(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 938 072**

(51) Int. CI.:

***C12Q 1/68*** (2008.01)
***G01N 33/53*** (2006.01)
***G01N 33/573*** (2006.01)
***G01N 33/569*** (2006.01)
***G01N 21/76*** (2006.01)
***C12N 15/10*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **09.05.2016 E 20216661 (7)**

(97) Fecha y número de publicación de la concesión europea: **30.11.2022 EP 3822365**

(54) Título: **Plataforma para el descubrimiento y análisis de agentes terapéuticos**

(30) Prioridad:

**11.05.2015 US 201562159710 P**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**04.04.2023**

(73) Titular/es:

**ILLUMINA, INC. (100.0%)**
**5200 Illumina Way**
**San Diego, CA 92122, US**

(72) Inventor/es:

**BOUTELL, JONATHAN MARK;**
**GOLYNSKIY, MISHA;**
**HE, MOLLY;**
**KELLINGER, MATTHEW WILLIAM;**
**PEISAJOVICH, SERGIO y**
**PREVITE, MICHAEL**

(74) Agente/Representante:

**ELZABURU, S.L.P**

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

## DESCRIPCIÓN

Plataforma para el descubrimiento y análisis de agentes terapéuticos

**Antecedentes**

Esta descripción se refiere generalmente al descubrimiento de agentes (p. ej., moléculas pequeñas, proteínas o células) que tienen usos médicos, agrícolas e industriales, y más específicamente a plataformas para seleccionar agentes candidatos para dichos usos.

Actualmente, el descubrimiento de fármacos por cribado de alto rendimiento utiliza múltiples etapas y plataformas para cribar grandes bibliotecas de moléculas pequeñas en busca de "pistas" o "aciertos" que muestren eficacia y potencia contra una diana particular, así como un perfil de toxicidad favorable contra un panel de dianas moleculares.

En la primera etapa del proceso, se criba una biblioteca de moléculas pequeñas con un alto rendimiento frente a una diana predeterminada. Estas bibliotecas suelen ser del orden de cientos de miles de compuestos. Las metodologías actuales de alto rendimiento se pueden automatizar. Como resultado, de la evolución de la robótica y la automatización, la instrumentación de alto costo puede cribar 100.000 moléculas/día utilizando volúmenes de reactivo de entrada en el rango de pocos microlitros. Los ensayos pueden ser homogéneos o heterogéneos, siendo los primeros relativamente simples y asequibles y los segundos más sensibles, aunque más complejos, lentos y costosos.

Si bien los ensayos iniciales de cribado de moléculas pequeñas son eficaces para identificar "aciertos" o "pistas", estos ensayos suelen ser solo una primera etapa. Típicamente, estas pistas o aciertos se criban adicionalmente con respecto a la genotoxicidad, toxicidad farmacológica, toxicidad celular y citotoxicidad para los organismos utilizando ensayos de perfilado contra dianas bien definidas para eliminar a los candidatos que puedan tener efectos clínicos adversos. Cada uno de estos ensayos se realiza de forma independiente utilizando una variedad de plataformas y ensayos en forma de kit que emplean flujos de trabajo independientes que se curan y se extraen datos para generar un consenso sobre la utilidad del acierto o la pista.

Los métodos de evolución de proteínas constituyen otro tipo de cribado. Estas metodologías han cubierto una amplia gama de permutaciones. Si bien el rendimiento, la velocidad y el bajo coste de estas metodologías han tendido a convertirlas en una opción líder para la evolución rápida de proteínas, los métodos a menudo son complejos y plantean desafíos. Por ejemplo, los cribados basados en emulsión plantean dificultades con la fusión/mezclado de gotitas (para el cribado multiplexado), dificultades con la ruptura de emulsiones (recuperación de componentes de interés), complicaciones que surgen de la variación de concentración debido a los tamaños variables de las gotitas y contaminación cruzada de gotitas.

Las terapias basadas en células tienen capacidades terapéuticas que son potencialmente ventajosas en comparación con las terapias basadas en moléculas pequeñas o basadas en proteínas. Entre las ventajas clave: las células pueden detectar señales externas, moverse a sitios específicos dentro del cuerpo e integrar múltiples estímulos y responder con comportamientos complejos (tales como la liberación de moléculas efectoras específicas). Alcanzar plenamente el potencial de las terapias basadas en células se beneficiará de la ingeniería precisa de las células terapéuticas para que sus "comportamientos" puedan controlarse tanto en el espacio como en el tiempo.

