

## **MÉTODOS Y COMPOSICIONES PARA EL PERFILADO DE ADN**

### SOLICITUDES RELACIONADAS

La presente solicitud reivindica los derechos de prioridad de la Solicitud Provisional n.º 62/103.524 presentada el 14 de enero del 2015, de la Solicitud Provisional de Estados Unidos n.º 62/043.060 presentada el 28 de agosto de 2014 y de la Solicitud Provisional de Estados Unidos n.º 61/940.942 presentada el 18 de febrero del 2014, cuyos contenidos se incorporan en el presente documento por referencia en su totalidad.

### 10 REFERENCIA AL LISTADO DE SECUENCIAS, TABLA O LISTADO DE PROGRAMAS INFORMÁTICOS

La presente solicitud se presenta junto con un listado de secuencias en formato electrónico. El listado de secuencias se proporciona como un archivo titulado ILLINC276WO\_Sequence\_Listing.TXT, creado el 13 de febrero de 2015, que tiene un tamaño de 55 Kb. La información en formato electrónico del listado de secuencias se incorpora en el presente documento por referencia en su totalidad.

### CAMPO DE LA DIVULGACIÓN

En el presente documento se proporcionan realizaciones relacionadas con métodos y composiciones para el perfilado de ADN. Algunas realizaciones se refieren a métodos de amplificación de secuencias diana de diversos tamaños en una sola reacción, seguidos por la secuenciación posterior de la biblioteca.

### ANTECEDENTES DE LA DIVULGACIÓN

25 Históricamente, el uso de un subconjunto de marcadores en un genoma

humano se ha utilizado para determinar la identidad personal, huella o perfil de ADN de un individuo. Estos marcadores incluyen localizaciones o loci de secuencias repetidas en tándem (STR) cortas y secuencias repetidas en tándem intermedias (ITR) que, en combinación, son útiles para distinguir un individuo de otro a nivel genético. El análisis de estos marcadores se ha estandarizado en el análisis de ADN encontrados en escenas de crímenes. Por ejemplo, en los Estados Unidos, se han combinado varias de estas secuencias repetidas para crear un Sistema de Índice Combinado de ADN (CODIS), que sirve como patrón de laboratorio para el perfilado de ADN en casos penales. Otros países han adaptado de manera similar un sistema patrón para el perfilado de ADN. Estos sistemas también se han utilizado para determinar relaciones de paternidad y familiares. Sin embargo, todos los sistemas actuales están basados en la separación por tamaño de estos loci repetidos sobre un sistema electroforético y, por tanto, están limitados al número de loci que pueden diferenciarse en dicho sistema. Por ejemplo, algunos de los sistemas comerciales actuales para el perfilado de ADN con fines forenses diferencian solo 16 marcadores debido a las limitaciones de los métodos de detección electroforéticos

#### SUMARIO DE LA DIVULGACIÓN

Las realizaciones se refieren a sistemas y a métodos que no están limitados en su contenido y que reúnen diferentes partes de información genética sobre un individuo para proporcionar un perfil de ADN más completo, integral, de un individuo. La presente divulgación describe métodos y composiciones que permiten obtener este perfil de un individuo, contribuyendo de esta manera a los campos de la genómica personal y forense.

Actualmente, el perfilado de ADN usa marcadores biológicos seleccionados para determinar la identidad de una muestra de ADN. Por ejemplo, el análisis más habitual para determinar un perfil de ADN es determinar el perfil de varias secuencias repetidas en tándem cortas (STR) encontradas en el genoma de un organismo. El análisis consiste en la amplificación de secuencias STR definidas que puede tener hasta 400 pb de longitud, que pueden diferenciarse por tamaño en un gel electroforético o usando electroforesis capilar (EC). La electroforesis se usa para detectar cambios de tamaño debidos a diferencias en el número de STR repetidas en un locus determinado y, por lo tanto, la longitud de los amplicones de PCR, que, para el sistema de EC, está comprendida entre 50 y 500 pb. Para ayudar a superar los límites impuestos por las metodologías de diferenciación por tamaño (es decir, no pueden diferenciarse STR de tamaños de amplicón solapantes), los métodos actuales de perfilado de ADN utilizan diferentes conjuntos de cebadores marcados de tal manera que los amplicones que solapan en tamaño pueden marcarse con diferentes colorantes fluorescentes, de forma que, después de la excitación, los espectros de emisión difieren, lo cual permite que los amplicones solapantes se diferencien usando diferencias en los espectros de excitación y emisión del colorante. Usando marcaje diferenciado, los métodos actuales permiten la multiplexación de 24 loci de STR diferentes usando 6 colorantes detectables de modo distinto en un proceso de perfilado de ADN.

Existen muchas limitaciones respecto a las metodologías de perfilado de ADN actuales. Como se ha mencionado anteriormente, los sistemas de diferenciación por tamaño limitan el número de loci que pueden determinarse

específicamente en un momento determinado. Otra limitación de los métodos establecidos para el perfilado de ADN es que el ADN a analizar algunas veces se degrada y el intervalo de tamaños de algunos de los marcadores no acomoda al ADN degradado, por ejemplo, los amplicones pueden ser más largos que el tamaño de los fragmentos del ADN degradado. Para el ADN degradado, los amplicones de 400 pb se consideran muy largos y pueden ocasionar la pérdida de amplificación de esos loci más largos. Cuando los analistas de ADN amplifican muestras de ADN degradado para identificar su perfil de STR, por ejemplo, una muestra encontrada en una escena de un crimen, a menudo no puede detectar todos los loci dando como resultado un perfil parcial, lo cual puede hacer que sea difícil o imposible hacer coincidir un sospechoso en una escena de un crimen con una muestra del crimen. Por lo general, con dichas muestras un analista de ADN tiene poca elección y, si queda alguna muestra, necesita realizar ensayos adicionales para identificar otros marcadores que puedan dar una pista sobre la identidad del individuo, tales como análisis de polimorfismos mononucleotídicos (SNP), mini-STR, o ADN mitocondrial (ADNmt). Sin embargo, en cada ensayo tienen que gastarse muestras valiosas sin certeza de éxito en la identificación final de un individuo. La FIG. 1A demuestra las posibles vías diferentes para la identificación de ADN, siendo todas ellas flujos de trabajo distintos y requiriendo todas ellas alícuotas de muestras valiosas. Cuando es necesario combinar uno o más flujos de trabajo sencillos y es posible que tengan que repetirse múltiples veces, entonces el proceso resultante no es tan sencillo o no se está haciendo un uso eficaz de una muestra valiosa.

Las realizaciones descritas en la presente solicitud proporcionan métodos,

composiciones y sistemas para determinar el perfil de ADN de un individuo u organismo mediante secuenciación de nueva generación (NGS), proporcionando de este modo una solución a los problemas y limitaciones de las metodologías actuales para el perfilado del ADN. La FIG. 1B muestra un  
5 flujo de trabajo a modo de ejemplo de los métodos desvelados en una realización. En el presente documento se desvelan métodos y composiciones para combinar una multitud de marcadores relevantes desde el punto de vista forense en un ensayo que incluye, pero sin limitación, repeticiones en tándem cortas (STR), repeticiones en tándem intermedias (ITR), polimorfismos  
10 mononucleotídicos informativos de identidad (iSNP), polimorfismos mononucleotídicos informativos de ascendencia (aSNP) y polimorfismos mononucleotídicos informativos de fenotipo (pSNP).

La presente divulgación describe ensayos que superan las limitaciones de las metodologías actuales para el perfilado de ADN. Las realizaciones  
15 desveladas proporcionan métodos y composiciones para la amplificación multiplex, preparación de bibliotecas y secuenciación de STR, ITR, iSNP, aSNP y pSNP combinados procedentes de una muestra de ácido nucleico en una sola reacción multiplex. Los métodos desvelados analizan una pluralidad de marcadores en un ensayo experimental con una manipulación mínima de  
20 muestras, usando bajas cantidades de ADN de muestra incluyendo ADN degradado. Algunas realizaciones descritas pueden utilizarse para perfiles de ADN de bases de datos y/o perfiles de ADN que pueden usarse para investigación criminal. Algunas realizaciones proporcionan métodos de PCR y composiciones desarrolladas para ser suficientemente sensibles como para  
25 detectar cantidades por debajo de un nanogramo de ADN. Además, los

parámetros de diseño de cebadores no convencionales permiten la utilización de una PCR altamente multiplexada para la identificación de STR, ITR y SNP en una relación multiplex. Para investigación criminal, los presentes métodos y composiciones incorporan identificadores de molécula única (UMI) que ayudan a eliminar, por ejemplo, errores en la PCR y secuenciación, tartamudeos (stutter) y similares de los resultados de secuenciación. Véase Kivioja *et al.*, *Nat. Meth.* 9, 72-74 (2012). Además, los resultados de los métodos y composiciones que se desvelan en el presente documento son compatibles con las bases de datos existentes.

Por lo tanto, las realizaciones desveladas en el presente documento proporcionan métodos para construir un perfil de ADN que comprende: proporcionar una muestra de ácido nucleico, amplificar la muestra de ácido nucleico con una pluralidad de cebadores que hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo mononucleotídico (SNP) y al menos una secuencia diana que comprende una repetición en tándem en una reacción multiplex para generar productos de amplificación, y determinar los genotipos de dicho al menos un SNP y de dicha al menos una repetición en tándem en los productos de amplificación, construyéndose de esta manera el perfil de ADN de la muestra de ácido nucleico.

En algunas realizaciones, los métodos comprenden generar una biblioteca de ácido nucleico a partir de los productos de amplificación. En algunas realizaciones, los métodos comprenden determinar las secuencias de la biblioteca de ácido nucleico. En algunas realizaciones, la muestra de ácido nucleico es de un ser humano. En algunas realizaciones, la muestra de ácido nucleico es de una muestra ambiental, una planta, un animal no humano, una

bacteria, una arquea, un hongo o un virus. En algunas realizaciones, el perfil de ADN se usa para uno o más de los siguientes: diagnóstico o pronóstico de enfermedades, identificación de biomarcadores de cáncer, identificación de anomalías genéticas o análisis de diversidad genética. En algunas

5 realizaciones, el perfil de ADN se usa para uno o más de los siguientes: creación de bases de datos, medicina forense, estudios de criminología, pruebas de paternidad o identificación de personas. En algunas realizaciones, dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico. En algunas realizaciones, cada uno

10 de la pluralidad de cebadores tiene una baja temperatura de fusión y/o tiene una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una temperatura de fusión que es menor de 60 grados centígrados. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una temperatura de fusión que es de aproximadamente 50

15 grados centígrados a aproximadamente 60 grados centígrados. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de aproximadamente 24 nucleótidos a aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la

20 pluralidad de cebadores comprende una secuencia de nucleótidos homopolimérica. En algunas realizaciones, la muestra de ácido nucleico se amplifica mediante reacción en cadena de la polimerasa (PCR). En algunas realizaciones, la muestra de ácido nucleico se amplifica en un tampón de amplificación que tiene una concentración de sal que está aumentada en

25 comparación con la concentración de sal de un tampón de amplificación usado

junto con cebadores diseñados convencionalmente. En algunas realizaciones, la sal comprende KCl, LiCl, NaCl, o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor de aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM. En algunas realizaciones, el SNP es un SNP de ascendencia, un SNP fenotípico, un SNP de identidad, o una combinación de los mismos. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 30 SNP. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 50 SNP. En algunas realizaciones, la repetición en tándem es una repetición en tándem corta (STR), una repetición en tándem intermedia (ITR), o una variante de las mismas. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 24 secuencias de repetición en tándem. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 60 secuencias de repetición en tándem. En algunas realizaciones, la muestra de ácido nucleico comprende de aproximadamente 100 pg a aproximadamente 100 ng de ADN. En algunas realizaciones, la muestra de ácido nucleico comprende de aproximadamente 10 pg a aproximadamente 100 pg de ADN. En algunas realizaciones, la muestra de ácido nucleico comprende de aproximadamente 5 pg a aproximadamente 10 pg de ADN. En algunas realizaciones, la muestra de ácido nucleico comprende ADN genómico. En algunas realizaciones, el ADN genómico procede de una

muestra forense. En algunas realizaciones, el ADN genómico comprende ADN degradado. En algunas realizaciones, se determina al menos el 50 % de los genotipos de dicho al menos un SNP y al menos una repetición en tándem. En algunas realizaciones, se determina al menos el 80 % de los genotipos de dicho al menos un SNP y al menos una repetición en tándem. En algunas realizaciones, se determina al menos el 90 % de los genotipos de dicho al menos un SNP y al menos una repetición en tándem. En algunas realizaciones, se determina al menos el 95 % de los genotipos de dicho al menos un SNP y al menos una repetición en tándem. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una o más secuencias marcadoras. En algunas realizaciones, dichas una o más secuencias marcadoras comprenden una etiqueta de cebador, una etiqueta de captura, una etiqueta de secuenciación, una etiqueta de identificador de molécula única, o una combinación de las mismas. En algunas realizaciones, dichas una o más secuencias marcadoras comprenden una etiqueta de cebador. En algunas realizaciones, dichas una o más secuencias marcadoras comprenden una etiqueta de identificador de molécula única.

Las realizaciones desveladas en el presente documento proporcionan métodos para construir una biblioteca de ácidos nucleicos que comprende: proporcionar una muestra de ácido nucleico, y amplificar la muestra de ácido nucleico con una pluralidad de cebadores que hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo mononucleotídico (SNP) y al menos una secuencia diana que comprende una secuencia de repetición en tándem en una reacción multiplex para generar productos de amplificación.

En algunas realizaciones, la muestra de ácido nucleico no está fragmentada antes de la amplificación. En algunas realizaciones, las secuencias diana no están enriquecidas antes de la amplificación. En algunas realizaciones, dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una o más secuencias marcadoras. En algunas realizaciones, dichas una o más secuencias marcadoras comprenden una etiqueta de cebador, una etiqueta de captura, una etiqueta de secuenciación, o una etiqueta de identificador de molécula única o una combinación de las mismas. En algunas realizaciones, los métodos incluyen amplificar los productos de amplificación con una segunda pluralidad de cebadores. En algunas realizaciones, cada uno de la segunda pluralidad de cebadores comprende una parte que corresponde a la etiqueta de cebador de la pluralidad de cebadores y una o más secuencias marcadoras. En algunas realizaciones, dichas una o más secuencias marcadoras de la segunda pluralidad de cebadores comprenden una etiqueta de captura, o una etiqueta de secuenciación, o una combinación de las mismas. En algunas realizaciones, los métodos incluyen añadir proteínas ligantes de unión a ADN monocatenario (SSB) a los productos de amplificación.

En algunas realizaciones, la muestra de ácido nucleico y/o los productos de amplificación se amplifican mediante relación en cadena de la polimerasa (PCR). En algunas realizaciones, la muestra de ácido nucleico y/o los productos de amplificación se amplifican en un tampón de amplificación que tiene una concentración de sal que está aumentada en comparación con la concentración de sal de un tampón de amplificación usado junto con cebadores

diseñados convencionalmente. En algunas realizaciones, la sal comprende KCl, LiCl, NaCl o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 5 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor de aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM.

Las realizaciones desveladas en el presente documento proporcionan una 10 biblioteca de ácido nucleico que comprende una pluralidad de moléculas de ácido nucleico, en la que la pluralidad de moléculas de ácido nucleico comprende al menos una secuencia de repetición en tándem flanqueada por un primer par de secuencias marcadoras y al menos una secuencia de polimorfismo mononucleotídico (SNP) flanqueada por un segundo par de 15 secuencias marcadoras. También se proporciona una biblioteca de ácido nucleico construida usando los métodos y composiciones desveladas en el presente documento. En algunas realizaciones, dicho al menos un SNP indica la descendencia o una característica fenotípica de la fuente de la pluralidad de moléculas de ácido nucleico.

20 Las realizaciones desveladas en el presente documento proporcionan una pluralidad de cebadores que hibridan específicamente con al menos una secuencia diana corta y al menos una secuencia diana larga en una muestra de ácido nucleico, en la que la amplificación de la muestra de ácido nucleico usando la pluralidad de cebadores en una sola reacción multiplex da como 25 resultado al menos un producto de amplificación corto y al menos un producto

de amplificación largo, en la que cada uno de la pluralidad de cebadores comprende una o más secuencias marcadoras.

En algunas realizaciones, la secuencia diana corta comprende un polimorfismo mononucleotídico (SNP) y la secuencia diana larga comprende una repetición en tándem. En algunas realizaciones, dichas una o más  
5 secuencias marcadoras comprenden una etiqueta de cebador, una etiqueta de captura, una etiqueta de secuenciación, una etiqueta de identificador de molécula única o una combinación de las mismas. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una baja temperatura de fusión  
10 y/o tiene una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una temperatura de fusión que es menor de 60 grados centígrados. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una temperatura de fusión que es de aproximadamente 50 grados centígrados a aproximadamente 60 grados  
15 centígrados. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de aproximadamente 24 nucleótidos a aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una secuencia de nucleótidos  
20 homopolimérica. En algunas realizaciones, la muestra de ácido nucleico se amplifica mediante reacción en cadena de la polimerasa (PCR). En algunas realizaciones, el SNP es un SNP de ascendencia, un SNP fenotípico, un SNP de identidad o una combinación de los mismos. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 30 SNP. En  
25 algunas realizaciones, la pluralidad de cebadores hibrida específicamente con

al menos 50 SNP. En algunas realizaciones, la repetición en tándem es una repetición en tándem corta (STR), una repetición en tándem intermedia (ITR), o una variante de las mismas. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 24 secuencias de repetición en tándem. En algunas realizaciones, la pluralidad de cebadores hibrida específicamente con al menos 60 secuencias de repetición en tándem

Las realizaciones desveladas en el presente documento proporcionan kits que comprenden al menos un medio de envase, en el que dicho al menos un medio de envase comprende una pluralidad de cebadores desvelados en el presente documento.

En algunas realizaciones, los kits incluyen un reactivo para una reacción de amplificación. En algunas realizaciones, el reactivo es un tampón de amplificación para la reacción en cadena de la polimerasa (PCR). En algunas realizaciones, el tampón de amplificación comprende una concentración de sal que está aumentada en comparación con la concentración de sal de un tampón de amplificación usado junto con cebadores diseñados convencionalmente. En algunas realizaciones, la sal comprende KCl, LiCl, NaCl o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor de aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La FIG. 1A y FIG. 1B muestran las diferencias en el A) flujo de trabajo actual para el perfilado de ADN en comparación con B) el de una realización a modo de ejemplo de la presente divulgación.

La FIG. 2 muestra una realización a modo de ejemplo de un método para  
5 crear una biblioteca útil para el perfilado de ADN.

La FIG. 3 muestra otra realización a modo de ejemplo de un método para la creación de una biblioteca útil para el perfilado de ADN.

La FIG. 4A, FIG. 4B, FIG. 4C y FIG. 4D son gráficos de líneas que ilustran resultados de electroferogramas de cómo un par de cebadores diseñado por  
10 métodos tradicionales y siguiendo protocolos de diseño de cebadores de PCR y limitaciones establecidas puede causar la amplificación no específica de dianas genómicas y el empeoramiento de la detección de amplicones deseados cuando se combinan con cebadores diseñados siguiendo los métodos de la presente divulgación; A) 10 pares de cebadores diseñados por  
15 los métodos de la presente divulgación dirigidos a locus de SNP, B) y D) los 10 cebadores más un par de cebadores adicional diseñado por métodos tradicionales que demuestran que el par de cebadores adicional interfiere con los 10 pares de cebadores durante la amplificación, y C) los 10 pares de cebadores más un par de cebadores adicional, en el que el par de cebadores  
20 adicional también se diseñó siguiendo los métodos de la presente divulgación, obteniéndose como resultado una amplificación satisfactoria de todos los SNP diana. El eje X es el tamaño de los fragmentos de la biblioteca (pb) y el eje Y son las unidades de fluorescencia (UF) de los picos amplificados de los fragmentos amplificados.

25 La FIG. 5A, FIG. 5B, FIG. 5C, FIG. 5D y FIG. 5E son que muestran

resultados a modo de ejemplo de un experimento que siguió el flujo de trabajo indicado en la FIG. 2, que se usó para identificar un panel de 56 STR y mezcla de 75 SNP informativos de identidad (iSNP), SNP informativos de ascendencia (aSNP) y SNP informativos de fenotipo (pSNP) en una reacción de  
5 amplificación multiplex y secuenciación a partir de una muestra. Se presentan resultados duplicados que demuestran la amplificación y secuenciación satisfactorias de los loci de STR del panel; A) diagrama de caja que demuestra el equilibrio intralocus para 25 STR heterocigotos del panel, B) diagrama de caja que demuestra el bajo tartamudeo para la mayor parte de los 56 loci de  
10 STR, C) diagrama de caja que demuestra el alcance de secuenciación para los loci de STR; D) diagrama de caja que demuestra el alcance de secuencia para los SNP y E) diagrama de caja que demuestra el equilibrio para 22 SNP heterocigotos del panel. La barra de error inferior indica el valor mínimo, la barra de error superior indica el valor máximo, la caja inferior indica el percentil  
15 25 y la caja superior indica el percentil 75, siendo la media la intersección entre las cajas inferior y superior.

La FIG. 6 muestra una serie de gráficos de barras que muestran las representaciones gráficas de loci de STR a modo de ejemplo del experimento de la FIG. 5. Las representaciones gráficas muestran diferentes validaciones  
20 alélicas para los STR en el panel de la FIG. 5.

La FIG. 7A, FIG. 7B, FIG. 7C, FIG. 7D y FIG. 7E son diagramas de cajas que muestran resultados a modo de ejemplo de un experimento que siguió el flujo de trabajo indicado en la FIG. 3, que se usó para identificar un panel de 26 STR y una mezcla de 94 iSNP, aSNP y pSNP en una reacción de amplificación  
25 multiplex y secuenciación a partir de una muestra. Se presentan resultados

duplicados que demuestran la amplificación y secuenciación satisfactoria de los STR del panel; A) diagrama de caja que demuestra equilibrio intralocus para 21 loci de STR heterocigotos del panel, B) diagrama de caja que demuestra el bajo tartamudeo para los 26 loci de STR (39 de los 47 alelos de los 26 loci no mostraron tartamudeo), C) diagrama de caja que demuestra el alcance de la secuenciación para los loci de STR (léanse los números normalizados usando los UMI), D) diagrama de caja que demuestra el alcance de la secuencia para los SNP y E) diagrama de caja que demuestra equilibrio para 21 iSNP heterocigotos del panel. La barra de error inferior indica el valor mínimo, la barra de error superior indica el valor máximo, la caja inferior indica el percentil 25 y la caja superior indica el percentil 75, siendo la media la intersección entre las cajas inferior y superior.

