las secuencias obtenidas de una muestra/individuo (las lecturas) mediante el uso de la computación en la nube conocida para este fin. Los archivos FASTQ generalmente se producen mediante procedimientos de secuenciación de próxima generación. Preferiblemente, para Illumina FASTQ se aplica un filtro de Q30 para el 90% de las bases. Se aplica un filtrado menos estricto para IonTorrent (las puntuaciones Q se definen como la propiedad que está relacionada logarítmicamente con la probabilidad de error de llamada base (P)2 como $Q = -10 \log_{10} P$. Por ejemplo, si Phred asigna una puntuación Q de 30 (Q30) a una base, esto equivale a la probabilidad de una llamada de base incorrecta (1 en 1000 veces. En este caso, la precisión de la llamada de base es del 99,9%). Esta etapa de filtrado permite obtener secuencias de alta calidad. Después de la etapa de filtrado, las secuencias de alta calidad se convierten preferiblemente en un formato FASTA y finalmente se recortan, preferiblemente 25 pares de bases, en uno y/o ambos extremos para eliminar la secuencia del cebador.

Las secuencias, preferiblemente las secuencias de alta calidad, obtenidas para cada región amplificada, se analizan para la medición cuantitativa del nivel de metilación utilizando las herramientas comunes utilizadas para este propósito, preferiblemente seleccionadas entre: BISMAR, BS-Seeeker2, BISM, BSPAT y QUMA. Preferiblemente, el FASTA para cada región amplificada se carga en esta herramienta.

Estas herramientas comparan la secuencia de tipo silvestre y la secuencia después del tratamiento con bisulfito y produce una medida cuantitativa del nivel de metilación. Por lo tanto, estas herramientas permiten medir el nivel de metilación de las CpG/islas de CpG que se encuentran en cada región amplificada.

Preferiblemente, las secuencias de referencia para cargar cada secuencia de alineación necesaria para estas herramientas se seleccionan de: las SEQ ID NO: 53-65.

Preferiblemente, el valor de hipometilación obtenido para la CpG hipometilada informada anteriormente de los genes hipometilados, preferiblemente seleccionado entre: TERT, MIR296 y GP1BB, se invierte de la siguiente manera: $r = 1 - x$, en la que x es el valor de hipometilación obtenido y r es el valor que se incluirá en el análisis discriminante lineal.

De acuerdo con una realización preferida, los datos del nivel de metilación pueden analizarse usando un análisis discriminante lineal, preferiblemente seleccionando al menos una, preferiblemente todas, las CpG seleccionadas entre: 19710956 para GP1BB; 98340854 para ZAP70; 241759621 para KIF1A; 1279758 para TERT; 80531799 para LRRTM1; 131781167 para NTM; 29886944 para MIR193; 122296586 para PARP15; 182322902 para ITGA4; 128564158 para FLI1; 9760888 para LINC0059; 15342885 para EPHX3; y 57392374 para MIR296.

Preferiblemente, para cuantificar el nivel de metilación de más de una secuencia con el fin de obtener un nivel de metilación total, se usa preferiblemente un análisis discriminante lineal (LDA) multiclase. Esto significa que, por ejemplo, cuando se mide el nivel de metilación de CpG de al menos una región genómica de más de un gen de interés, tal como el conjunto de genes descritos anteriormente, es aconsejable utilizar este algoritmo.

De acuerdo con el método de la presente invención, la función discriminante lineal que pondera la contribución de cada CpG metilada se calcula utilizando la siguiente fórmula:

$$y(r_i) = K + \sum c_i * r_i$$

En particular, el coeficiente discriminante para cada modificación epigenética (metilación) se basa en el número y tipo de regiones genómicas utilizadas durante el análisis.

El coeficiente discriminante (c_i) se calcula teniendo en cuenta los niveles (tasas) de cada metilación analizada de tal manera que se maximiza la diferencia entre muestras positivas y negativas.

La constante (K) es un número fijo que se suma a la función discriminante y se calcula de tal manera que se maximiza la diferencia entre muestras positivas y negativas.

El valor final de y se calcula después de sumar la constante al resumen del producto entre el umbral del nivel de metilación para al menos una CpG/isla de CpG y el coeficiente respectivo.

Para un agrupamiento correcto de cada muestra, es preferible un umbral que discrimina entre las muestras positivas y N negativas, pero no se limita a, uno calculado mediante un análisis de curva ROC.

Los valores por encima de cierto valor y se deben considerar positivos.

De acuerdo con una realización preferida, al utilizar el análisis discriminante lineal y una ROC, una muestra se considera positiva para HNSCC si:

- (a) usando la combinación de *TERT* humano y *KIF1A* - cuando y excede el umbral de -0,5306765 (AUC: 0,924) con una sensibilidad de 0,914 y 1 - especificidad de 0,123; y/o
- (b) usando la combinación de *LRRTM1*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *hTERT* - cuando y excede el umbral de -0,1582287 (AUC: 0,962) con una sensibilidad de 0,944 y 1 - especificidad de 0,085; y/o
- (c) *KIF1A*, *LRRTM1*, *ZAP70*, *miR193*, *miR296*, *hTERT* - cuando y excede el umbral de -0,1258801 (AUC: 0,962) con una sensibilidad de 0,917 y una 1 - especificidad de 0,07; y/o
- (d) *KIF1A*, *LRRTM1*, *FLI1*, *LINC00599*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *GP1BB*, *hTERT* - cuando y supera el umbral de 0,4430684 (AUC: 0,970) con una sensibilidad de 0,972 y 1 - especificidad de 0,056; y/o
- (e) *EPHX3*, *KIF1A*, *LRRTM1*, *FLI1*, *ITGA4*, *LINC00599*, *NTM*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *GP1BB*, *hTERT* - cuando y excede el umbral de 1.0582872 (AUC: 0.971) con sensibilidad de 0,917 y 1 - especificidad de 0.

El análisis discriminante lineal (LDA) genera una puntuación que pondera al menos una, preferiblemente todas, las CpG seleccionados de: 19710956 para *GP1BB*; 98340854 para *ZAP70*; 241759621 para *KIF1A*; 1279758 para *TERT*; 80531799 para *LRRTM1*; 131781167 para *NTM*; 29886944 para *MIR193*; 122296586 para *PARP15*; 182322902 para *ITGA4*; 128564158 para *FLI1*; 9760888 para *LINC0059*; 15342885 para *EPHX3*; y 57392374 para *MIR296*, del gen o los genes investigados.

Esta precisión cumple con los requisitos clínicos y, por lo tanto, el método de la invención es ideal para ser utilizado en lugar de y/o junto con los métodos de referencia actuales.

De acuerdo con un aspecto adicional de la presente invención, el método descrito anteriormente permite detectar la presencia de cancerización de campo, preferiblemente en pacientes resecados quirúrgicamente por HNSCC alrededor del sitio de intervención. En particular, la invención muestra que la presencia de la cancerización de campo con patrón de metilación alterado está relacionada con mal pronóstico y alto riesgo de reaparición o de tumores secundarios.

Un aspecto adicional de la presente invención se refiere a detectar la presencia de cancerización de campo en pacientes resecados quirúrgicamente por HNSCC alrededor del sitio de intervención. La presencia de cancerización de campo con patrón de metilación alterado se relaciona con mal pronóstico y alto riesgo de reaparición o de tumores secundarios.

Un aspecto adicional de la presente invención, se refiere a un kit para determinar la presencia y/o el riesgo de HNSCC, preferiblemente HG-SIL y/o OSCC en una muestra de prueba biológica, que contiene preferiblemente un cepillo exfoliante de mucosa oral obtenida de un individuo, que preferiblemente muestra OPML o lesiones sospechosas, comprendiendo dicho kit reactivos para la conversión con bisulfito de sodio y un conjunto de oligonucleótidos, enzimas

y tampones que amplifican específicamente al menos una de las secuencias SEQ ID NO: 53 - SEQ ID NO: 65 con al menos al menos un par de cebadores seleccionados del grupo que consiste en la SEQ ID NO: 1-52 y de la SEQ ID NO: 66-69.

5 Ejemplo

Para el establecimiento y las fases de validación del presente método, todos los pacientes consecutivos remitidos al Departamento de Ciencias Orales de la Universidad de Bolonia, desde enero de 2015 hasta julio de 2016.

10 Se excluyeron las lesiones con una etiología obvia, tal como traumatismos y ulceraciones aftosas infecciosas.

Todos los pacientes que presentaban lesiones orales que requerían una biopsia incisional con fines diagnósticos también se sometieron a cepillado oral de la misma lesión.

15 Las muestras de cepillado oral siempre se recogieron antes de la biopsia incisional para el diagnóstico histológico y la estadificación de cada lesión.

Los diagnósticos histológicos se realizaron siguiendo los criterios de la OMS (Thompson, 2006). La distinción entre HG-SIL y LG-SIL se hizo de acuerdo con la clasificación de Ljubjana de 2014 (Gale, Blagus et al., 2014).

20 Un total de 150 series de muestras de cepillado oral para el conjunto de datos de entrenamiento se componía de:

- 65 donantes sanos normales;
- 28 OSCC,
- 25 • 1 OSCC con rasgo sarcomatoide
- 6 HG-SIL,
- 35 mucosas normales alrededor del sitio de intervención de pacientes previamente tratados quirúrgicamente por un OSCC.

El conjunto de datos de validación estaba compuesto por:

- 30 • 2 OSCC;
- 3 PVL (leucoplasia verrugosa proliferativa)
- 14 OLP (liquen plano oral)
- 1 fibroma oral
- 35 • 20 donantes sanos normales

Método de cepillado oral

Se usó un citocepillo para recolectar células exfoliadas de la mucosa oral.

40 En las lesiones de OSCC y OPML, toda la superficie de las lesiones se cepilló suavemente de forma repetida cinco veces.

45 La recogida de células por cepillado se realizó siempre antes de la biopsia incisional y sin el uso de ningún anestésico local.

Después del cepillado, cada muestra del citocepillo se colocó en un tubo de 2 ml que contenía etanol absoluto para la conservación de las células.

50 Declaración de Ética

Todas las investigaciones clínicas se han realizado de acuerdo con los principios expresados en la Declaración de Helsinki. El estudio fue aprobado por el Comité de Ética local (número de estudio: 14092, número de protocolo: 899/CE). Toda la información relativa al material humano utilizado en este estudio se gestionó mediante códigos numéricos anónimos.

55 Purificación del ADN

El ADN de las muestras de cepillo exfoliante se purificó usando el kit de extracción de ADN completo MasterPure^{MR} (Epicenter, Madison, WI).

60 Tratamiento con bisulfito de ADN genómico

El tratamiento con bisulfito de sodio del ADN genómico se llevó a cabo con el kit EZ DNA Methylation-Lightning^{MR} (Zymo Research Europe, Freiberg, Alemania) de acuerdo con el protocolo del fabricante.

65

Fase de amplificación

Para la plataforma Illumina (MiSEQ), se generaron bibliotecas de amplicones de bisulfito específicas de locus multiplexando con cebadores etiquetados (vease la Tabla 1) usando ADN polimerasa Phusion U (ThermoFisher) 25 µl. La plataforma Illumina necesita un método diferente para evaluar amplicones, el enfoque Nextera^{MR} que requiere adaptadores en el extremo 5' como se describe en la Tabla 1. La inserción de un N o dos N en el medio garantiza la complejidad mínima para el agrupamiento. El código de barras se realizó utilizando el kit de índice Nextera^{MR}.