Los flujos de trabajo actuales en la ingeniería celular incluyen típicamente las siguientes etapas: (i) diseño de circuitos de señalización intracelular responsables de detectar, integrar y responder a los estímulos; (ii) ingeniería genética compleja para incorporar los genes responsables de mediar esas funciones en la célula; y (iii) métodos de cribado para identificar, entre todas las células de la biblioteca, aquellas que portan el complemento génico que realiza mejor la función deseada. Debido a la complejidad de los comportamientos celulares deseados, los métodos de cribado actuales no son ideales. Por ejemplo: los métodos basados en la clasificación de células activada por fluorescencia (FACS) tienen un alto rendimiento, pero solo observan instantáneas del comportamiento celular. La microscopía de fluorescencia, por otro lado, puede seguir la dinámica de los comportamientos celulares en detalle, pero tienen un bajo rendimiento.

El documento WO2015/031691A1 describe métodos, composiciones y kits para análisis de ácido nucleico multiplex de células individuales . Los métodos, composiciones y sistemas se describen para uso en la secuenciación masiva en paralelo de células individuales y para analizar miles de células simultáneamente. Los miles de células pueden comprender una población mixta de células (por ejemplo, células de diferentes tipos o subtipos, diferentes tamaños). El documento EP2805769A1 describe métodos para generar microcompartimentos que comprenden células individuales, identificadores que permiten rastrear el contenido de cada microcompartimento y reactivos para llevar a cabo un análisis de la célula o de sus componentes. También se refiere a métodos de alto rendimiento para analizar una pluralidad de células individuales en paralelo, incluyendo métodos para genómica de células individuales, transcriptómica o análisis de metilación.

Por tanto, existe la necesidad de plataformas y métodos para cribar moléculas pequeñas, proteínas, células y otros agentes en busca de propiedades beneficiosas. La presente descripción aborda esta necesidad y proporciona también otras ventajas.

**Breve resumen**

La presente descripción proporciona un método para cribar células, que comprende (a) proporcionar una pluralidad de células diferentes, en donde cada una de las células diferentes comprende una etiqueta de ácido nucleico que tiene una secuencia de etiqueta, estando unida la secuencia de etiqueta de ácido nucleico a cada célula de la pluralidad de células mediante unión covalente a un lípido de la membrana plasmática, un ácido graso de la membrana plasmática o una proteína en un lípido de la membrana plasmática de la célula, en donde cada secuencia de etiqueta comprende una secuencia única; (b) poner en contacto una mezcla de las células diferentes con un soporte sólido para formar una matriz de células unidas al soporte sólido; (c) cribar la matriz de células en el soporte sólido con relación a al menos una característica óptica, en donde la reacción de cribado comprende la detección de las células individuales que están unidas al soporte sólido; (d) secuenciar las secuencias de etiqueta de ácido nucleico que están unidas al soporte sólido; y (e) identificar al menos una célula en la matriz como una célula candidata basada en la característica óptica y la secuencia de etiqueta de la célula candidata.

También se describe en el presente documento un método para caracterizar agentes candidatos. El método puede incluir las etapas de (a) proporcionar una biblioteca de agentes candidatos, en la que cada agente candidato está unido a una etiqueta de ácido nucleico que tiene una secuencia de etiqueta; (b) poner en contacto la biblioteca de agentes candidatos con un soporte sólido para unir los agentes candidatos al soporte sólido, mediante lo cual se forma una matriz de agentes candidatos que incluye características individuales en el soporte sólido cada una correspondiente a un agente candidato individual de la biblioteca; (c) poner en contacto la matriz de agentes candidatos con un agente de cribado, en el que uno o más agentes candidatos de la matriz reaccionan con el agente de cribado; (d) detectar la matriz durante o después de poner en contacto la matriz con el agente de cribado, determinando de ese modo que al menos un agente candidato en la matriz reacciona con el agente de cribado; (e) secuenciar las etiquetas de ácido nucleico en la matriz para determinar la secuencia de las etiquetas que están unidas a cada uno de los agentes candidatos; e (f) identificar el al menos un agente candidato en la matriz que reacciona con el agente de cribado en base a la secuencia de la etiqueta que está unida al al menos un agente candidato.