La FIG. 8 muestra una serie de gráficos de barras que muestran las representaciones gráficas de loci de STR a modo de ejemplo del experimento de la FIG. 7. Las representaciones gráficas muestran diferentes validaciones alélicas para los STR en el panel de la FIG. 7.

La FIG. 9 representa gráficos de barras de muestras analizadas sin UMI y con UMI. El panel izquierdo de cada conjunto representa muestras analizadas sin UMI y el panel derecho para cada conjunto representa muestras analizadas con UMI. El eje X designa el número de repeticiones de la STR y el eje Y designa el número de recuento del alelo particular. Las líneas de error dentro de las barras separan el error de secuenciación (parte superior de la barra) de la secuencia correcta (parte inferior de la barra) dentro de la secuencia de STR.

Las FIG. 10A y 10B muestran resultados a modo de ejemplo de experimento en el que la proporción de ADN mujer:hombre fue de 90:10. A) Se

obtiene un subconjunto de validaciones de loci de STR para loci de STR cuando se usan métodos actuales de perfilado de ADN por electroforesis capilar y B) se obtienen varias validaciones de loci de STR para varios loci de STR cuando se usan los métodos de la presente solicitud. Tanto los métodos  
5 de EC como los métodos de la presente solicitud detectaron el nivel inferior de contaminación de ADN masculino.

La FIG. 11 muestra gráficos de barra que muestran que, en el experimento de la FIG. 9 se detectaron loci de STR específicos para el cromosoma Y, demostrando adicionalmente que la presente solicitud puede  
10 detectar ADN contaminante masculino y loci de STR específicos de ese ADN masculino, mientras que se necesitaría realizar dos experimentos con las metodologías actuales de EC actuales para hacer esto mismo.

La FIG. 12 es una tabla que muestra resultados de secuenciación de alto nivel a modo de ejemplo de un experimento usando 12 individuos de muestra y  
15 un individuo de referencia, que demuestra la coherencia de STR y validaciones de SNP entre dos copias.

La FIG. 13 es una tabla que muestra una estadística de población a modo de ejemplo del experimento mostrado en la FIG. 12.

La FIG. 14 es una tabla que muestra predicciones de fenotipo a modo de ejemplo basadas en el genotipo de pSNP del experimento mostrado en la FIG.  
20 12.

La FIG. 15 es un gráfico que muestra un mapeo de ascendencia a modo de ejemplo basado en el genotipo de aSNP del experimento mostrado en la FIG. 12.

25 La FIG. 16A, FIG. 16B, FIG. 16C, FIG. 16D y FIG. 16E son gráficos de

barras que muestran las representaciones gráficas a modo de ejemplo de loci de STR del experimento de la FIG. 12.

La FIG. 17A y FIG. 17B son gráficos de barras que muestran representaciones gráficas de SNP a modo de ejemplo del experimento de la  
5 FIG. 12.

La FIG. 18A y FIG 18B muestran representaciones de caja que muestran el equilibrio intralocus de loci de STR y SNP a modo de ejemplo del experimento de la FIG. 12.

La FIG. 19A y FIG. 19B son gráficos que muestran el análisis de  
10 tartamudeo de loci de STR a modo de ejemplo del experimento de la FIG. 12.

La FIG. 20 es una tabla que muestra heterocigotos isométricos a modo de ejemplo en loci de STR del experimento de la FIG. 12.

La FIG. 21 es un diagrama de bloques que muestra una representación gráfica de herencia a modo de ejemplo basada en variantes en la STR  
15 D8S1179 del experimento de la FIG. 12.

La FIG. 22 es un diagrama de bloques que muestra una representación gráfica de herencia a modo de ejemplo basada en variantes en la STR D13S317 del experimento de la FIG. 12.

La FIG. 23 es una tabla que muestra resultados de genotipado a modo de  
20 ejemplo usando ADN degradado.

La FIG. 24A y FIG. 24B muestran resultados de genotipado de STR a modo de ejemplo y el equilibrio intralocus a diferentes entradas de ADN.

La FIG. 25A y FIG. 25B muestran resultados de genotipado de SNP a modo de ejemplo y el equilibrio intralocus a entradas de ADN diferentes.

25 DESCRIPCIÓN DETALLADA

### Definiciones

Todas las patentes, solicitudes, solicitudes publicadas y otras publicaciones citadas en el presente documento se incorporan por referencia al material mencionado y en su totalidad. Si un término o frase se usa en el presente documento de tal manera que es contrario o incoherente de otro modo con una definición expuesta en las patentes, solicitudes, solicitudes publicadas y otras publicaciones que se incorpora en el presente documento por referencia, prevalece el uso en el presente documento sobre la definición que se incorporada en el presente documento por referencia.

10 Como se usa en el presente documento, las formas en singular “un”, “uno”, “una” y “el”, “la” incluyen las referencias en plural salvo que se indique lo contrario, expresamente o por el contexto. Por ejemplo, “un” dímero incluye uno o más dímeros, salvo que se indique lo contrario, expresamente o por el contexto.

15 Como se usa en el presente documento, las expresiones “perfil de ADN”, “huella genética” y “perfil genotípico” se usan indistintamente en el presente documento para referirse a variaciones alélicas en un conjunto de loci polimórficos, tales como una repetición en tándem, un polimorfismo mononucleotídico (SNP), etc. Un perfil de ADN es útil en la medicina forense para identificar a un individuo basándose en una muestra de ácido nucleico. El perfil de ADN como se usa en el presente documento también puede usarse para otras aplicaciones, tales como diagnóstico y pronóstico de enfermedades incluyendo cáncer, identificación de biomarcadores de cáncer, análisis de herencia, análisis de diversidad genética, identificación de anomalías genéticas, creación de bases de datos, medicina forense, estudios de

20

25

criminología, pruebas de paternidad, identificación de personas, etc.

Los términos “polinucleótido”, “oligonucleótido”, “ácido nucleico” y “molécula de ácido nucleico” se usan indistintamente en el presente documento para referirse a una forma polimérica de nucleótidos de cualquier longitud y pueden comprender ribonucleótidos, desoxirribonucleótidos, análogos de los mismos o mezclas de los mismos. Este término se refiere solo a la estructura primaria de la molécula. Por tanto, el término incluye ácido desoxirribonucleico (“ADN”) triple, doble y monocatenario, así como a ácido ribonucleico (“ARN”) triple, doble y monocatenario.

Como se usa en el presente documento, “identidad de secuencia” o “identidad” o “homología” en el contexto de dos secuencias de nucleótidos incluyen la referencia a los restos en las dos secuencias que son iguales cuando se alinean para su máxima correspondencia sobre una ventana de comparación específica. La parte de la secuencia de nucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir huecos) en comparación con la secuencia de referencia para conseguir el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que aparece el resto de base de ácido nucleico idéntico en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes entre el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para dar el porcentaje de identidad de secuencia.

Como se usa en el presente documento, “sustancialmente complementario o sustancialmente coincidente” significa que las dos secuencias de ácido nucleico tienen una identidad de secuencia de al menos

90 %. Preferentemente, las dos secuencias de ácido nucleico tienen al menos una identidad de secuencia de 95 %, 96 %, 97 %, 98 %, 99 % o 100 %. Como alternativa, “sustancialmente complementario” o “sustancialmente coincidente” significa que las dos secuencias de ácido nucleico pueden hibridar en  
5 condiciones de alta rigurosidad.

Se entiende que los aspectos y las realizaciones de la invención descritas en el presente documento incluyen aspectos y realizaciones “que consisten en” y/o “que consisten esencialmente en”.

Otros objetivos, ventajas y características de la presente invención  
10 resultarán evidentes a partir de la siguiente memoria descriptiva considerada en conjunto con los dibujos adjuntos.

#### Métodos para construir un perfil de ADN

Las metodologías establecidas para determinar el perfil de ADN están limitadas de diversas maneras. Por ejemplo, los métodos actuales detectan  
15 cambios de tamaño de los loci amplificados que difieren debido a cambios en las longitudes de secuencias repetidas en tándem encontradas en una muestra de ADN. Para las amplificaciones de STR multiplex para la visualización, las amplificaciones deben diseñarse para separar los diferentes tamaños de amplicones dentro de los límites de separación de tamaño del sistema  
20 electroforético, que para la EC es de aproximadamente 50-500 pb. De esta manera, únicamente puede visualizarse en un ensayo un número limitado de las secuencias repetidas. Por ejemplo, el kit de amplificación GLOBALFILER PCR (APPLIED BIOSYSTEMS) puede diferenciar supuestamente 24 loci de STR usando 6 colorantes diferentes. Adicionalmente, dichos métodos tienen  
25 inconvenientes cuando la muestra de ADN está degradada, como es habitual

en las muestras de ADN de una escena de un crimen, de tal manera que no es posible obtener productos de amplificación largos, lo cual tiene como resultado un perfil de ADN incompleto. Los métodos actuales algunas veces tampoco son suficientemente sensibles como para detectar pequeñas cantidades de ADN contaminante, de tal manera que puede no detectarse o no identificarse una mezcla mezclada, lo que podría ser crítico para un estudio de criminología. Por lo tanto, los métodos actuales pueden conducir a resultados incompletos que conducen a resultados no concluyentes, lo cual puede ser perjudicial para el perfilado del ADN.

Adicionalmente, las dianas actuales no incluyen información sobre la ascendencia de la muestra, rasgos fenotípicos tales como el posible color de ojos y otra información de la muestra individualizada. Algunas metodologías de secuenciación han intentado incluir la detección tanto de STR como de SNP. Por ejemplo, se ha intentado la preparación de bibliotecas seguida del enriquecimiento a medida para STR y SNP, sin embargo, no todos los STR quedan completamente cubiertos ya que los métodos de preparación de bibliotecas normalmente implican el corte de muestras que puede eliminar la secuencia diana. Adicionalmente, los métodos y protocolos de diseño de cebadores establecidos pueden proporcionar conjuntos de cebadores para la amplificación de secuencias largas (por ejemplo, STR) o secuencias cortas (por ejemplo, SNP), aunque no se han conseguido con éxito las combinaciones de ambas en una reacción.

La presente divulgación describe soluciones a los problemas y limitaciones de los sistemas de perfilado de ADN actuales. Los métodos y composiciones descritos en el presente documento permiten la combinación de

STR y SNP en un ensayo usando PCR para amplificar las dianas y generar bibliotecas para secuenciación. Mientras se desarrollaban los presentes ensayos, se descubrió inesperadamente que, por ejemplo, cuando se utilizaba un diseño de cebadores no convencional y contraintuitivo, tanto las SPR como

5 los SNP pueden amplificarse en una reacción que permite determinar la secuencia para todos los loci diana. Sorprendentemente, cuando se diseñan cebadores para amplificación usando parámetros contrarios al dogma actual en relación con el diseño de cebadores, se crearon cebadores que permitían amplificar regiones de STR más largas y amplificar las regiones de SNP más

10 cortas de una manera más o menos equilibrada, permitiendo de este modo la amplificación multiplex conjunta tanto de las STR como de los SNP.

Los métodos y composiciones desvelados en el presente documento para determinar el perfil de ADN de un organismo pueden usarse en cualquiera de los conjuntos de amplicones de diferente tamaño que se desee a partir de una

15 reacción de amplificación fuera del perfilado de ADN. Por ejemplo, si las dianas de interés para la PCR incluyen tanto regiones de genes largos como regiones de SNP cortas que puedan dar como resultado amplicones que varían en cuanto al tamaño de cientos a miles de pares de bases frente a amplicones de menos de 100 pares de bases, respectivamente, entonces los métodos y

20 composiciones descritos en el presente documento permitirían la amplificación simultánea exitosa del gen y dianas de SNP, lo cual no habría sido posible sin llevar a la práctica los métodos desvelados. Adicionalmente, los métodos y composiciones desvelados en el presente documento pueden aplicarse a cualquier organismo, por ejemplo, a seres humanos, primates no humanos,

25 animales, plantas, virus, bacterias, hongos y similares. De esta manera, los

presentes métodos y composiciones no solo son útiles para el perfilado de ADN (por ejemplo, análisis forense, paternidad, identificación de individuos, etc.) y para los seres humanos como genoma diana, sino que también se podrían usar para otras dianas tales como marcadores de cáncer y de enfermedades, marcadores de anomalías genéticas y/o cuando el genoma diana no sea un genoma humano.

Por lo tanto, las realizaciones desveladas en el presente documento proporcionan métodos para construir un perfil de ADN que comprende: proporcionar una muestra de ácido nucleico, amplificar la muestra de ácido nucleico con una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo mononucleotídico (SNP) y al menos una secuencia diana que comprende una repetición en tándem, y determinar los genotipos de dicho al menos un SNP y al menos una repetición en tándem en los productos de amplificación, construyendo de este modo el perfil de ADN de la muestra de ácido nucleico.

Los expertos en la materia apreciarán que puede usarse cualquier técnica adecuada para determinar los genotipos de las secuencias diana incluyendo, pero sin limitación, hibridación basada en matriz, secuenciación o similar. Por lo tanto, en algunas realizaciones, los métodos desvelados en el presente documento pueden comprender la generación de una biblioteca de ácidos nucleicos, tal como una biblioteca de secuenciación, a partir de los productos de amplificación, y determinar las secuencias de la biblioteca de ácidos nucleicos.

En algunas realizaciones, la presente divulgación proporciona métodos y composiciones para el perfilado de ADN que comprenden la identificación

simultánea de STR e iSNP, por ejemplo, para su uso en la creación de bases de datos de población o personales. En dichas bases de datos, no se requieren necesariamente los datos personales, ya que normalmente los individuos son conocidos. Sin embargo, si se desea información adicional, entonces pueden  
5 añadirse dianas de información adicionales para la identificación simultánea. Las repeticiones en tándem cortas son muy conocidas en la técnica y consisten en secuencias di- o trinucleotídicas repetidas. Las repeticiones en tándem intermedias se consideran normalmente secuencias repetidas de entre 4 y 7 secuencias de nucleótidos. Los SNP utilizados en el presente documento  
10 pueden ser de cualquier forma que pueda ofrecer una idea de las características físicas de una persona. Los ejemplificados en el presente documento son SNP que proporcionan pistas de la ascendencia o herencia (aSNP) y los que proporcionan pistas de características fenotípicas (SNP informativas de fenotipo). En los métodos descritos en el presente documento,  
15 un ensayo de perfil de ADN incluiría cualquier número de estos SNP en combinación con determinaciones de loci de STR e ITR.

Por ejemplo, la presente divulgación proporciona métodos y composiciones adicionales en los que, junto con las STR e iSNP, se incluyen dianas adicionales. Si se desea más información sobre un individuo, por  
20 ejemplo, cuando una muestra pertenece a un individuo o a un grupo de individuos desconocido como puede ser el caso en causas criminales, los otros marcadores de información se pueden añadir a las STR y iSNP, tales como SNP relacionados con la ascendencia (aSNP) y SNP relacionados con variantes fenotípicas (SNP informativos de fenotipo). La información adicional  
25 puede por tanto usarse para ayudar a los investigadores, por ejemplo, al

proporcionar perspectivas sobre la herencia de un individuo desconocido, el color de ojos, el color de pelo y similares. Por lo tanto, la adición de toda la información combinada puede proporcionar un perfil de ADN más completo de un individuo que no se conocía hasta ahora usando los métodos actuales de  
5 perfilado de ADN.

Los métodos y composiciones que se desvelan en el presente documento están diseñados para ser suficientemente sensibles para detectar cantidades inferiores a un nanogramo de moléculas de ácido nucleico. Adicionalmente, los métodos y composiciones desvelados en el presente documento pueden ser  
10 útiles para amplificar una muestra de ácido nucleico preparada que tiene moléculas de ácido nucleico de baja calidad, tal como ADN genómico degradado y/o fragmentado procedente de una muestra forense. La muestra de ácido nucleico puede ser una muestra purificada o un lisado que contenga ADN bruto, por ejemplo, procedente de un frotis bucal, papel, tejido u otro sustrato  
15 que pueda impregnarse con saliva, sangre u otros fluidos corporales. De esta manera, en algunas realizaciones, la muestra de ácido nucleico puede comprender pequeñas cantidades, o partes fragmentadas, de ADN, tal como ADN genómico. Por ejemplo, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es, es  
20 aproximadamente, o es menor que, 1 pg, 2 pg, 3 pg, 4 pg, 5 pg, 6 pg, 7 pg, 8 pg, 9 pg, 10 pg, 11 pg, 12 pg, 13 pg, 14 pg, 15 pg, 16 pg, 17 pg, 18 pg, 19 pg, 20 pg, 30 pg, 40 pg, 50 pg, 60 pg, 70 pg, 80 pg, 90 pg, 100 pg, 200 pg, 300 pg, 400 pg, 500 pg, 600 pg, 700 pg, 800 pg, 900 pg, 1 ng, 10 ng, 100 ng, o está en un intervalo definido por dos valores cualesquiera de estos, por ejemplo, de 10  
25 pg a 100 pg, de 10 pg a 1 ng, de 100 pg a 1 ng, de 1 ng a 10 ng, de 10 ng a

100 ng, etc. en algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es aproximadamente de 100 pg a aproximadamente 1 ng. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de  
5 ácido nucleico (por ejemplo, ADN genómico) que es más de aproximadamente 62,5 pg. En algunas realizaciones, en los procedimientos de fragmentación no se incluyen etapas de fragmentación adicionales, tales como ultrasonidos o digestión con endonucleasas.

En algunas realizaciones, los métodos y composiciones que se desvelan  
10 en el presente documento pueden determinar de manera satisfactoria los genotipos de una o más de las secuencias diana, por ejemplo, SNP, STR, etc., incluso con cantidades inferiores a un nanogramo de y/o muestras de ácido nucleico degradado. Por ejemplo, los métodos y composiciones desvelados en el presente documento pueden determinar satisfactoriamente el genotipo que  
15 es, es aproximadamente o es más de, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 %, o un intervalo entre dos cualesquiera de los valores anteriores, de las secuencias diana. En algunas realizaciones, los métodos y composiciones desvelados en el presente documento pueden determinar satisfactoriamente el genotipo de más de  
20 aproximadamente 50 %, 80 %, 90 %, 95 %, 98 % o más de las secuencias diana. En algunas realizaciones, los métodos y composiciones desvelados en el presente documento pueden conseguir un equilibrio intralocus de más de aproximadamente 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 %, o un intervalo entre dos cualesquiera de  
25 los valores anteriores, de las secuencias diana.

Para la investigación forense, la pluralidad de cebadores puede incorporar identificadores de molécula única (UNI) que ayudan a eliminar, por ejemplo, errores en la PCR y en la secuenciación, tartamudeos y similares de los resultados de secuenciación. Véase Kivioja *et al.*, mencionado anteriormente.

5 Como se ha indicado con más detalle en otras partes de esta divulgación, la inclusión de UMI en los cebadores también permite la identificación de variantes dentro de loci de repeticiones en tándem, aumentando de este modo la utilidad de los métodos y composiciones actuales para el perfilado de ADN y otros propósitos tales como análisis de herencia.

10 Por consiguiente, en algunas realizaciones, los genotipos de las secuencias de repetición en tándem desveladas en el presente documento pueden incluir variantes de secuencia dentro de los loci de repetición en tándem. Por lo tanto, un homocigoto para una repetición en tándem (por ejemplo, 13, 13 para D9S1122) usando el método tradicional puede  
15 identificarse como un heterocigoto isométrico basándose en variantes de secuencia dentro de la repetición en tándem. Como apreciarán los expertos en la materia, teniendo en cuenta las variantes de secuencia intralocus, se podría aumentar enormemente la utilidad de los métodos desvelados en el presente documento, por ejemplo, para análisis de herencia.

## 20 Métodos para construir una biblioteca de ácido nucleico

Las realizaciones desveladas en el presente documento proporcionan métodos de construcción de una biblioteca de ácidos nucleicos que comprende: proporcionar una muestra de ácido nucleico y amplificar la muestra de ácido nucleico con una pluralidad de cebadores que hibridan  
25 específicamente con al menos una secuencia diana que comprende un

polimorfismo mononucleotídico (SNP) y al menos una secuencia diana que comprende una secuencia de repetición en tándem.

Los métodos y composiciones desvelados en el presente documento están diseñados para ser suficientemente sensibles como para detectar  
5 cantidades inferiores a un nanogramo de moléculas de ácido nucleico. Adicionalmente, los métodos y composiciones desvelados en el presente documento pueden ser útiles para amplificar una muestra de ácido nucleico que consiste en moléculas de ácido nucleico de baja calidad, tales como ADN genómico degradado y/o fragmentado de una muestra forense. La muestra de  
10 ácido nucleico puede ser ADN purificado o un lisado que contiene ADN bruto, por ejemplo, procedente de un frotis bucal, papel, tejido u otro sustrato que pueda impregnarse con saliva, sangre u otros fluidos corporales. De esta manera, en algunas realizaciones, la muestra de ácido nucleico puede comprender bajas cantidades de ADN fragmentado, tal como ADN genómico.

15 Por ejemplo, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es, es de aproximadamente, o es menor que 1 pg, 2 pg, 3 pg, 4 pg, 5 pg, 6 pg, 7 pg, 8 pg, 9 pg, 10 pg, 11 pg, 12 pg, 13 pg, 14 pg, 15 pg, 16 pg, 17 pg, 18 pg, 19 pg, 20 pg, 30 pg, 40 pg, 50 pg, 60 pg, 70 pg, 80 pg, 90 pg, 100 pg, 200 pg, 300 pg, 400 pg, 500 pg, 600  
20 pg, 700 pg, 800 pg, 900 pg, 1 ng, 10 ng, 100 ng, o está en un intervalo definido por dos cualesquiera de estos valores, por ejemplo, de 10 pg a 100 pg, de 10 pg a 1 ng, de 100 pg a 1 ng, de 1 ng a 10 ng, de 10 ng a 100 ng, etc. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es de  
25 aproximadamente 100 pg a aproximadamente 1 ng. En algunas realizaciones,

la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es mayor de aproximadamente 62,5 pg. En algunas realizaciones, no se incluyen etapas de fragmentación adicionales, tales como ultrasonido o digestión con endonucleasas.