Las condiciones de ciclo para la PCR específica de la primera plantilla fueron: incubación inicial a 98 °C durante 2 min seguido de 1 ciclo de 98 °C 10 s, 62 °C 2 min, 72 °C 1 min, seguido de 35 ciclos a 98 °C durante 10 s, 62 °C durante 40 s y 72 °C durante 20 s. Se añadió una etapa de extensión final a 72 °C durante 5 min al final del último ciclo.

La PCR multiplex consistió en:

- Multiplex 1 *PARP15*; *ITGA4*; *LRRTM1*; *TERT*
- Multiplex 2 *NTM*; *MIR193a*; *KIF1A*
- Multiplex 3 *LINC00599*; *MIR137HG*; *ZAP70*; *EPHX3*
- Multiplex 4 *MIR296*; *FLI1N*; *GP1BB*

Las condiciones de reacción para los multiplex fueron: agua ultrapura: 8,8 µl, tampón 5x HF: 5 µl, betaína 5M: 5 µl, dNTP 10 mM: 0,5, cebadores 0,1 µM: 1,5 µl, MgCl₂ 50 mM: 0,5 µl, Phusion U: 0,2 µl, ADN tratado con bisulfito: 3,5 µl.

Los productos de amplificación de los multiplex 1-2 y 3-4 se combinaron y purificaron usando SPRI-AMPure XT (Agencourt-Beckman Coulter, Beverly, MA), se cuantificaron con el FluorometerQuantus^{MR} (Promega, Madison, WI) y luego se emplearon como plantilla (100 ng) para una segunda ronda de PCR (6 ciclos).

Las secuencias de códigos de barras específicas de la muestra (índices) y etiquetas enlazadoras universales (secuencias adaptadoras P5/P7) se añadieron en esta segunda PCR realizada con 6 ciclos (98 °C 1 min/98 °C 10 s/ 63 °C 30 s/ 72 °C 3 min).

La biblioteca de amplicones se purificó usando perlas AgencourtAMPure XP (Beckman Coulter, Krefeld, Alemania) y luego se cuantificaron con el FluorometerQuantus^{MR} (Promega, Madison, WI).

Las bibliotecas se diluyeron, se combinaron a 4 nM y se cargaron en una celda de Nano flujo en MiSEQ. Alternativamente, para el diseño experimental de secuenciación de amplicones para plataformas 454/IonTorrent, el diseño de 'cola universal' requiere dos conjuntos de pares de cebadores, utilizados en dos rondas sucesivas de PCR.

La primera ronda utiliza cebadores de fusión que se dirigen a las secuencias específicas de la plantilla (que definen los límites de los amplicones) fusionados a una secuencia universal que será la diana de los cebadores de la segunda ronda (Figura 1). Para mantener la direccionalidad de la secuenciación, se deben diseñar diferentes colas universales para los extremos del cebador A y el cebador B de los amplicones (Univ-A y Univ-B en la Figura 1-A).

En la segunda ronda, la secuencia Univ-A está dirigida por un cebador de fusión que está seguido por la secuencia del cebador A del sistema de secuenciación 454 más una secuencia MID para identificar la muestra; y lo mismo para la secuencia Univ-B, con Cebador B (Figura 1-A).

Como se informa en la Tabla I, la secuencia de cola A universal directa de elección es GTAATACGACGGTCAGT y la secuencia Universal B inversa con cola de elección es CAGGAAACAGCTATGAC.

Predicción *in silico* de la isla de CpG y el diseño del cebador

Para identificar supuestas islas de CpG en la región promotora se emplearon los siguientes genes: *ZAP70*, *ITGA4*, *KIF1A*, *PARP15*, *EPHX3*, *NTM*, *LRRTM1*, *FLI1*, *MIR193*, *LINC00599*, *MIR296*, *hTERT* y *GP1BB*, una secuencia genómica como la almacenada en el navegador del genoma Ensembl (<http://www.ensembl.org/index.html>), que incluye 1000 pb secuencia arriba del sitio ATG como secuencia de consulta. Se aplicó el diseño MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) para identificar las CpG y los mejores cebadores de elección (véase la Tabla 1).

En la segunda ronda de PCR, se reconoce que Universal A y B agregan los adaptadores específicos para 454 de la siguiente manera, incluido el adaptador a + clave + MID + Universal A o B:

ADPTUNIAFWMID2: CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA-GTAATACGACGGTCAGT (SEQ ID NO: 72)

ADPTUNIBRV MID5: CTATGCGCCTTGCCAGCCCGCTCAGATCAGACACG-CAGGAAACAGCTATGAC (SEQ ID NO: 73)

Se utilizó el mismo enfoque para la plataforma IonTorrent/Proton de la siguiente manera:

Cebador A: IonXpress 1: CCATCTCATCCCTGCGTGTCTCCGACTCAG-CTAAGGTAACGATGTAATACGA
CGGTCACT (SEQ ID NO: 74)

5 Cebador P: CCTCTCTATGGGCAGTCGGTGAT-CAGGAAACAGCTATGAC (SEQ ID NO: 75)

Análisis bioinformático:

Después del final del proceso, cada archivo de FASTQ obtenido después de la NGS se procesó de la siguiente manera:

- 10
- Usar Galaxy, la plataforma abierta basada en la web para investigación biomédica intensiva de datos (<https://usegalaxy.org>), se cargan ambos archivos FASTQ R1 y R2.
 - Usando Pear para fusionar lecturas de extremo emparejado (R1 directo y R2 inverso).
 - Usando FASTQ Groomer para hacer que el archivo FASTQ sea compatible con todos los formatos.

15

 - Filtrado por calidad: filtro para $Q30$ al 90% y > 100 pb de largo
 - Conversión de archivo FASTQ a FASTA
 - Recortar desde el principio hasta esta posición de 25 pb y eliminar todo desde esta posición hasta el final: -25 pb (Ignorar las líneas que comiencen con estos caracteres)
 - Descargar el archivo FASTA.

20

 - Luego, utilizar Perl para extraer lecturas de cada gen de los conjuntos analizados de la siguiente manera: (Por ejemplo, para extraer hTERT, elaborar la línea de comando: `grep-B 1 "GGAGAGTGTGGGGTGAGGTGGATAG" Nombre del archivo fuente.fasta>Nombre del archivo de salida.fasta`)
 - Por ejemplo, para el conjunto (e), una vez extraídos los 13 archivos, cargue cada uno en BISMA (<http://services.abc.uni-stuttgart.de/BDPC/BISMA/manualunique.php>) para calcular el porcentaje de CpG para cada CpG de cada gen.

25

 - Crear una tabla con el nivel de metilación de CpG dividiendo dos grupos, uno para muestras sanas normales (indicado como clase 0), el otro para OSCC y HG-SIL (clase 1) para cada CpG de cada gen; luego calcular la curva ROC para cada uno de las CpG evaluados y seleccionar la mejor CpG para cada gen; las Figuras 3A-C describen el análisis ROC y las tres mejores CpG para cada gen evaluado.

30

El comportamiento de cada CpG con respecto a su posición a lo largo del genoma humano se señala gráficamente mediante la herramienta graficadora de metilación (http://maplab.imppc.org/methylation_plotter/) en las Figuras 3A-C.

35 Se ha creado una tabla que incluye las mejores CpG para cada gen en la que la clase 0 = donantes normales y la clase 1 = OSCC.

Utilizando esta tabla para el análisis discriminante de LDA mediante la herramienta SPSS (seleccionando las siguientes opciones: intervalo variable: 0-1; Estadísticas: valores no estandarizados; Clasificaciones: calculo a partir de la dimensión de grupo; crear una tabla resumen; Guardar puntuación discriminante). Luego, haga un análisis de la curva ROC para encontrar el umbral correcto.

40 Ejemplo para el conjunto (a):

45 El conjunto (a) implica las siguientes CpG: 241759621 para KIF1A y 1279758 para TERT. El análisis discriminante lineal calculó los siguientes coeficientes:

Coefficientes de la función discriminante canónica

	Función
	1
KIF1A	5,338
TERT	0,707
(Constante)	-0,808

50
$$y = -0,808 + 5,338 * (\text{valor de KIF 1A}) + 0,707 * (1 - \text{valor hTERT})$$

Mediante el análisis de la curva ROC (Figura 4A), los valores más grandes de las variables de resultado de la prueba indican una evidencia más fuerte de un estado real positivo. El estado real positivo es 1.

55 Área bajo la curva

Variable o variables del resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Área	Error estándar ^a	Sig. asintótico ^b	Intervalo de confianza asintótico del 95%
------	-----------------------------	------------------------------	---

			Límite inferior	Límite superior
0,924	0,033	0,000	0,861	0,988

Por lo tanto, cuando el valor de y excede el umbral de -0,5306765, la muestra debe considerarse positiva y relacionada con la presencia de HG-SIL u OSCC.

5 Ejemplo para el conjunto (b)

El conjunto (b) implica los siguientes genes: *LRRTM1* (CpG 80531799), *PARP15* (CpG 122296586), *ZAP70* (CpG 98340854), *MIR193* (CpG 29886944), *MIR296* (CpG 57392374), *TERT* (CpG 1279758).

10 En este caso, el análisis discriminante lineal calculó los siguientes coeficientes:

Coeficientes de función discriminante canónica

	Función
	1
LRRTM1	2,422
PARP15	1,335
ZAP70	3,571
MIR193	3,073
MIR296	4,730
TERT	0,165
(Constante)	-3,470

15 Entonces el algoritmo será el siguiente: $y = -3,470 + 2,422 * (\text{valor de LRRTM1}) + 3,073 * (\text{valor de miR193}) + 4,730 * (1 - \text{valor de miR296}) + 0,165 * (1 - \text{valor de hTERT}) + 1,335 * (\text{valor de PARP15}) + 3,571 * (\text{valor de ZAP70})$.
Mediante el análisis de la curva ROC (Figura 4B), los valores más grandes de las variables del resultado de la prueba indican una evidencia más fuerte de un estado real positivo. El estado real positivo es 1.

Área bajo la curva

Variable o variables de resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Área	Error estándar ^a	Sig. asintótico ^b	Intervalo de confianza asintótico del 95%	
			Límite inferior	Límite superior
0,975	0,019	0,000	0,937	1,000
Bajo el supuesto no paramétrico b. Hipótesis nula: área verdadera = 0,5				

Coordenadas de la curva

30 Variable o variables de resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Positivo si es mayor o igual a ^a	Sensibilidad	1 - Especificidad
-0,1924607	0,971	0,077

35 El valor de corte más pequeño es el valor mínimo observado de la prueba menos 1, y el valor de corte más grande es el valor máximo observado de la prueba más 1. Todos los demás valores de corte son los promedios de dos valores de la prueba observados ordenados consecutivos.

Por tanto, cuando el valor y supera el umbral de 0,1924607, la muestra se considera positiva y está relacionada con la presencia de HG-SIL u OSCC.

40 Ejemplo para el conjunto (c)

El conjunto (c) implica los siguientes genes: *KIF1A* (CpG 241759621), *LRRTM1* (CpG 80531799), *ZAP70* (CpG 98340854), *MIR193* (CpG 29886944), *MIR296* (CpG 57392374), *TERT* (CpG 1279758).

45 Coeficientes de función discriminante canónica

	Función
	1
KIF1A	1,744
LRRTM1	1,844
ZAP70	3,656
MIR193	2,852
MIR296	5,441
TERT	0,228
(Constante)	-3,474

Mediante el análisis de la curva ROC (Figura 5A), los valores más grandes de las variables del resultado de la prueba indican una evidencia más fuerte de un estado real positivo. El estado real positivo es 1.

