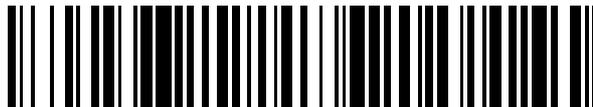


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 719 412**

51 Int. Cl.:

**C12N 15/09** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.02.2015 PCT/IB2015/050990**

87 Fecha y número de publicación internacional: **13.08.2015 WO15118513**

96 Fecha de presentación y número de la solicitud europea: **10.02.2015 E 15709555 (5)**

97 Fecha y número de publicación de la concesión europea: **19.12.2018 EP 3105324**

54 Título: **Control de sistemas NGS y métodos que los incluyen**

30 Prioridad:

**10.02.2014 GB 201402249**

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

**10.07.2019**

73 Titular/es:

**VELA OPERATIONS SINGAPORE PTE. LTD.  
(100.0%)  
50 Science Park Road, No. 05-07 The Kendall  
117406 Singapore, SG**

72 Inventor/es:

**WU, MENGCHU**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 719 412 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**

Control de sistemas NGS y métodos que los incluyen

**Campo de la invención**

5 La presente invención se refiere a controles para los ensayos de secuenciación de nueva generación. Dichos controles pueden servir como controles positivos y negativos y como control de extracción, respectivamente. La presente solicitud describe los correspondientes plásmidos, kits, sus usos y un método que implica el uso de controles de la presente invención.

**Antecedentes de la invención**

10 El uso de secuenciación de ácido nucleico se ha vuelto una herramienta fundamental en muchas áreas diagnósticas de la medicina moderna. En particular, los test genéticos basados en Secuenciación de Nueva Generación (NGS, del inglés "Next Generation Sequencing" están ganando aceptación rápidamente en el diagnóstico clínico. Un ejemplo es el área de oncología, donde la secuenciación de ácido nucleico se emplea para identificar, p.ej., si hay presentes mutaciones oncogénicas en un gen o si hay presentes en un genoma traslocalizaciones que inducen y/o indican cáncer. Adicionalmente, la secuenciación de ácido nucleico se emplea para detectar si un microorganismo patógeno  
15 (tal como p.ej., una bacteria o un virus) está presente en una muestra clínica, p.ej., en una muestra de tejido o una muestra de sangre procedente de un paciente humano. En este último método, se detectan las secuencias de ácido nucleico que no se observan en un sujeto humano, sino solo en el microorganismo.

20 Típicamente, la etapa de secuenciación real viene precedida de una reacción de amplificación para amplificar el ácido nucleico que va a ser secuenciado. Dicha reacción de amplificación se lleva a cabo normalmente mediante una reacción PCR; de esta manera, se usan cebadores específicos en la reacción de PCR, que se hibridan a secuencias por encima y por debajo de la secuencia que va a ser amplificada (es decir, los cebadores flanquean la secuencia que va a ser amplificada).

25 Debido al progreso en los últimos años, se han automatizado muchas etapas durante los métodos diagnósticos, que incluyen el flujo de trabajo NGS. Por ejemplo, la extracción de ácidos nucleicos de una muestra clínica, la reacción de amplificación, así como la propia reacción de secuenciación se llevan a cabo de forma automática. Los métodos NGS pueden dar como resultado la identificación y la determinación de secuencia de una secuencia diana, lo que puede indicar la presencia o la ausencia de una mutación oncogénica, o la presencia o ausencia de un ácido nucleico derivado de patógeno.

30 Sin embargo, no se puede descartar en un ensayo como el descrito anteriormente que determinadas etapas del ensayo se hayan llevado a cabo de forma inapropiada. Mientras que los ensayos convencionales que se basan en PCR (p.ej., qPCR) típicamente implican el uso de controles positivos y negativos para asegurar la calidad de los resultados, los ensayos basados en NGS generalmente no contienen dichos controles. Esto puede producir problemas en la correcta interpretación de los resultados de secuenciación.

35 Un ejemplo típico de ensayo llevado a cabo inadecuadamente es que una secuencia diana no sea detectada debido a que la etapa de extracción no se haya llevado a cabo de forma apropiada. Así, podría haber realmente presente un oncogén o un microorganismo infeccioso en la muestra aunque el resultado, es decir el producto de la reacción de secuenciación, sería que la reacción de secuenciación ha fallado, o que el análisis subsiguiente a la reacción de secuenciación clasificará dicho evento como resultado "inválido". Cuando la muestra está contaminada con ADN, el resultado del análisis de la reacción de secuenciación será que el resultado es un falso positivo o un falso negativo).  
40 Los controles proporcionados en la presente memoria, que incluyen un control negativo, permiten a los usuarios de métodos de secuenciación de alta capacidad (también denominado Secuenciación de Nueva Generación) reconocer los resultados potenciales de falso positivo y falso negativo. Los controles de sistema (SC) proporcionados en la presente memoria serán de ayuda para solventar fallos de secuenciación. Aunque el fallo de secuenciación no supone un riesgo para los pacientes, sigue siendo importante debido a los elevados costes asociados a los procesos NGS.  
45 Adicionalmente, el análisis de los datos de secuenciación obtenidos en NGS que incluyen controles de sistema apropiados según se proporcionan en la presente memoria es más fiable, más rápido y más preciso.

50 A fin de excluir una etapa de extracción inapropiada, se pueden usar controles de extracción. El control de extracción puede comprender la adición de las denominadas células de referencia intactas que comprenden una secuencia de ácido nucleico de control que habitualmente difiere de la secuencia que va a ser detectada. Dichas células típicamente son añadidas ("dopadas") a la muestra antes de la etapa de extracción automatizada (normalmente mediante la adición de las células en una concentración específica al tampón de lisis usado durante la extracción). La detección de una secuencia del control de extracción tras completar el ensayo indica que la etapa de extracción se ha llevado a cabo efectivamente de forma apropiada.

55 También se debe asegurar que las etapas de amplificación y secuenciación real funcionan en las condiciones aplicadas y que el límite de detección de los ácidos nucleicos diana es suficiente para realizar una afirmación clara en relación a la calidad y cantidad del resultado, es decir, para establecer que una secuencia específica está presente o ausente en una muestra con un determinado nivel de precisión.

Adicionalmente, se debe descartar cuidadosamente que ácidos nucleicos contaminantes conduzcan a resultados de falso positivo o falso negativo. Como se ha mencionado anteriormente, la presencia de ácidos nucleicos contaminantes puede, en el análisis de los datos NGS, dar la impresión de que una determinada secuencia diana (o secuencia de ácido nucleico de interés) está presente en la muestra de un paciente. Por otra parte, cuando la reacción de secuenciación NGS pretende determinar la presencia o ausencia de una secuencia diana natural y de una secuencia variante (p.ej., en un oncogén), la contaminación puede hacer que los resultados de secuenciación globales se desplacen a una posición en donde la versión natural o la variante estén sobre-representadas o infra-representadas. Esto también puede cambiar el resultado global del diagnóstico basado en NGS, p.ej., cuando se fija un nivel de expectación predeterminado.

Resumiendo lo anterior, los controles descritos en la presente memoria, que pueden designarse también como “Control de Sistemas” (SC), son requeridos para la evaluación diagnóstica que implica NGS, particularmente cuando los ensayos deben cumplir determinados requisitos reguladores de las respectivas autoridades reguladoras, tal como la FDA. Al mismo tiempo, dichos controles deberían estar disponibles fácilmente, es decir, poder obtenerse sin emplear técnicas de biología molecular complejas, deberían poder producirse a bajo coste, sin constituir un peligro para el entorno o el personal que los maneje. Adicionalmente, la presencia de controles no debería afectar negativamente a la sensibilidad de los ensayos diagnósticos. Éstos y otros objetivos se cumplen con la presente invención.

L. Jiang et al., Genome Research, Vol. 21, Nº 9 (2011), páginas 1543-1551, describe patrones sintéticos dopados para secuenciación de ARN.

Quail, M.A. et al., Genomics, Vol. 15 (2014) describe muestras dopadas y medios de diferenciación de códigos de barras para secuenciación Illumina.

Kovalic, D. et al., Plant Genome, Vol. 5, Nº3 (2012) describe el uso de NGS y bioinformática de análisis de secuencia de unión para lograr la caracterización molecular de cultivos.

Los inventores de la presente invención han tenido éxito, entre otros, en proporcionar métodos que implican controles de sistema para reacciones de Secuenciación de Nueva Generación (NGS) que pueden usarse como controles positivos y negativos, así como de extracción, en una reacción NGS.

Antes de pasar a describir con más detalle algunas de las realizaciones de la presente invención, se introducen las siguientes definiciones.

### Definiciones

Tal como se usa en la especificación y en las reivindicaciones, las formas singulares de “un” y “una” también incluyen los correspondientes plurales a menos que el contexto claramente dicte lo contrario.

Es necesario entender que el término “que comprende” no es limitativo. Para los propósitos de la presente invención, el término “que consiste en” se considera que es una realización preferida del término “que comprende”. Si de aquí en adelante se define que un grupo comprende al menos un número determinado de realizaciones, esto también pretende abarcar un grupo que preferiblemente consiste solo en dichas realizaciones. Cuando un control de ácido nucleico o una secuencia de ácido nucleico diana comprende un ácido nucleico (secuencia), significa que es posible la presencia de residuos adicionales en la secuencia que comprende la otra, p.ej., en secuencias de ácido nucleico de control que comprende plásmido.

Las expresiones “método diagnóstico basado en NGS específico de secuencia de ácido nucleico diana” o “método de detección de ácido nucleico diana basado en NGS” indican que dichos métodos comprenden Secuenciación de Nueva Generación, pero que son posibles etapas adicionales, p.ej., la extracción y/o purificación de ácidos nucleicos de una muestra dada, cualquier modificación del protocolo de NGS, siempre que la reacción de secuenciación se lleve a cabo según métodos conocidos en la técnica, y también etapas post-secuenciación, p.ej., el análisis, presentación de los datos obtenidos durante la secuenciación y el análisis o presentación de resultados proporcionados al final del análisis.

Cuando una secuencia de control y una secuencia de ácido nucleico son diferentes, esto significa que el ácido nucleico difiere al menos en un residuo de ácido nucleico. Frecuentemente, dichas secuencias proceden de genes diferentes, muy frecuentemente derivan de genes y organismos diferentes. Se prefiere que los ácidos nucleicos de control deriven de fuentes, o que hayan sido modificados de tal modo, que dichos controles no presenten ningún riesgo para el entorno, en particular para el personal del laboratorio que trate con ellos.

El término “detectar la presencia” tal como se usa en la presente memoria debe entenderse con el significado de “detectar la presencia o ausencia”.

Como se ha mencionado en el método reivindicado en la presente solicitud, una muestra que va a ser analizada en un método basado en secuenciación de alta capacidad o Secuenciación de Nueva Generación, es decir, NGS, comprende *potencialmente* un ácido nucleico que comprende una secuencia diana. Preferiblemente, el método de secuenciación de alta capacidad es un método de secuenciación de semiconductor iónico.