5        En algunas realizaciones, los métodos desvelados en el presente documento comprenden amplificación y preparación de bibliotecas en previsión de la secuenciación en paralelo posterior. Un ensayo puede incluir dos mezclas maestras de PCR, dos polimerasas termoestables, dos mezclas de cebadores y adaptadores de biblioteca. En algunas realizaciones, puede amplificarse una  
10 muestra de ADN para varios ciclos usando un primer conjunto de cebadores de amplificación que comprenden regiones específicas de diana y regiones marcadoras no específicas de diana y una primera mezcla maestra de PCR. La región marcadora puede ser cualquier secuencia, tal como una región marcadora universal, una región marcadora de captura, una región marcadora  
15 de amplificación, una región marcadora de secuenciación, una región marcadora de UMI y similar. Por ejemplo, una región marcadora puede ser el molde para los cebadores de amplificación utilizados en una segunda o posterior ronda de amplificación, por ejemplo, para la preparación de bibliotecas. En algunas realizaciones, los métodos comprenden añadir una  
20 proteína ligante de unión a ADN monocatenario (SSB) a los primeros productos de amplificación. Una alícuota de la primera muestra amplificada puede retirarse y amplificarse una segunda vez usando un segundo conjunto de cebadores de amplificación que son específicos para la región marcadora, por ejemplo, una región marcadora universal o una región marcadora de  
25 amplificación, de los primeros cebadores de amplificación que pueden

comprender una o más secuencias marcadoras adicionales, tales como etiquetas de secuencia específicas para uno o más flujos de trabajo de secuenciación aguas abajo, y la misma o segunda mezcla maestra de PCR. De esta manera, una biblioteca de la muestra de ADN original está lista para la  
5 secuenciación.

Un método alternativo comprendería la primera amplificación que se realiza en un volumen pequeño (por ejemplo, 15 ul) y, en lugar de transferir una alícuota a una nueva localización para una segunda ronda de amplificación, al tubo se le pueden añadir reactivos adicionales para realizar una segunda ronda  
10 de amplificación.

Una vez creada la biblioteca, esta puede purificarse y cuantificarse. En algunos ejemplos, la purificación puede realizar procesando la muestra a través de un sustrato tal como Perlas AMPURE XP (Beckman Couiter) que sirven para purificar los fragmentos de ADN de los componentes de reacción. Otro  
15 método podría ser la incorporación de un resto de purificación, tal como un resto de hapteno, en el segundo conjunto de cebadores de amplificación. Por ejemplo, si se incorporara una biotina en uno de los cebadores del segundo conjunto de cebadores de amplificación, entonces los fragmentos de la biblioteca podrían capturarse usando un resto de estreptavidina en una perla,  
20 por ejemplo. Utilizando la estrategia de captura, las bibliotecas también podrían normalizarse y cuantificarse usando Normalización Basada en Perlas (BBN). Sin embargo, las bibliotecas pueden purificarse y cuantificarse, o agruparse y cuantificarse si se están realizando reacciones múltiples, sin el uso de BBN. Por ejemplo, las bibliotecas podrían también cuantificarse mediante métodos  
25 electroforéticos en gel, BioAnalyzer, qPCR, métodos espectrofotométricos, kits

de cuantificación (por ejemplo, PicoGreen, etc.) y similares, como se sabe en la técnica. Después de la cuantificación, la biblioteca puede secuenciarse por secuenciación en paralelo.

En algunas realizaciones, el primer conjunto de cebadores de  
5 amplificación usado para amplificar un ADN diana se proporciona en una  
concentración tan ilimitada que cuando se añade una alícuota de la primera  
reacción de amplificación a un nuevo tubo y se añaden los reactivos de la  
segunda reacción de amplificación, hay una amplificación acumulada mínima o  
indetectable resultante del primer conjunto de cebadores de amplificación y no  
10 se requiere una etapa de limpieza entre la primera reacción de amplificación y  
la segunda reacción de amplificación. En algunos ejemplos, la concentración  
de los cebadores de amplificación para una primera PCR es, es de  
aproximadamente, o es menor que, 0,5 nM, 0,6 nM, 0,7 nM, 0,8 nM, 0,9 nM,  
1,0 nM, 1,5 nM, 2,0 nM, 3,0 nM, 4,0 nM, 5,0 nM, 6,0 nM, 7,0 nM, 8,0 nM, 9,0  
15 nM, 10,0 nM, 11,0 nM, 12,0 nM, o un intervalo entre cualquiera de estos  
valores, por ejemplo, de 0,5 nM a 1,0 nM, de 1,0 nM a 12 nM, de 0,8 nM a 1,5  
nM, etc. En algunas realizaciones, la concentración de los cebadores de  
amplificación para una primera PCR es de aproximadamente 0,9 nM a  
aproximadamente 10 nM.

20 La FIG. 2 muestra un flujo de trabajo a modo de ejemplo de los métodos  
descritos en el presente documento en una realización. Una secuencia de ADN  
genómico diana se amplifica usando un primer conjunto de cebadores que  
comprenden una región que flanquea la secuencia diana y regiones  
marcadoras de amplificación (que puede ser iguales o diferentes) dando como  
25 resultado amplicones que comprenden la secuencia diana y etiquetas en

ambos extremos. Una alícuota de los amplicones de la primera PCR se amplifica adicionalmente usando un segundo conjunto de cebadores específicos para las primeras secuencias marcadoras que comprenden adicionalmente secuencias de cebadores de secuenciación (secuencias adaptadoras i5 e i7), generándose así una biblioteca que comprende la secuencia de ADN diana flanqueada por secuencias usadas en secuenciación en paralelo, en este caso se utilizan secuencias i5 e i7 en este orden por métodos de síntesis divulgados por Illumina, Inc.

En la FIG. 3 se describe un ejemplo de un flujo de trabajo alternativo para determinar un perfil de ADN a partir de una muestra. En este ejemplo, se amplifica una diana de ADN con un primer par de cebadores que comprenden secuencias que flanquean la secuencia diana, secuencias etiqueta no diana (iguales o diferentes) y además secuencias identificadoras moleculares únicas o UMI, que comprenden bases aleatorias. Las UMI pueden usarse, por ejemplo, para reducir o eliminar, mediante la bioinformática, errores que se producen durante los procesos de preparación de la biblioteca (por ejemplo, artefactos de PCR o incorporaciones erróneas, etc.). El uso de los UMI puede ser importante para el perfilado de ADN, pero son de particular importancia para ayudar a eliminar errores cuando se secuencian muestras para investigación criminal. En este ejemplo, la primera ronda de amplificación se realiza durante 2 ciclos, que va seguida por la adición de una proteína ligante de unión a ADN monocatenario (SSB) e incubación a 37 grados centígrados durante 15 min seguido de una inactivación a 95 grados C/5 min que inactiva de un modo eficaz la amplificación adicional del primer conjunto de cebadores de amplificación durante la segunda ronda de amplificación. Aunque el

mecanismo se desconoce, se contempla que la adición de SSB une irreversiblemente los primeros cebadores de amplificación monocatenarios e impide que participen en las reacciones de amplificación posteriores. Después de la incubación con SSB, se añade un segundo conjunto de cebadores que  
5 comprenden etiquetas de secuencia y una segunda mezcla de PCR, obteniéndose la biblioteca de secuenciación.

#### Biblioteca de ácidos nucleicos

Las realizaciones desveladas en el presente documento proporcionan bibliotecas de ácido nucleico, que pueden usarse para secuenciación. En  
10 algunas realizaciones, las bibliotecas de ácido nucleico desveladas en el presente documento pueden comprender una pluralidad de moléculas de ácido nucleico, en las que la pluralidad de moléculas de ácido nucleico comprende al menos una secuencia de repetición en tándem flanqueada por un primer par de secuencias marcadoras y al menos una secuencia de polimorfismo  
15 mononucleotídico (SNP) flanqueada por un segundo par de secuencias marcadoras.

Como se indica en el presente documento, el tamaño de las moléculas de ácido nucleico puede variar enormemente usando los métodos y las composiciones desveladas en el presente documento. Los expertos en la  
20 materia apreciarán que las moléculas de ácido nucleico amplificadas a partir de una secuencia diana que comprenden una repetición en tándem (por ejemplo, STR) pueden tener un gran tamaño, mientras que las moléculas de ácido nucleico amplificadas a partir de una secuencia diana que comprende un SNP pueden tener un tamaño pequeño. Por ejemplo, las moléculas de ácido  
25 nucleico pueden comprender desde menos de cien nucleótidos a cientos o

incluso miles de nucleótidos. Por lo tanto, el tamaño de las moléculas de ácido nucleico puede tener un intervalo que esté entre dos valores cualesquiera de aproximadamente 50 pb, aproximadamente 60 pb, aproximadamente 70 pb, aproximadamente 80 pb, aproximadamente 90 pb, aproximadamente 100 pb, aproximadamente 110 pb, aproximadamente 120 pb, aproximadamente 130 pb, de aproximadamente 140 pb, aproximadamente 150 pb, aproximadamente 200 pb, aproximadamente 300 pb, aproximadamente 400 pb, aproximadamente 500 pb, aproximadamente 600 pb, aproximadamente 700 pb, aproximadamente 800 pb, aproximadamente 900 pb, aproximadamente 1 kb, o más. En algunas realizaciones, el tamaño mínimo de las moléculas de ácido nucleico puede tener una longitud que es, es de aproximadamente, o es menor que, 50 pb, 60 pb, 70 pb, 80 pb, 90 pb o 100 pb. En algunas realizaciones, el tamaño máximo de las moléculas de ácido nucleico puede tener una longitud que es, es de aproximadamente, o es mayor que 100 pb, 150 pb, 200 pb, 250 pb, 300 pb, 350 pb, 400 pb, 450 pb, 500 pb o 1 kb.

Para la generación de grupos, los fragmentos de la biblioteca se inmovilizan en un sustrato, por ejemplo, un portaobjetos, que comprende secuencias de oligonucleótidos homólogas para capturar e inmovilizar los fragmentos de la biblioteca de ADN. Los fragmentos de la biblioteca de ADN inmovilizados se amplifican usando metodologías de amplificación en grupo como ilustran las divulgaciones de las Patentes de Estados Unidos n.º 7.985.565 y 7.115.400, cuyos contenidos se incorporan en el presente documento por referencia en su totalidad. Los materiales incorporados de las Patentes de Estados Unidos n.º 7.985.565 y 7.115.400 describen métodos de amplificación de ácido nucleico en fase sólida que permiten que los productos

de amplificación se inmovilicen sobre un soporte sólido para formar matrices compuestas por grupos o “colonias” de moléculas de ácido nucleico inmovilizadas. Cada grupo o colonia sobre dicha matriz se forma a partir de una pluralidad de cadenas polinucleotídicas inmovilizadas idénticas y una pluralidad de cadenas polinucleotídicas complementarias inmovilizadas idénticas. Las matrices formadas de este modo se denominan generalmente “matrices agrupadas”. Los productos de reacciones de amplificación en fase sólida, tales como los descritos en las Patentes de Estados Unidos n.º 7.985.565 y 7.115.400 se denominan estructuras “en puente” formadas por hibridación de pares de cadenas polinucleotídicas inmovilizadas y cadenas complementarias inmovilizadas, estando ambas cadenas inmovilizadas sobre el soporte sólido en el extremo 5', preferentemente mediante una unión covalente. Las metodologías de amplificación en grupo son ejemplos de métodos en los que un molde de ácido nucleico inmovilizado se usa para producir amplicones inmovilizados. También pueden usarse otras metodologías adecuadas para producir amplicones inmovilizados a partir de fragmentos de ADN inmovilizados producidos de acuerdo con los métodos proporcionados en el presente documento. Por ejemplo, pueden formarse uno o más grupos o colonias mediante PCR en fase sólida si uno o los dos cebadores de cada par de cebadores de amplificación están inmovilizados. Sin embargo, los métodos descritos en el presente documento no se limitan a ninguna metodología de preparación de secuenciación o plataforma de secuenciación particular y pueden abordarse con otros métodos de preparación de plataforma de secuenciación en paralelo y plataformas de secuenciación asociadas.

## 25 Cebadores

Las realizaciones desveladas en el presente documento proporcionan una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana corta y al menos una secuencia diana larga en una muestra de ácido nucleico, donde la amplificación de la muestra de ácido nucleico usando la pluralidad de cebadores en una sola reacción multiplex da como resultado al menos un producto de amplificación corto y al menos un producto de amplificación largo, donde cada uno de la pluralidad de cebadores comprende una o más secuencias marcadoras. También se desvela en el presente documento una pluralidad de cebadores que tienen las secuencias expuestas en las Tablas 1-2.

Para la amplificación multiplex de secuencias diana largas (por ejemplo, STR, ITR) y secuencias diana cortas (por ejemplo, SNP), se diseñan cebadores que permitan la amplificación equilibrada en todos los tipos diana. Los métodos y composiciones desvelados en el presente documento pueden usarse para la amplificación de múltiples secuencias diana de repetición en tándem en una sola reacción multiplex. Por ejemplo, la pluralidad de cebadores puede hibridar específicamente con varias secuencias de repetición en tándem que es, es de aproximadamente, o es de más de 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100, o un intervalo entre dos valores cualesquiera, tal como de 4 a 12, de 10 a 24, de 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 24 secuencias de repetición en tándem. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 60 secuencias de repetición en tándem. Los métodos y composiciones desvelados en el presente documento pueden usarse para amplificar múltiples secuencias diana de SNP en una sola

reacción. Por ejemplo, la pluralidad de cebadores puede hibridar específicamente con un número de secuencias de SNP, que es, es de aproximadamente, o es más de 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100 o un intervalo entre dos valores cualesquiera, tal como de 4 a 12, de 10 a 24, de 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 30 secuencias de SNP. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 50 secuencias de SNP.

Durante la experimentación se descubrió que las secuencias diana de SNP cortas se amplificaban preferentemente con respecto a las secuencias diana de STR más largas cuando se usaban cebadores que se diseñaban siguiendo los criterios establecidos y el conocimiento del diseño de cebadores satisfactorio. Adicionalmente, al menos en el flujo de trabajo de secuenciación por síntesis en el que se generan grupos y se secuencian los propios grupos (por ejemplo, cuando se sigue la secuenciación por síntesis (SBS, desvelado en otras partes del presente documento) asociada con los secuenciadores Illumina Inc.) también se produjo amplificación de grupo preferente de los fragmentos de SNP de la biblioteca más corta. Para solucionar estas dos desviaciones, se necesitó una nueva estrategia para el diseño de cebadores que permitiera una amplificación equilibrada entre las secuencias diana de SNP cortas y las secuencias diana de STR largas.

Una de las estrategias incluyó cebadores de diseño para amplificaciones de STR. Con las STR, las secuencias repetidas a menudo se encuentran embebidas dentro de regiones repetidas más grandes; por lo tanto, el diseño de cebadores específicos para la amplificación de STR puede ser problemático.

Adicionalmente, las STR y sus regiones flanqueantes son algunas veces ricas en AT. En un ejemplo, se diseñaron cebadores para las regiones problemáticas usando una estrategia de diseño contraria a la convencional y a los criterios de diseño de PCR bien establecidos. Los criterios establecidos para el diseño de

5 cebadores de PCR indica que, entre otros criterios, 1) la longitud óptima para los cebadores es de 18-22 nucleótidos, 2) la  $T_f$  debe estar en el intervalo de 55-58 grados C, 3) el contenido de GC debe ser de aproximadamente 40-60 %, 4) y deben evitarse repeticiones de regiones dinucleotídicas AT repetidas, siendo el máximo <4 repeticiones dinucleotídicas AT. Se diseñaron cebadores que

10 eran más largos que los cebadores de PCR típicos, por ejemplo, de 23-35 nucleótidos de longitud en lugar de 18-22 nucleótidos, tenían bajas temperaturas de fusión ( $T_f$ ), por ejemplo, de aproximadamente 54 grados centígrados en lugar de aproximadamente 58 grados C, y los cebadores eran ricos en AT, tres parámetros que, según enseñan los criterios de PCR

15 establecidos convencionales, deben evitarse para un diseño de cebadores óptimo. En efecto, se diseñaron cebadores no óptimos. Sorprendentemente, se descubrió que estos cebadores largos, ricos en AT, con baja  $T_f$  multiplexaban las STR mejor que los cebadores cortos, de alta  $T_f$  y bajo contenido de AT. Sin ligarse a ninguna teoría, se contempla que los cebadores más cortos que se

20 diseñaron siguiendo los criterios de diseño de PCR establecidos podrían formar dímeros que tendrían altas temperaturas de fusión y, por tanto, se formaron dímeros eficazmente en condiciones de PCR normales, mientras que los cebadores más largos, con baja  $T_f$ , podrían formar dímeros en condiciones de  $T_f$  realmente bajas y, por tanto, no serían estables para la formación de

25 dímeros, permitiendo de este modo una mayor participación de los cebadores

más largos, de baja  $T_m$ , en condiciones de amplificación normales en comparación con los cebadores cortos, con alta  $T_f$  (por ejemplo, 18-22 nucleótidos,  $T_f$  de 60 grados C, contenido de CG al 50 %).

Los cebadores más largos, con baja  $T_f$ , ricos en AT, para la amplificación de STR después se multiplexaron con los cebadores más cortos, de alta  $T_f$ , de diseño convencional, que se dirigían a SNP. Sin embargo, las reacciones de amplificación multiplex de nuevo no fueron satisfactorias a la hora de proporcionar una amplificación equilibrada tanto de STR como de SNP en una reacción multiplex. Se contempló que, quizá aplicando el diseño de cebadores no convencional para amplificar dianas no problemáticas, por ejemplo, para amplificar las dianas de SNP, se podría llegar a amplificaciones multiplex satisfactorias. De esta manera, para el diseño de cebadores para SNP se aplicaron los mismos criterios usados para diseñar cebadores no óptimos para STR (largos, baja  $T_m$ , rico en AT). Sorprendentemente, los nuevos cebadores diseñados dieron como resultado un mejor equilibrio entre la amplificación de STR y SNP en una reacción multiplex.

La FIG. 4 muestra ejemplos de la interacción entre los cebadores diseñados convencionales y no convencionales en una reacción multiplex. En la FIG. 4<sup>a</sup>, una reacción multiplex de 10 dianas de SNP muestra la amplificación esperada en el intervalo deseado de aproximadamente 200-350 pb para la biblioteca. Los cebadores usados para amplificar los 10 SNP en el multiplex se diseñaron para ser más largos, tener menor  $T_f$  y ser más ricos en AT de lo que se aconseja por los criterios de diseño de cebadores de PCR establecidos. Cuando se diseña un 11<sup>o</sup> par de cebadores usando los criterios de diseño de PCR establecidos, es decir, los cebadores son cortos, tienen una

alta T<sub>f</sub> y no son ricos en AT, y se añaden a los 10 pares, el multiplex resultante muestra amplificación no específica del ADN diana. Como se observa en las FIG. 4B y 4D, la adición de un 11º par de cebadores diseñados de forma convencional interfiere con el 10 plex de pares de cebadores no convencionales y da como resultado una amplificación multiplex no satisfactoria de los SNP diana. Sin embargo, la adición de un 11º par de cebadores que también está diseñado de forma no convencional siguiendo los mismos criterios que el 10 plex de pares de cebadores da como resultado la amplificación satisfactoria de las dianas de SNP (FIG. 4C)

10 Por consiguiente, en algunas realizaciones, cada uno de la pluralidad de cebadores tiene una baja temperatura de fusión, por ejemplo, menor de 60 grados centígrados o de aproximadamente 50 grados centígrados a aproximadamente 60 grados C, y/o tiene una longitud de al menos 24 nucleótidos, por ejemplo, de aproximadamente 24 nucleótidos a  
15 aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una secuencia nucleotídica homopolimérica.

En algunos ejemplos, los pares diseñados de forma no convencional comprenden secuencias que flanquean las STR y SNP diana, y secuencias no molde adicionales. Las secuencias adicionales pueden ser, por ejemplo, secuencias marcadoras que tienen una función durante la preparación de bibliotecas o para las metodologías de secuenciación. Por ejemplo, una secuencia marcadora puede ser una secuencia de captura tal como un resto de hapteno que puede capturarse por un resto compañero inmovilizado para  
25 purificar fragmentos de la biblioteca. Un ejemplo de un resto de hapteno es la

biotina que puede capturarse por estreptavidina para aislar los fragmentos de biblioteca de los componentes de la reacción y similares. Una secuencia marcadora también podría ser una secuencia de amplificación, por ejemplo, que es complementaria a un cebador de amplificación y que se usa en una o  
5 más reacciones de amplificación. Las FIG. 2 y 3 muestran ejemplos de secuencias marcadoras que se usan en una segunda ronda de amplificación después de una primera ronda de amplificación. Una secuencia marcadora también podría ser una etiqueta de secuencia. Las FIG. 2 y 3 también muestran ejemplos de etiquetas de secuencia, el adaptador i5 y el adaptador i7 se usan  
10 en la secuenciación como cebadores de hibridación, de generación de grupos y de secuenciación durante las reacciones de secuenciación por síntesis como se describe en el presente documento. Otro ejemplo de una secuencia marcadora es un identificador de molécula única o UMI, como se muestra en la FIG. 3.

15 Un UMI comprende un tramo aleatorio de nucleótidos que puede usarse durante la secuenciación para corregir los errores de PCR y secuenciación, añadiendo de este modo una capa adicional corrección de errores a los resultados de secuenciación. Los UMI podrían ser, por ejemplo, de 3-10 nucleótidos de longitud, sin embargo, el número dependerá de la cantidad de  
20 ADN de entrada. Por ejemplo, si se usa un 1 ng de ADN para dirigirse a aproximadamente 250 sitios, entonces se espera que se necesiten aproximadamente 350 copias x 250 dianas, es decir aproximadamente 90.000 UMI diferentes. Si se utiliza más ADN, por ejemplo 10 ng, entonces podrían necesitarse aproximadamente 1 millón de UMI diferentes. Todos los duplicados  
25 de PCR procedentes de la misma reacción PCR tendrían la misma secuencia

de UMI, por lo tanto, los duplicados pueden compararse y cualquier error en la secuencia tal como sustituciones, deleciones, inserciones de una sola base (es decir, tartamudeo en PCR) pueden excluirse de los resultados de la secuenciación por medio de la bioinformática. Los identificadores de molécula  
5 única también pueden usarse en el análisis de una muestra mezclada. Las muestras mezcladas, por ejemplo, una muestra de ADN femenino que está contaminada con ADN masculino puede desconvolucionarse para indicar las contribuciones de ADN tanto masculino como femenino usando secuencias de UMI. Por ejemplo, podría haber un total de cuatro números repetidos diferentes  
10 para dos ADN mezclados; sin embargo, podría haber menos de cuatro si la mezcla de las dos muestras comparte alelos en un locus particular. Estos alelos compartidos pueden distinguirse y pueden determinarse porcentajes aproximados usando los UMI para determinar el número de alelos diferentes en la población inicial de moléculas de ADN. Por ejemplo, las moléculas iniciales  
15 podrían contarse y, si estuviera presente un contribuidor minoritario, por ejemplo, al 5 %, entonces el 5 % de los UMI identificaría un genotipo y el 95 % identificaría un segundo genotipo. Después de la PCR, si uno de los alelos (o quizás más) se segara después de la amplificación, entonces no se observaría la proporción 5:95. Sin embargo, usando los UMI, una proporción sesgada  
20 podría corregirse después de condensar duplicados de PCR usando la detección de UMI y corrección. Esto es importante cuando se trata de diferenciar un artefacto de tartamudeo de PCR y un verdadero contribuidor minoritario.