Área bajo la curva

[Variable o variables del resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Área	Error estándar ^a	Sig. asintótico ^b	Intervalo de confianza asintótico del 95%	
			Límite inferior	Límite superior
0,973	0,019	0,000	0,936	1,000

$y = -3,474 + 1,844 * (\text{valor de LRRTM1}) + 2,852 * (\text{valor de miR193}) + 5,441 * (1 - \text{valor de miR296}) + 3,656 * (\text{valor de ZAP70}) + 1,744 * (\text{valor de KIF1A}) + 0,228 * (1 - \text{valor de hTERT})$

Por lo tanto, cuando el valor de y supera el umbral de -0,2394214 (sensibilidad: 0,943; 1 - especificidad: 0,077), la muestra debe considerarse positiva y relacionada con la presencia de HG-SIL o OSCC.

Ejemplo para el conjunto (d)

El conjunto (d) implica los siguientes genes: *KIF1A* (CpG 241759621), *LRRTM1* (CpG 80531799), *FLI1* (CpG 128564158), *LINC00599* (CpG 9760888), *PARP15* (CpG 122296586), *ZAP70* (CpG 98340854), *MIR193* (CpG 29886944), *MIR296* (CpG 57392374), *GP1BB* (CpG 19710956), *hTERT* (CpG 1279758).

Coefficientes de función discriminante canónica

	Función
	1
KIF1A	1,585
LRRTM1	1,631
ZAP70	1,996
MIR193	0,625
MIR296	8,743
TERT	-0,083
FLI1	-2,630
LINC00599	-0,443
PARP15	1,995
GP1BB	3,899
(Constante)	-3,977

Mediante el análisis de la curva ROC (Figura 5B), los valores más grandes de las variables del resultado de la prueba indican una evidencia más fuerte de un estado real positivo. El estado real positivo es 1.

Área bajo la curva

Variable o variables del resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Área	Error estándar ^a	Sig. asintótico ^b	Intervalo de confianza asintótico del 95%	
			Límite inferior	Límite superior
0,973	0,026	0,000	0,922	1,000

$y = - 3,977 - 2,630 * (\text{valor de FLI1}) + 3,899 * (1 - \text{valor de GP1BB}) + 1,631 * (\text{valor de LRRTM1}) - 0,443 * (\text{valor de LINC00599}) + 0,625 * (\text{valor de miR193}) + 8,743 * (1 - \text{valor de miR296}) + 1,995 * (\text{valor de PARP15}) + 1,996 * (\text{valor de ZAP70}) + 1,585 * (\text{valor de KIF1A}) - 0,083 * (1 - \text{valor de hTERT})$

- 5 Por lo tanto, cuando el valor de y supera el umbral de 0,6760154 (sensibilidad: 0,971; 1 - especificidad: 0,015), la muestra debe considerarse positiva y relacionada con la presencia de HG-SIL o OSCC.

Ejemplo para el conjunto (e)

- 10 El conjunto (e) implica los siguientes genes con sus CpG: 19710956 para GP1BB; 98340854 para ZAP70; 241759621 para KIF1A; 1279758 para TERT; 80531799 para LRRTM1; 131781167 para NTM; 29886944 para MIR193; 122296586 para PARP15; 182322902 para ITGA4; 128564158 para FLI1; 9760888 para LINC0059; 15342885 para EPHX3; y 57392374 para MIR296.

- 15 Coeficientes de la función discriminante canónica

	Función
	1
KIF1A	1,789
ZAP70	1,983
GP1BB	3,801
LRRTM1	1,428
TERT	-0,265
PARP15	2,338
FLI1	-3,471
NTM	-0,637
LINC00599	-0,036
EPHX3	-0,415
ITGA4	2,276
MIR193	0,894
MIR296	7,938
(Constante)	-3,954

Mediante el análisis de la curva ROC (Figura 5D), los valores más grandes de las variables del resultado de la prueba indican una evidencia más fuerte de un estado real positivo. El estado real positivo es 1.

- 20 Área bajo la curva

Variable o variables del resultado de la prueba: puntuaciones discriminantes de la función 1 para el análisis 1

Área	Error estándar ^a	Sig. asintótico ^b	Intervalo de confianza del 95% asintótico	
			Límite inferior	Límite superior
0,973	0,026	0,000	0,921	1,000

- 25 $y = - 3,954 - 0,415 * (\text{valor de EPHX3}) - 3,471 * (\text{valor de FLI1}) + 3,801 * (1 - \text{valor de GP1BB}) + 2,276 * (\text{valor de ITGA4}) + 1,428 * (\text{valor de LRRTM1}) - 0,036 * (\text{valor de LINC00599}) + 0,894 * (\text{valor de miR193}) + 7,938 * (1 - \text{valor de miR296}) - 0,637 * (\text{valor de NTM}) + 2,338 * (\text{valor de PARP15}) + 1,983 * (\text{valor de ZAP70}) + 1,789 * (\text{valor de KIF1A}) - 0,265 * (1 - \text{valor de hTERT})$

- 30 Por lo tanto, cuando el valor de y supera el umbral de 0,8732193 (sensibilidad: 0,971; 1 - especificidad: 0), la muestra debe considerarse positiva y relacionada con la presencia de HG-SIL u OSCC.

Pruebas de intervalo múltiple usando la puntuación de la combinación "e":

- 35 Se ha calculado una prueba de intervalo múltiple de Duncan que evalúa la puntuación final de pacientes de diferentes grupos (65 donantes normales, 29 con OSCC, 6 con HGSIL, 30 de mucosa distante en pacientes con OSCC) usando la combinación "e" como sigue:

- 40 $y = - 3,954 - 0,415 * (\text{valor de EPHX3}) - 3,471 * (\text{valor de FLI1}) + 3,801 * (1 - \text{valor de GP1BB}) + 2,276 * (\text{valor de ITGA4}) + 1,428 * (\text{valor de LRRTM1}) - 0,036 * (\text{valor de LINC00599}) + 0,894 * (\text{valor de miR193}) + 7,938 * (1 - \text{valor de miR296}) - 0,637 * (\text{valor de NTM}) + 2,338 * (\text{valor de PARP15}) + 1,983 * (\text{valor de ZAP70}) + 1,789 * (\text{valor de KIF1A}) - 0,265 * (1 - \text{valor de hTERT})$

Esta prueba identifica una diferencia estadística entre el grupo OSCC frente a donantes sanos y frente a la mucosa contralateral. Además, la prueba de intervalo múltiple muestra una diferencia estadística entre el grupo de HGSIL y los donantes sanos y la mucosa contralateral:

5 La Tabla II que indica para cada grupo la media de la puntuación:

patología	Recuento	Media	Grupos homogéneos
Donantes sanos	65	-1,586	C
Mucosa distante en pacientes con OSCC	30	-0,444399	B
OSCC	29	2,66856	A
HG SIL	6	2,71892	A

Tabla III que indica la significancia para cada comparación de grupos:

Contraste	Sig.	Diferencia	Límites +/-
OSCC - HG SIL		-0,0503633	0,983538
OSCC - mucosa distante en pacientes OSCC	*	3,11296	0,571081
OSCC - donantes sanos	*	4,25456	0,489711
HG SIL - mucosa distante en pacientes con OSCC	*	3,16332	0,980724
HG SIL - donantes sanos	*	4,30492	0,935683
Mucosa distante en pacientes con OSCC-donantes sanos	*	1,1416	0,484034
* denota una diferencia estadísticamente significativa.			

10 Con estos datos, la prueba de intervalo múltiple para puntuaciones generadas por LDA (combinación "e") identifica una diferencia significativa entre el grupo de OSCC y los donantes sanos y la mucosa contralateral. Además, la prueba de intervalo múltiple mostró una diferencia significativa entre el grupo HGSIL y los donantes sanos y la mucosa contralateral, lo que demuestra que OSCC y HGSIL tienen un comportamiento epigenético diferente al de los donantes normales y la mucosa contralateral normal.

15 Validación de algoritmos en una cohorte independiente

20 Teniendo en cuenta el conjunto de datos de validación, se detectaron todos los donantes normales por debajo del valor umbral, así como un fibroma oral y 12 de 14 con OLP; por el contrario, los dos OLP restantes, todos los PVL y todos los OSCC fueron positivos como se esperaba.

La prueba de Kruskal-Wallis (Figura 5) mostró una diferencia significativa entre grupos para las puntuaciones generadas con el análisis discriminante lineal (LDA) ($T = 78,8587$, $p < 0,05$).

REIVINDICACIONES

1. Método para detectar y/o para el pronóstico de un carcinoma de células escamosas de cabeza y cuello (HNSCC) que comprende las etapas de:

- (i) purificar el ADN genómico de una muestra aislada obtenida de un individuo;
 - (ii) tratar el ADN genómico con bisulfito, preferiblemente bisulfito de sodio;
 - (iii) amplificar el ADN genómico, después de la etapa (ii) utilizando al menos un par de cebadores, permitiendo dicho par de cebadores la amplificación de al menos una secuencia de la región del ADN genómico de cada uno de los genes comprendidos en el siguiente conjunto de genes:
(b) *LRRTM1*, *PARP15*, *ZAP70*, *miR193*, *miR296*, *hTERT* y, opcionalmente, de al menos uno de los siguientes genes: *FLI1*, *LINC00599*, *GP1BB*, *EPHX3*, *ITGA4*, *NTM*, *KIF1A*, en los que dicha secuencia comprende al menos una CpG y/o una isla de CpG;
 - (iv) medir el nivel de metilación de la CpG y/o isla de CpG que cae en la secuencia amplificada de acuerdo con la etapa (ii) y que pertenece al siguiente conjunto de genes:
LRRTM1, *PARP15*, *ZAP70*, *miR193*, *miR296*, *hTERT* y, opcionalmente, de al menos uno de los siguientes genes: *FLI1*, *LINC00599*, *GP1BB*, *EPHX3*, *ITGA4*, *NTM*, *KIF1A*;
- en los que la hipermetilación de al menos una CpG y/o una isla CpG que cae en la secuencia de al menos uno de los siguientes genes de los conjuntos: *EPHX3*, *KIF1A*, *LRRTM1*, *FLI1*, *ITGA4*, *LINC00599*, *NTM*, *PARP15*, *ZAP70*, *miR193* y/o la hipometilación de la al menos una CpG y/o una isla de CpG que cae en la secuencia de al menos uno de los siguientes genes de los conjuntos: *miR296*, *GP1BB*, *hTERT* es (son) indicativos de una muestra positiva para HNSCC en la que la hipermetilación y/o la hipometilación se (son) evalúan en comparación con el nivel de metilación medido en una muestra aislada de un individuo sano.

2. Método de acuerdo con la reivindicación 1, en el que el HNSCC es cualquier tumor que se desarrolle en la boca, nariz y garganta, preferiblemente es un carcinoma oral de células escamosas (OSCC), o un precursor de OSCC, preferiblemente una lesión intraepitelial escamosa de alto grado (HG -SIL), o una lesión premaligna oral (OPML), o un campo de cancerización, en el que dicho OPML se selecciona preferiblemente de: leucoplasia, eritroplasia, lesión palatina por fumar cigarrillos al revés, liquen plano oral, fibrosis submucosa oral (SMF) y lupus eritematoso discoide.

3. Método de acuerdo con la reivindicación 1 o 2, en el que la muestra biológica se selecciona entre: un cepillado de mucosa oral o hisopo bucal de ADN, plasma, saliva o cualquier muestra que contenga células epiteliales preferiblemente de la cavidad oral.

4. Método de acuerdo con una cualquiera de las reivindicaciones 1-3, en el que dicha región de ADN genómico se selecciona de: una región reguladora de un gen, preferiblemente el promotor del gen, la 5'UTR, la 3'UTR y las regiones codificantes, preferiblemente exones y/o intrones y/u orillas.