En el contexto de la presente invención el término “ácido nucleico” se refiere a un polímero natural de desoxirribonucleótido o de ribonucleótido, en forma de cadena sencilla o doble. El ácido nucleico en particular puede ser ADN de cadena doble y ARN de cadena sencilla.

5 El término “secuencia” tal como se usa en la presente memoria se refiere a la presencia secuencial de las bases en un polímero de desoxirribonucleótido o de ribonucleótido, en donde una base que se da en un polímero de desoxirribonucleótido se selecciona del grupo que consiste en A, T, G y C y una base que se da en un polímero de ribonucleótido se selecciona del grupo que consiste en A, U, G y C. Por tanto, una secuencia de bases en un polímero de desoxirribonucleótido puede ser, p.ej., GGAAGCAAGCCT, mientras que una secuencia de bases en un polímero de ribonucleótido puede ser, p.ej., GGAAUCGAU.

10 Tal como se usa en la presente memoria, el término “muestra” se refiere a cualquier muestra biológica procedente de un sujeto humano o veterinario que puede ser evaluada para determinar la presencia de un ácido nucleico que comprende una secuencia diana. Las muestras pueden incluir tejidos obtenidos de cualquier órgano, tal como, por ejemplo, tejido pulmonar; y fluidos obtenidos de cualquier órgano tal como, por ejemplo, sangre, plasma, suero, fluido linfático, fluido sinovial, fluido cerebroespinal, fluido amniótico, sangre de cordón amniótico, lágrimas, saliva, y lavados nasofaríngeos. Como se ha enumerado anteriormente, las muestras también pueden derivar de una región específica del cuerpo, p.ej., el tracto respiratorio; las muestras del tracto respiratorio incluyen frotis de garganta, lavados de garganta, frotis nasales y especímenes del tracto respiratorio inferior.

15 La muestra puede derivar de un sujeto humano o veterinario. Por consiguiente, un “paciente” puede ser un sujeto humano o veterinario. Si se hace referencia a una “muestra clínica”, esto indica que la muestra procede de un paciente que se sospecha que porta un ácido nucleico que comprende una secuencia diana.

20 Una muestra de control que se usa en métodos según la presente invención se prepara mediante la adición de ácidos nucleicos de control a una muestra, p.ej., cuando la muestra es material FFPE de un órgano humano, la muestra de control comprende material FFPE de la misma fuente, pero adicionalmente comprende un ácido nucleico de control añadido a la muestra. El ácido nucleico de control puede añadirse a la muestra de control antes o después de la extracción, dependiendo de la etapa que debería controlarse, p.ej., la extracción y/o la secuenciación, etc.

25 Tal como se usa en la presente memoria, el término “amplificación” se refiere a procedimientos mediados por enzimas que son capaces de producir miles de millones de copias del ácido nucleico diana. Los ejemplos de procedimientos de amplificación de diana mediados por enzimas conocidos en la técnica incluyen la PCR.

30 “Extraer ácidos nucleicos” significa que cualesquier ácidos nucleicos presentes en un vial son aislados de cualquier ruido de fondo celular, particularmente son aislados de células intactas o tejidos. Preferiblemente, los ácidos nucleicos también son lavados durante el proceso y, opcionalmente, son concentrados. Después de la extracción, todos los restos celulares o de tejidos no relacionados con los ácidos nucleicos han sido eliminados. Los métodos de extracción típicos pueden incluir el uso de tampón de lisis hipotónico, calor y/o detergentes, y son conocidos por el especialista en la técnica.

35 El término “secuenciación” se usa en la presente memoria con su significado habitual en biología molecular. De esta manera, se determina la existencia secuencial exacta de bases en una secuencia de ácido nucleico.

El término “microorganismo” tal como se usa en la presente memoria se usa en su más amplio significado. Así, un microorganismo puede ser cualquier tipo de bacteria, arquea, protozoo, hongo y virus. Se menciona explícitamente que los virus entran dentro de la definición de un “microorganismo” tal como se usa en la presente memoria.

40 El término “oncogén” se usa en la presente memoria en su significado común en biología molecular y oncología, respectivamente. Por tanto, existen, p.ej., mutaciones conocidas en genes, que convierten un gen “normal o natural” en oncogénico, es decir, inductor de cáncer; los ejemplos a este respecto son mutaciones que hacen que las quinasas sean constitucionalmente activas de tal modo que constantemente se señalizan señales específicas (p.ej., señales inductoras de crecimiento) y se inician los correspondientes procesos. “Oncogenes” tal como se usa en la presente memoria también se refiere a traslocalizaciones intra- o inter-cromosómicas que dan como resultado también en situaciones inductoras de cáncer.

Una “secuencia diana” tal como se refiere en la presente memoria es una secuencia en el ácido nucleico, cuya presencia es detectada en los métodos según la presente invención; una “secuencia diana” es característica del ácido nucleico específico, cuya presencia es detectada.

45 Tal como se usa en la presente memoria, el ácido nucleico que está siendo secuenciado se denomina ácido nucleico diana (o “diana”). Los ácidos nucleicos diana incluyen, aunque sin limitación, ADN tal como, aunque sin limitación, ADN genómico, ADN mitocondrial, ADNc y otros similares, y ARN tal como, aunque sin limitación, ARNm, miARN, y otros similares. El ácido nucleico diana puede derivar de cualquier fuente que incluye fuentes naturales o fuentes sintéticas. Los ácidos nucleicos pueden ser productos de PCR, cósmidos, plásmidos, miembros o especies de bibliotecas naturales o sintéticas, y otros similares. La invención no pretende estar limitada en este aspecto. El ácido nucleico puede proceder de fuentes animales o patógenas, incluyendo, sin limitación, mamíferos tales como humanos, y microbios tales como bacterias, virus, hongos, parásitos y micobacterias. En algunas realizaciones, el ácido nucleico

no es un ácido nucleico vírico. El ácido nucleico diana puede obtenerse a partir de cualquier fluido o tejido corporal, incluyendo, aunque sin limitación, sangre, saliva, fluido cerebroespinal ("CSF"), piel, pelo, orina, heces, y mucus. El ácido nucleico diana también puede derivar, sin limitación, de una muestra ambiental (tal como una muestra de agua), una muestra de comida, o una muestra forense, la muestra puede ser una muestra fresca (p.ej., material de biopsia sometido directamente a extracción de ácido nucleico), o una muestra que haya sido tratada para permitir el almacenamiento, p.ej., una muestra que ha sido fijada en formalina y/o embebida en parafina (muestras FFPE, del inglés "formalin-fixed and/or paraffin-embedded"). Los métodos según la presente invención son particularmente adecuados para muestras clínicas, p.ej. muestras FFPE.

De forma general, los ácidos nucleicos diana están ligados a secuencias en uno o ambos de los extremos 5' y 3'. Estas secuencias adaptadoras comprenden sitios de cebador de secuenciación (es decir, sitios en los cuales se hibridará el cebador de secuenciación) para ser usados en los métodos de secuenciación de la invención.

En algunas realizaciones, las dianas sometidas a amplificación, como se discute más adelante, tienen la misma o similar longitud (p.ej., una variación del 5-10% entre dianas). En algunas realizaciones, dicha variación puede mantenerse tan pequeñas como sea posible a fin de asegurar que todas las plantillas se aplican uniformemente.

Los productos amplificados pueden inmovilizarse sobre la superficie del soporte (p.ej. una superficie de vidrio) en una variedad de formas. Por ejemplo, el proceso de amplificación puede llevarse a cabo en disolución y el producto final se une a continuación a la superficie del soporte. El producto de amplificación puede unirse al soporte sólido en su extremo 5' o en su extremo 3'. La unión puede ser mediante hibridación a un ácido nucleico que está inmovilizado sobre la superficie del soporte, o puede ser mediante interacción de grupos funcionales del extremo del producto de amplificación con grupos funcionales de la superficie del soporte. Los ejemplos incluyen el uso de ADN marcado con biotina o biotina dual (Margulides et al. *Nature* 437: 376 (2005)) con superficies de soporte recubiertas con estreptavidina/avidina/neutravidina, DIG (digoxigenina) y anticuerpos anti-DIG o fragmentos de anticuerpo, fluoresceína y anticuerpos anti-fluoresceína o fragmentos de anticuerpo (Gore et al. *Nature* 442, 836-9 (2006)), o a través del uso de agentes reticulantes heterofuncionales tales como succinimidil propionato biotinilado-PEG que pueden acoplarse por ejemplo a vidrio funcionalizado con aminas y usarse para inmovilizar ADN marcado con biotina a través de sándwich de estreptavidina (es decir, una interacción ácido nucleico biotina estreptavidina/avidina/neutravidina-biotina soporte sólido).

Las plantillas se pueden referir como que están inmovilizadas aleatoriamente sobre la superficie. Esto significa que las plantillas no están colocadas sobre la superficie del soporte sólido en base a la secuencia. Sin embargo, se localizan sobre el soporte sólido de un modo que asegura que cada plantilla está rodeada por un área (y, por tanto, un volumen) que no estará ocupado por otra plantilla durante las reacciones de incorporación mediadas por polimerasa y/o durante la expresión de la plantilla. Esto es, en algunos casos, las plantillas se posicionan en la superficie a una distancia suficiente unas de otras para evitar cualquier interacción entre las plantillas.

El soporte sólido se refiere al elemento al cual se une la plantilla o sobre el que se inmoviliza, y puede estar compuesto por cualquier material, que incluye, aunque sin limitación, vidrio u otro material basado en sílice, plástico u otro material basado en polímero, sin embargo siempre que el material sea relativamente inerte para la plantilla, el cebador, la polimerasa, dNTPs, y otros componentes usados en la reacción de secuenciación y lavado. El soporte sólido puede ser rígido o no. Puede ser poroso. Puede ser continuo o no. En algunas realizaciones, el soporte sólido es un portaobjetos de vidrio. En algunas realizaciones, el soporte es una pluralidad de bolitas o partículas (tal como micropartículas) que están inmovilizadas a su vez sobre un soporte sólido. Dichas bolitas pueden ser porosas. El soporte puede ser una malla. En algunas realizaciones, el propio soporte sólido es un detector o sensor, tal como un generador de imágenes de contacto, aunque sin limitación.

Debe entenderse que una pluralidad de plantillas, tanto idénticas como diferentes, pueden unirse al soporte sólido, siempre que cada miembro de la pluralidad esté separado una distancia suficiente de los otros miembros, de tal modo que no se produzcan solapamientos entre plantillas.