Un cebador de los presentes métodos puede comprender una o más  
25 secuencias marcadoras. Las secuencias marcadoras pueden ser una o más

secuencias cebadoras que no son homólogas a la secuencia diana, pero por ejemplo pueden usarse como moldes para una o más reacciones de amplificación. La secuencia marcadora puede ser una secuencia de captura, por ejemplo, una secuencia de hapteno tal como biotina que puede usarse para

5 purificar amplicones de los componentes de la reacción. Las secuencias marcadoras pueden ser secuencias tales como secuencias adaptadoras que son ventajosas para capturar los amplicones de la biblioteca sobre un sustrato, por ejemplo, para la amplificación en puente en previsión de la secuencia por tecnologías de síntesis como se describe en el presente documento.

10 Adicionalmente, las secuencias marcadoras pueden ser etiquetas de identificadores de molécula única de normalmente entre, por ejemplo, 3-10 nucleótidos compuestos por tramo aleatorio de nucleótidos que puede usarse para corregir errores durante los métodos de preparación de la biblioteca y/o de secuenciación.

15 Adicionalmente, es ventajoso que una reacción PCR multiplexada contenga cebadores oligonucleotídicos para sustancialmente todas las dianas agrupadas en una mezcla. Sin embargo, como se desvela en el presente documento, los oligonucleótidos son más largos, de forma no característica, que los cebadores diseñados usando parámetros tradicionales. La adición

20 adicional de secuencias marcadoras a los cebadores, tales como la adición de UMI que añaden una secuencia génica específica diana, crean secuencias de cebadores incluso más largas. En algunas realizaciones, puede añadirse glicina betaína (aproximadamente 1,5 M) a la pluralidad de cebadores. Por ejemplo, en algunas realizaciones, los tampones de amplificación usados en

25 las reacciones de amplificación con cebadores no convencionales como se

desvela en el presente documento, comprenden una concentración de betaína que es, es de aproximadamente o es más de, 100 mM, 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 700 mM, 800 mM, 900 mM, 1 M, 1,2 M, 1,3 M, 1,4 M, 1,5 M, 1,6 M, 1,7 M, 1,8 M, 1,9 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M, 8 M, 9 M, 10 M, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 500 mM a 2 M, de 1 M a 1,5 M, etc. De esta manera, la mezcla de cebadores como se describe en el presente documento complementada con betaína, por ejemplo, a aproximadamente 1,5 M, sería ventajosa cuando se llevan a la práctica los métodos de la presente divulgación. En algunas realizaciones, puede añadirse glicerol a la pluralidad de cebadores. Por ejemplo, en algunas realizaciones, los tampones de amplificación usados en las reacciones de amplificación con cebadores no convencionales como se desvela en el presente documento comprenden una concentración de glicerol que es, es de aproximadamente, o es mayor de, 100 mM, 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 700 mM, 800 mM, 900 mM, 1 M, 1,2 M, 1,3 M, 1,4 M, 1,5 M, 1,6 M, 1,7 M, 1,8 M, 1,9 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M, 8 M, 9 M, 10 M, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 500 mM a 2 M, de 1 M a 1,5 M, etc. Por lo tanto, una mezcla de cebadores como se describe en el presente documento complementada con glicerol, por ejemplo, a aproximadamente 1,5 M, sería ventajosa cuando se llevan a la práctica los métodos de la presente divulgación.

En algunas realizaciones, también pueden modificarse tampones asociados con el diseño de cebadores no convencional usado en los métodos de amplificación de la presente divulgación. Por ejemplo, en algunas realizaciones, las concentraciones de sales, tal como KCl, LiCl, NaCl o una

combinación de las mismas, del tampón de amplificación están incrementadas en comparación con la concentración de sal de un tampón de amplificación usado junto con cebadores diseñados convencionalmente. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento, comprenden una concentración de KCl que es, es de aproximadamente, o es más de, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de KCl que es de aproximadamente 145 mM. En algunas realizaciones, los tampones de amplificación usados en las reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de LiCl que es, es de aproximadamente, o es más de, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de LiCl que es de aproximadamente 145 mM. En algunas realizaciones, los tampones de amplificación usados en reacciones

de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de NaCl que es de, es de aproximadamente, o es más de, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 5 200 mM, 250 mM, 300 mM, 400 mM, 500 mM, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en las reacciones de amplificación con cebadores no convencionales como se desvela en el presente documento comprenden una concentración de NaCl que 10 es de aproximadamente 145 mM.

En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento pueden comprender  $\text{MgSO}_4$ ,  $\text{MgCl}_2$  o una combinación de los mismos.

## 15 Kits

Ciertas realizaciones desveladas en el presente documento proporcionan kits que comprenden al menos un medio de envase, en el que dicho al menos un medio de envase comprende una pluralidad de cebadores como se desvela en el presente documento. En algunas realizaciones, el medio de envase 20 puede ser un tubo, un pocillo, una placa de microtitulación, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con un número de secuencias de repetición en tándem que es, es de aproximadamente, o es más de 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100, o un intervalo entre dos cualesquiera de estos valores, tal 25 como de 4 a 12, de 10 a 24, de 30 a 100, etc. En algunas realizaciones, la

pluralidad de cebadores puede hibridar específicamente con al menos 24 secuencias de repetición en tándem. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 60 secuencias de repetición en tándem. Los métodos y composiciones desvelados en el presente documento pueden usarse para amplificar múltiples secuencias diana de SNP en una sola reacción. Por ejemplo, la pluralidad de cebadores puede hibridar específicamente con un número de secuencias de SNP, que es, es de aproximadamente, o es más de 4, 6, 8, 10, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100, o un intervalo entre dos valores cualesquiera, tal como de 4 a 12, de 10 a 24, de 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 30 secuencias de SNP. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 50 secuencias de SNP.

En algunas realizaciones, dicho al menos un medio de envase comprende un tampón de amplificación. En algunas realizaciones, los tampones asociados con el diseño de cebadores no convencional usado en los métodos de amplificación de la presente divulgación también pueden modificarse. Por ejemplo, en algunas realizaciones las concentraciones de sales, tales como KCl, LiCl, NaCl o una combinación de las mismas, del tampón de amplificación está incrementada en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de manera convencional. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de KCl, NaCl o LiCl que es, es de aproximadamente, o es más de 60 mM, 70 mM, 80

mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM, o un intervalo entre dos cualesquiera de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de  
5 amplificación usados en las reacciones de amplificación con cebadores no convencionales como se desvelan en el presente documento comprenden una concentración de KCl, NaCl o LiCl que es de aproximadamente 145 mM.

En algunas realizaciones, los tampones de amplificación usados en las reacciones de amplificación con cebadores no convencionales como se  
10 desvelan en el presente documento pueden comprender  $\text{MgSO}_4$ ,  $\text{MgCl}_2$  o una combinación de los mismos.

#### Métodos de secuenciación

Los presentes métodos no están limitados a ninguna plataforma de secuenciación particular, sin embargo, se ejemplifican en el presente  
15 documento con respecto a SBS, o secuenciación por síntesis, tipos de secuenciación en paralelo. Son técnicas particularmente aplicables aquellas en las que se unen ácidos nucleicos a localizaciones fijas en una matriz de tal manera que sus posiciones relativas no cambian y en las que se obtienen imágenes repetidas de la matriz. Son particularmente aplicables los ejemplos  
20 en los que se obtienen imágenes en diferentes canales de color, por ejemplo, coincidiendo con los diferentes marcadores usados para diferenciar un tipo de base nucleotídica de otro.

Las técnicas de SBS generalmente implican la extensión enzimática de una cadena de ácido nucleico naciente mediante la adición iterativa de  
25 nucleótidos frente a una cadena molde. En los métodos tradicionales de SBS,

puede proporcionarse un solo monómero nucleotídico a un nucleótido diana en presencia de una polimerasa en cada paso. Sin embargo, en los métodos descritos en el presente documento, pueden proporcionarse más de un tipo de monómero nucleotídico a un ácido nucleico diana en presencia de una polimerasa en un paso.

Las técnicas de SBS pueden utilizar monómeros nucleotídicos que tienen un resto marcador o que carecen de un resto marcador. Por consiguiente, pueden detectarse eventos de incorporación basándose en una característica del marcador, tal como la fluorescencia del marcador; una característica del monómero nucleotídico tal como el peso molecular o la carga; un subproducto de incorporación del nucleótido, tal como la liberación de pirofosfato; o similar. En algunos ejemplos en los que están presentes dos o más nucleótidos diferentes en un reactivo de secuenciación, los nucleótidos diferentes pueden diferenciarse entre sí o, como alternativa, los dos o más marcadores diferentes pueden ser indiferenciables en las condiciones de las técnicas de detección que se estén usando. Por ejemplo, los diferentes nucleótidos presentes en un reactivo de secuenciación pueden tener diferentes marcadores y pueden diferenciarse usando parámetros ópticos apropiados como se ejemplifica por los métodos de secuenciación desarrollados por Solexa (ahora Illumina, Inc.).

Algunos ejemplos incluyen técnicas de pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) según se incorporan nucleótidos particulares a la cadena naciente (Ronaghi, M., Kaxamohamed, S., Pettersson, B., Uhlen, M. y Nyren, P. (1996) "Real-time DNA sequencing using detection of pyrophosphate release". *Analytical Biochemistry* 242(1), 84-9; Ronaghi, M. (2001) "Pyrosequencing sheds light on

DNA sequencing". Genome Res. 11(1), 3-11; Ronaghi, M, Uhlen, M. y Nyren, P. (1998) "A sequencing method based on real-time pyrophosphate". Science 281(5375), 363; Patente de Estados Unidos n.º 6.210.891; Patente de Estados Unidos n.º 6.258.568 y Patente de Estados Unidos n.º 6.274.320, cuyas

5 divulgaciones se incorporan en el presente documento por referencia en su totalidad). En la pirosecuenciación, el PPI liberado puede detectarse al convertirse inmediatamente en adenosintrifosfato (ATP) por la ATP sulfurilasa, y el nivel de ATP generado se detecta a través de los fotones producidos por luciferasa. Los ácidos nucleicos a secuenciar pueden unirse a características

10 en una matriz y pueden obtenerse imágenes de la matriz al capturar las señales quimioluminiscentes que se producen debido a la incorporación de nucleótidos en las características de la matriz. Puede obtenerse una imagen después de tratar la matriz con un tipo de nucleótido particular (por ejemplo, A, T, C o G). Las imágenes obtenidas después de la adición de cada tipo de

15 nucleótido serán diferentes en función de las características en la matriz que se detecten. Estas diferencias en la imagen reflejan el diferente contenido de secuencia de las características sobre la matriz. Sin embargo, las localizaciones relativas de cada característica permanecerán sin cambios en las imágenes. Las imágenes pueden guardarse, procesarse y analizarse

20 usando los métodos expuestos en el presente documento. Por ejemplo, las imágenes obtenidas después del tratamiento de la matriz con cada tipo de nucleótido diferente pueden manipularse de la misma manera que se ejemplifica en el presente documento para imágenes obtenidas a partir de diferentes canales de detección para métodos de secuenciación basados en

25 terminadores reversibles.

En otro ejemplo de SBS, la secuenciación del ciclo se realiza por la adición por etapas de nucleótidos terminadores reversibles que contienen, por ejemplo, un marcador colorante escindible o fotoblanqueable como se describe, por ejemplo, en el documento WO 04/018497 y en la Patente de Estados Unidos n.º 7.057.026, cuyas divulgaciones se incorporan en el presente documento por referencia. Este enfoque se está comercializando por Solexa (ahora Illumina Inc.), y también se describe en los documentos WO 91/06678 y WO 07/123.744, cada uno de los cuales se incorpora en el presente documento por referencia. La disponibilidad de terminadores marcados con fluorescencia en los que puede invertirse la terminación y puede escindirse el marcador fluorescente facilita la secuenciación eficaz de la terminación reversible cíclica (CRT). Las polimerasas también pueden comodificarse por ingeniería para incorporar eficazmente y extenderse a partir de estos nucleótidos modificados. Se describen sistemas de SBS y métodos a modo de ejemplo adicionales que pueden utilizarse con los métodos y sistemas descritos en el presente documento en la Publicación de Solicitud de Patente de Estados Unidos n.º 2007/0166705, Publicación de Solicitud de Patente de Estados Unidos n.º 2006/0188901, Patente de Estados Unidos n.º 7.057.026, Publicación de Solicitud de Patente de Estados Unidos n.º 2006/0240439, Publicación de Solicitud de Patente de Estados Unidos n.º 2006/0281109, Publicación PCT n.º WO 05/065814, Publicación de Solicitud de Patente de Estados Unidos n.º 2005/0100900, Publicación PCT n.º WO 06/064199, Publicación PCT n.º WO 07/010.251, Publicación de Solicitud de Patente de Estados Unidos n.º 2012/0270305 y Publicación de Solicitud de Patente de Estados Unidos n.º 2013/0260372, cuyas divulgaciones se incorporan en el presente documento

por referencia en su totalidad.

Algunos ejemplos pueden utilizar la detección de cuatro nucleótidos diferentes usando menos de cuatro marcadores diferentes. Por ejemplo, puede realizarse SBS utilizando métodos y sistemas descritos en los materiales  
5 incorporados de la publicación de Solicitud de Patente de Estados Unidos n.º 2013/0079232. Como un primer ejemplo, un par de tipos de nucleótidos puede detectarse a la misma longitud de onda, pero diferenciarse basándose en la diferencia en la intensidad de un miembro del par en comparación con el otro, o basándose en un cambio en un miembro del par (por ejemplo, mediante  
10 modificación química, modificación fotoquímica o modificación física) que hace que aparezca o desaparezca una señal aparente en comparación con la señal detectada para el otro miembro del par. Como un segundo ejemplo, pueden detectarse tres de cuatro tipos de nucleótidos diferentes en condiciones particulares, mientras que un cuarto tipo de nucleótido carece de un marcador  
15 que sea detectable en estas condiciones, o que se detecta mínimamente en estas condiciones (por ejemplo, detección mínima debido a fluorescencia de fondo, etc.). La incorporación de los tres primeros tipos de nucleótidos en un ácido nucleico puede determinarse basándose en la presencia de sus señales respectivas y la incorporación del cuarto tipo de nucleótido en el ácido nucleico  
20 puede determinarse basándose en la ausencia de detección o la detección mínima de cualquier señal. Como un tercer ejemplo, un tipo de nucleótido puede incluir uno o más marcadores que se detectan en dos canales diferentes, mientras que otros tipos de nucleótidos se detectan en no más de uno de los canales. Las tres configuraciones a modo de ejemplo anteriormente  
25 mencionadas no se consideran mutuamente excluyentes y pueden usarse en

diversas combinaciones. Una realización a modo de ejemplo que combina los tres ejemplos es el método de SBS basado en fluorescencia que usa un primer tipo de nucleótido que se detecta en un primer canal (por ejemplo, dATP que tiene un marcador que se detecta en el primer canal cuando se excita con una primera longitud de onda de excitación), un segundo tipo de nucleótido que se detecta en un segundo canal (por ejemplo, dCTP que tiene un marcador que se detecta en el segundo canal cuando se excita con una segunda longitud de onda de excitación), un tercer tipo de nucleótido que se detecta tanto en el primer como en el segundo canal (por ejemplo, dTTP que tiene al menos un marcador que se detecta en ambos canales cuando se excita con la primera y/o segunda longitudes de onda de excitación) y un cuarto tipo de nucleótido que carece de marcador que no se detecta en ningún canal o se detecta mínimamente (por ejemplo, dGTP que no tiene marcador).

Adicionalmente, como se describe en los materiales incorporados de la Publicación de Solicitud de Patente de Estados Unidos n.º 2013/0079232, pueden obtenerse datos de secuenciación usando un solo canal. En dichos enfoques denominados de secuenciación de un solo colorante, el primer tipo de nucleótido está marcado pero el marcador se retira después de generarse la primera imagen, y el segundo tipo de nucleótido se marca solo después de generar la primera imagen. El tercer tipo de nucleótido conserva su marcador tanto en la primera como en la segunda imágenes, y el cuarto tipo de nucleótido permanece sin marcar en ambas imágenes.

Algunos ejemplos pueden utilizar secuenciación por técnicas de ligamiento. Dichas técnicas utilizan ADN ligasa para incorporar oligonucleótidos e identificar la incorporación de dichos oligonucleótidos. Los oligonucleótidos

tienen normalmente diferentes marcadores que se correlacionan con la identidad de un nucleótido particular en una secuencia con la que hibridan los oligonucleótidos. Al igual que con otros métodos de SBS, pueden obtenerse imágenes después del tratamiento de una matriz de características de ácido nucleico con los reactivos de secuenciación marcados. Cada imagen mostrará características de ácidos nucleicos que han incorporado marcadores de un tipo particular. En las diferentes imágenes estarán presentes o ausentes diferentes características debido al diferente contenido en la secuencia de cada característica, pero la posición relativa de las características permanecerá sin cambios en las imágenes. Las imágenes obtenidas con métodos de secuenciación basados en ligamiento pueden guardarse, procesarse y analizarse como se expone en el presente documento. Se describen métodos y sistemas de SBS a modo de ejemplo que pueden utilizarse con los métodos y sistemas descritos en el presente documento en la Patente de Estados Unidos n.º 6.969.488, Patente de Estados Unidos n.º 6.172.218, y Patente de Estados Unidos n.º 6.306.597, cuyas divulgaciones se incorporan en el presente documento por referencia en su totalidad.

Algunos ejemplos pueden utilizar secuenciación por nanoporos (Deamer, D. W. y Akeson, M. "Nanopores and nucleic acids: prospects for ultrarapid sequencing". Trends Biotechnol. 18, 147-151 (2000); Deamer, D. y D. Branton, "Characterization of nucleic acids by nanopore analysis". Acc. Chem. Res. 35: 817-825 (2002); Li, J., M. Gershow, D. Stein, E. Brandin, y J. A. Golovchenko, "DNA molecules and configurations in a solid-state nanopore microscope" Nat. Mater. 2: 611-615 (2003), cuyas divulgaciones se incorporan en el presente documento por referencia en su totalidad). En dichas realizaciones, el ácido

nucleico diana pasa a través de un nanoporo. El nanoporo puede ser un poro sintético o una proteína de membrana biológica, tal como la  $\alpha$ -hemolisina. A medida que el ácido nucleico diana pasa a través de nanoporo, cada par de bases pueden identificarse midiendo fluctuaciones en la conductancia eléctrica del poro. (Patente de Estados Unidos n.º 7.001.792; Soni, G. V. y Meller, "A. Progress toward ultrafast DNA sequencing using solid-state nanopores". Clin. Chem. 53, 1996-2001 (2007); Healy, K. "Nanopore-based single-molecule DNA analysis". Nanomed. 2, 459-481 (2007); Cockroft, S, L, Chu, J., Amorin, M. y Ghadiri, M. R. "A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution". J. Am. Chem. Soc. 130, 818-820 (2008), cuyas divulgaciones se incorporan en el presente documento por referencia en su totalidad). Los datos obtenidos de la secuenciación por nanoporos pueden guardarse, procesarse y analizarse como se expone en el presente documento. En particular, los datos pueden tratarse como una imagen de acuerdo con el tratamiento a modo de ejemplo de imágenes ópticas y otras imágenes que se exponen en el presente documento.

Algunos ejemplos pueden utilizar métodos que implican la monitorización en tiempo real de la actividad ADN polimerasa. Pueden detectarse incorporaciones de nucleótidos a través de interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa portadora de fluoróforo y nucleótidos marcados con  $\gamma$  fosfato como se describe, por ejemplo, en la Patente de Estados Unidos n.º 7.329.492 y en la Patente de Estados Unidos n.º 7.211.414 (cada una de las cuales se incorpora en el presente documento por referencia) o las incorporaciones de nucleótidos pueden detectarse con guías de onda de modo cero como se describe, por

ejemplo, en la Patente de Estados Unidos n.º 7.315.019 (que se incorpora en el presente documento por referencia) y usando análogos de nucleótidos fluorescentes y polimerasas modificadas por ingeniería como se describe, por ejemplo, en la Patente de Estados Unidos n.º 7.405.281 y en la Publicación de Solicitud de Patente de Estados Unidos n.º 2008/0108082 (cada una de los cuales se incorpora en el presente documento por referencia). La iluminación puede estar restringida a un volumen a escala zeptolétrica alrededor de una polimerasa conectada a la superficie de tal manera que la incorporación de nucleótidos marcados con fluorescencia puede observarse con un fondo bajo (Levene, M. J. *et al.* "Zero-mode waveguides for single-molecule analysis at high concentrations". *Science* 299, 682-686 (2003); Lundquist, P. M. *et al.* "Parallel confocal detection of single molecules in real time". *Opt. Lett.* 33, 1026-1028 (2008); Korlach, J. *et al.* "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nano structures". *Proc. Natl. Acad. Sci. USA* 105, 1176-1181 (2008), cuyas divulgaciones se incorporan en el presente documento por referencia en su totalidad). Las imágenes obtenidas con dichos métodos pueden guardarse procesarse y analizarse como se expone en el presente documento.

Algunas realizaciones de SBS incluyen la detección de un protón liberado después de la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede usar un detector eléctrico y técnicas asociadas que están disponibles en el mercado en Ion Torrent (Guilford, CT, a Life Technologies subsidiary) o métodos y sistemas de secuenciación descritos en los documentos US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; o US

2010/0282617 A1, cada uno de los cuales se incorpora en el presente documento por referencia. Los métodos expuestos en el presente documento para la amplificación de ácidos nucleicos diana usando exclusión cinética pueden aplicarse fácilmente a sustratos usados para detectar protones. Más  
5 específicamente, los métodos expuestos en el presente documento pueden usarse para producir poblaciones clonales de amplicones que se usan para detectar protones.