También se describe en la presente memoria un método para producir una matriz de proteínas. El método puede incluir las etapas de (a) proporcionar una biblioteca de moléculas de ADNc que están unidas a un soporte sólido; (b) amplificar las moléculas de ADNc en el soporte sólido para formar grupos, en el que cada grupo incluye múltiples copias de una molécula de ADNc particular de la biblioteca; (c) transcribir las múltiples copias en los grupos para producir múltiples moléculas de ARNm unidas a cada uno de los grupos; y (d) traducir las moléculas de ARNm en los grupos para producir múltiples proteínas unidas a cada uno de los grupos.

También se describe en la presente memoria un método para producir una matriz de proteínas que incluye las etapas de (a) proporcionar una biblioteca de moléculas de ARNm, en la que las moléculas de ARNm individuales en la biblioteca incluyen una secuencia diana y una secuencia de etiqueta, (b) derivar una primera sub-biblioteca de la biblioteca, la primera sub-biblioteca que incluye ácidos nucleicos que tienen las secuencias de etiqueta o complementos de las mismas, en la que los ácidos nucleicos están unidos a características individuales en un soporte sólido, (c) derivar una segunda sub-biblioteca de la biblioteca, la segunda sub-biblioteca que incluye ácidos nucleicos que tienen las secuencias diana y las secuencias de etiqueta o complementos de las mismas; (d) poner en contacto la segunda sub-biblioteca con la primera sub-biblioteca, uniendo así los ácidos nucleicos de la segunda sub-biblioteca al soporte sólido mediante la hibridación de las secuencias de etiqueta y los complementos de las mismas; y (e) traducir las secuencias diana en el soporte sólido para producir una matriz de proteínas unidas a las características individuales.

También se describe la presente memoria una matriz que incluye (a) un soporte sólido; (b) una biblioteca de diferentes moléculas de ADNc unidas al soporte sólido, en la que cada molécula de ADNc diferente está unida a una característica individual en el soporte sólido, y en la que cada característica incluye múltiples copias de una molécula de ADNc particular (c) moléculas de ARNm unidas a las moléculas de ADNc, en las que cada una de las moléculas de ADNc es complementaria a la respectiva molécula de ARNm unida; y (d) moléculas de proteínas unidas a las moléculas de ARNm, en las que cada una de las moléculas de proteína está codificada por la respectiva molécula de ARNm unida.

También se describe la presente memoria una matriz que incluye (a) una biblioteca de moléculas de ARNm, en la que las moléculas de ARNm individuales en la biblioteca comprenden una secuencia diana y una secuencia de etiqueta, (b) un soporte sólido que comprende ácidos nucleicos que tienen complementos de las secuencias de etiqueta, en las que los ácidos nucleicos están unidos a características individuales en un soporte sólido, en el que las secuencias de etiqueta de las moléculas de ARNm individuales se hibridan con las respectivas secuencias de etiqueta complementarias en las características individuales del soporte sólido, y en el que se unen proteínas derivadas de la traducción de las moléculas de ARNm a las respectivas moléculas de ARNm.

**Breve descripción de los dibujos**

La Fig. 1A a la Fig. 1F muestran las etapas de un proceso para el cribado de alto rendimiento de agentes candidatos.

La Fig. 2 muestra un proceso para la síntesis y el cribado de alto rendimiento de agentes proteicos candidatos.

La Fig. 3A y la Fig. 3B muestran un proceso para etiquetado, síntesis y cribado de alto rendimiento de agentes proteicos candidatos.

La Fig. 4 muestra la exposición de un agente de cribado a células que están unidas a una celda de flujo y gráficas de fluorescencia frente al tiempo esperado de diversas respuestas celulares al agente de cribado.

La Fig. 5 muestra células clasificadas en micropocillos individuales y luego etiquetadas con etiquetas específicas de micropocillos.

La Fig. 6 muestra la decodificación de ubicaciones de células en un soporte sólido utilizando perlas codificadas.