Típicamente, la plantilla debe estar unida a un grupo funcional observable (o detectable) en su extremo libre. Dicho grupo funcional pretende representar el extremo libre de la plantilla y por tanto su posición y movimiento en la dirección de la fuerza indica la longitud de la plantilla. El grupo funcional observable puede ser cualquier número de grupos funcionales y la invención no está limitada por su naturaleza. La naturaleza del grupo funcional observable dictará el tipo de sensor o detector adecuado para observar (o detectar o monitorizar) cambios en la longitud de la plantilla. En algunas realizaciones importantes, el grupo funcional observable es una bolita tal como una micropartícula, e incluso más particularmente, tal como una partícula magnética.

Los grupos funcionales pueden unirse a la plantilla a través de una variedad de métodos y empleando una variedad de interacciones, que incluyen, aunque sin limitación, interacciones no covalentes tales como biotina/estreptavidina, DIG/anti-DIG, y pares de unión fluorescente/anti-fluorescente, así como interacciones covalentes, tales como las discutidas en la presente memoria en relación a la inmovilización covalente de plantillas (o cebadores) sobre superficies de soporte.

El soporte sólido es parte de una célula de flujo, o está adyacente a ella. Tal como se usa en la presente memoria,

una célula de flujo es una cámara que tiene un puerto de entrada y uno de salida a través de los cuales avanza un fluido. El soporte sólido al cual está unido la plantilla puede estar debajo, encima o al lado de la célula de flujo, dependiendo de la posición del sistema de detección usado para observar la plantilla. El soporte sólido puede ser una pared de la célula de flujo que incluye una pared de fondo, una pared lateral y una pared superior.

5 Como puede apreciarse, la secuenciación precisa y rápida de la plantilla depende de la extensión y la velocidad con la que los nucleótidos no incorporados son eliminados del sistema. De este manera, es importante una eliminación rápida y completa (o casi completa) de los nucleótidos no incorporados. El sistema de microfluidos también debe diseñarse para maximizar el lavado dando como resultado potencialmente menores volúmenes de lavado y una menor duración del lavado.

10 La eliminación de nucleótidos no incorporados también se puede ver facilitada en parte o totalmente mediante el uso de apirasa, que degrada los dNTPs no incorporados y los vuelve inadecuados para una incorporación posterior. La apirasa puede fluir libremente, añadida al tampón de lavado, e introducida en la célula de flujo una vez que la incorporación de cualquier tipo de trifosfato de nucleótido ha cesado (indicado por el cese de cualquier movimiento por encima del ruido de fondo del grupo funcional detectable del extremo de la plantilla). Alternativa o adicionalmente, la apirasa puede fijarse o inmovilizarse dentro de la célula de flujo tal como por ejemplo a la superficie del soporte sólido (al cual la plantilla también está fijada o inmovilizada). Esto puede ocurrir mediante el uso de un ligando a fin de hacer más accesible a la enzima y para eliminar cualquier impedimento estérico relativo a una estrecha proximidad a la superficie. La apirasa se puede unir a una variedad de ligandos que difieren en longitud. De este modo, la apirasa puede estar presente en una variedad de corrientes de flujo dentro de la célula de flujo, que incluyen aquellas más próximas a las paredes y aquellas más próximas, o en el centro, de las corrientes de flujo. Tal como se ha discutido anteriormente, son las corrientes de flujo próximas a las paredes las que avanzan con una baja velocidad y los dNTPs no incorporados presentes en estas corrientes de flujo tienen menos probabilidad de ser eliminados. La inclusión de apirasa en dichas corrientes de flujo debería mejorar la eliminación de estos dNTPs. Esto aumentará la probabilidad de que los cambios en la longitud de la plantilla sean resultado de la incorporación de un dNTP recién introducido en la célula de flujo, más que de un dNTP residual y no incorporado que permanece en la célula de flujo después del lavado.

En algunos aspectos de la invención, los métodos de secuenciación se refieren a secuenciación por reacciones de síntesis. Esto significa que determinar la secuencia de un primer ácido nucleico requiere la síntesis de un segundo ácido nucleico que usa el primero como plantilla. De este modo, la secuencia del segundo ácido nucleico se determina a partir del orden y el número de dNTPs incorporados, y la secuencia del primer ácido nucleico se determina como el complemento de la secuencia del primer ácido nucleico. Los métodos de la invención detectan la incorporación de dNTP mediante un cambio en la longitud de la plantilla y no observando directamente la adición del dNTP al ácido nucleico que está siendo sintetizado. Como resultado, el dNTP puede ser dNTP natural (es decir, dNTP que carece de cualquier modificación que incluye cualquier marca detectable exógena tal como un fluoróforo). Como debería quedar claro a partir de esta descripción, los métodos de secuenciación de la invención también requieren que la plantilla permanezca intacta. Algunos aspectos de la invención implican métodos de secuenciación que se describen como que tienen lugar en ausencia de fluorescencia o de un modo no fluorescente. Estas caracterizaciones significan que los métodos se pueden llevar a cabo sin detección de fluorescencia, particularmente sin detección de fluorescencia de cada dNTP incorporado. Las realizaciones de estos métodos, por tanto, pueden emplear dNTPs naturales que no han sido modificados por la adición de un fluoróforo exógeno. Estas caracterizaciones no excluyen, sin embargo, la posibilidad de que el grupo funcional observable conjugado al extremo libre de la plantilla sea el mismo fluorescente. En este último caso, los cambios de longitud de la plantilla se pueden visualizar a través de la fluorescencia del grupo funcional observable, más que de cualquier fluorescencia procedente de dNTP incorporado individualmente.

De forma similar, también debe entenderse que los métodos de secuenciación provistos en la presente memoria son capaces de detectar la incorporación de nucleótidos detectando el propio grupo funcional observable (p.ej., como es posible con un generador de imágenes de contacto CMOS). De esta manera, en algunas realizaciones, los grupos funcionales observables son detectados directamente y sin la necesidad de un evento mediado por enzima. Un ejemplo de incorporación de nucleótido detectada enzimáticamente es la pirosecuenciación acoplada a detección mediada por sulfúrilasa y luciferasa del pirofosfato inorgánico liberado. (Véase, Leamon y Rothberg, Chemical Reviews, "Cramming More Sequencing Reactions onto Microreactor Chips", 2006). Así, los aspectos de la invención son referidos como métodos no enzimáticos (o como detección de la incorporación de nucleótidos por vía no enzimática) ya que la incorporación de nucleótidos puede ser detectada en ausencia de señales generadas por enzimas.

En diversas realizaciones, un análisis de interés particular son los iones hidrógeno, y los sistemas ISFET a gran escala según la presente descripción están configurados específicamente para medir el pH. En otras realizaciones, las reacciones químicas que están siendo monitorizadas pueden referirse a procesos de síntesis de ADN, u a otros procesos químicos y/o biológicos, y los sistemas chemFET pueden configurarse específicamente para medir el pH o uno o más análisis adicionales que proporcionan información relevante relativa a un proceso químico particular de interés. En varios aspectos, los sistemas chemFET se fabrican usando tecnologías de procesamiento CMOS convencionales, y son configurados particularmente para facilitar la adquisición rápida de datos a partir del sistema entero (escaneando todos los píxeles para obtener las correspondientes señales de salida de píxel). Un sistema de secuenciación preferido es el Sistema Ion PGM, sin embargo, también se contemplan otros sistemas de secuenciación basados en la detección de protones. Por ejemplo, los sistemas de pirosecuenciación y la secuenciación-por-síntesis ilumina también son opciones. Con respecto a la detección y medida de análisis, debería apreciarse que en diversas

realizaciones discutidas con más detalle a continuación, uno o más analitos medidos mediante un sistema chemFET según la presente descripción pueden incluir cualquiera de una variedad de sustancias químicas que proporcionen información relevante en relación a un proceso químico o procesos químicos de interés (p.ej., unión de múltiples cadenas de ácido nucleico, unión de un anticuerpo a un antígeno, etc.). En algunos aspectos, la capacidad para medir los niveles o concentraciones de uno o más analitos, además de simplemente detectar la presencia de un analito, proporciona información valiosa en relación al proceso o procesos químicos. En otros aspectos, la mera detección de la presencia de un analito o analitos de interés puede proporcionar información valiosa. El método de secuenciación más preferido implica el uso del Sistema Ion Torrent's PGM, es decir, métodos de secuenciación NGS que implican secuenciación de semiconductor iónico.

En otro aspecto, la invención se refiere a un método para secuenciar ácidos nucleicos que comprende fragmentar un ácido nucleico plantilla para generar una pluralidad de ácidos nucleicos fragmentados, uniendo una cadena de cada uno de la pluralidad de ácidos nucleicos fragmentados individualmente a partículas para generar una pluralidad de partículas que tiene cada una un ácido nucleico fragmentado de cadena sencilla unida a la misma, administrando la pluralidad de partículas que tienen un ácido nucleico fragmentado de cadena sencilla unido a un sistema chemFET que tiene una cámara de reacción separada para cada sensor del área, y en donde solo una partícula es situada en cada cámara de reacción, y llevar a cabo una reacción de secuenciación simultáneamente en la pluralidad de cámaras.

La invención contempla llevar a cabo una pluralidad de diferentes reacciones de secuenciación simultáneamente dentro de la misma célula de flujo o sobre el mismo soporte sólido. Cada reacción de secuenciación produce información sobre una plantilla inmovilizada sobre el soporte sólido. El número de plantillas que pueden secuenciarse en un único experimento dependerá de la longitud esperada de la plantilla y del área de soporte sólido. Por lo tanto, dependiendo de la realización, se pueden inmovilizar sobre un soporte sólido al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900 o al menos 1000 plantillas y de este modo secuenciarse simultáneamente. En otras realizaciones adicionales, se pueden secuenciar simultáneamente 100-500, 100-750, 100-1000, 500-1000, 600-1000, 700-1000, 800-1000, 900-1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000, 5000-10000, o más plantillas.

La reacción de secuenciación se lleva a cabo incorporando dNTPs en una cadena de ácido nucleico recién sintetizada que se hibrida a la plantilla. La cadena recién sintetizada puede derivar de un cebador que está unido a la plantilla o de otra molécula a partir de la cual puede proceder la extensión mediada por polimerasa.

En un ejemplo no limitativo, la reacción de secuenciación puede comenzar poniendo en contacto las plantillas con cebadores en las condiciones que permiten su hibridación, y poner en contacto los híbridos plantilla/cebador con polimerasas. Dicha puesta en contacto puede producirse antes, durante y/o después de la inmovilización sobre el soporte sólido. En una realización importante, se produce después de la inmovilización sobre el soporte sólido.

Una vez que los cebadores y las polimerasas están unidos a la plantilla, se hacen pasar por la célula de flujo ciclos repetidos de reactivos. Cuando el flujo de reactivo contiene un nucleótido que es complementario al nucleótido de la plantilla que está en directamente por debajo del extremo 3' del cebador, la polimerasa incorporará el dNTP. Si las posiciones contiguas aguas debajo de la plantilla están ocupadas por nucleótidos idénticos (referido en la presente memoria como homopolímero), la polimerasa incorporará un número idéntico de dNTPs complementarios. Dicha incorporación cesará cuando el dNTP en flujo no sea complementario al siguiente nucleótido disponible sobre la plantilla. La cantidad de dNTP fluida y el tiempo de dicho flujo excederán respectivamente el número de bases complementarias en la plantilla y el tiempo necesario para incorporar todos los posibles dNTPs.