Los métodos de SBS anteriores pueden realizarse ventajosamente en formatos multiplex de tal manera que se manipulan simultáneamente múltiples  
10 ácidos nucleicos diana diferentes. En realizaciones particulares, diferentes ácidos nucleicos diana pueden tratarse en un recipiente de reacción común o en una superficie de un sustrato particular. Esto permite un suministro conveniente de reactivos de secuenciación, la retirada de reactivos que no han reaccionado y la detección de eventos de incorporación de una manera  
15 multiplex. En realizaciones que usan ácidos nucleico diana unidos a la superficie, los ácidos nucleicos diana pueden estar en un formato de matriz. En un formato de matriz, los ácidos nucleico diana pueden estar normalmente unidos a una superficie de una manera espacialmente diferenciable. Los ácidos nucleico diana pueden estar unidos mediante unión covalente directa, unión a  
20 una perla o a otra partícula o unión a una polimerasa o a otra molécula que está unida a la superficie. La matriz puede incluir una sola copia de un ácido nucleico diana en cada sitio (también denominado característica) o en cada sitio o característica pueden estar presentes múltiples copias que tienen la misma secuencia. Pueden producirse copias múltiples mediante métodos de  
25 amplificación tales como amplificación en puente o PCR en emulsión como se

describe con detalle más adelante.

Los métodos de la presente divulgación utilizan la tecnología Illumina, Inc. para la secuenciación de las bibliotecas de perfil de ADN creadas mediante la puesta en práctica de los métodos descritos en el presente documento. En los  
5 ejemplos descritos en el presente documento se usó el instrumento de secuenciación MiSeq para el agrupamiento y secuenciación. Sin embargo, como se ha indicado anteriormente y entiende un experto en la materia, los métodos de la presente invención no están limitados por el tipo de plataforma de secuenciación utilizada.

## 10 Ejemplos

Los siguientes ejemplos desvelan varios métodos y materiales para el perfilado de ADN. Estos métodos y materiales pueden modificarse conservando al mismo tiempo la esencia y el alcance de la invención. Dichas modificaciones resultarán evidentes para los expertos en la materia tras la  
15 consideración de esta divulgación o la puesta en práctica de los métodos desvelados en el presente documento. Por consiguiente, no se pretende que estos métodos o materiales se limiten a los ejemplos específicos desvelados en el presente documento, sino que incluyan todas las modificaciones y alternativas que se encuentren dentro del alcance y esencia de la presente  
20 divulgación.

### Ejemplo 1-diseño de cebador no convencional

El programa de diseño informático DesignStudio de Illumina, Inc. (San Diego, CA) se modificó y usó para el diseño de cebadores. Un experto en la materia entenderá, por supuesto, que también pueden usarse programas de  
25 diseño de cebadores alternativos tales como Primer3 y restablecerse los

parámetros por defecto para imitar el intento de los parámetros modificados para el diseño de cebadores. Los parámetros normalmente se restablecen en el archivo config.xml que viene con el programa informático, sin embargo, esto puede diferir cuando se usan diferentes programas informáticos y es una práctica consultar los materiales específicos para acceder a los parámetros por defecto para cada programa informático. En el programa informático de diseño de cebadores pueden restablecerse los siguientes parámetros:

- 1) La longitud mínima deseada del amplicón se restablece a >60<
- 2) La longitud máxima deseada del amplicón se restablece a >120<
- 10 3) La estrecha separación de candidatos se restablece a >3< (el valor por defecto es de 30 pb)
- 4) La sonda con % GC Máx. se restablece a >60< para permitir un mayor número de tramos con repeticiones ricas en AT
- 5) La Tm media se restablece a >57< (por defecto 59 C) para reducir la Tm media
- 15 6) La Tm máxima se restablece a >60< (por defecto 73)
- 7) La Tm mínima se restablece a >51< (por defecto 55)
- 8) La longitud media de sonda se restablece a >28< (por defecto 27)
- 9) La longitud máxima de sonda se restablece a >38< (por defecto 30)
- 20 10) La longitud mínima de sonda se restablece a >25< (por defecto 22)

Para el diseño de cebadores de SPN, el intervalo hasta la diana para el extremo 3' del cebador se estableció para ser "pequeño", para mantener a los cebadores a una distancia de aproximadamente 1 pb del SNP establecido como diana. Una vez restablecidos todos los parámetros, el programa de diseño de cebadores puede hacerse funcionar sobre la secuencia para

determinar los candidatos de pares de cebadores que se encontrarán en estos nuevos parámetros. Por ejemplo, un usuario del programa informático puede generar una lista de dianas que dice al programa informático donde mirar en el genoma para diseñar los cebadores. En el presente ejemplo, las regiones

5 establecidas como dianas se copiaron y se pegaron en la aplicación gráfica de interfaz de usuario, con el programa informático DesignStudio usado para orientar y dirigir el diseño de cebadores. Una vez introducidas las regiones establecidas como dianas en el programa, el programa se dirigió a Crear un Archivo de Diseño (Create a Design File) para comenzar la herramienta y crear

10 los diseños de los cebadores. En el presente ejemplo, el principal resultado fue un archivo .txt que incluía las secuencias de cebadores y/o algunas de las regiones contenían fallos y eran “no diseñables”, punto en el cual necesitaban redefinirse y volverse a procesar las secuencias establecidas como diana. El programa informático usado en este experimento proporcionó los cebadores

15 diseñados que se mapearon sobre la secuencia que se especificó como la región establecida como diana. Después de los parámetros restablecidos, se diseñaron cebadores que no seguían los criterios convencionales para el diseño de cebadores para amplificación; pero que permitían amplificación multiplex de STR largas y SNP cortos.

20 Los ejemplos de cebadores dirigidos a STR diseñados ventajosos en los métodos desvelados en el presente documento incluyen los indicados en la Tabla 1. Los ejemplos de cebadores dirigidos a SNP ventajosos en los métodos desvelados en el presente documento incluyen los expuestos en la Tabla 2.

Tabla 1-cebador dirigido a STR sin etiquetas y tamaños de amplicón

| SEQ ID NO | CEBADOR LOCUS STR | EJEMPLOS DE CEBADORES STR SIN ETIQUETAS | TAMAÑO DE AMPLICÓN |
|-----------|-------------------|-----------------------------------------|--------------------|
| 1         | AmelPP_F_T        | CCCTGGGCTCTGTAAAGAA                     | 106, 112           |
| 2         | AmelPP_R_Sm       | ATCAGAGCTTAAACTGGGAAGCTG                |                    |
| 3         | CSF1PO_F1_T       | ACAGTAACTGCCTTCATAGATAG                 | 117                |
| 4         | CSF1PO_R1_Sm      | GTGTCAGACCCCTGTTCTAAGTA                 |                    |
| 5         | D5S818_F2_T       | TGATTTTCCTCTTTGGTATCCTTATGTAAT          | 112                |
| 6         | D5S818_R2_Sm      | ACAACATTGTATCTTTATCTGTATCCT             |                    |
| 7         | D8S1179_F1_T      | TTTGTATTTCAATGTGTACATTCGTATC            | 110                |
| 8         | D8S1179_R1_Sm     | ACCTATCCTGTAGATTATTTTCACGTG             |                    |
| 9         | D18S51_F1_T       | CTCTGAGTGACAAATTGAGACCTT                | 184                |
| 10        | D18S51_R1_T       | TTAACTTCTCTGGTGTGTGGAGATG               |                    |
| 11        | D19S433_F1_T      | TTTGGTGCAACCCATTACCCG                   | 188                |
| 12        | D19S433_R1_Sm     | AGGAGGTTGAGGCTGCAAAA                    |                    |
| 13        | D7S820_F2_T       | CACCAAATATTGGTAATTAAATGTTTACTATAGAC     | 167                |
| 14        | D7S820_R2_Sm      | TAAAGGGTATGATAGAACACTTGTC               |                    |
| 15        | D16S539_F2_T      | CAAAGGCAGATCCCAAGCTCT                   | 160                |
| 16        | D16S539_R2_Sm     | TGTGTGTGCATCTGTAAGCAT                   |                    |
| 17        | D3S1358_F2_T      | TGGTGTGTATTCCCTGTGCC                    | 170                |
| 18        | D3S1358_R2_Sm     | GCAGTCCAATCTGGGTGACA                    |                    |
| 19        | D10S1248_F1_T     | CCAATCTGGTCACAAACATATTAATGAA            | 148                |
| 20        | D10S1248_R1_Sm    | TTTCCCTTGTCTTGTTATTAAAGGAAC             |                    |
| 21        | TH01_F1_T         | TTCCCATTTGGCCTGTTCCCTC                  | 112                |
| 22        | TH01_R1_Sm        | CTGTACACAGGGCTTCCGAG                    |                    |
| 23        | FGA_F2_T          | GCTGAATGATTTGTCTGTAATTG                 | 188                |
| 24        | FGA_R2_Sm         | GAATCACAGATTAACTGTAACCAAAATAAAATTAG     |                    |
| 25        | D61043_F1_T       | CAATAGTGTGCAAGGATGGGTG                  | 175                |

|    |                  |                                    |         |
|----|------------------|------------------------------------|---------|
| 26 | D61043_R1_Sm     | TCTGTGGTCTCCAGCTTAC                |         |
| 27 | TPOX_F1_T        | CTTAGGGAACCTCACTGAATG              | 77      |
| 28 | TPOX_R1_Sm       | GTCTTGTGACGCTTATTTGC               |         |
| 29 | D13S317_F2_T     | TTGGGTTGAGCCATAGGCAG               | 162     |
| 30 | D13S317_R2_Sm    | GCATCCGTGACTCTCTGGAC               |         |
| 31 | D21S11_F1_T      | GTTATGGGACTTTCTCAGTCTCCAT          | 226     |
| 32 | D21S11_R3_Sm     | GAGACTAATAGGAGGTAGATAGACTGG        |         |
| 33 | D12S391_F1_T     | GAGACTGTATTAGTAAGGCTTCTC           | 253     |
| 34 | D12S391_R2_Sm    | CCTGGACTGAGCCATGCTCC               |         |
| 35 | D151656_F2_T     | CAGTCTGTGTTAGTCAGGATTC             | 173     |
| 36 | D151656_R1_Sm    | TCAAGGTCACACTGTGTGATGT             |         |
| 37 | D9S1122_F3_T     | CTTCTGAAAGCTTCTAGTTTACCT           | 120     |
| 38 | D9S1122_R2_Sm    | TTGCTTATTTGTGGGGTATTTCA            |         |
| 39 | PentaE_F1_T      | AAGAATTCCTTATTTGGGTTATTAATTG       | 362     |
| 40 | PentaE_R1_Sm     | AAATTGTGGACAGGTGCGGT               |         |
| 41 | D17S1301_F2_T    | CCATGTAAAAATACATGCATGTGTTTATATAC   | 142     |
| 42 | D17S1301_R2_Sm   | TGATTAAGAAAGTGAAGGTAAAAATGTGTATAAC |         |
| 43 | D2S441_F2_T      | CCAAATGTTATGATTAATCTTTTAAATGGAGC   | 160     |
| 44 | D2S441_R3_Sm     | GTAACAAGGGCTACAGGAATCATGAG         |         |
| 45 | D4S2408_F3_T     | TCACTCACTGAAATGACTGAAAAATAG        | 302     |
| 46 | D4S2408_R9_Sm    | AGGTACATAACAGTCAATAGAAAAG          |         |
| 47 | D2S1338_F2_T     | GAGTTATTCAGTAAGTTAAAGGATTGCAG      | 162     |
| 48 | D2S1338_R2_Sm    | GGGAGCCAGTGGATTGGAAACAG            |         |
| 49 | PentaD_F3_T      | GCATGGTGAGGCTGAAGTAG               | 268     |
| 50 | PentaD_R1_Sm     | CTAACCTATGGTCATAACGATTTTT          |         |
| 51 | vWA_F3_T         | GATGATAAGAATAATCAGTATGTGACTTGG     | 160     |
| 52 | vWA_R3_Sm        | ATAGGTTAGATAGAGATAGGACAGATGATA     |         |
| 53 | SE33_F1_T        | CCCTACCGCTATAGTAACCTTGC            | 380     |
| 54 | SE33_R2_Sm       | CACGTCTGTAATCCAGCTCCTA             |         |
| 55 | D20S482_F3_T     | GGAAGCGTGTACTAGAGTTCTTCAG          | 145     |
| 56 | D20S482_R2_Sm    | GGACAGCCTCCATATCCACATG             |         |
| 57 | DXS10074_F1_T    | TTCTACTGCCTCCACCTTATTG             | 212     |
| 58 | DXS10074_R1_Sm   | TTTATGGTCTCAGTGCCCTCAGA            |         |
| 59 | DXS10103_F1_Sm   | TCATAATCACATATCATGAGC              | 177     |
| 60 | DXS10103_R1_T    | AAACAGAACCAAGGGGAATGAA             |         |
| 61 | DXS10135_F1_T    | TGAAACTAAAGTCAAATGGGGCTAC          | 268     |
| 62 | DXS10135_R1_Sm   | TAAGGGGTGACACCTCTCTGGATA           |         |
| 63 | DXS8377_F2_Sm    | CCCAGCCTACATCTACCACTTCATG          | 276     |
| 64 | DXS8377_R2_T     | CTAATGTTCTGATGGACCTTTGGAAAGC       |         |
| 65 | DXS7423_F1_Sm    | GTCYCCAGTACCAGCTAGCTTAG            | 191     |
| 66 | DXS7423_R1_T     | TCTCCCAACCTGCCCTTTATCA             |         |
| 67 | DXS8378_F1_Sm    | TTTGGGTGACACAGTGGCT                | 442     |
| 68 | DXS8378_R1_T     | TTGATCAACACAGGAGGTTTGACC           |         |
| 69 | HPRTB_F1_Sm      | TATACCACTTTGATGTTGACACTAGTTTAC     | 213     |
| 70 | HPRTB_R1_T       | CCTGTCTATGGTCTCGATTCAAT            |         |
| 71 | DXS10148_F3_Sm   | TGCATGACAGAGGGAGATTCT              | 256     |
| 72 | DXS10148_R3_T    | AGAGGGGAAATAGTAGAATGAGGATG         |         |
| 73 | DXS7132_F3_Sm    | GCCAACTCTATTAGTCAACGTTT            | 204     |
| 74 | DXS7132_R4_T     | CTGGTCTCTAGCTCACATACAGT            |         |
| 75 | DYF387S1ab_F2_T  | TTTACCCTTAACAAGAAAAAAGAAGAA        | 227,231 |
| 76 | DYF387S1ab_R2_Sm | CAGTGTGAGAAGTGTGAGAAGTGC           |         |
| 77 | DYS385a_b_F1_T   | GACACCATGCCAAACAACAAC              | 260,248 |
| 78 | DYS385a_b_R1_Sm  | ATCTATCTATTCGAATTACATAGTCC         |         |
| 79 | DYS389I_II_F3_T  | TCATTATACCTACTCTGTATCCAACCTCTC     | 183,303 |
| 80 | DYS389I_II_R3_Sm | GGAACACAATTATCCCTGAGTAGCAG         |         |

|     |                |                                              |     |
|-----|----------------|----------------------------------------------|-----|
| 81  | DYS390_F2_T    | GGTAGCATAATAGAAATTTATGAGTGGG                 | 318 |
| 82  | DYS390_R2_Sm   | GAAGACAGACITCAATATCACAGAACATCG               |     |
| 83  | DYS391_F1_T    | GTGTATCTATTTCATTCAATCATAACCCC                | 143 |
| 84  | DYS391_R1_Sm   | CTCCCTGGTTGCAAGCAATTGCC                      |     |
| 85  | DYS438_F1_T    | CCAAAATTAGTGGGGAATAGTTGAAC                   | 149 |
| 86  | DYS438_R2_Sm   | GTCGAGATCACACCATTGCATTTC                     |     |
| 87  | DYS439_F1_T    | GCCTGGCTTGGAAITCTTTTACCC                     | 195 |
| 88  | DYS439_R1_Sm   | TTTAAGTCTTTAATCTATCTTGAATTAATAGATTC          |     |
| 89  | DYS481_F1_T    | CTTTAAGAGGAGTCTGCTAAAAGGAATG                 | 144 |
| 90  | DYS481_R3_Sm   | TCACCAGAAGGTTGCAAGAC                         |     |
| 91  | DYS505_F1_T    | TCTGGCGAAGTAACCCAAAC                         | 174 |
| 92  | DYS505_R1_Sm   | TGGAGTCAGTTCACCAGAAGG                        |     |
| 93  | DYS522_F2_T    | GGAACCAAGTGAGAGCCG                           | 306 |
| 94  | DYS522_R2_Sm   | CTCAGAGTGCTGAACCCAG                          |     |
| 95  | DYS533_F2_T    | GTATTTATTCATGATCAGTTCTTAACCAACC              | 206 |
| 96  | DYS533_R2_Sm   | CTACCTAATATTTATCTATATCATTCTAATTATGCTCTT<br>C |     |
| 97  | DYS549_F1_T    | CTCTAAAGGTTTTTTTTGGTGGCATAAG                 | 222 |
| 98  | DYS549_R1_Sm   | GATTAATACAACAAAAATTTGGTAATCTGAAA             |     |
| 99  | DYS570_F1_T    | CAACCTAAGCTGAAATGCAGATATTC                   | 170 |
| 100 | DYS570_R1_Sm   | GTTATGAAACGTAAAAATGAATGATGACTAG              |     |
| 101 | DYS576_F2_T    | GCAGTCTCATTTCCTGGAGATGAAGG                   | 191 |
| 102 | DYS576_R1_Sm   | CTTGGGCTGAGGAGTTCAATC                        |     |
| 103 | DYS612_F2_T    | GCCAGTAAGAATAAAATTACAGCATGAAG                | 287 |
| 104 | DYS612_R2_Sm   | GAATAATCTACCAGCAACAATGGCT                    |     |
| 105 | DYS635_F4_T    | TGCCCCAATGGAATGCTCTCT                        | 274 |
| 106 | DYS635_R2_Sm   | GCTCCATCTCAAACAACAAAAACACAAAAAATG            |     |
| 107 | DYS643_F2_T    | GGGTCAATTGAACCTCATGCTCTG                     | 170 |
| 108 | DYS643_R1_Sm   | CCCCCAAAATTCTACTGAAGTAAA                     |     |
| 109 | Y_GATAH4_F2_T  | TAACAGGATAAAATCACCTATCTATGTAT                | 175 |
| 110 | Y_GATAH4_R2_Sm | GCTGAGGAGAATTTCCAAATTTA                      |     |

Tabla 2-cebadores dirigidos a SNP

| SEQ ID NO | CEBADOR SNP           | EJEMPLOS DE CEBADORES SNP SIN ETIQUETAS |
|-----------|-----------------------|-----------------------------------------|
| 111       | rs10092491_iSNPI_T_F2 | CCCGCAAACAACTAGGATAAAATCTCTA            |
| 112       | rs1015250_iSNPI_T_F   | CGACATGGGAAATGTCAGATCATAAGAC            |
| 113       | rs1024116_iSNPI_T_F2  | CCAGGGAGTGAAAAATCCTTTTATCATC            |
| 114       | rs1028528_iSNPI_T_F2  | GAGGATGAAGGTTAGAGCCAGACCT               |
| 115       | rs1029047_iSNPI_T_F2  | TGTGGAATAAACTGAAGGCTAAAGAAAA            |
| 116       | rs1031825_iSNPI_T_F2  | CAAGCCCTATGCCAAGGATATAACAATG            |
| 117       | rs10488710_iSNPI_T_F  | GAGGTTTTACTGTATTAGGAGTTCCAC             |
| 118       | rs10495407_iSNPI_T_F  | CAGATGTGAGATGATAAATTCGTTCTCC            |
| 119       | rs1058083_iSNPI_T_F   | TTGTTCTTCTCCATCCCAITTCACCC              |
| 120       | rs10773760_iSNPI_T_F  | CTTGTACATTCCTTATCTGCTATGTGG             |
| 121       | rs1294331_iSNPI_T_F2  | CTCTCTTTGGAGTTTTATGTGTTGCTAC            |
| 122       | rs12997453_iSNPI_T_F  | CTCTGATGATGTGCAAGAAAGGTAGGTA            |
| 123       | rs13182883_iSNPI_T_F  | TCAGACTATGTTTTAAGGAGACTATGAGG           |
| 124       | rs13218440_iSNPI_T_F  | CTAAGTATCTACCAATGTGCTACGTACC            |
| 125       | rs1335873_iSNPI_T_F   | CACGTGGATGATATGGTTTCTCAAGG              |
| 126       | rs1336071_iSNPI_T_F2  | AGCACCTATATATTATACCTGAAAGCAT            |
| 127       | rs1355366_iSNPI_T_F   | CCCATGATTTCTTGTGGTGAGAATTC              |
| 128       | rs1357617_iSNPI_T_F   | CACCTCTGTACTTTAATTTGACTTCCC             |