La Fig. 7 muestra la decodificación de ubicaciones de células en un soporte sólido usando etiquetas de ácido nucleico.

La Fig. 8 muestra la transferencia de etiquetas de ácidos nucleicos desde una célula a un sitio en un soporte sólido y la secuenciación de la etiqueta de ácido nucleico en el soporte sólido.

La Fig. 9 muestra células etiquetadas por la unión de un par de ácidos nucleicos a ácidos grasos en la membrana celular.

La Fig. 10 muestra la captura de una célula etiquetada con ácido nucleico en la superficie de una celda de flujo.

La Fig. 11 muestra el desprendimiento de células de la superficie de una celda de flujo mediante la escisión por endonucleasa de restricción de una etiqueta de ácido nucleico.

La Fig. 12 muestra la identificación de células en la superficie de una celda de flujo basada en la ubicación de las etiquetas decodificadas en los recipientes de los que se obtuvieron las células.

**Descripción detallada**

En el presente documento se describen aparatos y métodos para el cribado de alto rendimiento de agentes candidatos, tales como candidatos para uso terapéutico. Las realizaciones particulares expuestas en la presente memoria usan de forma beneficiosa las técnicas y aparatos de secuenciación de ácidos nucleicos. Una ventaja de usar técnicas y aparatos de secuenciación como se muestra en la presente memoria es la capacidad de ordenar espacialmente un gran número de agentes candidatos diferentes de manera que los ácidos nucleicos que están asociados con cada agente candidato puedan distinguirse y de tal manera que la respuesta de cada agente candidato a un estímulo particular se puede detectar individualmente en el soporte sólido.

En realizaciones particulares, se proporciona una pluralidad de agentes candidatos diferentes en una biblioteca, en la que cada agente candidato está unido (o asociado de otro modo) con una etiqueta de ácido nucleico única. Los agentes candidatos son células. Estos agentes candidatos etiquetados se pueden unir a un soporte sólido de modo que los miembros individuales formen una matriz de agentes candidatos separados espacialmente. Luego, la matriz de agentes candidatos puede cribarse mediante exposición a un agente de cribado (u otro estímulo) y las reacciones del agente candidato pueden detectarse en ubicaciones resueltas espacialmente en el soporte sólido. La ubicación de uno o más "aciertos" se puede identificar basándose en la detección resuelta espacialmente de señales esperadas, deseadas o únicas en la matriz. También se puede llevar a cabo una reacción de secuenciación en la matriz (ya sea antes o después de una etapa de cribado) con el fin de identificar cada etiqueta con respecto a su ubicación en la matriz. Un agente candidato puede identificarse correlacionando la ubicación del "acierto" con la identidad de la etiqueta en esa ubicación.

Los agentes candidatos basados en células contienen ácidos nucleicos que pueden secuenciarse para identificar células individuales en una población. Además, las secuencias de ácido nucleico pueden evaluarse para determinar características útiles de células individuales. De nuevo, el uso de etiquetas asignadas *a priori* puede permitir distinguir las características de las células sin la necesidad de secuenciar otros contenidos de la célula.

Aunque los métodos y aparatos de esta descripción se ejemplifican en la presente memoria con respecto al cribado de agentes candidatos para la función terapéutica, se entenderá que se pueden cribar otras características funcionales o estructurales. Por ejemplo, los métodos y aparatos se pueden utilizar para cribar toxicidad, uso agrícola (p. ej., pesticidas, factores de crecimiento, hormonas, etc.), uso industrial (p. ej., catalizadores, tintes, plásticos, etc.), nutrición (p. ej., sabores, conservantes, etc.), limpieza medioambiental o similares. Generalmente, los métodos y aparatos pueden usarse para cribar funciones biológicas o no biológicas.