Cabe destacar que la incorporación de los dNTPs complementarios se produce en más de uno de los cebadores ligados. Más preferiblemente, la incorporación se produce al menos en el 10%, al menos el 25%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90% o en todos los cebadores ligados. El porcentaje de cebadores puede depender del número de copias de diana en la plantilla. Para algunas realizaciones, la incorporación se produce al menos 30, al menos 35, al menos 40, al menos 45, al menos 50, al menos 60, al menos 70, al menos 80, al menos 90, al menos 100 o más cebadores por plantilla individual. Debe entenderse que la invención contempla la incorporación de dNTPs en el máximo número de cebadores hibridados sobre una plantilla dada a fin de aumentar la relación señal-ruido a través del aumento de la magnitud del cambio de longitud que se produce (tanto si es un aumento como una reducción de longitud).

Como parte de la reacción de secuenciación, se ligará un dNTP (o "se incorporará" tal como se usa en la presente memoria) al 3' de la cadena recién sintetizada (o al extremo 3' del cebador de secuenciación en el caso del primer dNTP incorporado) si su nucleótido complementario está presente en la misma localización en el ácido nucleico plantilla. La incorporación del dNTP introducido convierte una región de cadena sencilla de la plantilla en una región de cadena doble, y esta conversión se ve reflejada entonces en un cambio de longitud de la plantilla bajo tensión. El cambio de longitud se detecta determinando y monitorizando la posición del grupo orgánico observable (p.ej., una partícula) localizado en el extremo libre de la plantilla. Por lo tanto, si la posición de la partícula no cambia después de cualquier paso de flujo dado, entonces no se han incorporado dNTPs y se puede concluir que el flujo a través de dNTP no era complementario al siguiente nucleótido disponible en la plantilla. Si se detecta un cambio en la posición del grupo funcional, entonces el flujo a través de dNTP era complementario y se incorporó en la cadena recién sintetizada. Los dNTPs pueden fluir en cualquier orden siempre que el orden sea conocido y preferiblemente se mantiene

constante a lo largo del experimento de secuenciación.

Un ciclo de secuenciación típico para algunos aspectos de la invención puede incluir el lavado de la cámara de flujo (y los pocillos) con tampón de lavado, la medida de la posición del grupo funcional observable unido al extremo del ácido nucleico plantilla, la introducción de una primera especie dNTP (p.ej., dATP) en la cámara de flujo en presencia de polimerasa, la medida de la posición del grupo funcional observable, el flujo a través de apirasa opcionalmente en tampón de lavado, el flujo a través de tampón de lavado, la introducción de una segunda especie dNTP en presencia de polimerasa, etc. Este proceso continúa hasta que los 4 dNTP (es decir, dATP, dCTP, dGTP y dTTP) han fluido a través de la cámara y se ha permitido que se incorporen en las cadenas recién sintetizadas. Este ciclo de 4 nucleótidos se puede repetir cualquier número de veces, incluyendo, aunque sin limitación, 10, 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 o más veces. El número de ciclos vendrá dado por la longitud de la diana que esté siendo secuenciada y por la necesidad de reponer los reactivos de reacción, en particular las reservas de dNTP y de tampones de lavado. De esta manera, la longitud de secuencia que puede determinarse usando los métodos de la invención puede ser de al menos 50 nucleótidos, al menos 100 nucleótidos, al menos 200 nucleótidos, al menos 300 nucleótidos, al menos 400 nucleótidos, al menos 500 nucleótidos, al menos 600 nucleótidos, al menos 700 nucleótidos, al menos 800 nucleótidos, al menos 900 nucleótidos, hasta, e incluyendo, 1000 nucleótidos, 1500 nucleótidos, 2000 nucleótidos o más nucleótidos.

Las polimerasas adecuadas pueden ser ADN polimerasas, ARN polimerasas o subunidades de las mismas, siempre que dichas subunidades sean capaces de sintetizar una nueva cadena de ácido nucleico en base a la plantilla y partiendo del cebador hibridado. Un ejemplo de una subunidad de polimerasa adecuada es la exo-versión del fragmento Klenow de ADN polimerasa I de *E. coli* que carece de actividad de exonucleasa 3' a 5'. Otras polimerasas adecuadas incluyen T4 exo-, Terminator y Bst polimerasas. La polimerasa puede estar libre en disolución (y puede estar presente en disoluciones de lavado y/o de dNTP) o puede estar fijada sobre el soporte sólido, una o más de las paredes de la célula de flujo, la plantilla, o los cebadores.

Debe entenderse que los métodos de secuenciación proporcionados en la presente memoria tienen una serie de aplicaciones que incluyen sin limitación la determinación de la secuencia de nucleótidos parcial o completa de un ácido nucleico (o una colección de ácidos nucleicos como los que existen en un genoma, que incluyen genomas de mamífero y más particularmente genomas humanos), determinar la presencia o la ausencia de un ácido nucleico en una muestra (como puede ser útil, por ejemplo, en métodos diagnósticos o forenses), determinar si el ácido nucleico comprende una mutación o variación en la secuencia (tal como por ejemplo una variación alélica que incluye un único polimorfismo de nucleótido), determinar si un ácido nucleico conocido ha sufrido una mutación que da como resultado la generación de una nueva especie (tal como puede ser la causa subyacente de microorganismos resistentes a antibióticos), determinar la presencia de un organismo modificado genéticamente o de ácidos nucleicos diseñados genéticamente, determinar si existen diferencias, y cuáles son, entre dos muestras (tal como, por ejemplo, entre tejido normal y tejido enfermo), determinar qué régimen terapéutico será el más eficaz para tratar a un sujeto que presenta una afección particular, como puede determinarse a partir de la constitución genética del sujeto, y obtener el genotipo (p.ej., analizando una o más localizaciones genéticas para determinar por ejemplo el estatus de portador). En algunas de estas realizaciones, la secuencia de nucleótidos determinada usando los métodos de la invención puede compararse con una secuencia conocida o de referencia a fin de orientar la secuencia obtenida y/o para identificar diferencias entre las dos. Esto puede ayudar a identificar la variación genética y la mutación. La secuencia conocida o de referencia puede ser una secuencia determinada previamente (por ejemplo, que resulta de la secuenciación genómica completa de una especie).

Los métodos descritos en la presente memoria también pueden usarse para ayudar a la identificación y tratamiento de la afección. Por ejemplo, los métodos se pueden usar para identificar una secuencia asociada a una afección particular o para identificar una secuencia que se usa para diagnosticar la ausencia de una afección particular. Las muestras que van a ser analizadas pueden proceder de cualquier sujeto, incluyendo humanos. La afección puede ser cáncer o una infección.

Los métodos también se pueden usar para identificar una secuencia asociada a una respuesta positiva a un agente. El método puede comprender la secuenciación de ADN a partir de una pluralidad de sujetos que han exhibido una respuesta positiva y a partir de una pluralidad de sujetos que han exhibido una respuesta negativa a un agente usando uno o más métodos de secuenciación proporcionados en la presente memoria, e identificar una secuencia común en la pluralidad de sujetos que han exhibido una respuesta positiva o de los sujetos que han exhibido una respuesta negativa de que esta secuencia no está presente en otra pluralidad de sujetos. Preferiblemente, el sujeto es un mamífero, y más preferiblemente un humano.

Los métodos descritos en la presente memoria pueden automatizarse de tal modo que las reacciones de secuenciación se llevan a cabo con medios robotizados. Adicionalmente, los datos de secuenciación obtenidos de un detector o un sensor pueden ser alimentados a un ordenador personal, un asistente digital personal, un teléfono móvil, un sistema de videojuegos, o una televisión, de tal modo que un usuario puede monitorizar el progreso de las reacciones de secuenciación de forma remota.

También se describen kits que comprenden los diversos reactivos necesarios para llevar a cabo las reacciones de amplificación y/o secuenciación y las instrucciones de uso según los métodos establecidos en la presente memoria.

Los métodos proporcionados en la presente memoria dependen de la detección de nucleótidos individuales en cada copia de una diana de la plantilla. El límite de resolución depende de la resolución del sistema de detección usado.

5 Aunque en la presente memoria se han descrito e ilustrado varias realizaciones inventivas, los especialistas en la técnica concebirán fácilmente una variedad de medios y/o estructuras adicionales para llevar a cabo la función y/o para obtener los resultados y/o una o más de las ventajas descritas en la presente memoria, y cada una de dichas variaciones y/o modificaciones se considera que entran dentro del alcance de las realizaciones inventivas descritas en la presente memoria. De forma más general, los especialistas en la técnica apreciarán fácilmente que todos los parámetros, dimensiones, materiales y configuraciones descritos en la presente memoria pretenden ser ilustrativos y que los parámetros, dimensiones, materiales y/o configuraciones reales dependerán de la aplicación o aplicaciones específicas para la(s) cual(es) se usan las enseñanzas de la invención. Los especialistas en la técnica reconocerán, o serán capaces de determinar usando no más que experimentación rutinaria, muchos equivalentes a las realizaciones inventivas específicas descritas en la presente memoria. Por tanto, debe entenderse que las anteriores realizaciones se presentan meramente a modo de ejemplo y que, dentro del alcance de las reivindicaciones anexas y de los equivalentes a las mismas; las realizaciones de la invención se pueden llevar a la práctica de forma distinta a lo descrito y reivindicado específicamente. Las realizaciones inventivas de la presente descripción están dirigidas a cada característica, sistema, artículo, material, kit y/o método individual descrito en la presente memoria. Adicionalmente, cualquier combinación de dos o más de dichas características, sistemas, artículos, materiales, kits y/o métodos, si dichas características, sistemas, artículos, materiales, kits y/o métodos no son mutuamente inconsistentes, queda incluida dentro del alcance inventivo de la presente descripción.

20 Todas las definiciones, tal como se definen y usan en la presente memoria, debería entenderse que prevalecen sobre las definiciones de diccionario, definiciones en documentos incorporados a modo de referencia, y/o los significados ordinarios de los términos definidos.

#### Aspectos de la invención

25 La presente invención proporciona un nuevo método para detectar la presencia de un ácido nucleico que comprende una secuencia diana en una muestra, que comprende Secuenciación de Nueva Generación, en donde se usa un sistema de controles, es decir, control positivo y/o negativo, que preferiblemente no afecta a la sensibilidad del método de detección de ácido nucleico.

30 La presente invención se basa en el uso de ácidos nucleicos de control, también referidos como Control de Sistemas (SC) que pueden añadirse ("dopado") a los recipientes de reacción que contienen material de muestra que va a ser analizado en un ensayo diagnóstico (aunque también en investigación). Los controles se definen con más detalle en las reivindicaciones anexas.