|     |           |            |                               |
|-----|-----------|------------|-------------------------------|
| 129 | rs1382387 | ISNPI_T_F  | GTITTTCTTCATTCCTCATGTTGTGTAC  |
| 130 | rs1413212 | ISNPI_T_F  | CACTCTTCTGAATCTGGTCAACAAC     |
| 131 | rs1454361 | ISNPI_T_F  | CAAGTTATATCATAGAGTCTACGACCCC  |
| 132 | rs1463729 | ISNPI_T_F  | CTGCAACTATCAGTCTCTGCCCTTATTC  |
| 133 | rs1493232 | ISNPI_T_F  | GATGTGTCTCAAACGTGTTATTGTGAGG  |
| 134 | rs1498553 | ISNPI_T_F  | GAACCTATTATCCAGAGACCTGTTCTC   |
| 135 | rs1523537 | ISNPI_T_F  | CATAATACAACCTGTCCTTGGAGTTACT  |
| 136 | rs1528460 | ISNPI_T_F  | GTGACCAGTAGTTCTATGAGCAAGTATG  |
| 137 | rs159606  | ISNPI_T_F  | CCACATTGTATGGTTTTTAGGCACCATG  |
| 138 | rs1736442 | ISNPI_T_F  | CTAATAAGTGGGACAGTTAAGAGAAGGC  |
| 139 | rs1821380 | ISNPI_T_F  | CAAGACAAGCGATTGAAAGAAGTGGAT   |
| 140 | rs1865510 | ISNPI_T_F  | CCTGTCAATCTTCTACCAAGAGGGTAA   |
| 141 | rs1979255 | ISNPI_T_F  | GAATCATAGCTTGTGTGTGTCAGGG     |
| 142 | rs2016276 | ISNPI_T_F  | GAATTACAAGTATTGTCATCCAGCCT    |
| 143 | rs2040411 | ISNPI_T_F  | GACCAACTTGGCTTTAACAGATGCAAAAT |
| 144 | rs2046361 | ISNPI_T_F2 | TCCCTACCTTTAAGACTTTTCTATTGG   |
| 145 | rs2056277 | ISNPI_T_F2 | CATTATCTCGTCATCTTCCCTGTCTTG   |
| 146 | rs2076848 | ISNPI_T_F  | GCATCAAAATCACCAGTGAAATATTGA   |
| 147 | rs2107612 | ISNPI_T_F  | ATGAGTACATTATTCAACTGTTTTGGAG  |
| 148 | rs2111980 | ISNPI_T_F  | CAGCCATGTTGTAAACATTTTACGGTC   |
| 149 | rs214955  | ISNPI_T_F  | GCACATCTAAGAACTGGTGATTCTATC   |
| 150 | rs221956  | ISNPI_T_F  | GCTAGAAAAAGCTGAGATAGCTGTGAAG  |
| 151 | rs2342747 | ISNPI_T_F  | CCTTGAAGCTCATTCTTTGTGTCCC     |
| 152 | rs2399332 | ISNPI_T_F  | CTGGACACCAGACCAAAAAACAATAACC  |
| 153 | rs251934  | ISNPI_T_F  | GTAATTAGAGGGCAGTGAGGCTTTTAA   |
| 154 | rs279844  | ISNPI_T_F  | CTCCAGAAGCTACTGGGATAITTAATAG  |
| 155 | rs2830795 | ISNPI_T_F  | TGAGCCAAATCAGCAATATAATAGGACT  |
| 156 | rs2831700 | ISNPI_T_F  | CCTAGAACCACAAATATCTGTCTTGGC   |
| 157 | rs2920816 | ISNPI_T_F2 | CCATTGATTCTCTACAGTTCTGCAGGTA  |
| 158 | rs321198  | ISNPI_T_F  | CTCCACACTTTATACAGGTGAATCTGA   |
| 159 | rs338882  | ISNPI_T_F  | CATTTTCTCTCTCTCTCTCTCACCTTC   |
| 160 | rs354439  | ISNPI_T_F  | GCCTCTCTTCCCTTATGTATCTCTCTC   |
| 161 | rs3780962 | ISNPI_T_F  | GGCTTTTGAAGAAAAACACTAACCTGTC  |
| 162 | rs430046  | ISNPI_T_F  | CACCTATGGGCCTCTCTTATTTCTCC    |
| 163 | rs4364205 | ISNPI_T_F  | CATTTGATAGCCATTGGGTGTTTCCA    |
| 164 | rs445251  | ISNPI_T_F  | CCATCACACTATCCTGACATGAACAAAT  |
| 165 | rs4606077 | ISNPI_T_F  | GAAGATTTCATCCCAAGTGAAGCAC     |
| 166 | rs560681  | ISNPI_T_F  | GCACTTCATAAAGAATCAGTCAGGATGC  |
| 167 | rs6444724 | ISNPI_T_F  | GGAGAATCAGGAAATAGTCACTTCTTAC  |
| 168 | rs6811238 | ISNPI_T_F  | CATTTGACCTTCTAGCCAAATGAAGTAC  |
| 169 | rs7041158 | ISNPI_T_F  | GGAAATTTCTGAGAATAACATTGCCTCTC |
| 170 | rs717302  | ISNPI_T_F  | CATATGTTGGGGAGCTAAACCTAATGA   |
| 171 | rs719366  | ISNPI_T_F  | CACTGTGACCACAGCATCTTTTAATCTC  |
| 172 | rs722098  | ISNPI_T_F2 | GGGTAAAGAAATATTCAGCACATCCAAA  |
| 173 | rs722290  | ISNPI_T_F  | GAGTATCCCTTATCTAAATGCTGGTCC   |
| 174 | rs727811  | ISNPI_T_F  | CTTTTCTCTTACCGGAACCTCAACGAC   |
| 175 | rs729172  | ISNPI_T_F  | CCTCATTAAATATGACCAAGGCTCCTCTG |
| 176 | rs733164  | ISNPI_T_F  | TGACTCTAATTGGGGATGTGGTAATTAG  |
| 177 | rs735155  | ISNPI_T_F  | GACCTAACCTGGAGAAAACCGGAGA     |
| 178 | rs740598  | ISNPI_T_F  | GTTCCTCTCTCTGAACCTTTGTCTCAG   |
| 179 | rs740910  | ISNPI_T_F  | GCAAAACACAAAGATAGGTTCCAGTTT   |
| 180 | rs763869  | ISNPI_T_F  | CATATCAAGTGCTTTCTGTTGACATTTG  |
| 181 | rs8037429 | ISNPI_T_F  | CTGAAAAGTGTACGTAAGAGGTCATTG   |
| 182 | rs8078417 | ISNPI_T_F  | CATCTGAGTGTGAGAAGAGCCTCAA     |
| 183 | rs826472  | ISNPI_T_F2 | CCAGCAAAAACCTCTTTCTCCAGTAA    |

|     |            |      |   |    |                               |
|-----|------------|------|---|----|-------------------------------|
| 184 | rs873196   | ISNP | T | F  | GCTAGGAAAGTTTCTCTCTGTTTCA     |
| 185 | rs876724   | ISNP | T | F  | GAATATCTATGAGCAGGAGTTAGCAG    |
| 186 | rs891700   | ISNP | T | F2 | CTAATCAGTGTCACTATGTGTGAGCTAT  |
| 187 | rs901398   | ISNP | T | F  | CATCATACAGACTCAAGGAGCTTAGCTG  |
| 188 | rs907100   | ISNP | T | F  | CTTCCAAGCCTTGGAAAACACAGAAAA   |
| 189 | rs914165   | ISNP | T | F  | GTACCTTATAAATCACGGAGTGCAGAC   |
| 190 | rs917118   | ISNP | T | F  | CAAGTGGTAAGAGATGACTGAGGTCAA   |
| 191 | rs938283   | ISNP | T | F  | CTTCTTCTTAGAAGGACACTGTCAG     |
| 192 | rs964681   | ISNP | T | F  | GTTATGGAGGATTGGTAAGAACCAGAG   |
| 193 | rs987640   | ISNP | T | F  | GAGCTGTTTAAAGGGTAAAGGGGTAGTTA |
| 194 | rs9905977  | ISNP | T | F  | GCAGACAAAACCATGACAATGATCTTAG  |
| 195 | rs993934   | ISNP | T | F  | CCCATGATGAACAGTTTGCACATAATG   |
| 196 | rs9951171  | ISNP | T | F  | CTCAATTTCTTGTCCCTGCTTTCATG    |
| 197 | rs10092491 | ISNP | S | R2 | TTAGAAATCCAGATAGAGCTAAACTG    |
| 198 | rs1015250  | ISNP | S | R  | GTTAGGAAAGAACCAGGTGTTTT       |
| 199 | rs1024116  | ISNP | S | R2 | GCAAAAGTAAATACAAAGGCATCTTT    |
| 200 | rs1028528  | ISNP | S | R2 | CAATGCAAAAGAAAGTCTTACTCGAC    |
| 201 | rs1029047  | ISNP | S | R2 | CAITTCCTAAACTCTAAACAAACATTTG  |
| 202 | rs1031825  | ISNP | S | R2 | GGTCCCTAACCTATTAAATTTAATGAG   |
| 203 | rs10488710 | ISNP | S | R  | GACTTCAATTTATGTCAGCATTTAAAA   |
| 204 | rs10495407 | ISNP | S | R  | CCTCTGGTTCATTGGATTCTCATTTG    |
| 205 | rs1058083  | ISNP | S | R  | TCTCCATGAAACTTGGGTTAATTTTG    |
| 206 | rs10773760 | ISNP | S | R  | TGCTGGAGTTCGTCAAATTCAG        |
| 207 | rs1294331  | ISNP | S | R2 | GTAGCATAAAACATCCAAAAATCAAT    |
| 208 | rs12997453 | ISNP | S | R  | TGCTTTAAAGATACAGGTATCTGTATTAC |
| 209 | rs13182883 | ISNP | S | R  | CTCTCCGTTACTTCTCTCTGCTTT      |
| 210 | rs13218440 | ISNP | S | R  | GATCCTGAGATTCACCTCTAGTCCT     |
| 211 | rs1335873  | ISNP | S | R  | CCGTACCAGGTACCTAGCTATGTACT    |
| 212 | rs1336071  | ISNP | S | R2 | CTTCTGTTTGTCCATCTGAAATTC      |
| 213 | rs1355366  | ISNP | S | R  | CAAAGTTAAGTATCACCATCCAGCTGG   |
| 214 | rs1357617  | ISNP | S | R  | ATAGGGATAGCTGATAAGAAACATGACC  |
| 215 | rs1382387  | ISNP | S | R  | CTAATAAGACGCTGCATCTGCCCA      |
| 216 | rs1413212  | ISNP | S | R  | TCCAGGAGACATTTGTTTATATAAGTGA  |
| 217 | rs1454361  | ISNP | S | R  | AGACACTTTTCAGTATCCATTTAGAAAC  |
| 218 | rs1463729  | ISNP | S | R  | GTTTCACATGTGCATGCTTTGGGT      |
| 219 | rs1493232  | ISNP | S | R  | CCAAAGCTATTCTCTCTTTGGGTGC     |
| 220 | rs1498553  | ISNP | S | R  | GAAAGTTCACTTCAGATGTTCAAAGCC   |
| 221 | rs1523537  | ISNP | S | R  | GGGTTTCAGTCTGCAACAAGATCTTG    |
| 222 | rs1528460  | ISNP | S | R  | TGGAGATCAATATTAGCCTTAACATAT   |
| 223 | rs159606   | ISNP | S | R  | GACTGTTTCTCATCCGTATTATTGT     |
| 224 | rs1736442  | ISNP | S | R  | AACACACAGAAACATCAAGCTGAGC     |
| 225 | rs1821380  | ISNP | S | R  | TTCTGACATTCCTCTCTCTATCTG      |
| 226 | rs1886510  | ISNP | S | R  | TATGACGCTGGATTTCACAACAAC      |
| 227 | rs1979255  | ISNP | S | R  | CAGAGACTATGGATGGTATTAGGTCAA   |
| 228 | rs2016276  | ISNP | S | R  | ACTTTGTGTGGCTGAGAGAGAGAAA     |
| 229 | rs2040411  | ISNP | S | R  | TGAGTGTCTCTGTATTTCTTACTCTAAG  |
| 230 | rs2046361  | ISNP | S | R2 | ATTTTGGTCAATTTGACACTTCACC     |
| 231 | rs2056277  | ISNP | S | R2 | GGTGTAGGGAGACAGGCATGAATG      |
| 232 | rs2076848  | ISNP | S | R  | TGAAACTTTTCAACTCTCTACCGCC     |
| 233 | rs2107612  | ISNP | S | R  | GTTAAATTTGCCACTAATTATGTGTTT   |
| 234 | rs2111980  | ISNP | S | R  | AACGATCTATGCAAGCAAGATCTTTG    |
| 235 | rs214955   | ISNP | S | R  | GATGCTTGCAACAACAGACTGAAAAGG   |
| 236 | rs221956   | ISNP | S | R  | GTCGTGTGTCTCTGAGATGATGAATG    |
| 237 | rs2342747  | ISNP | S | R  | GGGAGGAAGAAAACAGAGAGTCTTGA    |
| 238 | rs2399332  | ISNP | S | R  | AGTTTGTGGTCTCTTTGAGAAGTATC    |

|     |            |      |    |    |                                |
|-----|------------|------|----|----|--------------------------------|
| 239 | rs251934   | SNP  | S  | R  | GGCAGATGAAGTAGATATCTGGCTG      |
| 240 | rs279844   | SNP  | S  | R  | GTTCAAGTCAATTTTGACCAGATATT     |
| 241 | rs2830795  | SNP  | S  | R  | AGACATAGGACACACCATTTTATTGTCT   |
| 242 | rs2831700  | SNP  | S  | R  | TCAAAATATTTGGCTAAACTATTGCCGG   |
| 243 | rs2920816  | SNP  | S  | R2 | CTGGAGTTATTAAATAATTGGATTATAGC  |
| 244 | rs321198   | SNP  | S  | R  | TTACCTGTTTCCCTTTGTGATCCAC      |
| 245 | rs338882   | SNP  | S  | R  | ACCAAGTCAAGAGCTCTGAGAGACAT     |
| 246 | rs354439   | SNP  | S  | R  | ACAGTGAATGATATTGAGAATATTGTGC   |
| 247 | rs3780962  | SNP  | S  | R  | GAACAAGGTCAAGATATCAGCTTACC     |
| 248 | rs430046   | SNP  | S  | R  | AGGTCATACAATGAATGGTGTGATGT     |
| 249 | rs4364205  | SNP  | S  | R  | ATCCACCATGAGAAATATATCCACAA     |
| 250 | rs445251   | SNP  | S  | R  | ACAATTCAAATTAATGTAAAACTGCAAGTG |
| 251 | rs4606077  | SNP  | S  | R  | TAGTTCTAGTGTGGGATCTGACTCC      |
| 252 | rs560681   | SNP  | S  | R  | GAACATCTGTTGAGGTTTCTCTCCATC    |
| 253 | rs6444724  | SNP  | S  | R  | GAAAGGACTAAATGTTGAACACTGGT     |
| 254 | rs6811238  | SNP  | S  | R  | TGTGTGTTTAAAGCCAGGTTTGT        |
| 255 | rs7041158  | SNP  | S  | R  | GATGGACTGGAACCTGAGGATTTTCA     |
| 256 | rs717302   | SNP  | S  | R  | AGCTTTAGAAAGGCATATCGTATTAACG   |
| 257 | rs719366   | SNP  | S  | R  | TTATAGTGAGTAAAGGACAGGCCCC      |
| 258 | rs722098   | SNP  | S  | R2 | ACACATCTGTTGACAGTAATGAAATATCC  |
| 259 | rs722290   | SNP  | S  | R  | GTTTAAACTGGATACCATCCCCAAGAC    |
| 260 | rs727811   | SNP  | S  | R  | ATGAGATTGCTGGGAGATGCAGATG      |
| 261 | rs729172   | SNP  | S  | R  | CACATTTCCCTCTTGCAGTTACATAC     |
| 262 | rs733164   | SNP  | S  | R  | GACAAGCCTCGCTTGAGTTTCTTT       |
| 263 | rs735155   | SNP  | S  | R  | TGTGAGAGTGTCACCGAATCAACG       |
| 264 | rs740598   | SNP  | S  | R  | AAATAGCAATGGCTCGTCTATGGTTAG    |
| 265 | rs740910   | SNP  | S  | R  | TGCTAAGTAAGGTGAGTGGTATAATCA    |
| 266 | rs763869   | SNP  | S  | R  | ATAAATATGATGTGGCTACTCCCTCAT    |
| 267 | rs8037429  | SNP  | S  | R  | GCTACACCTCCATAGTAATAATGTAAGAG  |
| 268 | rs8078417  | SNP  | S  | R  | TGAAGCAGCTAGAGAACTCTGTACGT     |
| 269 | rs826472   | SNP  | S  | R2 | TTTTGTCTCTGTTATATTAGTCACTACTC  |
| 270 | rs873196   | SNP  | S  | R  | ATAGCCCTGCATTCAAATCCCAAGTG     |
| 271 | rs876724   | SNP  | S  | R  | TCCATTTTATACCACTGCCTGAAG       |
| 272 | rs891700   | SNP  | S  | R2 | GCAGTAAACATTTTCATCAAATTTCCA    |
| 273 | rs901398   | SNP  | S  | R  | TCTGGGTGCAAACTAGCTGAATATCAG    |
| 274 | rs907100   | SNP  | S  | R  | GAAATCTGGAGGC AATTCATGATGCC    |
| 275 | rs914165   | SNP  | S  | R  | ATACAATGATGATCACACGGGACCCCT    |
| 276 | rs917118   | SNP  | S  | R  | CCATGAAGATGGAGTCAACATTTTACA    |
| 277 | rs938283   | SNP  | S  | R  | TCCTAACCCCTAGTACGTTAGATGTG     |
| 278 | rs964681   | SNP  | S  | R  | GAGGTGATTCTCTGTGAGGAACGTCG     |
| 279 | rs987640   | SNP  | S  | R  | GTACATTCACCTTAACAGGCTCTCTTCC   |
| 280 | rs9905977  | SNP  | S  | R  | AATTCATGAGCTGGTGTCCAAGGAG      |
| 281 | rs993934   | SNP  | S  | R  | ATAACAGTCTCCAGAGTATATTAGCTTAG  |
| 282 | rs9951171  | SNP  | S  | R  | GTTCTCTGGGATGCAACATGAGAG       |
| 283 | rs10497191 | aSNP | T  | F  | GAAAGGATGAAGAGGGTGGATATTGGAG   |
| 284 | rs1079597  | aSNP | T  | F  | CCAAACCTCATCTCTTACCTGGATT      |
| 285 | rs11652805 | aSNP | T  | F  | GTCCAAAGTCAAGTGCAAGTATAGTTGG   |
| 286 | rs1229984  | aSNP | T  | F  | ACAATCTTTCTGAACTGAACAGCTTC     |
| 287 | rs12439433 | aSNP | T  | F  | CAAAGGAAGGCATTTCTAATGATCTTC    |
| 288 | rs12498138 | aSNP | T  | F  | CTTGCCTTGTCTTTCTCTTCAGGGAA     |
| 289 | rs12913832 | pSNP | NU | T  | CTGCTTCAAGTGATATAAACTCAGAT     |
| 290 | rs1426654  | aSNP | T  | F  | CCTAGGAAAGCAGTAACATTCAGGAG     |
| 291 | rs1462906  | aSNP | T  | F  | GCAATTTGTTCACTTTTAGTTCTGTAGC   |
| 292 | rs1572018  | aSNP | T  | F  | GGCTAATATGCATGTGTTCTATGCTCT    |
| 293 | rs16891982 | aSNP | T  | F  | CAGAGTTTCTATCTACGAAAGAGGAGT    |

|     |                         |                                |
|-----|-------------------------|--------------------------------|
| 294 | rs174570_aSNP1_T_F      | ATCCTAGACCTCCAGGTGGAATGATC     |
| 295 | rs17642714_aSNP1_T_F    | CTTGGCTGTCTCAATATTTTGGAGTAAG   |
| 296 | rs1800414_aSNP1_T_F     | GAGTAAATGAGCTGTGGTTTCTCTCTTA   |
| 297 | rs1834619_aSNP1_T_F     | CTTCCATGTGGACCCCTTAAACATTACAG  |
| 298 | rs1876482_aSNP1_T_F     | GCATAGTGAGCTGTTGATAGAGCTTTTG   |
| 299 | rs1919550_aSNP1_T_F     | CTAGAACAAAATCATTGGCTCTCCTAGT   |
| 300 | rs192655_aSNP1_T_F      | GTCCTGGTGAGTACTGGCTGAATGTAAA   |
| 301 | rs200354_aSNP1_T_F      | CCAGAGGATGCTGCTAAACATTCTACAA   |
| 302 | rs2024566_aSNP1_T_F     | GCTCATGCCTGGAATTCACCTTTATTTT   |
| 303 | rs2042762_aSNP1_T_F     | CTAACTAGACATTTGGGCCACCTTACTT   |
| 304 | rs2166624_aSNP1_T_F     | GTCATGGTGCTATAGAATGTACAGGT     |
| 305 | rs2196051_aSNP1_T_F     | CCCTCTCAAGTTTGTGAGCAAAATATCAC  |
| 306 | rs2238151_aSNP1_T_F     | CTCTATCTTGCTGCAATGGACTTTCC     |
| 307 | rs260690_aSNP1_T_F      | CCTAGAAACAGATTTTGAAGGGCTCTTG   |
| 308 | rs2814778_aSNP1_T_F     | AAATGAGGGGCATAGGGATAAGGGA      |
| 309 | rs310644_aSNP1_T_F      | CCTAGAAATCTGATACGTTATCCTATGA   |
| 310 | rs3737576_aSNP1_T_F     | AGGAGAGATATATTCAACATGAACCCAA   |
| 311 | rs3811801_aSNP1_T_F     | GAACATCTCTGACCAGAAATTTCCAGTA   |
| 312 | rs3823159_aSNP1_T_F     | GTGTAGTGAAATCTTAGACTTAGGTAA    |
| 313 | rs3916235_aSNP1_T_F     | AATACATGAAAAGTAATACATGGGGCA    |
| 314 | rs4471745_aSNP1_T_F     | ATTAAATGTTTACTTCTATCTACAAGGA   |
| 315 | rs4833103_aSNP1_T_F     | CATTTTGTGAAATGCAAGGGCAAAATCT   |
| 316 | rs4891825_aSNP1_NU_T_F  | GCTGAGAGGCTTAATTCATCAAGATGA    |
| 317 | rs4918664_aSNP1_NU_T_F  | CCCCTCTAACTTAGTTTATGAGGAG      |
| 318 | rs6754311_aSNP1_T_F     | GTAACACATCTCTTTGGGAAGCTAGC     |
| 319 | rs6990312_aSNP1_NU_T_F  | CTTAGCTTCAGTGAAAATGGTTCCTCTC   |
| 320 | rs7226659_aSNP1_NU_T_F  | CTTTCTTAGCTCTCTCCATTCTCTTC     |
| 321 | rs7326934_aSNP1_NU_T_F  | GTCATGCAGTGCTTCACTGAGGATTAT    |
| 322 | rs735430_aSNP1_NU_T_F   | CTCTATCTGCTCAGAGCCTGCTTAAAG    |
| 323 | rs7554936_aSNP1_NU_T_F  | GGAAAGGATACAGTGTTGAGCAAGATAG   |
| 324 | rs7657799_aSNP1_NU_T_F  | GCCAACTTGATTCTCTTCAAAATGCTTG   |
| 325 | rs7722456_aSNP1_T_F     | AGATGGGGTTTACCATGTTTCCAG       |
| 326 | rs798443_aSNP1_T_F      | GTACAGTAGTTAGTTCCAGACTGATGA    |
| 327 | rs7997709_aSNP1_T_F     | GTAATAATCTAACTGTGTTTCCCTCAGT   |
| 328 | rs870347_aSNP1_T_F      | GAACCAAAAGGAATTAAGAGACTAGGGG   |
| 329 | rs917115_aSNP1_T_F      | CTGCTTTACGGCTTCTCTTCTCTTC      |
| 330 | rs10497191_aSNP1_S_R    | CCCACATCCTTCCATTATAGGCAA       |
| 331 | rs1079597_aSNP1_S_R     | TACATGATCCTAAGGGCAGCAGGAA      |
| 332 | rs11652805_aSNP1_S_R    | GTTTGGTGATCCTCTTTCTCTCTC       |
| 333 | rs1229984_aSNP1_S_R     | GACTGTAGTCACCCCTCTCTCAACA      |
| 334 | rs12439433_aSNP1_S_R    | AGAGTGAAATACATAGAAAAGAACTTAAAG |
| 335 | rs12498138_aSNP1_S_R    | ATTGCGAGAAACAGATAAATATTGAAG    |
| 336 | rs12913832_pSNP1_NU_S_R | ACAGGAACAAAGAAATTTGTTCTCATGG   |
| 337 | rs1426654_aSNP1_S_R     | CCTTGGATTGTCTCAGGATGTTGCA      |
| 338 | rs1462906_aSNP1_S_R     | CTGGGATGTTGTTTGGCTTTGTG        |
| 339 | rs1572018_aSNP1_S_R     | ATTGCTAGTACACTAATGGATATATGTGAG |
| 340 | rs16891982_aSNP1_S_R    | GAATAAAGTGAGGAAAACACGGAGTTG    |
| 341 | rs174570_aSNP1_S_R      | GAGAGAGGCAGAAAGGAGGGATGAA      |
| 342 | rs17642714_aSNP1_S_R    | TACTCTGTCTTCAGTAGCTGTTCTTGG    |
| 343 | rs1800414_aSNP1_S_R     | TTAGACTCACCAAGATCAAGATGAATGC   |
| 344 | rs1834619_aSNP1_S_R     | ATCTCAATAAAGCTGTTCAAAACAGAAAG  |
| 345 | rs1876482_aSNP1_S_R     | TAAAGAAAATGCCATGGGCTGTACCC     |
| 346 | rs1919550_aSNP1_S_R     | ATTGTGAGCAGAACAGAGTGTAGTG      |
| 347 | rs192655_aSNP1_S_R      | ATTCTTTGATAGCTCACGAAATTTCCC    |
| 348 | rs200354_aSNP1_S_R      | AAAATGAGACCTCGTATCTTTGCAGC     |