Los métodos y aparatos mostrados en la presente memoria proporcionan una ventaja en la capacidad de realizar mediciones cuantitativas durante el cribado de alto rendimiento. Por ejemplo, las plataformas de secuenciación disponibles comercialmente, tales como las comercializadas por Illumina, Inc. (San Diego, CA) incluyen ópticas de precisión que tienen un rango dinámico relativamente amplio para cuantificar señales fluorescentes. Una ventaja adicional es la capacidad de seguir la dinámica temporal de una reacción de cribado en múltiples ubicaciones en una matriz de agentes candidatos. Por el contrario, los métodos de clasificación fluídica que se utilizan en muchos cribados tradicionales proporcionan solo una instantánea de un agente candidato que pasa, lo que limita la medición a un solo punto de tiempo. Los métodos y aparatos mostrados en la presente memoria proporcionan un alto rendimiento a la par, o en algunos casos mejor, que las técnicas de clasificación fluídica, pero con el beneficio añadido de las mediciones basadas en el tiempo de los resultados del cribado. Una ventaja adicional del aparato y los métodos

mostrados en la presente memoria es la capacidad de detectar resultados de cribado y detectar etiquetas de una manera que permite la correlación espacial de las etiquetas con los resultados de cribado con el fin de identificar "aciertos" del cribado.

Se entenderá que los términos utilizados en la presente memoria adquieren su significado habitual en la técnica relevante, a menos que se especifique lo contrario. A continuación, se muestran varios términos utilizados en la presente memoria y sus significados.

Tal y como se usa en la presente memoria, el término "amplicón", cuando se usa en referencia a un ácido nucleico, significa el producto de copiar el ácido nucleico, en el que el producto tiene una secuencia de nucleótidos que es igual o complementaria a al menos una porción de la secuencia de nucleótidos del ácido nucleico. Un amplicón se puede producir mediante cualquiera de una variedad de métodos de amplificación que usan el ácido nucleico, o un amplicón del mismo, como molde, incluyendo, por ejemplo, extensión por polimerasa, reacción en cadena de la polimerasa (PCR), amplificación por círculo rodante (RCA), amplificación por desplazamiento múltiple (MDA), extensión por ligación o reacción en cadena de ligación. Un amplicón puede ser una molécula de ácido nucleico que tiene una única copia de una secuencia de nucleótidos particular (p. ej., un producto de PCR) o múltiples copias de la secuencia de nucleótidos (p. ej., un producto concatemérico de RCA). Un primer amplicón de un ácido nucleico diana es típicamente una copia complementaria. Los amplicones posteriores son copias que se crean, después de la generación del primer amplicón, a partir del ácido nucleico diana o de un amplicón. Un amplicón posterior puede tener una secuencia que es sustancialmente complementaria al ácido nucleico diana o sustancialmente idéntica al ácido nucleico diana.

Tal y como se usa en la presente memoria, el término "matriz" se refiere a una población de características o sitios que se pueden diferenciar entre sí según la ubicación relativa. Las diferentes moléculas u otros ítems que se encuentran en diferentes sitios de una matriz se pueden diferenciar entre sí según las ubicaciones de los sitios en la matriz. Un sitio individual de una matriz puede incluir una o más moléculas (u otros ítems) de un tipo particular. Por ejemplo, un sitio puede incluir una única molécula de ácido nucleico que tiene una secuencia particular o un sitio puede incluir varias moléculas de ácido nucleico que tienen la misma secuencia (y/o secuencia complementaria, de la misma). Los sitios de una matriz pueden ser características diferentes ubicadas en el mismo sustrato. Las características ejemplares incluyen, sin limitación, pocillos en un sustrato, perlas (u otras partículas) en o sobre un sustrato, proyecciones de un sustrato, crestas en un sustrato o canales en un sustrato. Los sitios de una matriz pueden ser sustratos separados, cada uno con una molécula diferente (u otro ítem). Se pueden identificar diferentes ítems unidos a sustratos separados según las ubicaciones de los sustratos sobre un soporte sólido al que están asociados los sustratos o según las ubicaciones de los sustratos en un líquido o gel. Las matrices ejemplares en las que se encuentran sustratos separados sobre un soporte sólido incluyen, sin limitación, aquellas que tienen perlas en los pocillos.