5. Método de acuerdo con una cualquiera de las reivindicaciones 1-4, en el que la etapa de amplificación (iii) implica 1) una primera etapa de amplificación específica de plantilla para amplificar al menos una secuencia de la región de ADN genómico; y 2) una segunda etapa de amplificación para codificar con barras los productos de amplificación específicos de la plantilla y la identificación de la muestra.

6. Método de acuerdo con una cualquiera de las reivindicaciones 1-5, en la que la etapa de amplificación específica de la plantilla se realiza usando al menos un par de cebadores de fusión, comprendiendo dicho par de cebadores de fusión un cebador directo y otro inverso para amplificar la región de interés del ADN genómico, en el que cada cebador del par es un cebador de fusión que comprende una porción específica de la plantilla que se hibrida con al menos una porción de dicha al menos una secuencia de la región genómica y una segunda porción que es una secuencia universal.

7. Método de acuerdo con una cualquiera de las reivindicaciones 1-6, en el que la segunda etapa de amplificación implica el uso de al menos un par de cebadores de fusión, comprendiendo dicho par de cebadores de fusión un cebador directo y uno inverso en el que cada cebador del par comprende, comenzando desde el extremo 5' de la secuencia del cebador: 1) una secuencia adaptadora específica para la plataforma de secuenciación utilizada, preferiblemente dicha plataforma de secuenciación es Illumina y/o IonTorrent; 2) una secuencia de código de barras específica de la muestra, preferiblemente un identificador múltiplex (MID) o un índice; y 3) una etiqueta enlazadora universal complementaria a la secuencia universal del par cebador usado en la primera etapa de amplificación.

8. Método de acuerdo con una cualquiera de las reivindicaciones 1-7, en el que al menos un par cebador de la etapa (iii) o el par de cebadores de fusión de la reivindicación 6 se selecciona de:

- SEQ ID NO: 1 y 2 y/o SEQ ID NO: 27 y 28 para amplificar *PARP15* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
- SEQ ID NO: 3 y 4 y/o SEQ ID NO: 29 y 30 para amplificar *ITGA4* preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o

- SEQ ID NO: 5 y 6 y/o SEQ ID NO: 31 y 32 para amplificar NMT preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 7 y 8 y/o SEQ ID NO: 33 y 34 para amplificar ZAP70 preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - 5 – SEQ ID NO: 9 y 10 y/o SEQ ID NO: 35 y 36 para amplificar MIR193A preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 11 y 12 y/o SEQ ID NO: 37 y 38 para amplificar EPHX3 preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - 10 – SEQ ID NO: 13 y 14 y/o SEQ ID NO: 39 y 40 para amplificar KIF1A preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 15 y 16 y/o SEQ ID NO: 41 y 42 para amplificar LINC00599, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 17 y 18 y/o SEQ ID NO: 43 y 44 para amplificar FLI1, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - 15 – SEQ ID NO: 19 y 20 y/o SEQ ID NO: 45 y 46 para amplificar GP1BB, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 21 y 22 y/o SEQ ID NO: 47 y 48 para amplificar MIR296, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - 20 – SEQ ID NO: 23 y 24 y/o SEQ ID NO: 49 y 50 para amplificar LRRTM1, preferiblemente con las plataformas IonTorrent e Illumina respectivamente; y/o
 - SEQ ID NO: 25 y 26 y/o SEQ ID NO: 51 y 52 para amplificar TERT humano, preferiblemente con las plataformas IonTorrent e Illumina respectivamente.
9. Método de acuerdo con cualquiera de las reivindicaciones 1-8, en el que al menos una secuencia de la región de ADN genómico se selecciona de SEQ ID NO: 53-65 en la que:
- SEQ ID NO: 53 es la región de PARP15 que abarca desde 122296564 hasta 122296725 de la hebra + del cromosoma 3 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 54 es la región de ITGA4 que se abarca desde 182322887 hasta 182323053 de la hebra + del cromosoma 2 con respecto a las coordenadas de hg19; y/o
 - 30 – SEQ ID NO: 55 es la región de NTM que abarca desde 131781042 hasta 131781185 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 56 es la región de ZAP70 que abarca desde 98340750 hasta 98340885 de la hebra + del cromosoma 12 con respecto a las coordenadas de hg19; y/o
 - 35 – SEQ ID NO: 57 es la región de MIR193A que abarca desde 29886860 hasta 29887068 de la hebra + del cromosoma 17 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 58 es la región de EPHX3 que abarca desde 15342826 hasta 15343049 de la hebra - del cromosoma 19 con respecto a las coordenadas de hg19; y/o
 - 40 – SEQ ID NO: 59 es la región de KIF1A que abarca desde 241759586 hasta 241759727 de la hebra - del cromosoma 2 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 60 es la región de LINC00599 que abarca desde 9760739 hasta 9760890 de la hebra - del cromosoma 8 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 61 es la región de FLI1 que abarca desde 128564020 hasta 128564160 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19; y/o
 - 45 – SEQ ID NO: 62 es la región de GP1BB que abarca desde 19710829 hasta 19710960 de la hebra + del cromosoma 22 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 63 es la región de MIR296 que abarca desde 57392355 hasta 57392545 de la hebra - del cromosoma 20 con respecto a las coordenadas de hg19; y/o
 - 50 – SEQ ID NO: 64 es la región de LRRTM1 que abarca desde 80531676 hasta 80531807 de la hebra - del cromosoma 2 con respecto a las coordenadas de hg19; y/o
 - SEQ ID NO: 65 es la región de TERT humano (hTERT) que abarca desde 1279743 hasta 1279851 de la hebra - del cromosoma 21 con respecto a las coordenadas de hg19.
10. Método de acuerdo con una cualquiera de las reivindicaciones 1-9, en la que al menos una CpG/isla de CpG se selecciona de:
- Una CpG/isla de CpG que cae en la región genómica de PARP15 SEQ ID NO: 53 que tiene coordenadas desde 122296564 hasta 122296725 de la hebra + del cromosoma 3 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 122296564, 122296586, 122296613, 122296617, 122296623, 122296630, 122296637, 122296645, 122296649, 122296656, 122296663, 122296665, 122296671, 122296677, 122296680, 122296692, 122296708, 122296710, 122296717, 122296723 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 122296586; y/o
 - Una CpG/isla de CpG que cae en la región genómica de ITGA4 SEQ ID NO: 54 que tiene coordenadas desde 182322887 hasta 182323053 de la hebra + del cromosoma 2 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 182322890, 182322898, 182322902, 182322908, 182322915, 182322937,

- 182322956, 182322962, 182322980, 182322985, 182323004, 182323016, 182323024, 182323028 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 182322902; y/o
- 5 - Una CpG/isla de CpG que cae en la región genómica de NTM SEQ ID NO: 55 que tiene las coordenadas desde 131781042 hasta 131781185 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 131781046, 131781058, 131781060, 131781080, 131781082, 131781084, 131781098, 131781110, 131781113, 131781122, 131781144, 131781151, 131781158, 131781167, 131781183 y combinación de las mismas, más preferiblemente la coordenada de la CpG/isla de CpG es 131781082 y/o 131781167; o y/o
- 10 - Una CpG/isla de CpG que cae en la región genómica de ZAP70 SEQ ID NO: 56 que tiene coordenadas desde 98340750 hasta 98340885 de la hebra + del cromosoma 12 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 98340751, 98340762, 98340764, 98340766, 98340772, 98340775, 98340794, 98340801, 98340809, 98340813, 98340816, 98340823, 98340828, 98340834, 98340844, 98340854, 98340859, 98340869, 98340877, 98340881 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es y/o 98340751 y/o 98340854; y/o
- 15 - Una CpG/isla de CpG que cae en la región genómica de MIR193A SEQ ID NO: 57 que tiene coordenadas desde 29886860 hasta 29887068 de la hebra + del cromosoma 17 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 29886868, 29886870, 29886875, 29886881, 29886885, 29886890, 29886903, 29886910, 29886913, 29886936, 29886939, 29886944, 29886949, 29886951, 29886964, 29886967, 29886971, 29886983, 29886992, 29886995, 29887001, 29887008, 29887015, 29887045, 29887049, 29887063 y combinaciones de las mismas, más preferiblemente, la coordenada de CpG/isla de CpG es 29886870 y/o 29886944; y/o
- 20 - Una CpG/isla de CpG que cae en la región genómica de EPHX3 SEQ ID NO: 58 que tiene coordenadas desde 15342826 hasta 15343049 de la hebra - del cromosoma 19 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 15342895, 15342889, 15342885, 15342875, 15342869, 15342866, 25 15342864, 15342855, 15342851, 15342848, 15342845, 15342840, 15342838, 15342831, 15343002, 15343000, 15342991, 15342988, 15342985, 15342983, 15342976, 15342962, 15342960, 15342934, 15342928, 15342918, 15342916, 15342913, 15342906 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 15342840 y/o 15342885; y/o
- 30 - Una CpG/isla de CpG que cae en la región genómica de KIF1A SEQ ID NO: 59 que tiene coordenadas desde 241759586 hasta 241759727 de la hebra - del cromosoma 2 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 241759726, 241759722, 241759719, 241759717, 241759715, 241759702, 241759697, 241759688, 241759686, 241759681, 241759676, 241759671, 241759665, 241759663, 241759656, 241759651, 241759649, 241759646, 241759643, 241759640, 241759632, 241759621, 241759618, 241759614, 241759602, 241759596, 241759588 y combinaciones de las mismas, más preferiblemente, la coordenada de CpG/isla 35 de CpG es 241759686 y/o 241759681 y/o 241759651 y/o 241759621; y/o
- Una CpG/isla de CpG que cae en la región genómica de LINC00599 SEQ ID NO: 60 que tiene las coordenadas desde 9760739 hasta 9760890 de la hebra - del cromosoma 8 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 9760888, 9760881, 9760878, 9760870, 9760856, 9760852, 9760848, 9760839, 9760832, 9760822, 9760820, 9760815, 9760811, 9760809, 9760805, 9760803, 40 9760777, 9760764, 9760759, 9760752 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 9760888; y/o
- Una CpG/isla de CpG que cae en la región genómica de FLI1 SEQ ID NO: 61 que tiene coordenadas desde 128564020 hasta 128564160 de la hebra + del cromosoma 11 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 128564022, 128564033, 128564051, 128564063, 45 128564065, 128564068, 128564078, 128564089, 128564106, 128564134, 128564158 y combinaciones de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 128564158 y/o 128564051; y/o
- Una CpG/isla de CpG que cae en la región genómica de GP1BB SEQ ID NO: 62 que tiene coordenadas desde 19710829 hasta 19710960 de la hebra + del cromosoma 22 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 19710839, 19710842, 19710847, 19710855, 19710864, 19710880, 50 19710890, 19710902, 19710909, 19710916, 19710919, 19710937, 19710944, 19710946, 19710948, 19710952, 19710954, 19710956 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 19710956 y/o 197109946; y/o
- Una CpG/isla de CpG que cae en la región genómica de MIR296 SEQ ID NO: 63 que tiene coordenadas desde 57392355 hasta 57392545 de la hebra - del cromosoma 20 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 57392543, 57392538, 57392536, 57392517, 57392501, 57392469, 57392452, 57392440, 57392437, 57392430, 57392419, 57392394, 57392386, 57392374, 57392363 y combinación 55 de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 57392374 y/o 57392363; y/o
- Una CpG/isla de CpG que cae en la región genómica de LRRMT1 SEQ ID NO: 64 que tiene coordenadas desde 80531676 hasta 80531807 de la hebra - del cromosoma 2 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 80531806, 80531803, 80531799, 80531796, 80531789, 80531786, 60 80531783, 80531780, 80531776, 80531763, 80531759, 80531756, 80531752, 80531749, 80531745, 80531742, 80531738, 80531719, 80531702, 80531697, 80531695, 80531677 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 80531799 y/o 80531697; y/o
- Una CpG/isla de CpG que cae en la región genómica de TERT (hTERT) SEQ ID NO: 65 que tiene coordenadas desde 1279743 hasta 1279851 de la hebra - del cromosoma 21 con respecto a las coordenadas de hg19, preferiblemente al menos una coordenada seleccionada de: 1279847, 1279838, 1279829, 1279775, 1279758, 65

1279746 y combinación de las mismas, más preferiblemente la coordenada de CpG/isla de CpG es 1279758 y/o 1279775 y/o 1279747.