35 Fue sorprendente descubrir que los ácidos nucleicos de control de la presente invención pueden usarse en métodos diagnósticos basados en NGS y permitir la determinación de si la extracción y las etapas siguientes del método NGS, p.ej., la amplificación de ácido nucleico, la preparación de biblioteca y la secuenciación, cumplen estándares de calidad fiables, y que el flujo de trabajo de la NGS es completamente funcional. Al mismo tiempo, los ácidos nucleicos de control de la presente invención permiten monitorizar si se ha producido o no una contaminación de ADN. Estos controles son idealmente adecuados como controles de calidad requeridos para cumplir los requisitos reguladores obligatorios. También se observó de forma sorprendente que solo se requieren cantidades muy pequeñas de los ácidos nucleicos de control para alcanzar los objetivos fijados anteriormente. Se observó que dichas cantidades bajas de ácidos nucleicos de control no influyen negativamente a la extracción, la amplificación y la reacción de secuenciación (u otras etapas de la NGS) de ácidos nucleicos que derivan de muestras que van a ser investigadas.

Típicamente se lleva a cabo una reacción de control negativo en un recipiente de reacción separado, en donde solo se añade tampón a dicho vial específico, en lugar de una muestra. La presencia de la secuencia diana en el control negativo indica que las muestras han sido contaminadas.

45 Se describen controles de ácido nucleico para un método diagnóstico de la invención basado en Secuenciación de Nueva Generación específico de una secuencia de ácido nucleico diana, en donde dicho control de ácido nucleico comprende al menos una secuencia de control de ácido nucleico que es diferente de la secuencia de ácido nucleico diana. Es posible, no obstante, que el control comprenda más secuencias de control de ácido nucleico, que pueden encontrarse por separado o en la misma molécula de ácido nucleico.

50 La secuencia de control de ácido nucleico puede derivar de cualquier organismo, microorganismo o parte del mismo, p.ej., un órgano, siempre que la secuencia de control sea diferente de las secuencias diana del organismo, microorganismo u órgano de los cuales se derive la secuencia diana. Por ejemplo, cuando se analiza una muestra de un paciente humano, la secuencia de control puede derivarse de una fuente no humana para asegurar la identificación específica y al mismo tiempo evitar cualquier reactividad cruzada de los materiales usados.

55 En un aspecto adicional, el control de ácido nucleico usado en los métodos de la invención puede seleccionarse de un grupo que comprende ácidos nucleicos víricos, bacterianos, fúngicos y derivados de patógenos, o ácidos nucleicos derivados de animales, plantas, o ácidos nucleicos que derivan de un órgano o parte del cuerpo que sea diferente al órgano o parte del cuerpo que es analizada para determinar la presencia o la ausencia de dicha secuencia de ácido

nucleico diana.

En un aspecto de la invención, el control de ácido nucleico usado en los métodos de la invención se selecciona entre ácidos nucleicos víricos, p.ej., entre ácidos nucleicos que derivan de virus animales o vegetales. La secuencia de ácido nucleico se usa generalmente de un modo que se descarta el potencial infeccioso para que no exista riesgo para el personal que maneja los controles o el entorno.

En un aspecto adicional, la secuencia de control de ácido nucleico usada en los métodos de la invención se selecciona entre ácidos nucleicos que derivan del Virus del Mosaico de Tabaco (TMV). Estas secuencias pueden ser secuencias naturales, es decir, tal como se encuentran en la naturaleza, o las secuencias pueden estar modificadas, es decir, las secuencias pueden contener mutaciones, adiciones, eliminaciones, inversiones de residuos de ácido nucleico. Es posible que las secuencias mutantes sirvan como ácidos nucleicos de control, p.ej., los ácidos nucleicos de TMV que han sido modificados de tal modo que portan un cambio trazable en comparación con la secuencia natural. Es posible mutar el TMV mediante mutaciones, adiciones, eliminaciones o inversiones de residuos de ácido nucleico. Los ácidos nucleicos de TMV usados como controles de ácido nucleico son ejemplos de secuencias no humanas adecuadas. Una ventaja de dichas secuencias no humanas, p.ej., ácidos nucleicos de TMV, es que dichas secuencias pueden servir como controles negativos cuando las secuencias de ácido nucleico derivan de un ser humano.

En un aspecto adicional de la invención, el control de ácido nucleico usado en los métodos de la invención comprende al menos dos secuencias de control de ácido nucleico diferentes. Dichas secuencias pueden diferir en al menos un residuo de ácido nucleico, pero es posible también modificar más residuos.

En otro aspecto, el control usado en los métodos de la invención comprende al menos una secuencia de aminoácidos natural y al menos un mutante de dicha secuencia de ácido nucleico natural, p.ej., secuencias no humanas naturales y mutantes, tales como secuencias de TMV. Una ventaja de usar una mezcla de secuencias de ácido nucleico natural y mutante es que dichas mezclas pueden servir como material de referencia para Vocación de Variante y para proporcionar una tasa de mutación definida. Esto aumenta enormemente la sensibilidad de los métodos según la presente invención. Sin material de referencia, la vocación de tasa de mutación (en los ácidos nucleicos diana, p.ej., en oncogenes) sería menos fiable.

La secuencia de control modificada puede o no estar mutada artificialmente, es decir, la mutación puede producirse en la naturaleza o no, siempre que la modificación pueda ser trazada mediante secuenciación en un método basado en NGS. Es deseable que los resultados del ensayo de NGS permitan decidir que los controles de ácido nucleico han sido añadidos a una muestra que se sospecha que contiene un ácido nucleico diana, o que los ácidos nucleicos de control han sido usados sin muestra, p.ej., actuando como controles internos negativos en un análisis basado en NGS de ácidos nucleicos. En este último caso, la presencia de secuencia(s) de control y la ausencia de las secuencias diana según se determina en una reacción de secuenciamiento basada en NGS dada, son indicativas de la ausencia de contaminación con material derivado de la muestra, o con material derivado de cualquier otra fuente, p.ej., ácidos nucleicos arrastrados desde análisis previos, del entorno, etc.

El control de ácido nucleico usado en los métodos de la invención puede presentar una única mutación de nucleótido, una mutación de más de un nucleótido, una inversión, eliminación o adición de residuos de ácido nucleico.

En un aspecto adicional, se proporciona una mezcla de ácidos nucleicos de control usados en los métodos de la invención. La mezcla se puede usar en un método diagnóstico basado en Secuenciación de Nueva Generación específico de secuencia de ácido nucleico diana y comprende al menos dos controles de ácido nucleico en una relación predeterminada. Por ejemplo, si un método o ensayo diagnóstico basado en NGS está dirigido a detectar variantes a un determinado nivel de frecuencia de variante (p.ej., 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2%, 1% o por debajo de 1%) se mezclan secuencias de control diferentes para reflejar dicho nivel de frecuencia variante. Por ejemplo, cuando se debería detectar un nivel de frecuencia variante del 5%, se mezcla un ácido nucleico de control (por ejemplo, TMV natural) con 5% de otro ácido nucleico de control (por ejemplo, TMV que porta una modificación, p.ej., una mutación puntual individual, una mutación triple, una inversión, eliminación, o similar). La mezcla resultante de 95% de secuencia de ácido nucleico de control natural y 5% de ácido nucleico de control mutante puede usarse para verificar que el método basado en NGS es capaz de detectar correctamente el nivel de frecuencia variante. El especialista en la técnica sabe cómo determinar con precisión la cantidad de ácidos nucleicos variantes, es decir, la cantidad de ácidos nucleicos variantes (o mutantes/modificados) que deberían mezclarse con los ácidos nucleicos naturales que actúan como controles en los métodos de la presente invención o en mezclas inventivas.

También se describen kits adecuados para la detección de ácidos nucleicos diana en métodos diagnósticos basados en la Secuenciación de Nueva Generación específicos de secuencia. Dichos kits comprenden al menos uno de los controles de ácido nucleico usados en los métodos de la invención como se ha definido anteriormente, o una mezcla de control de ácido nucleico como se ha detallado anteriormente. Adicionalmente, de forma opcional los kits comprenden reactivos adicionales para los métodos diagnósticos basados en Secuenciación de Nueva Generación, p.ej., reactivos químicos, folletos de instrucciones, y similares.

En la presente memoria también se proporcionan métodos diagnósticos basados en Secuenciación de Nueva Generación, que comprenden extraer el material de ácido nucleico de una muestra y una muestra de control. La

muestra de control se prepara antes de la extracción. Estos métodos comprenden además someter el material de ácido nucleico extraído a un método de Secuenciación de Nueva Generación, p.ej., secuenciación de semiconductor iónico. En algunos aspectos, los métodos diagnósticos basados en Secuenciación de Nueva Generación proporcionados en la presente memoria, comprenden una muestra de control que contiene material de muestra suplementado con cualquiera de los controles de ácido nucleico o las mezclas de control de ácido nucleico mencionadas anteriormente.

En otros aspectos adicionales de la invención, el método diagnóstico basado en Secuenciación de Nueva Generación está dirigido a la detección de la presencia o la ausencia de secuencias diana derivadas de un agente infeccioso o de un oncogén. Por supuesto, es posible analizar más de una diana en un experimento de NGS. Adicionalmente, los métodos NGS pueden usar muestras obtenidas de una o más fuentes, p.ej., de un paciente humano o de muchos pacientes humanos. También es posible usar material de muestra procedente de una fuente y actuar sobre diferentes ácidos nucleicos diana diferentes. Se lleva al cabo un análisis NGS con una muestra procedente de una fuente que sirve como muestra de control, es decir, la muestra ha sido dopada con al menos uno de los controles de ácido nucleico o mezclas anteriores. Cuando se usan muestras procedentes de más de una fuente, o si se analiza una muestra procedente de una fuente con respecto a diferentes dianas, para cada fuente puede haber un control, o para cada diana puede haber un control.

Por consiguiente, los métodos diagnósticos basados en Secuenciación de Nueva Generación según la invención permiten la detección y el análisis de la presencia o ausencia de una secuencia diana derivada de un agente infeccioso o de un oncogén. La presencia de secuencias de control de ácido nucleico posteriores a la reacción de secuenciación es indicativa de una extracción de ácido nucleico de las muestras llevada a cabo con éxito.

Como se ha indicado antes, los métodos diagnósticos basados en Secuenciación de Nueva Generación de la presente invención pueden comprender además un control de contaminación, es decir un control negativo. Con este fin, un ácido nucleico de control que no fue añadido a material de muestra es sometido al flujo de trabajo de NGS. El control de contaminación comprende al menos un control de ácido nucleico como el definido anteriormente o una mezcla de control de ácido nucleico como la definida anteriormente. El método diagnóstico basado en Secuenciación de Nueva Generación comprende someter el control de contaminación a una reacción de secuenciación específica de ácido nucleico diana, en donde la ausencia de detección de una secuencia específica de ácido nucleico diana en el control de contaminación y la detección de la presencia de una secuencia específica de ácido nucleico de control son indicativas de la ausencia de contaminación.