|     |                        |                               |
|-----|------------------------|-------------------------------|
| 349 | rs2024566_aSNP1_S_R    | AAATGCAGAACTGCCAAAAGAAACCC    |
| 350 | rs2042762_aSNP1_S_R    | GAGAACTCTGTAATGCCAGGGTCTG     |
| 351 | rs2166624_aSNP1_S_R    | ATGGATTCTGTTTCAGACATCTAATT    |
| 352 | rs2196051_aSNP1_S_R    | ATCACTAGAAAGAAAAGAGTTCCTATT   |
| 353 | rs2238151_aSNP1_S_R    | GAAGTTTAAAGAGTGGGAACATGGGG    |
| 354 | rs260690_aSNP1_S_R     | CTACGTAAGCAAAAATGATCACCAC     |
| 355 | rs2814778_aSNP1_S_R    | AACCTGATGGCCCTCATTAGTCTT      |
| 356 | rs310644_aSNP1_S_R     | CACCAGATTCTAGGAATAGCATGTGAG   |
| 357 | rs3737576_aSNP1_S_R    | AAGAGCATAGTGAGGGTTAGACCT      |
| 358 | rs3811801_aSNP1_S_R    | CTTTATTTAGTGTAGAGATCAGTCTC    |
| 359 | rs3823159_aSNP1_S_R    | TGAGTCTCTTACCTAATCTTGGTTGTC   |
| 360 | rs3916235_aSNP1_S_R    | AATCCAAAGCAACTCTCTTTTACCAC    |
| 361 | rs4471745_aSNP1_S_R    | TTTACTGGAACCTGATTTTGTGGA      |
| 362 | rs4233103_aSNP1_S_R    | TGCCACTGATATATCAGTACCTGAGT    |
| 363 | rs4291825_aSNP1_NU_S_R | ACAATCTCAATCCCCCTTAATGTTTC    |
| 364 | rs4918664_aSNP1_NU_S_R | GTGGGCAGAGAGAGTAAGAGAACCT     |
| 365 | rs6754311_aSNP1_S_R    | CAAACCAGATTCTGGCAGAAATAGTTAGC |
| 366 | rs6990312_aSNP1_NU_S_R | CTTCTCTCCATCTCTCTCTCCAC       |
| 367 | rs7226659_aSNP1_NU_S_R | AGATCAAGGGATCTGTGGGACAATAAC   |
| 368 | rs7326934_aSNP1_NU_S_R | GGGGAGTGATTCTCAAGCATCTGATT    |
| 369 | rs735480_aSNP1_NU_S_R  | CATGAGTTTGAGGTAAGATGAAGGAGA   |
| 370 | rs7554936_aSNP1_NU_S_R | TCTCTCTCATCTAGTGAATGCCATC     |
| 371 | rs7657799_aSNP1_NU_S_R | GGGTGATGATCTACCTTGAGGTATA     |
| 372 | rs7722456_aSNP1_S_R    | CTCAAGGCCCTGGGTCTGAAATTAC     |
| 373 | rs798443_aSNP1_S_R     | ACATCTCCAGTTAATAATTCCACTAAC   |
| 374 | rs7997709_aSNP1_S_R    | TGGATTGCTCAACAAATAGTCTAAAA    |
| 375 | rs870347_aSNP1_S_R     | CATGCGACATCCAGGTAGCTAAAATAC   |
| 376 | rs917115_aSNP1_S_R     | ATGGATAAAATGGAACCTTCAAGAGAA   |
| 377 | rs12203592_pSNP1_T_F   | GTTTTATGTAAAGCTTCGTCATATGGCT  |
| 378 | rs12821256_pSNP1_T_F   | GTTCCAACCTAGTCATAAAGTTCCTGG   |
| 379 | rs12896399_pSNP1_T_F   | GGGTCTGATGTTGTATTGATGAGGAAG   |
| 380 | rs1393350_pSNP1_T_F    | CCTAACAGAAAGTCACTGTTGTATCTG   |
| 381 | rs1800407_pSNP1_T_F    | TCACTCTGGCTTGACTCTCTCTGTG     |
| 382 | rs2378249_pSNP1_T_F    | GGCTGGTTTCAGTCTGGAGACTTATTT   |
| 383 | rs2402130_pSNP1_T_F    | CTTCACTCGATGACGATGATGATGAT    |
| 384 | rs4959270_pSNP1_T_F    | GACAATAACAGCCACAAAGGATGGAAAAG |
| 385 | rs1805009_pSNP1_T_F    | GAACCAGACCACACAATATCACCAC     |
| 386 | rs28777_pSNP1_T_F      | TCTACCTCTTTGATGTCCCTTCGATAG   |
| 387 | rs16891982_pSNP1_T_F   | CAGAGTTTCTCATCTACGAAAGAGGAGT  |
| 388 | rs683_pSNP1_T_F        | CCCAGCTTTGAAAAGTATGCCTAGAACT  |
| 389 | rs12913832_pSNP1_T_F   | CTGCTTCAAGTGTATATAAACTCACAGT  |
| 390 | rs12203592_pSNP1_S_R   | TTGTTTCATCCACTTTGGTGGGTAAAAG  |
| 391 | rs12821256_pSNP1_S_R   | TAATTAAGCTCTGTGTTTAGGGTTTTT   |
| 392 | rs12896399_pSNP1_S_R   | CAATCTTTGTTCTTTAGGTCAATATAT   |
| 393 | rs1393350_pSNP1_S_R    | TACTCTTCTCAGTCCCTTCTCTGC      |
| 394 | rs1800407_pSNP1_S_R    | TGAGACAGAGCATGATGATCATGGC     |
| 395 | rs2378249_pSNP1_S_R    | GCACAAGTCTAGGAACACTTTTGAC     |
| 396 | rs2402130_pSNP1_S_R    | GAAGTATTTGAACCATACGGAGCCC     |
| 397 | rs4959270_pSNP1_S_R    | TGAGGAACACATCCAAACTATGACAC    |
| 398 | rs1805009_pSNP1_S_R    | TTTCTCGCCCTCATCATCTGCAATG     |
| 399 | rs28777_pSNP1_S_R      | TCAGTTGATTTTCATGTGATCCTCACAG  |
| 400 | rs16891982_pSNP1_S_R   | GAATAAAGTGAGGAAAACACGGAGTTG   |
| 401 | rs683_pSNP1_S_R        | ATTACCTTCTTCTAATACAAGCATATG   |
| 402 | rs12913832_pSNP1_S_R   | ACAGGAACAAAGAAATTTGTTCTCATGG  |

### Ejemplo 2-perfilado de ADN para tipos de bases de datos

Este ejemplo describe un experimento que sigue el flujo de trabajo de la FIG. 2. Este ejemplo no utiliza UMI, ya que podría suponerse que las muestras obtenidas fuesen de individuos cuya identidad ya se conoce.

- 5 Para este experimento, las STR se multiplexan con iSNP como se indica en la Tabla 3.

Tabla 3-SNP informativas de identidad y STR

| SNP informativos de identidad |           |           |           |           |
|-------------------------------|-----------|-----------|-----------|-----------|
| rs1005533                     | rs1357617 | rs2076848 | rs4530059 | rs763869  |
| rs10092491                    | rs1360288 | rs2107612 | rs4606077 | rs8037429 |
| rs1015250                     | rs1382387 | rs2111980 | rs560681  | rs8078417 |
| rs1024116                     | rs1413212 | rs214955  | rs576261  | rs826472  |
| rs1028528                     | rs1454361 | rs221956  | rs6444724 | rs873196  |
| rs1029047                     | rs1463729 | rs2269355 | rs6811238 | rs876724  |
| rs1031825                     | rs1490413 | rs2342747 | rs6955448 | rs891700  |
| rs10488710                    | rs1493232 | rs2399332 | rs7041158 | rs901398  |
| rs10495407                    | rs1498553 | rs251934  | rs717302  | rs907100  |
| rs1058083                     | rs1523537 | rs279844  | rs719366  | rs914165  |
| rs10773760                    | rs1528460 | rs2830795 | rs722098  | rs917118  |
| rs10776839                    | rs159606  | rs2831700 | rs722290  | rs938283  |
| rs1109037                     | rs1736442 | rs2920816 | rs727811  | rs964681  |
| rs1294331                     | rs1821380 | rs321198  | rs729172  | rs987640  |
| rs12997453                    | rs1886510 | rs338882  | rs733164  | rs9905977 |
| rs13182883                    | rs1979255 | rs354439  | rs735155  | rs993934  |
| rs13218440                    | rs2016276 | rs3780962 | rs737681  | rs9951171 |
| rs1335873                     | rs2040411 | rs430046  | rs740598  |           |
| rs1336071                     | rs2046361 | rs4364205 | rs740910  |           |
| rs1355366                     | rs2056277 | rs445251  | rs7520386 |           |

STR autosómicas

|         |          |          |          |             |
|---------|----------|----------|----------|-------------|
| D1S1656 | CSF1PO   | vWA      | D21S11   | D4S2408     |
| D2S441  | D7S820   | D13S317  | TPOX     | D17S1301    |
| D2S1338 | D8S1179  | Penta E  | SE33     | D9S1122     |
| D3S1358 | D10S1248 | D16S539  | Penta D  | D6S1043     |
| FGA     | TH01     | D18S51   | D22S1045 | Amelogenina |
| D5S818  | D12S391  | D19S433  | D20S482  |             |
| X STRs  |          |          |          |             |
| DXS8378 | DXS8377  | DXS10101 | DXS10148 | DXS10146    |
| DXS7132 | DXS10135 | DXS10134 | DXS10079 |             |

|            |          |         |          |             |
|------------|----------|---------|----------|-------------|
| HPRTB      | DXS10074 | DXS7423 | DXS10103 |             |
| Y STRs     |          |         |          |             |
| DYS456     | DYS393   | DYS437  | DYS533   | DYS449      |
| DYS389I/II | DYS391   | DYS438  | DYS518   | DYS522      |
| DYS390     | DYS439   | DYS448  | DYS570   | DYS505      |
| DYS458     | DYS635   | DYS576  | DYS643   | DYS627      |
| DYS19      | DYS392   | DYS481  | DYS460   | DYF387S1a/b |
| DYS385a/b  | YGATAH4  | DYS549  | DYS612   |             |

Por supuesto, podrían añadirse SNP y STR adicionales a la lista anterior. Los ejemplos de otras posibles dianas incluyen, pero sin limitación, los marcadores encontrados en la Tabla 4.

Tabla 4-ejemplos de STR y SNP adicionales para la multiplexación

| SNP informativos de identidad |           |           |           |           |
|-------------------------------|-----------|-----------|-----------|-----------|
| rs1004357                     | rs1554472 | rs2567608 | rs521861  | rs9606186 |
| rs1019029                     | rs1872575 | rs2811231 | rs5746846 | rs985492  |
| rs1027895                     | rs2073383 | rs2833736 | rs590162  | rs9866013 |
| rs10500617                    | rs2175957 | rs315791  | rs6591147 |           |
| rs10768550                    | rs2255301 | rs3744163 | rs689512  |           |
| rs12480506                    | rs2270529 | rs4288409 | rs7205345 |           |
| rs13134862                    | rs2272998 | rs464663  | rs7229946 |           |
| rs1358856                     | rs2291395 | rs4789798 | rs7704770 |           |
| rs1410059                     | rs2292972 | rs4796362 | rs8070085 |           |
| rs1478829                     | rs2503107 | rs4847034 | rs9546538 |           |
| STR autosómicas               |           |           |           |           |
| D1S1677                       | D3S4529   | D18S853   | D10S1435  |           |
| D11S4463                      | D6S1017   | D14S1434  | D5S2500   |           |
| D1S1627                       | D1GATA113 | D2S1776   |           |           |

Se diseñaron cebadores que contenían una secuencia de cebador de PCR específica de gen en el extremo 3' y una secuencia marcadora adaptadora en el extremo 5'. En este experimento, los cebadores directos contenían la secuencia marcadora para los adaptadores i5 de TriSeq Custom Amplicon y los cebadores inversos contenían la secuencia marcadora para los adaptadores i7 del kit de ARN TruSeq Small. Las etiquetas pueden usarse como sitios de cebadores de amplificación, así como sitios de cebadores de secuenciación.

Secuencia marcadora de adaptador i5      5' TACACGACGCTCTTCCGATCT3' (SEQ ID NO: 403)

Secuencia marcadora de adaptador i7      5' CTTGGCACCCG AGAATTCCA3' (SEQ ID NO: 404)

Para equilibrar la amplificación entre las STR y los SNP en el multiplex, se modificaron los parámetros de diseño de cebador para los SNP como se describe en el Ejemplo 1. El conjunto original de cebadores de SNP diseñados usando el Design Studio de Illumina eran cebadores de PCR clásicos - secuencias cortas con altas temperaturas de fusión y poca o ninguna estructura secundaria. Se usó Design Studio para diseñar sondas de TruSeq Custom Amplicon y para crear el complemento inverso de la sonda corriente abajo para elaborar el cebador PCR inverso. Sin embargo, estos cebadores, no se multiplexaron bien y un mal cebador podría transformar el ensayo de bueno a malo (por ejemplo, todos dímeros de cebadores y ningún producto) (FIG. 4). En un intento por crear mejores cebadores para la multiplexación, se usó un Primer3 (shareware) que contenía una característica de biblioteca de cebado erróneo. Se descubrió que los cebadores diseñados por Primer3 se comportaban incluso peor en el ensayo multiplex que los cebadores de Design Studio. Sorprendentemente, se estaban generando datos que mostraban que

los cebadores de STR se multiplexaban bien. Se observó que los pares de cebadores mal diseñados dirigidos a dianas de STR no causaban fallos en el ensayo multiplex como lo hacían los cebadores de SNP. Los cebadores de STR son largos, ricos en AT, y tienen temperaturas de fusión bajas, contrario a lo que se considera un “buen” cebador.

Los cebadores de SNP se rediseñaron siguiendo los parámetros del Ejemplo 1. Los cebadores se mezclaron entre sí para todas las dianas. Para este ejemplo, se mezclaron pares de cebadores para 56 STR con pares de cebadores para 75 iSNP, aSNP y SNP informativos de fenotipo. Se midió polimerasa (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR y se añadieron los cebadores. La mezcla se transfirió con una pipeta a los pocillos de una placa de PCR, pero la amplificación también podría realizarse en tubos, etc. Se añadió ADN a la placa como ADN purificado en un volumen de 15 microlitros, sin embargo, también podrían usarse extractos lisados de muestras de sangre o bucales procedentes de frotis o papel de filtro no tratado, o directamente de muestras de sangre o de boca en tarjetas FTA, etc. Para este experimento, se utilizó ADN 2800M de control purificado a 1 ng y 100 pg. Las reacciones se sometieron a PCR durante un número de ciclos determinado (en el caso del ejemplo, 25 ciclos) siguiendo el protocolo:

|       |   |        |                    |
|-------|---|--------|--------------------|
| 95 °C | - | 3 min  |                    |
| 96 °C | - | 1 min  |                    |
| 54 °C | - | 2 min* | *aumento 0,5 °C/s  |
| 68 °C | - | 45 s** | **aumento 0,2 °C/s |
| 60 °C | - | 30 min |                    |

|      |   |               |  |
|------|---|---------------|--|
| 4 °C | - | Mantenimiento |  |
|------|---|---------------|--|

Después de los ciclos, las placas se retiraron del termociclador. La reacción se llevó a 50 microlitros con polimerasa (Kapa HiFi, Kapa Biosystems), mezcla maestra de PCR que contenía todos los componentes necesarios para la PCR y un par de adaptadores (un adaptador i7 y un adaptador i5). Se realizó una segunda ronda de PCR durante un número de ciclos determinado (10 ciclos en el caso del ejemplo) para generar las bibliotecas de secuenciación, siguiendo el protocolo:

|       |   |               |
|-------|---|---------------|
| 98 °C | - | 30 s          |
| 98 °C | - | 15 s          |
| 66 °C | - | 30 s          |
| 72 °C | - | 1 min         |
| 72 °C | - | 5 min         |
| 10 °C | - | mantenimiento |

10

Después de los ciclos, la placa que contenía las bibliotecas completas se retiró del termociclador. En este momento, las muestras pueden agruparse en volumen y purificarse como una sola muestra usando perlas magnéticas (SPR1) por ejemplo. Las muestras también pueden purificarse individualmente. El conjunto o las bibliotecas individuales pueden cuantificarse usando un método basado en qPCR, usando un Fragment Analyzer o BioAnalyzer, o usando PicoGreen y un lector de placa (como en el caso del ejemplo). Un experto en la materia sabrá la gran cantidad de opciones para la cuantificación

15

de bibliotecas. Si las bibliotecas se purifican individualmente, pueden normalizarse a 2 nM en cada concentración y agruparse en volumen.

Los conjuntos de bibliotecas purificadas se desnaturalizaron, diluyeron, agruparon y secuenciaron en el instrumento de secuenciación MiSeq con un  
5 proceso de secuenciación de 350 ciclos y las dos lecturas de índices. Después de la secuenciación, las muestras se desmultiplexaron de acuerdo con las secuencias adaptadoras y se analizaron a través del canal Forensics Genomics (Illumina, Inc.). Las lecturas de STR se separaron de las lecturas de SNP y se analizaron independientemente. Las STR se analizaron usando el algoritmo  
10 descrito en una solicitud de patente anterior (PCT/US2033/30867, incorporada en el presente documento por referencia en su totalidad). El número (o números) repetido(s) y cualquier variación de secuencia se indicaron junto con los números de lectura. Los SNP se analizaron usando un manifiesto y las validaciones se notificaron junto con los números de lectura. Se calcularon el  
15 equilibrio relativo entre alelos (Mín/Máx%), el equilibrio entre locus (% VC), las tasas de error y las tasas de tartamudeo para los loci de STR. Los resultados para las STR en el multiplex de bases de datos inicial se muestran en la FIG. 5A-C. El equilibrio (equilibrio medio 80 %), tartamudeo (~3 %) y las tasas de error (menos de 5 %) cumplen los requisitos de entrada del diseño para los loci  
20 incluidos en este ejemplo. El % VC (~142 %) se calculó usando los 56 loci. A pesar de que los cebadores utilizados muestran equilibrio entre loci, se espera una mayor optimización de cebadores para mejorar el equilibrio entre loci. Las validaciones para los loci conocidos coinciden con los resultados publicados para 2800M. Los resultados para los SNP se muestran en la FIG. 5D-E. El  
25 alcance, validaciones de alelos, tartamudeo y otros artefactos para los 56 loci

de STR en el multiplex grande se muestran en la FIG. 6. Estos gráficos imitan los electroferogramas generados por tecnología EC. Las barras son análogas a picos para el alelo definido (eje X) y los recuentos de lectura (eje Y) son análogos a UFR. El alcance para los SNP era en cualquier parte de 10-2500X  
5 dependiendo del SNP, sin embargo, cada SNP que se multiplexó se contó y proporcionó validaciones precisas.

### Ejemplo 3-perfilado de ADN para estudios de criminología

Este ejemplo describe un experimento que sigue el flujo de trabajo de la FIG. 3. Este ejemplo incorpora UMI en los cebadores, ya que podría suponerse  
10 que las muestras obtenidas proceden de individuos cuya identidad aún se desconoce.

Para este experimento, los STR se multiplexaron con iSNP, aSNP y SNP informativos de fenotipo como se observa en la Tabla 5.