5 11. Método de acuerdo con una cualquiera de las reivindicaciones 1-9, en el que después de la etapa de amplificación (iii), las secuencias amplificadas se secuencian, preferiblemente utilizando una plataforma de secuenciación de próxima generación (NGS), más preferiblemente dicha plataforma NGS es Illumina y/o IonTorrent.

10 12. Método de acuerdo con una cualquiera de las reivindicaciones 1-11, en el que las secuencias se filtran aplicando un filtro de error que varía desde Q30 para al menos el 90% de las bases para la plataforma Illumina y preferiblemente Q20 o menos para otras plataformas.

15 13. Método de acuerdo con una cualquiera de las reivindicaciones 1-12, donde la etapa (iv) se realiza utilizando una herramienta que permite una medición cuantitativa del nivel de metilación en base a comparar una secuencia genómica humana completa y la secuencia después del tratamiento con bisulfito, preferiblemente dicha herramienta. siendo seleccionado entre: BISMARCK, BS-Seeeker2, BISMA, BSPAT y QUMA.

20 14. Método de acuerdo con una cualquiera de las reivindicaciones 1-13, en el que se usa un Análisis de discriminación lineal multiclase para medir el nivel de metilación total de al menos una CpG/isla de CpG de al menos una región de ADN genómico de más de un gen de interés, preferiblemente el conjunto de genes de la reivindicación 1.

15. Método de acuerdo con una cualquiera de las reivindicaciones 1-14, en el que después del Análisis de discriminación lineal se obtiene una curva ROC para definir un umbral capaz de determinar la muestra positiva para HNSCC, preferiblemente la muestra positiva para OSCC/HG-SIL.