En algunas realizaciones de los métodos de la invención, dicha muestra es una muestra clínica, en donde dicha muestra clínica preferiblemente es de un sujeto humano. Puede preferirse que dicha muestra clínica sea una muestra de tejido o una muestra de fluido corporal. Dicha muestra puede, p.ej., ser una muestra de tejido obtenida del tracto respiratorio, el tracto gastrointestinal o un trasplante tras realizar el trasplante. Adicionalmente, dicha muestra puede en particular ser una muestra de fluido corporal seleccionada del grupo que consiste en sangre, plasma, suero, fluido linfático y saliva.

En otra realización preferida, el ácido nucleico diana puede ser un ácido nucleico procedente de un microorganismo. Preferiblemente, dicho microorganismo se selecciona del grupo que consiste en bacterias, arqueas, protozoos, hongos y virus. En una realización particularmente preferida, dicho microorganismo es una bacteria seleccionada del grupo que consiste en *Streptococos* de Grupo A, miembros de Complejo de *Mycobacterium tuberculosis* que incluyen *Mycobacterium tuberculosis*, *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium canettii* y *Mycobacterium microti*, *Salmonella* entérica spp., *Clostridium difficile*, enterococos resistentes a vancomicina y *Staphylococcus aureus* resistente a metilicina. En otra realización particularmente preferida, dicho microorganismo es un virus seleccionado del grupo que consiste en adenovirus, virus de la gripe, virus de la gripe aviar A (H7N9), coronavirus de Síndrome Respiratorio de Oriente Medio, norovirus, virus BK, citomegalovirus, virus de Epstein-Barr, virus de herpes simplex 1, virus de herpes simplex 2, virus de Varicella-Zoster, enterovirus, VIH-1, virus de la hepatitis B, virus de la hepatitis C, virus del Dengue y virus Chikungunya.

En algunas realizaciones de los métodos de la invención, el ácido nucleico diana puede ser un oncogén, preferiblemente un oncogén humano, p.ej., un oncogén seleccionado del grupo que consiste en BCR-ABL mayor, BCR-ABL menor, BCR-AML1 ETO, PML-RARA, mutantes BRAF V600, mutantes KRAS, mutantes NRAS, EGFR, KIT o PIK3CA. En realizaciones, los métodos de la presente invención pueden usarse por tanto para detectar la presencia de uno o más oncogenes en una muestra.

Los ácidos nucleicos de control usados en los métodos de la invención pueden usarse en métodos diagnósticos basados en NGS en cantidades que permitan la detección específica y/o el nivel de detección esperados. La presencia de controles en reacciones de muestra no debería influir negativamente en el resultado. Eso significa que no debería haber interferencias de los ácidos nucleicos de control con la extracción, amplificación y secuenciación, o cualesquier otras etapas del método NGS, en términos de calidad suficiente y/o cantidad suficiente de los ácidos nucleicos diana, lo que podría conducir a interpretaciones falsas de los resultados.

Dependiendo de la sensibilidad requerida para una secuencia diana específica en términos de frecuencia variante (vf), se elige la cantidad de controles de la presente invención. Por ejemplo, cuando se desea detectar el 5% de frecuencia

variante (vf), se pueden usar 5 fg del control de sistemas de la invención (construidos con una frecuencia variante del 5%) para proporcionar una cobertura de la diana de más de 1.000 veces. Si la sensibilidad deseada cambia a vf del 1%, es posible usar 20 fg de los controles de la invención (construidos con una frecuencia variante del 1%) para proporcionar una cobertura de la diana de 10.000 veces. Puede ser suficiente que de 0,1 a 1.000 fg, preferiblemente de 0,1 a 100 fg, más preferiblemente de 1,0 a 50,0 fg, de 1,0 a 25 fg, más preferiblemente de 1,0 a 10,0 fg, p.ej., 5 fg de ácido nucleico de control, estén presentes en una reacción NGS. Por ejemplo, uno puede añadir aproximadamente de 100 fg a 1.000 fg de las muestras. Sin embargo, dependiendo de la fuente del material de diana, la cantidad correcta se puede adaptar. Con este fin, se puede someter material de referencia procedente de diferentes fuentes, o material que ha sido tratado físicamente para parecerse al material fuente deseado, a un análisis por valoración de las cantidades de ácido nucleico de control para determinar la cantidad adecuada requerida en cada ensayo específico. Por ejemplo, la extracción de ácidos nucleicos de un material vegetal o de tejidos animales puede requerir diferentes pretratamientos. Dependiendo del rendimiento esperado o de la pérdida de ácidos nucleicos procedentes de diferentes fuentes, la alimentación cuantitativa de material de control puede ser diferente.

### Figuras

La Figura 1 proporciona una visión general de un método NGS inventivo según la invención. A, Se llevaron a cabo ocho reacciones en esta representación esquemática. En el primer pocillo de reacción (1), solo se añade un control de sistemas (SC) positivo, es decir, esta reacción no contiene ningún material de muestra. Los pocillos de reacción (2 a 8) contienen material de muestra FFPE y SC añadido. También se muestran los dispositivos usados en el flujo de trabajo automatizado. En B y C, se muestran los resultados de la reacción de secuenciación. Aunque el control positivo y las secuencias de SC añadidas en los pocillos (1) y (2 a 8) fueron detectados según lo deseado, las secuencias diana no estaban presentes en el pocillo (1). Por otro lado, la reacción de secuenciación de oncogenes presentes en las muestras FFPE de los pocillos (2) a (8) se llevó a cabo con éxito. El control de sistemas usado en el pocillo (1) fue una mezcla de plásmidos que comprenden las secuencias de ácido nucleico no humano naturales y plásmidos que comprenden mutaciones en dicha secuencia de ácido nucleico no humano. Los plásmidos se mezclaron de tal modo que las mutaciones estaban presentes con una frecuencia variante del 5%, y el resto se completó con plásmido que comprende secuencias de ácido nucleico no humano natural. Como material FFPE de referencia, se usaron líneas celulares obtenidas de Horizon Discovery Ltd., R.U. Estas líneas celulares fueron diseñadas para portar mutaciones conocidas presentes a frecuencias variantes definidas (2-33%).

La Figura 2 muestra los resultados de experimentos de NGS con código de barras que comprenden ADN-FFPE con 10, 5, 1, 0,5, 0,25 y 0,1 fg de SC añadidos en cada reacción PCR (AmpliSeq), junto con una muestra de ADN similar con 100 fg de SC añadidos a la muestra al comienzo del flujo de trabajo ("workflow") automatizado. A, gráfico de barras de cobertura de SC frente a cantidades conocidas de SC añadido (manualmente) a cada reacción de secuenciación antes de AmpliSeq (Life Technologies). El ADN FFPE usado para esta parte del ensayo no debe estar dopado con SC al inicio del flujo de trabajo. La cobertura de SC resultante fue proporcional a los niveles de SC añadido a cada reacción AmpliSeq. B, La representación de cobertura de SC frente a niveles de SC mostró una relación lineal entre cobertura de SC y nivel de SC. Aquí, la cobertura de SC se normalizó respecto a la cobertura media de todos los amplicones de la muestra. Por interpolación (línea discontinua roja), la muestra FFPE con 100 fg de SC añadidos al comienzo de las etapas de extracción producirá una cobertura de SC comparable a la de una muestra con 5 fg de SC añadidos en la reacción AmpliSeq. Este ensayo demostró que SC puede ser recuperado después del flujo de trabajo automatizado, y que se puede estimar su nivel en la reacción de secuenciación. Por tanto, se puede usar SC como control añadido. (Nótese que todas las muestras usadas para este experimento de secuenciación presentaban entradas de ADN coincidentes, excepto para SC).

La Figura 3 muestra los mismos experimentos con código de barras descritos en la anterior Figura 2. El gráfico de barras mostrado aquí es una representación de la cobertura media de amplicones diana frente a niveles variantes de SC añadidos en cada muestra. Aquí, el aumento de SC desde 0,1 hasta 10 fg no suprime la cobertura media en las muestras, demostrando que a estos niveles (de 0,1 a 10 fg) el SC no afecta negativamente a la amplificación de los amplicones diana.

La Figura 4 muestra un experimento con SC fijo (5 fg añadidos en cada reacción AmpliSeq) con alimentaciones de ADN variadas (0,1-5 ng). Aquí, la cobertura de SC fue inversamente proporcional a la cobertura promediada de dianas. Se representa la cobertura de SC no normalizada y normalizada frente a la alimentación de ADN (A y B, respectivamente). Los datos demuestran que los niveles de SC varían según la alimentación de ADN en una muestra, y por tanto puede ser útil como punto de control para bajas alimentaciones de ADN.

La Figura 5 muestra el pocillo de control negativo (pocillo n°1) que solo contiene el SC. Estos datos demuestran la utilidad del SC como control positivo y control negativo (NTC). A, Se detectaron las secuencias correspondientes al SC, pero no las correspondientes a secuencias diana humanas. B, Se recuperaron las tres mutaciones en SC tras Secuenciación de Nueva Generación. El amplicón correspondiente al SC se denota como EC\_pUC97.

La Figura 6 comprende la Tabla 1 que muestra la tabla lógica implicada en el uso de Control de Sistemas (SC) como concepto de calidad (QC) para la NGS. Un experimento de secuenciación exitoso requiere la recuperación de secuencias para el SC añadido en cada muestra, y los de SC en el pocillo (1) (SC solo). Adicionalmente, no debería haber secuencias que mapeen a dianas humanas en el pocillo (1). De este modo, el pocillo (1) actúa como control

5 externo y como control positivo global para el experimento, así como control no de plantilla (NTC). El SC añadido actúa como control positivo de la extracción y en la muestra para cada muestra. Para una muestra exitosa, es de esperar que los amplicones diana esperados sean detectados con éxito. El fallo en la recuperación del SC en el pocillo (1) indica un fallo importante del flujo de trabajo y el experimento debería considerarse no válido. Los amplicones diana, incluso cuando son secuenciados con éxito, pueden no ser fiables. Para una muestra que falló en la generación de amplicones diana, pero superó los criterios para el SC solo del pocillo (1), la presencia de SC añadido indica que el flujo de trabajo está operativo, pero la muestra no tiene suficiente ADN ampliable. La ausencia de añadido indica la presencia de inhibidores en la muestra.

10 Debe entenderse que, aunque la invención ha sido descrita en combinación con las realizaciones descritas en la presente memoria, la descripción precedente, así como los ejemplos incluidos a continuación, pretenden ilustrar y no limitar el alcance de la invención. Otros aspectos, ventajas y modificaciones dentro del alcance de la invención serán evidentes para los especialistas en la técnica a la cual pertenece la invención.