Tabla 5-STR y SNP de estudios de criminología

| SNP informativos de identidad |           |            |            |             |
|-------------------------------|-----------|------------|------------|-------------|
| rs1005533                     | rs1357617 | rs2076848  | rs4530059  | rs763869    |
| rs10092491                    | rs1360288 | rs2107612  | rs4606077  | rs8037429   |
| rs1015250                     | rs1382387 | rs2111980  | rs560681   | rs8078417   |
| rs1024116                     | rs1413212 | rs214955   | rs576261   | rs826472    |
| rs1028528                     | rs1454361 | rs221956   | rs6444724  | rs873196    |
| rs1029047                     | rs1463729 | rs2269355  | rs6811238  | rs876724    |
| rs1031825                     | rs1490413 | rs2342747  | rs6955448  | rs891700    |
| rs10488710                    | rs1493232 | rs2399332  | rs7041158  | rs901398    |
| rs10495407                    | rs1498553 | rs251934   | rs717302   | rs907100    |
| rs1058083                     | rs1523537 | rs279844   | rs719366   | rs914165    |
| rs10773760                    | rs1528460 | rs2830795  | rs722098   | rs917118    |
| rs10776839                    | rs155606  | rs2831700  | rs722290   | rs938283    |
| rs1109037                     | rs1736442 | rs2920816  | rs727811   | rs964681    |
| rs1294331                     | rs1821380 | rs321198   | rs729172   | rs987640    |
| rs12997453                    | rs1886510 | rs338882   | rs733164   | rs9905977   |
| rs13182883                    | rs1979255 | rs354439   | rs735155   | rs993934    |
| rs13218440                    | rs2016276 | rs3780962  | rs737681   | rs9951171   |
| rs1335873                     | rs2040411 | rs430045   | rs740598   |             |
| rs1336071                     | rs2046361 | rs4354205  | rs740910   |             |
| rs1355360                     | rs2056277 | rs445251   | rs7520336  |             |
| STR autosómicas               |           |            |            |             |
| D1S1656                       | CSF1PO    | vWA        | D21S11     | D4S2408     |
| D2S441                        | D7S820    | D13S317    | TPOX       | D17S1301    |
| D2S1338                       | D8S1179   | Penta E    | SE33       | D9S1122     |
| D3S1358                       | D10S1248  | D16S539    | Penta D    | D6S1043     |
| FGA                           | TH01      | D18S51     | D22S1045   | Amelogenin  |
| D5S818                        | D12S391   | D19S433    | D20S482    |             |
| STR X                         |           |            |            |             |
| DXS8378                       | DXS8377   | DXS10101   | DXS10148   | DXS10146    |
| DXS7132                       | DXS10135  | DXS10134   | DXS10079   |             |
| HPRTB                         | DXS10074  | DXS7423    | DXS10103   |             |
| STR Y                         |           |            |            |             |
| DYS456                        | DYS393    | DYS437     | DYS533     | DYS449      |
| DYS389I/II                    | DYS391    | DYS438     | DYS518     | DYS522      |
| DYS300                        | DYS430    | DYS448     | DYS570     | DYS505      |
| DYS458                        | DYS635    | DYS576     | DYS643     | DYS627      |
| DYS19                         | DYS392    | DYS481     | DYS460     | DYF387S1a/b |
| DYS385a/b                     | YGATAH4   | DYS549     | DYS612     |             |
| SNP informativos de fenotipo  |           |            |            |             |
| N29insA                       | rs1805006 | rs1110400  | rs12203592 | rs2378249   |
| rs11547464                    | rs1805007 | rs28777    | rs1042602  | rs12896399  |
| rs885479                      | rs1805009 | rs16891982 | rs1800407  | rs1393350   |
| rs1805008                     | Y152OCH   | rs12821256 | rs2402130  | rs683       |

|                                 |            |           |            |           |
|---------------------------------|------------|-----------|------------|-----------|
| rs1805005                       | rs2228479  | rs4959270 | rs12913832 |           |
| SNP informativos de ascendencia |            |           |            |           |
| rs10497191                      | rs17642714 | rs2238151 | rs4471745  | rs7554936 |
| rs1079597                       | rs1800414  | rs2593595 | rs459920   | rs7657799 |
| rs11652805                      | rs1834619  | rs260690  | rs4833103  | rs7722456 |
| rs1229984                       | rs1871534  | rs2814778 | rs4891825  | rs798443  |
| rs12439433                      | rs1876482  | rs310644  | rs4918664  | rs7997709 |
| rs12498138                      | rs1919550  | rs3737576 | rs671      | rs870347  |
| rs12913832                      | rs192655   | rs3811801 | rs6754311  | rs917115  |
| rs1426654                       | rs200354   | rs3814134 | rs6990312  | rs9522149 |
| rs1462906                       | rs2024566  | rs3823159 | rs7226659  |           |
| rs1572018                       | rs2042762  | rs3827760 | rs7251928  |           |
| rs16891982                      | rs2166624  | rs3916235 | rs7326934  |           |
| rs174570                        | rs2196051  | rs4411548 | rs735480   |           |

Este ejemplo incluye UMI para los cebadores de STR. Para estos ejemplos, solamente los cebadores de STR contienen UMI, sin embargo, los

5 cebadores tanto de STR como de SNP incluirían UMI si se desea y esa opción no se excluye de la práctica. Para este ejemplo, sin embargo, solo los cebadores de STR incorporan UMI para propósitos demostrativos. Se introdujeron identificadores de molécula única durante dos ciclos de PCR. (FIG. 3). En primer lugar, igual que en el Ejemplo 2, los cebadores de PCR contenían

10 una secuencia cebadora de PCR específica de gen en el extremo 3' y una secuencia marcadora adaptadora en el extremo 5', lo mismo que para las secuencias marcadoras usadas en el Ejemplo 2 para secuencias de i5 e i7. En este experimento, las UMI se colocan entre la secuencia del cebador específica de gen y la secuencia marcadora. En el caso de este ejemplo, hubo cinco

15 bases aleatorias usadas para el UMI tanto en los cebadores directos como en los inversos. Los cebadores se mezclaron entre sí para todas las dianas. La mezcla de cebadores comprendía 26 pares de cebadores de STR autosómica y 86 pares de cebadores de SNP (incluyendo 92 SNP). Se añadió polimerasa

- (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR, y se añadieron los cebadores. La mezcla se transfirió con una pipeta a los pocillos de una placa de PCR. Se añadió ADN a la placa como ADN purificado, óptimamente 1 ng. Como en el
- 5 Ejemplo 2, el ADN purificado del control 2800M se ensayó a 1 ng. La mezcla de reacción multiplex se sometió a dos ciclos de PCR siguiendo el protocolo:

|       |               |                 |
|-------|---------------|-----------------|
| 98 C  | 3 min         |                 |
| <hr/> |               |                 |
| 98 C  | 2 min         |                 |
| 54 C  | 12 min        | aumento 0,2 C/s |
| 72 C  | 4 min         | 2 ciclos        |
| <hr/> |               |                 |
| 4 C   | mantenimiento |                 |

- Después del ciclado, las muestras se retiraron del termociclador y se
- 10 añadió proteína de unión a ADN monocatenario (SSB) de *E. coli* a la reacción. Se contempló que la SSB reduce los dímeros de cebadores por los cebadores específicos de gen marcados no habituales e impide una mayor amplificación de estos cebadores. La SSB se incubó con la muestra en hielo, como alternativa también podría usarse temperatura ambiente o incubación a 37 °C.
- 15 Después de esta incubación, se añadió polimerasa (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR, y la mezcla maestra se añadió a la muestra con un par de adaptadores (adaptadores i7 e i5) y se sometió a un número de ciclos determinado (en este ejemplo 34 ciclos), siguiendo el protocolo:

|      |               |
|------|---------------|
| 95 C | 3 min         |
| 95 C | 30 s          |
| 66 C | 30 s          |
| 72 C | 1 min         |
| 72 C | 5 min         |
| 10 C | mantenimiento |

Las muestras se purificaron con perlas SPRI y las bibliotecas individuales pudieron cuantificarse usando un método basado en qPCR usando un Fragment Analyzer (como en el caso del ejemplo) o BioAnalyzer, o usando

5 PicoGreen y un lector de placa. Las bibliotecas se normalizaron a 2 nM para cada concentración y se agruparon en volumen.

Los grupos de bibliotecas purificadas se desnaturalizaron, diluyeron, agruparon y secuenciaron usando el MiSeq con un proceso de secuenciación de 350 x 100 ciclos y las dos lecturas de índices. Después de la secuenciación,

10 los datos se determinaron como se indica en el Ejemplo 2. Sin embargo, dado que los cebadores contienen UMI, los UMI se usaron para colapsar los datos usando duplicados de PCR para retirar la secuenciación y errores y artefactos de la PCR. Los SNP se analizaron usando un manifiesto y las validaciones se notificaron junto con los números de lectura. Se calcularon el equilibrio relativo

15 entre alelos (%Min/Máx), equilibrio entre loci (% VC), tasas de error, y tasas de tartamudeo (STR, únicamente). Los resultados para el multiplex inicial de la investigación se muestran en la FIG. 7A-E. La cobertura, validaciones de alelo, tartamudeo y otros artefactos para los 26 loci de STR en el multiplex grande se muestran en la FIG. 8. Estos gráficos imitan a los electroferogramas generados

por EC. Las barras son análogas a picos y los recuentos de lectura son análogos a UFR. La cobertura para los SNP era cualquier sitio de 10-5500X dependiendo del SNP, sin embargo, cada SNP que se multiplexó, se contó y proporcionó resultados útiles.

5        Un resultado que se generó por estos estudios fue que el tartamudeo resultó ser un artefacto de PCR. Esta hipótesis se ha propuesto por muchos investigadores (y se ha indicado deslizamiento de la polimerasa en cánceres de colon humanos), pero esto no se ha demostrado para los ensayos de medicina forense. Los UMI pueden utilizarse para demostrar que el tartamudeo es  
10       efectivamente un artefacto de PCR. Los productos con  $n+1$  o  $n-1$  repeticiones tienen los mismos UMI que los productos con el número de repeticiones correcto (FIG. 9). En la FIG. 9A, cada locus muestra los resultados sin corrección de UMI en comparación con la FIG. 9B, en la que se realiza corrección de UMI. Como se demuestra, sin corrección de UMI, el equilibrio  
15       entre los alelos no es tan bueno como cuando se realiza corrección de UMI. Además, hay considerablemente más tartamudeo de lo que aparente sin corrección de UMI. La parte de la barra por encima de la línea entre barras representa el error de secuenciación, mientras que por debajo la línea representa la secuencia correcta dentro de la secuencia de STR. El error se  
20       reduce enormemente con corrección de UMI. Por ejemplo, el locus SE33 tiene errores que se eliminan con corrección de UMI. La corrección de errores puede ser extremadamente importante para la investigación criminal para proporcionar el perfilado de ADN más preciso posible.

#### Ejemplo 4-perfilado de ADN usando 12 individuos de muestra

25       Métodos y materiales

Se ensayó el ADN de 12 individuos de muestra (Muestra n.º: 1, 3, 4, 5, 6, 7, 10, 13, 14, 15, 16, 17) y un genoma de referencia (2800M) siguiendo el flujo de trabajo de la FIG. 3. Este experimento incorpora UMI en los cebadores de STR como se describe en el Ejemplo 3. Se analizaron dos copias de cada muestra con el Kit ForenSeq DNA Signature Library Prep de Illumina® en el secuenciador MiSeq. Se usó un ng de ADN para cada replicación usando la Mezcla de Cebadores de ADN B: mezcla de muestras recogidas, que contiene cebadores para 61 STR más Amelogenina, 95 SNP informativos de identidad, 56 SNP informativos de la ascendencia, 22 SNP informativos de fenotipo (también se usaron 2 SNP de ascendencia para la predicción del fenotipo).

#### Parámetros por defecto

STR: umbral analítico = 6,5 %; umbral de interpretación = 15 %. SNP: umbral analítico = 3 %; umbral de interpretación = 15 %.

Las validaciones de secuenciación de alto nivel, como la cobertura y validaciones de loci para el perfilado de ADN de los 12 individuos de muestra se muestran en la FIG. 12. Como puede observarse, cada locus se cubrió con al menos 100.000 lecturas en ambas replicaciones. Solo dos muestras dieron como resultado una validación de STR fracasada (1 de 61). Los 173 SNP fueron validaciones satisfactorias en todos los individuos en ambas replicaciones. Las validaciones de STR de muestra de dos individuos de muestra se muestran en la FIG. 16. Las validaciones de SNP de muestra de dos individuos de muestra se muestran en la FIG. 17. La FIG. 13 muestra la estadística de la población, tal como la probabilidad de coincidencia aleatoria (RMP *random match probability*) del National Institute of Standards and Technology (NIST) auto-STR, frecuencia de halotipo con 95% de confianza de

NIST Y-STR, RMP de iSNP de dbSNP y RMP de las STR de la base de datos U.S. Y-STR.

Se predijeron los fenotipos, tal como color de ojos y color de pelo de los 12 individuos de la muestra y el individuo de referencia basándose en el genotipo de pSNP en el experimento, y se compararon con los fenotipos 5 autonotificados (FIG. 14). Se observa un alto grado de correlación entre los fenotipos predichos y notificados.

La ascendencia de los 12 individuos de la muestra se predijo usando el genotipo de 56 aSNP en el experimento. Se calcularon las puntuaciones de PCA1 y PCA3 de cada individuo de la muestra, y se representaron frente a 10 muestras de referencia en una representación de la ascendencia. Como se muestra en la FIG. 15, la ascendencia de los individuos de la muestra puede predecirse basándose en la localización en la representación de la ascendencia. Se incluyeron catorce puntos centroides en la representación de la ascendencia (círculos). Basándose en el punto centroide más cercano, se 15 predijo la ascendencia de cada individuo de la muestra.

El experimento del perfilado de ADN también mostró un alto nivel de equilibrio intralocus tanto en los loci de STR como en los loci de SNP, como puede observarse en la FIG. 18, y un bajo nivel de tartamudeo, como puede 20 observarse en la FIG. 19.

Seis de los 12 individuos más 2800M tienen al menos un locus heterocigoto isométrico, que se muestra en la FIG. 20. Un locus heterocigoto isométrico se define como un STR que tiene el mismo número de repeticiones, dos secuencias diferentes, que están igualmente equilibradas. Usando la 25 información sobre las variantes en la STR D8S1179, el alelo 13 de la muestra

15 se asignó a la abuela, muestra 17 (FIG. 21). Se usó información variante similar en la STR D13S317 para asignar los alelos de la muestra 15. Sin embargo, en este caso no puede averiguarse el origen de ningún alelo (FIG. 22).

## 5 Ejemplo 5-perfilado de ADN para el uso en investigación, medicina forense o paternidad

Este ejemplo se basa en el flujo de trabajo descrito en ForenSeq™ DNA Signature Prep Guide (Illumina, San Diego, CA), cuyo contenido se incorpora por referencia en su totalidad en el presente documento.

10 En este ejemplo puede usarse ADN purificado o lisado en bruto. En el caso del ADN purificado, cada muestra de 1 ng se diluyó a 0,2 ng/μl con agua sin nucleasa. Para el lisado en bruto, cada muestra de 2 μl se diluyó con 3 μl de agua sin nucleasa. Se constituyó una mezcla maestra para ocho o más reacciones. Para cada reacción, se añadieron 5,4 μl de ForenSeq PCR1  
15 Reaction Mix, 0,4 μl de ForenSeq Enzyme Mix y 5,8 μl de mezcla de cebador de ADN (A o B) en un tubo de microcentrífuga de 1,5 ml. Se transfirieron 10 μl de Mezcla Maestra a cada pocillo de la placa de PCR y se añadió el ADN o lisado. La mezcla de reacción múltiplex se sometió a PCR siguiendo el protocolo:

20 98 grados centígrados durante 3 min.

8 ciclos de:

96 grados centígrados durante 45 s.

80 grados centígrados durante 30 s.

25 54 grados centígrados durante 2 min, con modo de aumento específico

68 grados centígrados durante 2 min, con modo de aumento específico

10 ciclos de:

96 grados centígrados durante 30 s.

5 68 grados centígrados durante 3 min, con modo de ascenso específico

68 grados centígrados durante 10 min.

Mantenido a 10 grados centígrados.

Después de los ciclos, las muestras se retiraron del termociclador. Se  
10 añadieron ForenSeq PCR2 Reaction Mix a las muestras con un par de adaptadores (adaptadores i7 e i5) y la mezcla se sometió a 15 ciclos siguiendo el protocolo:

98 grados centígrados durante 30 s.

15 ciclos de:

15 98 grados centígrados durante 20 s.

66 grados centígrados durante 30 s.

68 grados centígrados durante 90 s.

68 grados centígrados durante 10 min.

Mantenido a 10 grados centígrados.

20 Las muestras se purificaron con perlas de purificación de muestra, y las bibliotecas se normalizaron y agruparon en volumen. Las bibliotecas agrupadas se diluyeron en tampón de hibridación (HT1), al que se añadió Control de Secuenciación Humana (HSC) y se termodesnaturalizó en la preparación para la secuenciación.

25 Ejemplo 6-genotipado con ADN degradado

La FIG. 23 muestra resultados de genotipado usando ADN cortado y/o tratado con DNasa que representa ADN degradado. Como se muestra, más de 50 % de los loci de STR y SNP se validaron correctamente con ADN cortado. Se consiguió una probabilidad de coincidencia aleatoria (RMP) de  $10^{-19}$  con DNA de menos de 100 pb. También se predijo la ascendencia correcta usando ADN degradado.

#### Ejemplo 7-sensibilidad de genotipado

Las FIGS. 24 y 25 muestran resultados de sensibilidad de genotipado a niveles de entrada de ADN subnanogramicos de 7,82 pg a 1 ng. Como se muestra, el 100 % de los alelos se validaron satisfactoriamente a una entrada de ADN de 125 ng tanto para STR como para SNP. Más del 50 % de los alelos se validaron satisfactoriamente a una entrada de ADN tan baja como de 7,82 pg. El equilibrio intralocus fue mayor del 70 % para la mayoría de los loci a un ADN de entrada de 1 ng.

Todos los números que expresan cantidades de ingredientes, condiciones de reacción y similares usados en la memoria descriptiva deben entenderse como modificados en todos los casos por el término “aproximadamente”. Por consiguiente, a menos que se indique lo contrario, los parámetros numéricos establecidos en el presente documento son aproximaciones que pueden variar dependiendo de las propiedades deseadas que se pretenden obtener. Cuando menos, y sin intentar limitar la aplicación de la doctrina de equivalentes al alcance de ninguna reivindicación en ninguna solicitud que reivindique los derechos de prioridad de la presente solicitud, cada parámetro numérico debe interpretarse teniendo en cuenta el número de dígitos significativos y los enfoques de redondeo habituales.

Todas las referencias citadas en el presente documento incluyendo, pero sin limitación, las solicitudes publicadas y no publicadas, patentes y referencias de bibliografía, se incorporan en el presente documento por referencia en su totalidad y por la presente constituyen una parte de esta memoria descriptiva.

- 5 En caso de que las publicaciones y patentes o solicitudes de patente incorporadas por referencia contradigan la divulgación contenida en la memoria descriptiva, debe entenderse que la memoria descriptiva suplanta a y/o prevalece sobre cualquier material contradictorio.

- 10 La mención de las publicaciones o documentos anteriores no pretende ser una admisión de que cualquiera de ellos sea una técnica anterior pertinente, ni constituye ninguna admisión en cuanto al contenido o fecha de estas publicaciones o documentos.

- 15 Aunque la presente invención se ha descrito con detalle en relación con realizaciones de la misma con referencia a los dibujos que se adjuntan, cabe observar que varios cambios y modificaciones serán obvios para los expertos en la materia. Dichos cambios y modificaciones deben considerarse incluidos dentro del alcance de la presente invención. Debe entenderse que las diversas realizaciones de la invención se han presentado solamente a modo de ejemplo y no a modo de limitación. Del mismo modo, los diversos diagramas pueden
- 20 representar un ejemplo de configuración arquitectónica o de otro tipo para la invención, que se realiza para ayudar a entender las características y funcionalidad que pueden incluirse en la invención. La invención no se limita a las arquitecturas o configuraciones de ejemplo ilustradas, sino que puede implementarse usando una diversidad de arquitecturas y configuraciones
- 25 alternativas. Adicionalmente, aunque la invención se describe anteriormente en

relación con diversas realizaciones e implementaciones a modo de ejemplo, debe entenderse que las diversas características y funcionalidad descritas en una o más de las realizaciones individuales no están limitadas en su aplicabilidad a la realización particular con la cual se describen. En su lugar, pueden aplicarse, en solitario o en alguna combinación, a una o más de las diversas realizaciones de la invención, se describan o no dichas realizaciones, y se presenten o no dichas características como parte de una realización descrita. Por tanto, la amplitud y el alcance de la invención no deben limitarse por ninguna de las realizaciones a modo de ejemplo descritas anteriormente.

Los términos y frases que se usan en este documento, y sus realizaciones, salvo que se indique expresamente lo contrario, deben interpretarse en sentido amplio en lugar de en sentido limitante. Como ejemplos de lo anterior: el término “que incluye” debe interpretarse en el sentido de que “incluye sin limitación” o similares; el término “ejemplo” se utiliza para proporcionar casos a modo de ejemplo del tema en discusión, no como una lista exhaustiva o limitante de los mismos; y adjetivos tales como “convencional”, “tradicional”, “normal”, “estándar”, “conocido” y términos de significado similar, no deben interpretarse como una limitación del artículo descrito a un período de tiempo determinado, o a un artículo disponible a partir de un momento dado. En su lugar, debe entenderse que estos términos incluyen tecnologías convencionales, tradicionales, normales o estándar que pueden estar disponibles, pueden conocerse ahora, o pueden conocerse en cualquier momento en el futuro. Del mismo modo, no debe entenderse que un grupo de elementos unidos mediante la conjunción “y” requiera que todos y cada uno de esos elementos estén presentes en la agrupación, sino más bien

debe considerarse “y/o” a menos que sea evidente por el contexto o se indique expresamente lo contrario. Del mismo modo, no debe interpretarse que un grupo de elementos unidos por la conjunción “o” requiera la exclusividad mutua entre ese grupo, sino que también debe considerarse “y/o” a menos que sea

5 evidente por el contexto o se indique expresamente lo contrario. Además, aunque los artículos, elementos o componentes de la invención pueden describirse o reivindicarse en singular, se contempla que el plural está dentro del alcance de la misma a menos que se indique explícitamente la limitación al singular. Por ejemplo, “al menos uno” puede referirse a uno solo o a una

10 pluralidad y no está limitado a ninguno. La presencia, en algunos casos, de palabras y frases amplias tales como “uno o más”, “al menos”, “pero sin limitación”, u otras frases similares, no quiere decir que pretenda o requiera el caso más restrictivo cuando están ausentes dichas frases amplias.