Fig. 1-A

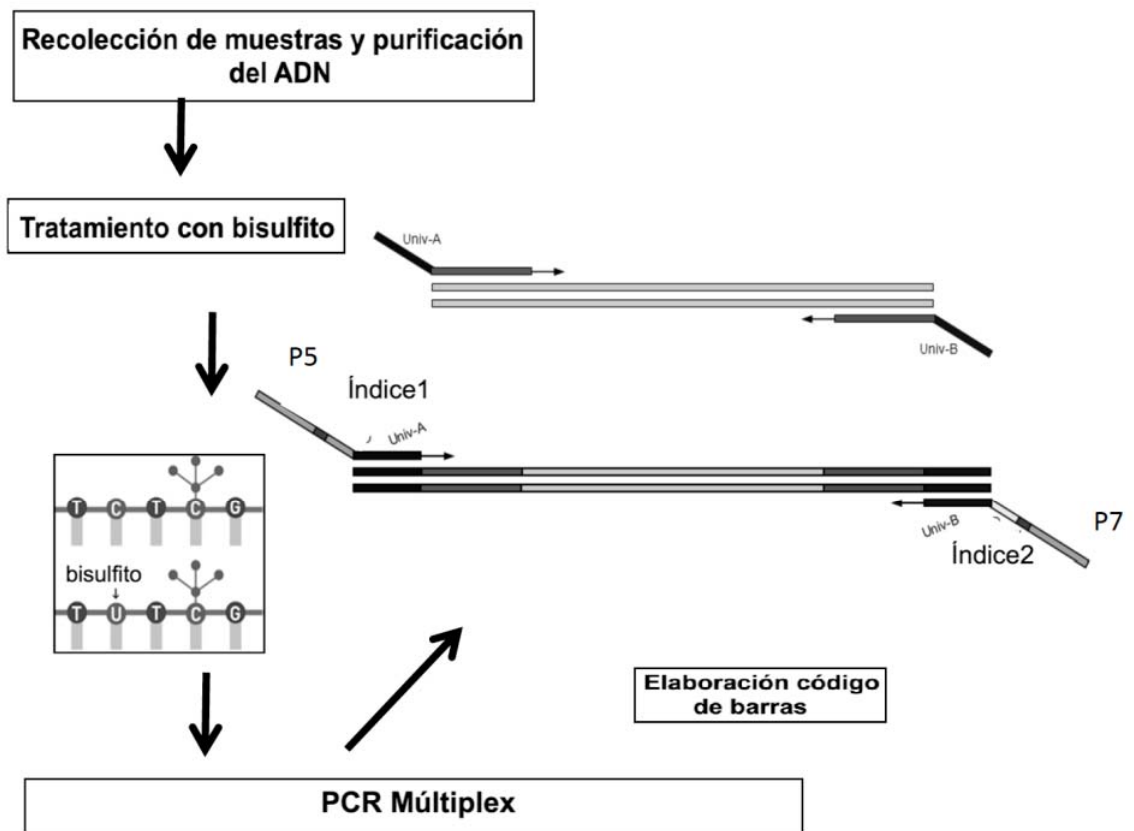


Fig. 1-B

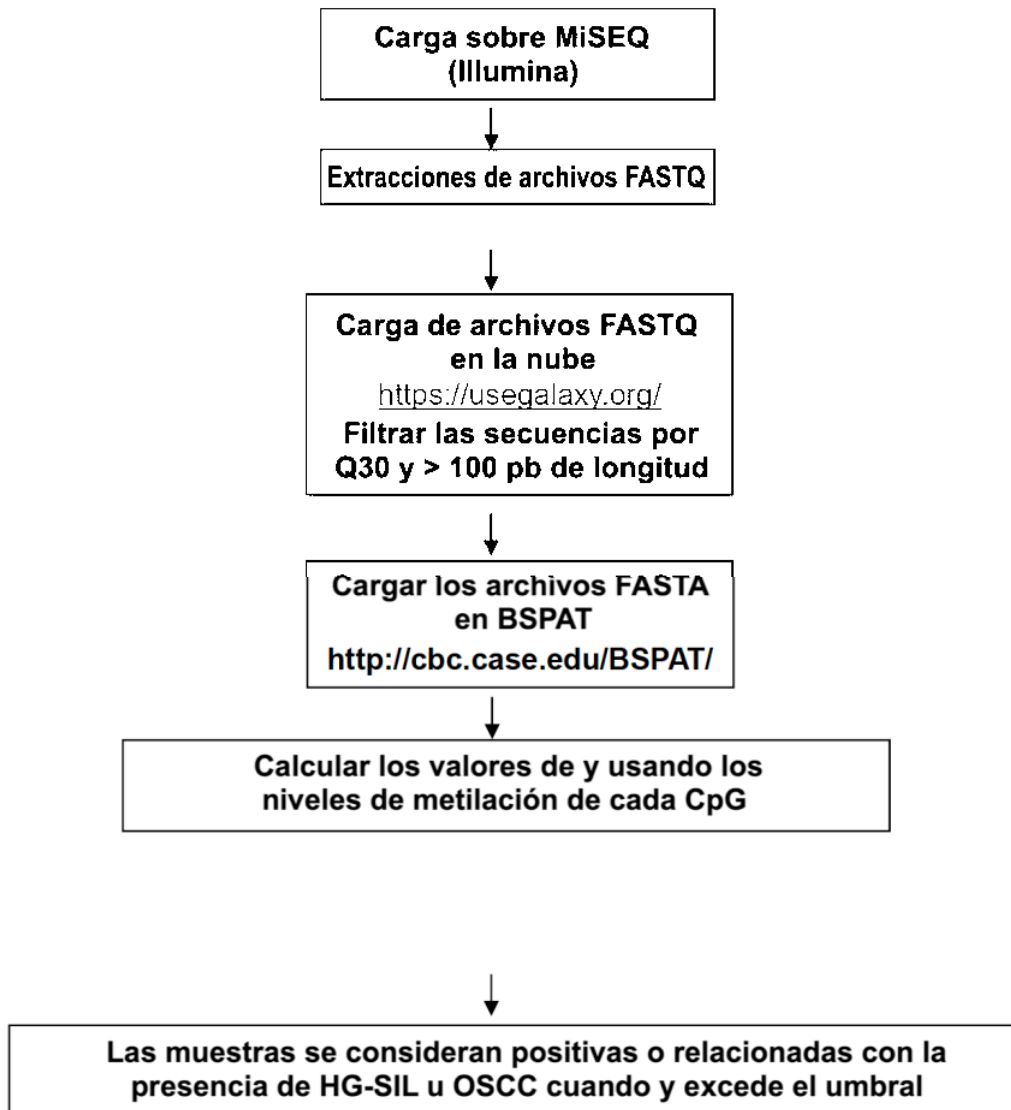


Fig. 2A

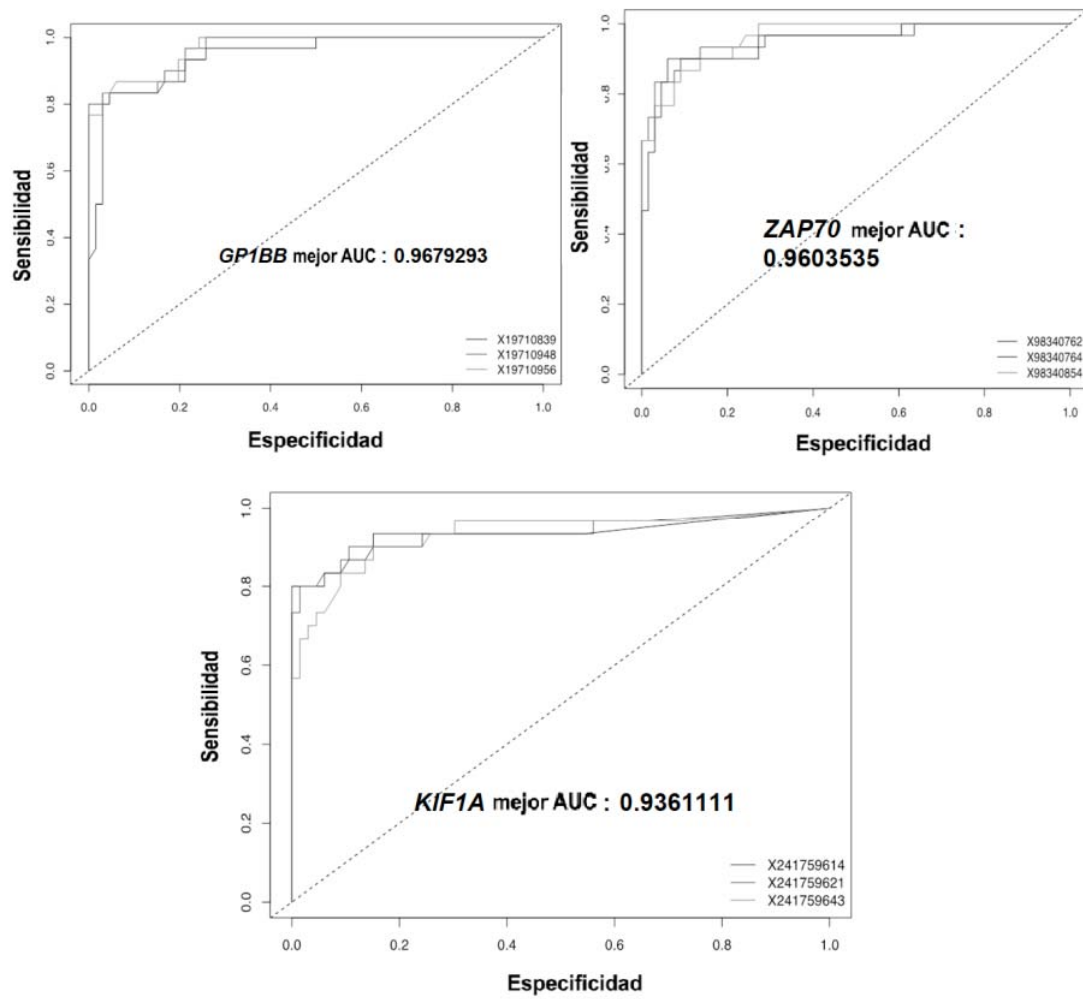


Fig. 2B

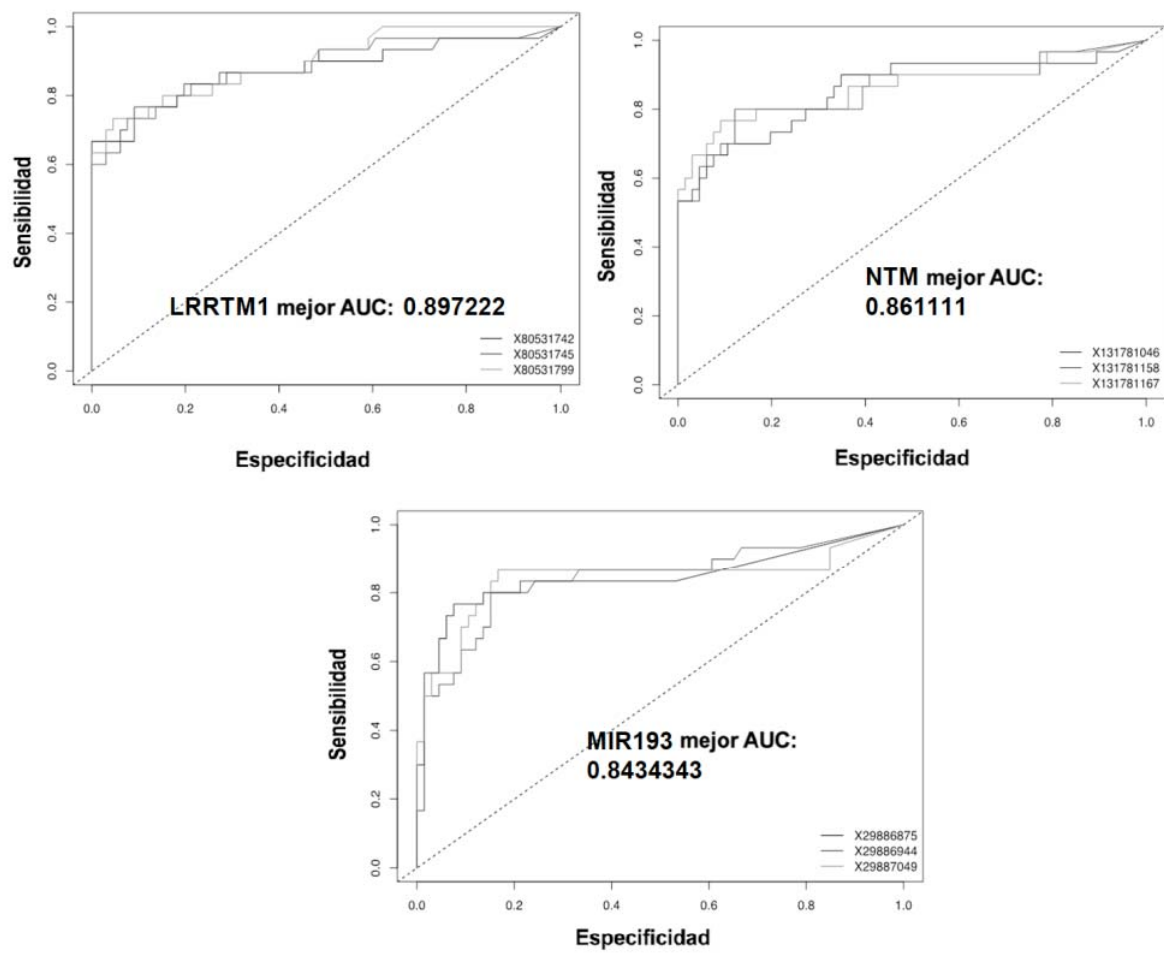