Tal y como se usa en la presente memoria, el término "adjunto" se refiere al estado de dos cosas que están juntas, sujetas, adheridas, conectadas o unidas entre sí. Por ejemplo, un analito, tal como un ácido nucleico, puede unirse a un material, tal como un gel o un soporte sólido, mediante un enlace covalente o no covalente. Un enlace covalente se caracteriza por compartir pares de electrones entre átomos. Un enlace no covalente es un enlace químico que no implica compartir pares de electrones y puede incluir, por ejemplo, enlaces de hidrógeno, enlaces iónicos, fuerzas de van der Waals, interacciones hidrófilas e interacciones hidrófobas. En algunas realizaciones, la unión puede producirse mediante un agente de captura. Un agente de captura puede incluir, por ejemplo, un anticuerpo, receptor, ácido nucleico, ligando, lectina, carbohidrato, avidina, biotina o análogos de los mismos.

Tal y como se usa en la presente memoria, el término "agente candidato" pretende significar un ítem que se sospecha que tiene una estructura o función particular. Los ítems ejemplares incluyen, pero no están limitados a, moléculas, células y componentes subcelulares. Las moléculas pueden ser opcionalmente moléculas biológicamente activas tales como proteínas, aminoácidos, ácidos nucleicos (p. ej., ADN o ARN), nucleótidos, polisacáridos, sacáridos, metabolitos, vitaminas, cofactores enzimáticos o similares. Otros agentes candidatos incluyen macrociclos, péptidos cíclicos, moléculas fusionadas (p. ej., fusión de ácido nucleico-proteína) o construcciones presentadas (p. ej., péptidos en fagos). Las funciones ejemplares que se puede sospechar que tiene un agente candidato incluyen, pero no están limitadas a, activación de otro agente, inhibición de otro agente, modificación química de otro agente, degradación de otro agente, síntesis de otro agente, en donde el otro agente puede ser opcionalmente uno cualquiera o más de los ítems ejemplificados anteriormente como agentes candidatos. La estructura de un agente candidato puede ser cualquier estructura conocida o sospechada para los ítems anteriores u otros ítems conocidos en la técnica.

Tal y como se usa en la presente memoria, el término "diferente", cuando se usa en referencia a ácidos nucleicos, significa que los ácidos nucleicos tienen secuencias de nucleótidos que no son iguales entre sí. Dos o más ácidos nucleicos pueden tener secuencias de nucleótidos que son diferentes en toda su longitud. Alternativamente, dos o más ácidos nucleicos pueden tener secuencias de nucleótidos que son diferentes a lo largo de una parte sustancial de su longitud. Por ejemplo, dos o más ácidos nucleicos pueden tener porciones de secuencias de nucleótidos diana que son diferentes para las dos o más moléculas mientras que también tienen una porción de secuencia universal que es la misma en las dos o más moléculas. El término se puede aplicar de manera similar a proteínas que se pueden distinguir como diferentes entre sí basándose en diferencias en la secuencia de aminoácidos.

Tal y como se usa en la presente memoria, el término "cada uno", cuando se usa en referencia a una colección de ítems, pretende identificar un ítem individual en la colección, pero no se refiere necesariamente a todos los ítems de

la colección. Pueden ocurrir excepciones si la descripción explícita o el contexto dicta claramente lo contrario.

Tal y como se usa en la presente memoria, el término "extender", cuando se usa en referencia a un ácido nucleico, pretende significar la adición de al menos un nucleótido u oligonucleótido al ácido nucleico. En realizaciones particulares, pueden añadirse uno o más nucleótidos al extremo 3' de un ácido nucleico, por ejemplo, mediante catálisis por polimerasa (p. ej., ADN polimerasa, ARN polimerasa o transcriptasa inversa). Se pueden utilizar métodos químicos o enzimáticos para añadir uno o más nucleótidos al extremo 3' o 5' de un ácido nucleico. Pueden añadirse uno o más oligonucleótidos al extremo 3' o 5' de un ácido nucleico, por ejemplo, mediante métodos químicos o enzimáticos (p. ej., catálisis por ligasa). Un ácido nucleico se puede extender de una manera dirigida por molde, por lo que el producto de la extensión es complementario a un ácido nucleico molde que se hibrida con el ácido nucleico que se extiende.