15 Los siguientes ejemplos se presentan para proporcionar a los especialistas en la técnica una descripción completa y la descripción de cómo preparar y usar las composiciones de la invención. Los ejemplos están planteados como ejemplos no limitativos de la invención. Aunque se ha realizado el esfuerzo para asegurar la precisión con respecto a variables tales como cantidades, temperatura, etc., deben tomarse en consideración errores experimentales y desviaciones. A menos que se indique lo contrario, las partes son partes en peso, la temperatura está en grados centígrados, y la presión es la atmosférica, o próxima a ella. Todos los componentes se obtuvieron comercialmente, a menos que se indique lo contrario.

20

**REIVINDICACIONES**

- 5 1. Un método diagnóstico basado en Secuenciación de Nueva Generación que comprende extraer material de ácido nucleico de una muestra y al menos una muestra de control, en donde la muestra de control se prepara antes de la extracción, que comprende además someter al material de ácido nucleico extraído a un método de Secuenciación de Nueva Generación, en donde la muestra de control comprende material suplementado con un control de ácido nucleico, en donde dicho control de ácido nucleico comprende al menos una secuencia de control de ácido nucleico que es diferente de la secuencia de ácido nucleico diana, y en donde dicho control comprende al menos una secuencia de ácido nucleico diana, y en donde dicho control comprende al menos una secuencia de ácido nucleico natural y al menos un mutante de dicha secuencia de ácido nucleico natural.
- 10 2. El método diagnóstico basado en Secuenciación de Nueva Generación según la reivindicación 1, en donde dicha secuencia de control de ácido nucleico se selecciona entre ácidos nucleicos no derivados de humano, o se selecciona entre ácidos nucleicos víricos.
- 15 3. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 y 2, en donde dicha secuencia de control de ácido nucleico se selecciona entre ácidos nucleicos que derivan del Virus del Mosaico del Tabaco.
- 20 4. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 3, que comprende al menos dos controles de ácido nucleico como los definidos en una cualquiera de las reivindicaciones 1 a 3 en una relación predeterminada.
- 25 5. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 4, en donde dicho método es adecuado para la detección de la presencia o ausencia de secuencias derivadas de un agente infeccioso o de un oncogén.
- 30 6. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 5, en donde la presencia o la ausencia de una secuencia diana derivada de un agente infeccioso o de un oncogén y la presencia de al menos una secuencia de control de ácido nucleico posterior a la reacción de secuenciamiento son indicativas de una extracción de ácido nucleico completada a partir de las muestras.
- 35 7. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 6, que además comprende un control de contaminación.
- 40 8. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 7, en donde el control de contaminación comprende al menos un control de ácido nucleico como el definido en las reivindicaciones 1 a 3, en donde el control de contaminación no comprende material de muestra.
- 45 9. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 8, que comprende someter al control de contaminación a una reacción de secuenciación específica de ácido nucleico diana.
10. El método diagnóstico basado en Secuenciación de Nueva Generación según una cualquiera de las reivindicaciones 1 a 9, que comprende someter al control de contaminación a una reacción de secuenciación específica de ácido nucleico diana, en donde la ausencia de detección de una secuencia específica de ácido nucleico diana en el control de contaminación y la detección de la presencia de una secuencia específica de ácido nucleico de control son indicativas de la ausencia de contaminación.
11. El método diagnóstico basado en Secuenciación de Nueva Generación o el método de detección de ácido nucleico diana basado en Secuenciación de nueva Generación según una cualquiera de las reivindicaciones 1 a 10, en donde dicho ácido nucleico de control está presente en una cantidad de 0,1 a 10 fg después de la extracción de los ácidos nucleicos de las muestras.
12. El método diagnóstico basado en Secuenciación de Nueva Generación o el método de detección de ácido nucleico diana basado en Secuenciación de nueva Generación según una cualquiera de las reivindicaciones 1 a 11, en donde la muestra es una muestra clínica derivada de un sujeto humano, y en donde el ácido nucleico de control es un plásmido que comprende secuencias naturales y/o mutantes de un virus no humano, seleccionado entre virus animales que no infectan humanos, virus vegetales, o en donde el ácido nucleico de control ha sido diseñado artificialmente y no tiene un equivalente en la naturaleza.