Fig. 2D

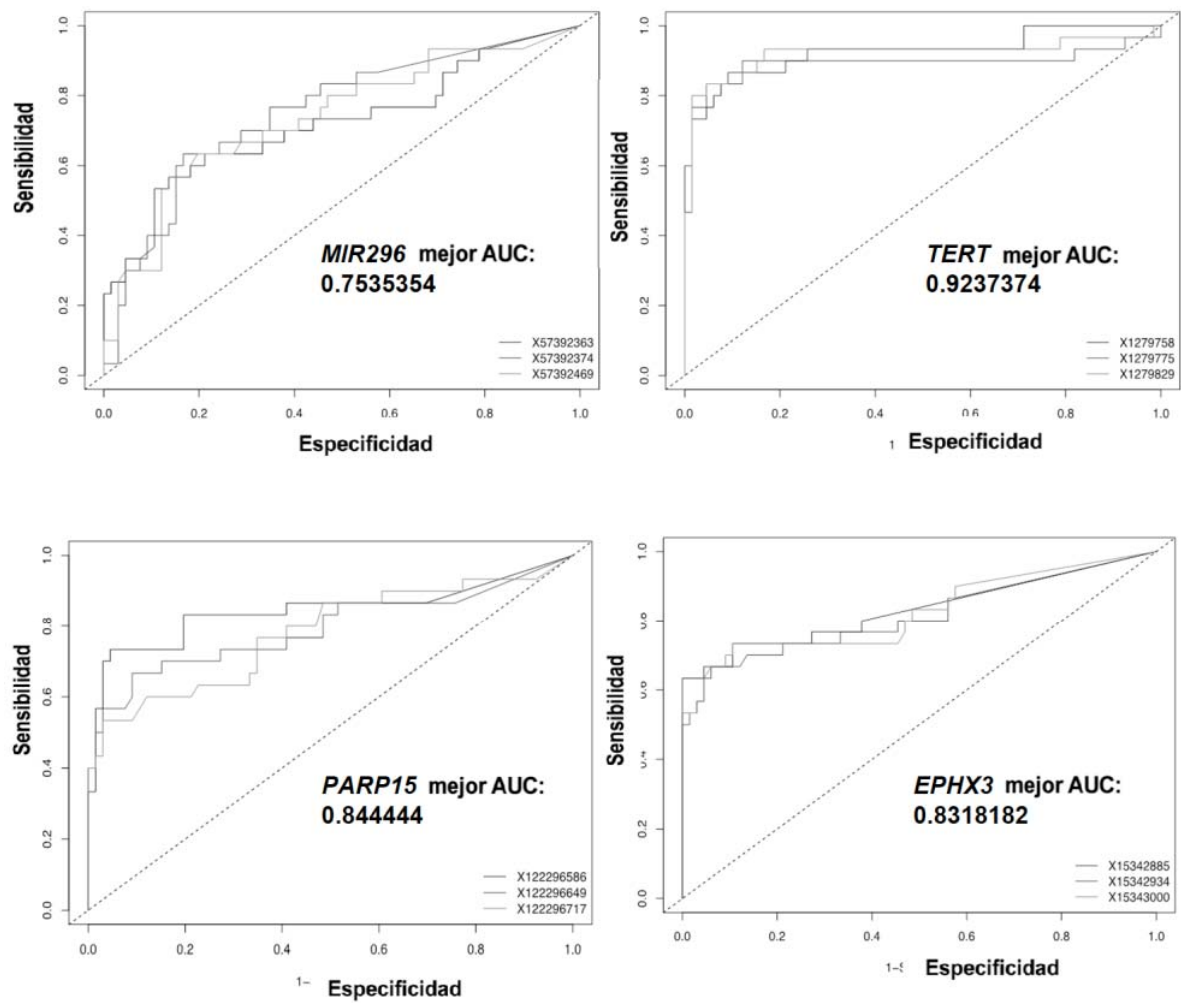


Fig. 2D

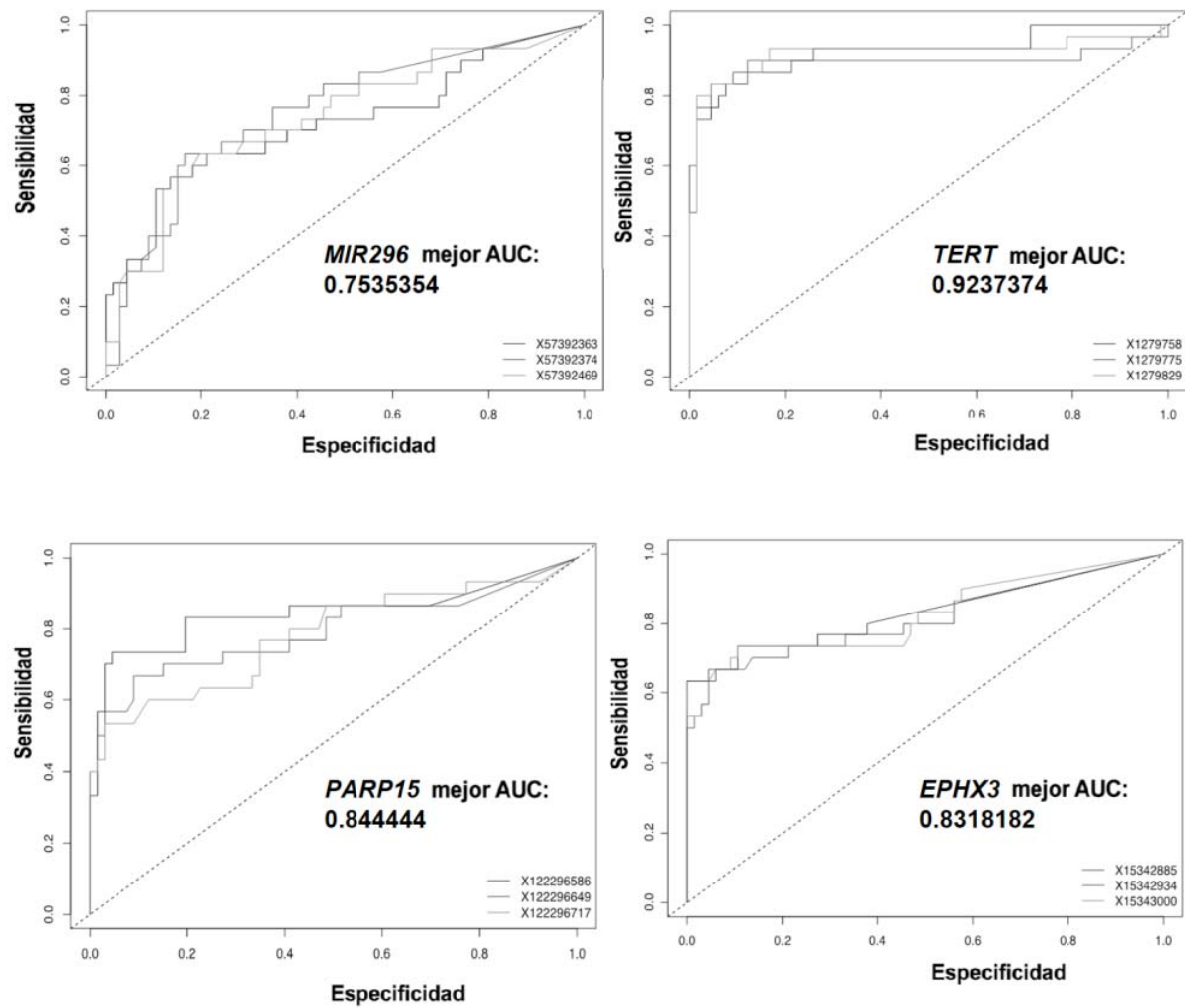


Fig. 3A

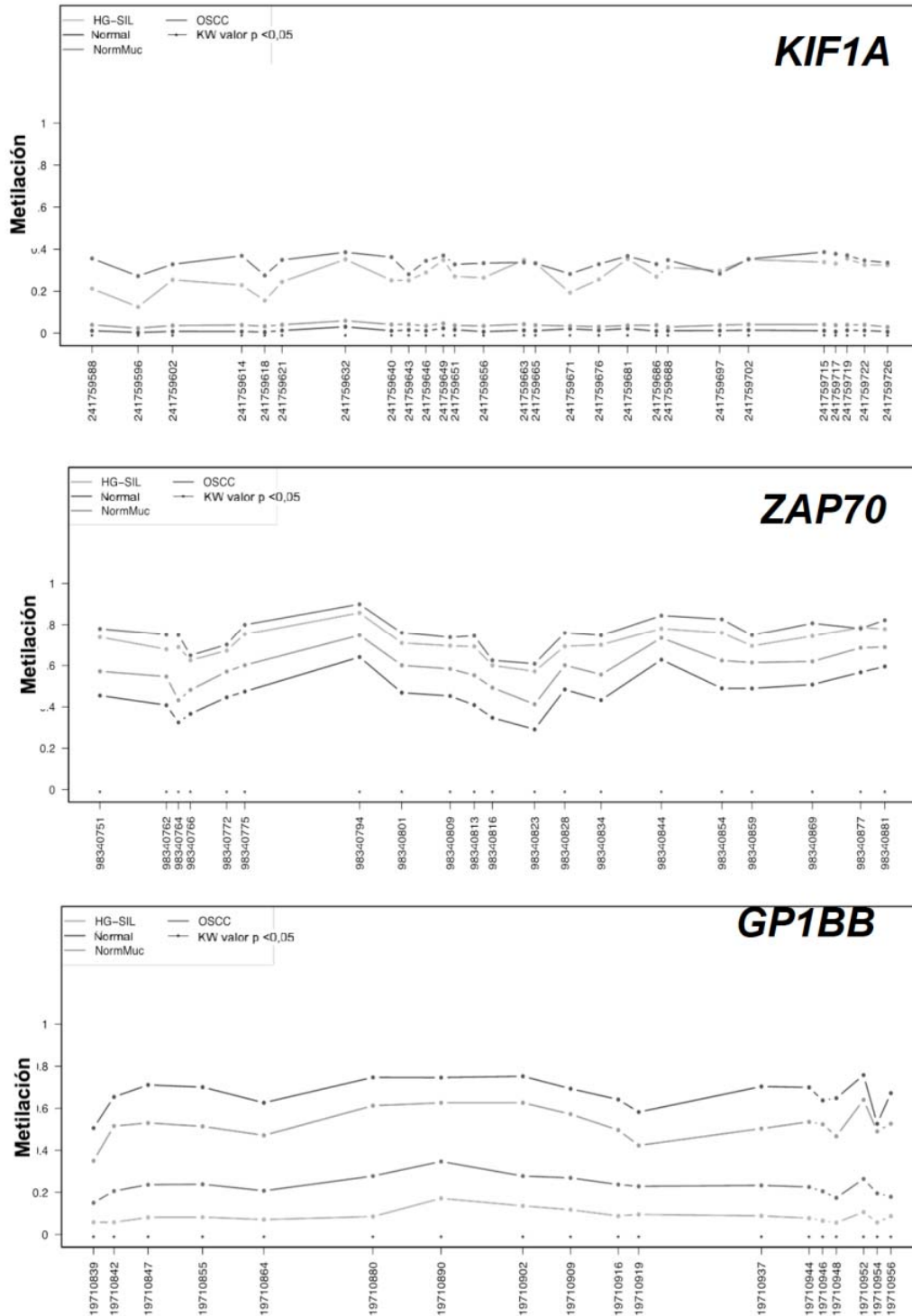


Fig. 3B

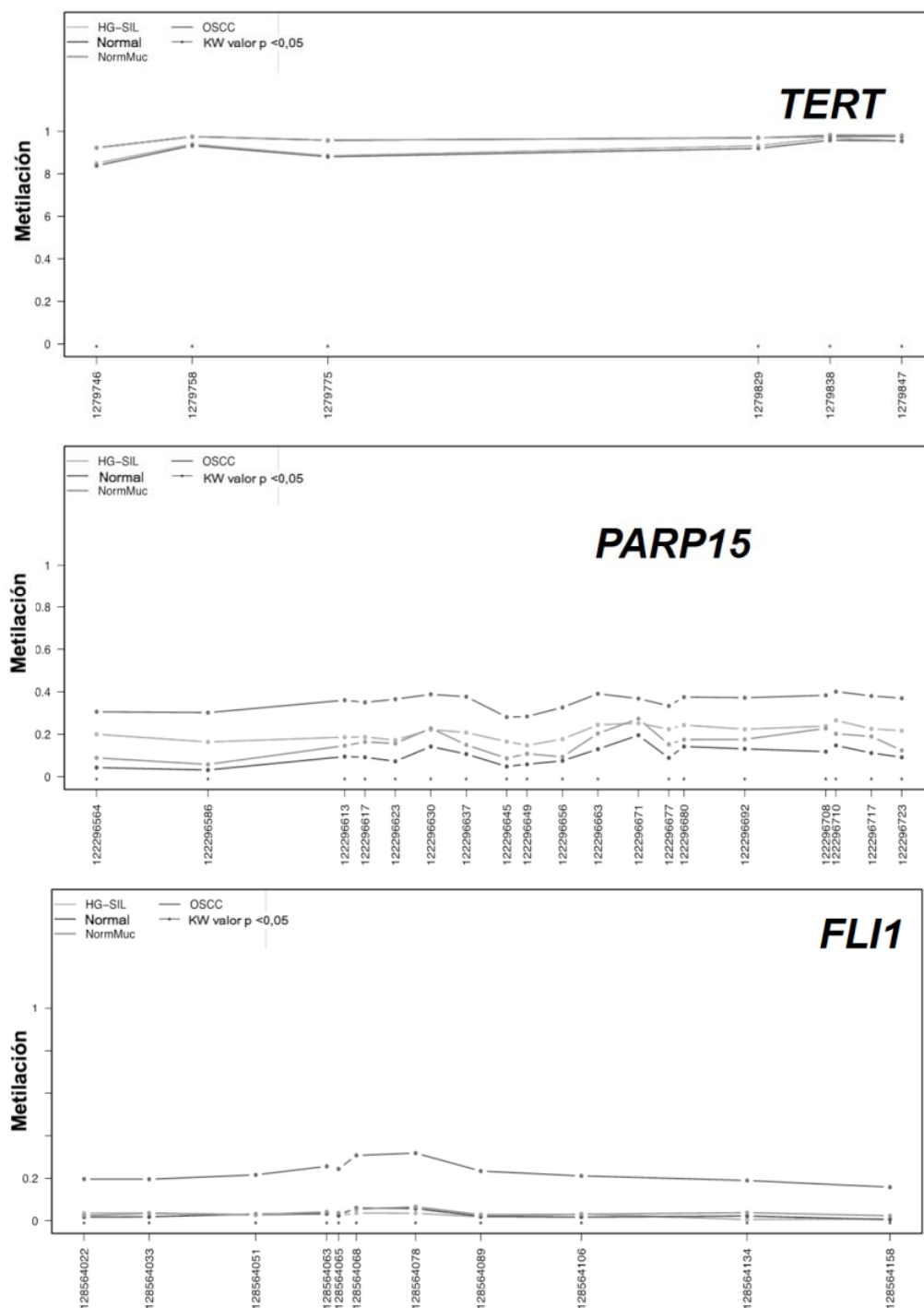