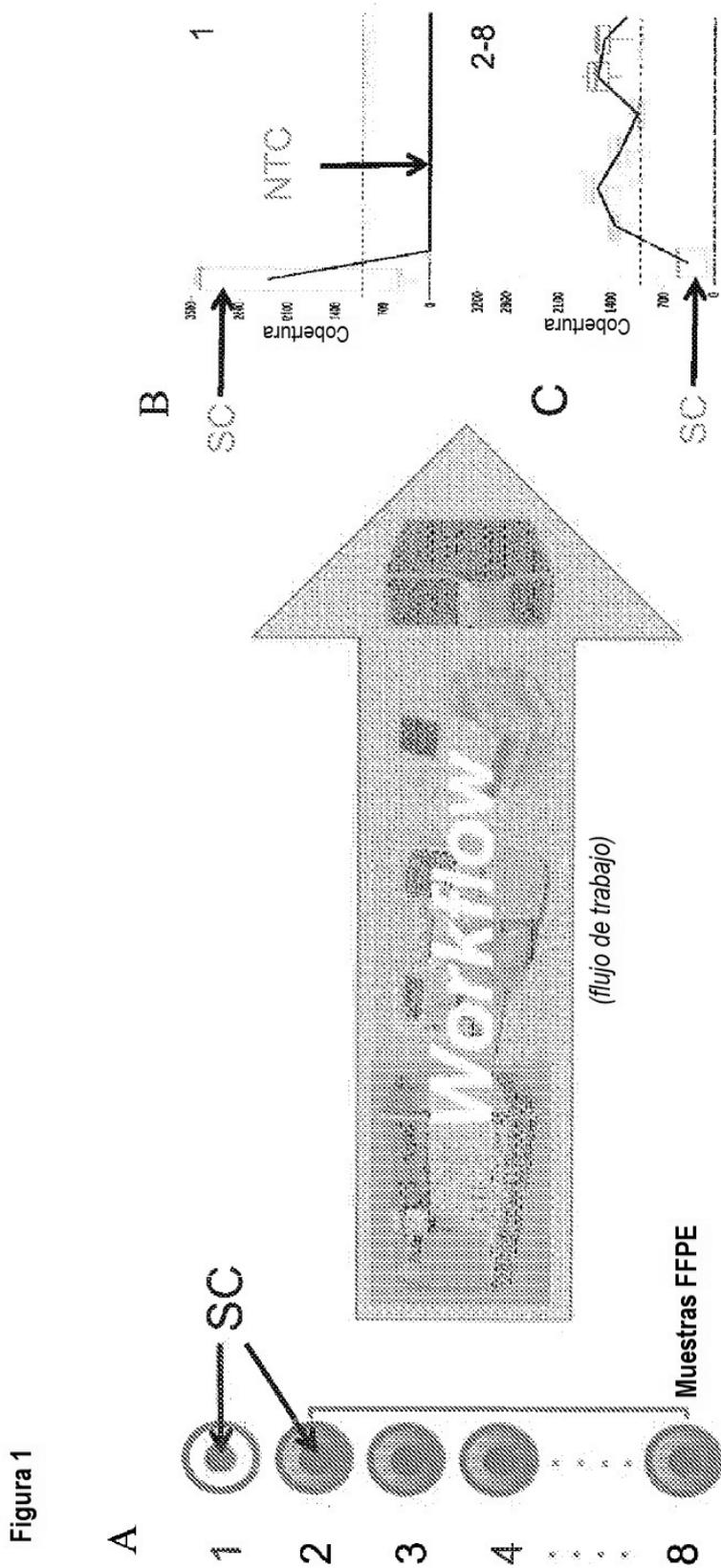


Fig. 2

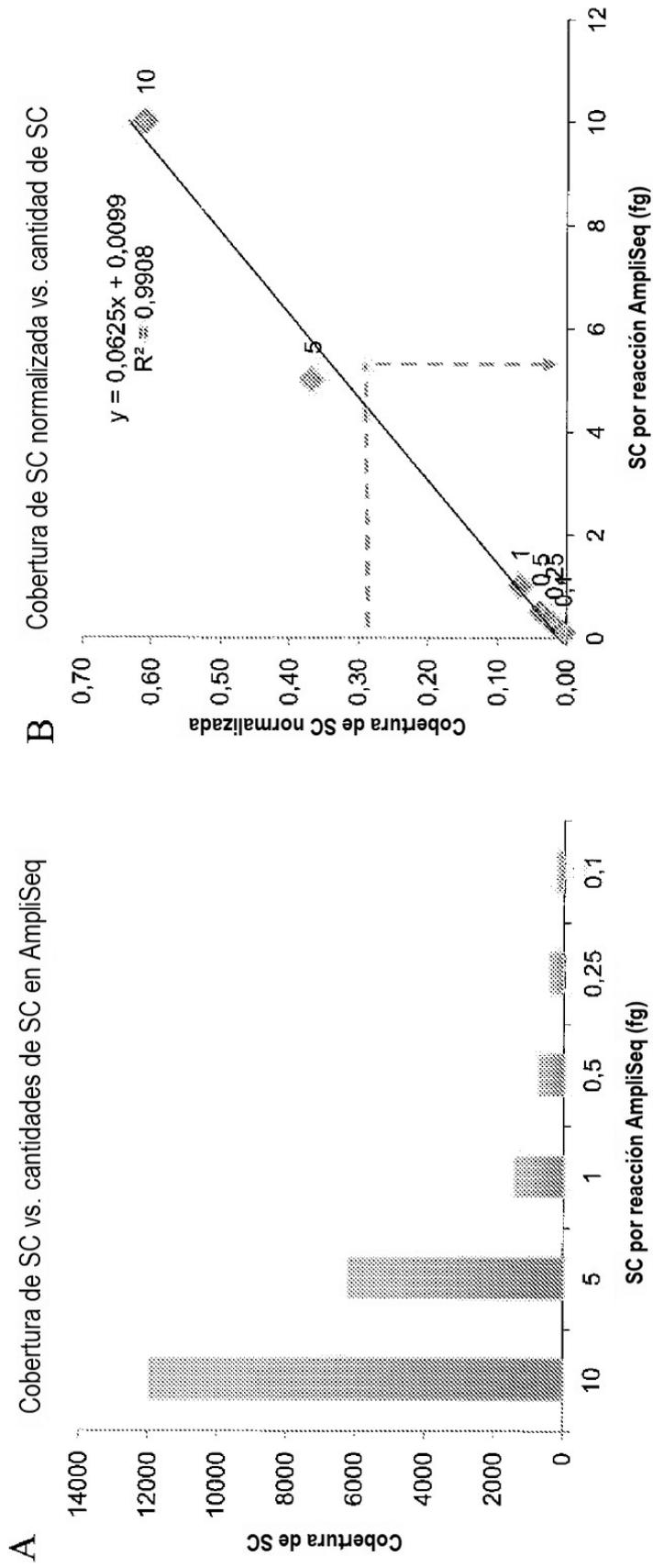


Fig. 3

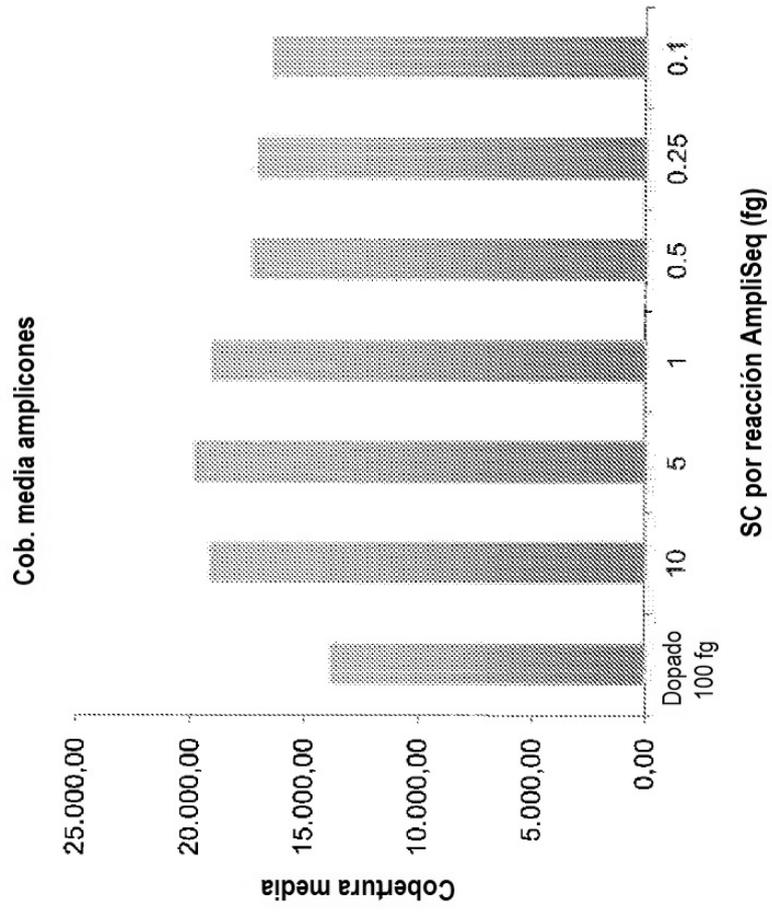


Fig. 4 A

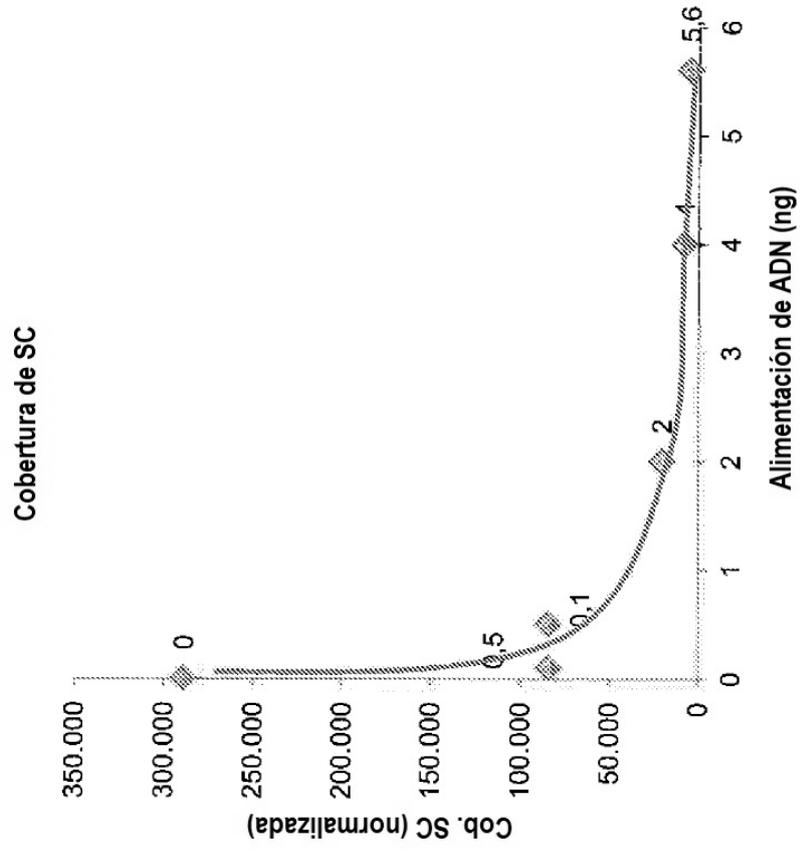


Fig. 4 B

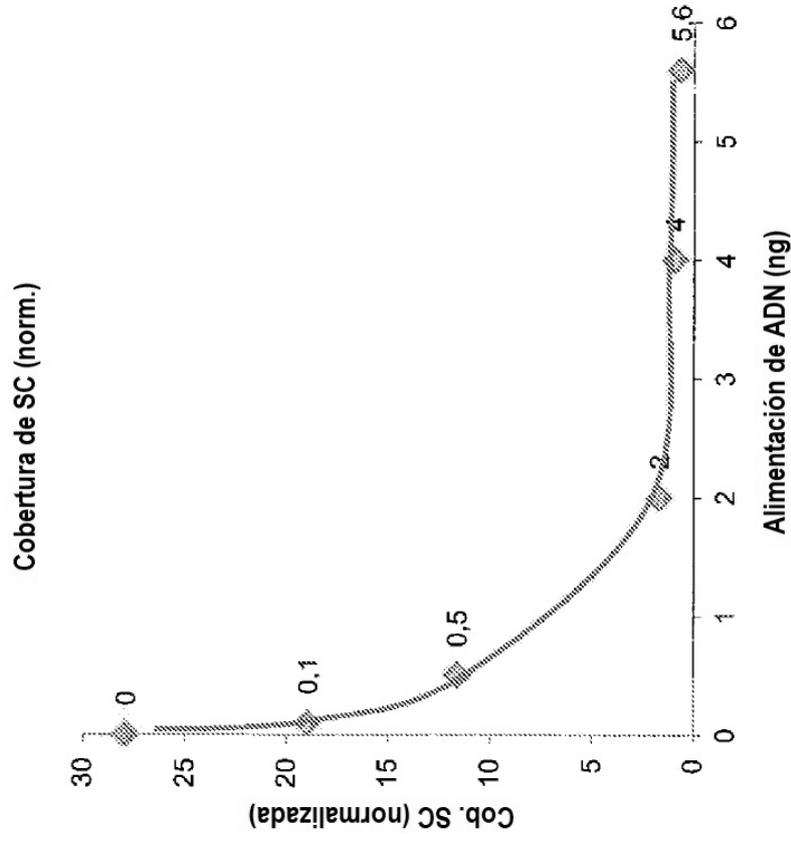
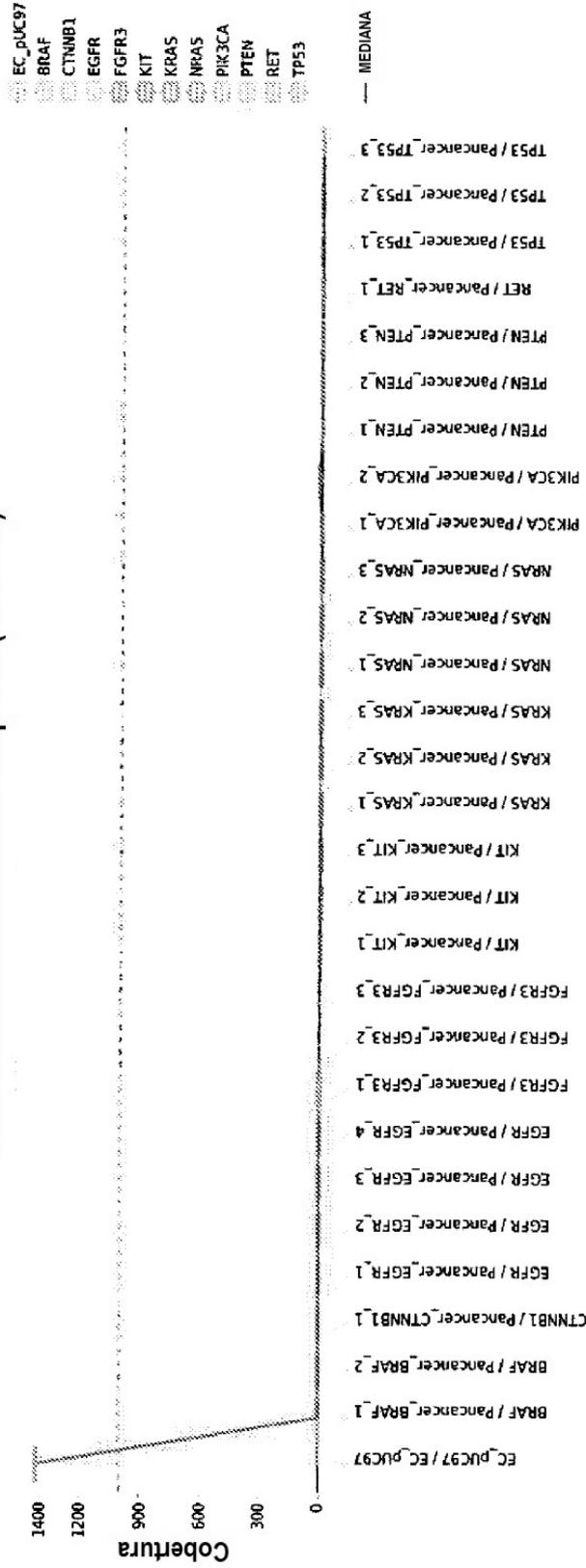


Fig. 5

Distribución de cobertura de amplicón (resumen)



Experado	5%	5%	5%
Diana	EC_pUC97	EC_pUC97	EC_pUC97
Mutación	c.8G>A	delAAA	c.44_45CA>AG
35_1_Extraido	10.60%	10.10%	9.60%
NTC	1,374	1,410	1,406

Figura 6 – Tabla 1

SC añadido	SC solo	Muestra	Conclusión
+/-	+	+	Experimento con éxito
+/-	+	- (algunos amplicones)	Eliminaciones en ADN tumoral, inhibición de PCR en muestra. Es posible la llamada a variante en amplicones con éxito.
+	+	- (todos los amplicones)	Muestra de baja calidad (ADN < LOD)
-	+	- (todos los amplicones)	Fuerte inhibición de PCR en la muestra
-	-	-	Fallo en el flujo de trabajo general
+/-	- (sin amplicón de TMV)	+	Posible fallo en el flujo de trabajo general, los resultados no son fiables
+/-	- (errores de llamada de variante)	+	Posible falso positivo / falso negativo en llamadas de variante
+/-	- (amplicones humanos)	+	Posible contaminación