Fig. 3C

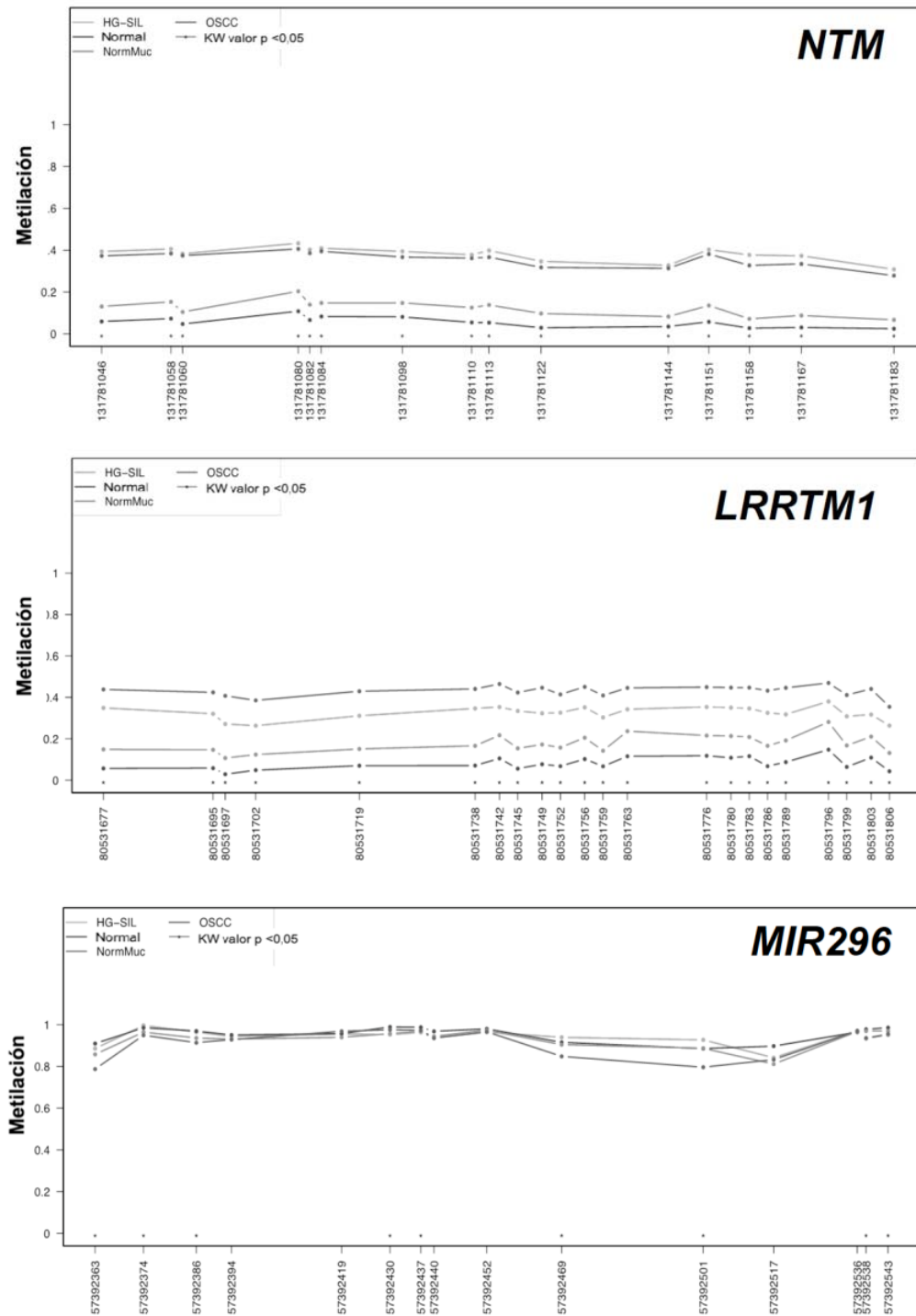


Fig. 3D

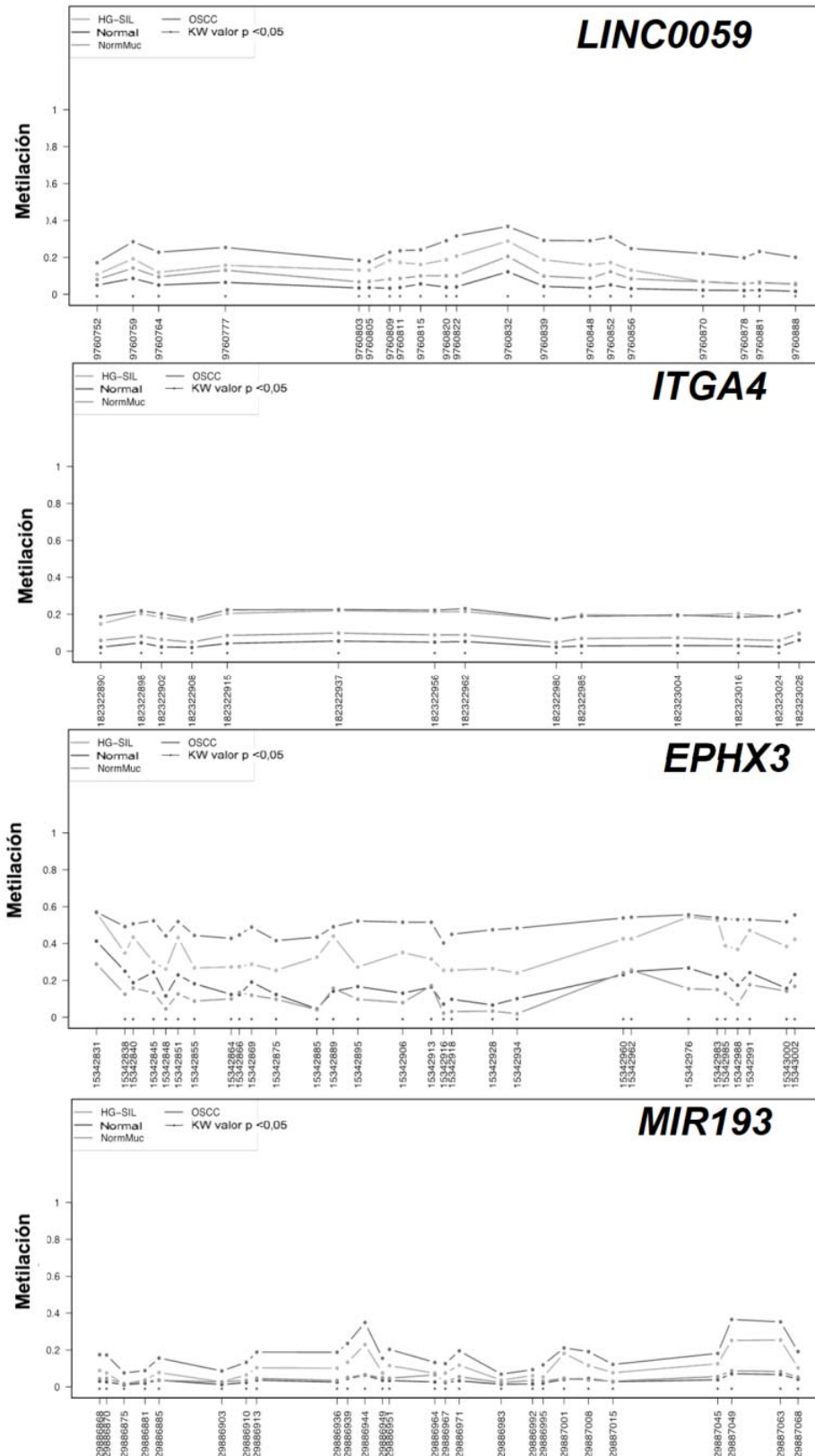


Fig. 4A

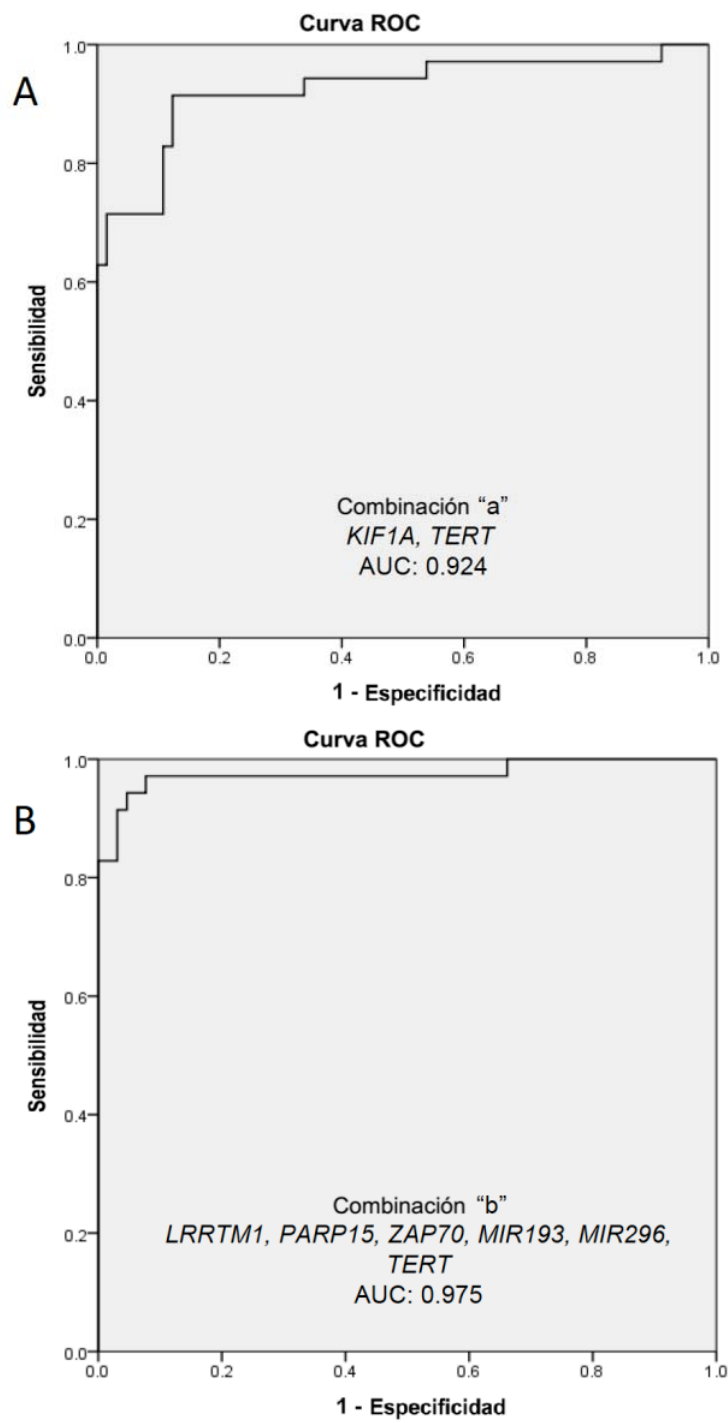


Fig. 5

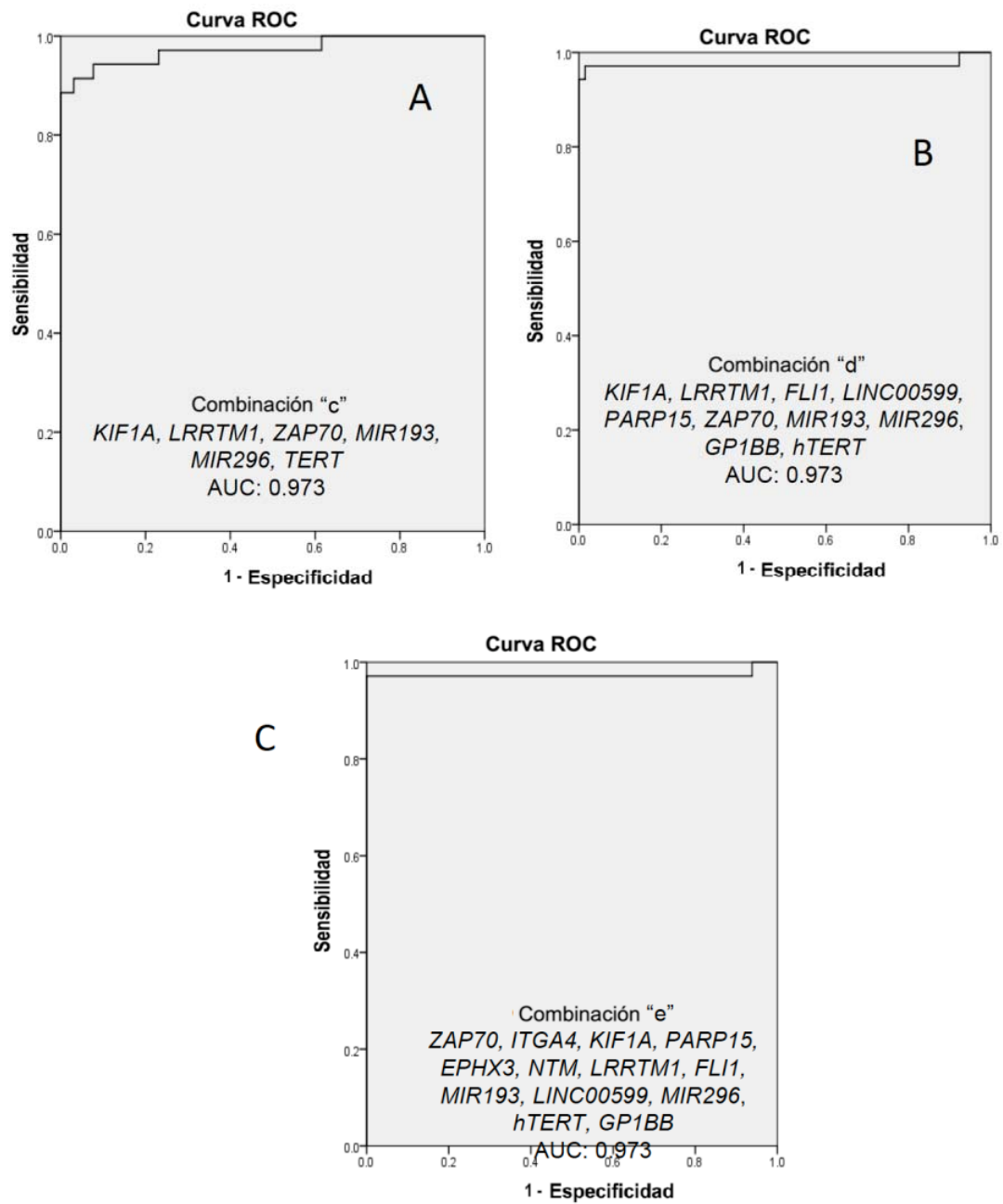


Fig. 6