Tal y como se usa en la presente memoria, el término "característica" significa una ubicación en una matriz para una especie particular de molécula o célula. Una característica puede contener solo una molécula (o célula) o puede contener una población de varias moléculas (o células) de la misma especie. En algunas realizaciones, las características están presentes en un soporte sólido antes de unir una molécula o célula. En otras realizaciones, la característica se crea mediante la unión de una molécula o célula al soporte sólido. Las características de una matriz son típicamente discretas. Las características discretas pueden ser contiguas o pueden tener espacios entre sí. El tamaño de las características y/o el espaciado entre las características puede variar de manera que las matrices pueden ser de densidad alta, densidad media o densidad más baja. Las matrices de densidad alta se caracterizan por tener sitios separados por menos de aproximadamente 15 µm. Las matrices de densidad media tienen sitios separados por aproximadamente 15 a 30 µm, mientras que las matrices de densidad baja tienen sitios separados por más de 30 µm. Una matriz útil en la presente memoria puede tener, por ejemplo, sitios que están separados por menos de 100 µm, 50 µm, 10 µm, 5 µm, 1 µm o 0,5 µm. Puede usarse un aparato o método de la presente descripción para detectar una matriz a una resolución suficiente como para distinguir sitios en las densidades o intervalos de densidad anteriores.

Tal y como se usa en la presente memoria, el término "celda de flujo" significa un recipiente que tiene una cámara donde se puede llevar a cabo una reacción, una entrada para suministrar reactivos a la cámara y una salida para eliminar reactivos de la cámara. En algunas realizaciones, la cámara está configurada para la detección de la reacción que ocurre en la cámara. Por ejemplo, la cámara puede incluir una o más superficies transparentes que permitan la detección óptica de matrices, moléculas marcadas ópticamente o similares en la cámara. Las celdas de flujo ejemplares incluyen, pero no están limitadas a, las utilizadas en un aparato de secuenciación de ácidos nucleicos tales como las celdas de flujo para las plataformas Genome Analyzer®, MiSeq®, NextSeq® o HiSeq® comercializadas por Illumina, Inc. (San Diego, CA); o para la plataforma de secuenciación SOLiD™ o Ion Torrent™ comercializada por Life Technologies (Carlsbad, CA). También se describen celdas de flujo ejemplares y métodos para su fabricación y uso, por ejemplo, en WO 2014/142841 A1; Pub. Solic. Pat. de EE. UU. No. 2010/0111768 A1 y Pat. de EE. UU. No. 8.951.781.

Tal y como se usa en la presente memoria, el término "biblioteca" pretende significar una colección que incluye varios ítems diferentes. Los ítems de una colección pueden diferir en estructura y/o función. Por ejemplo, la colección puede incluir ácidos nucleicos que tienen diferentes secuencias de nucleótidos, o la colección puede incluir proteínas que tienen diferente estructura primaria (es decir, secuencia de aminoácidos), estructura secundaria, estructura terciaria o estructura cuaternaria. Sin embargo, se entenderá que puede haber cierta redundancia de ítems en una biblioteca. Por ejemplo, pueden estar presentes múltiples copias de un ácido nucleico o proteína particular en una biblioteca que, no obstante, incluye una gran variedad de diferentes ácidos nucleicos o proteínas. Los tipos ejemplares de ítems que pueden estar en una biblioteca incluyen los mostrados en la presente memoria con respecto a agentes candidatos o agentes de cribado.

Tal y como se usa en la presente memoria, el término "luminiscente" significa emitir radiación de cuerpo frío. Se pretende que el término sea distinto de incandescencia, que es la radiación emitida por un material como resultado del calor. Generalmente, la luminiscencia se produce cuando una fuente de energía desplaza un electrón de un átomo fuera de su estado fundamental de energía más baja a un estado excitado de energía más alta; luego, el electrón devuelve la energía en forma de radiación para que pueda volver a su estado fundamental. Un tipo de ítem luminiscente particularmente útil es el que emite radiación de cuerpo frío cuando la radiación de excitación proporciona energía. Dichos ítems se denominan "fluorescentes" o "fotoluminiscentes". La fluorescencia o fotoluminiscencia se puede percibir como la emisión de radiación de un ítem a una longitud de onda que es el resultado de irradiar el ítem a otra longitud de onda.

Tal y como se usa en la presente memoria, los términos "ácido nucleico" y "nucleótido" pretenden ser consistentes con su uso en la técnica e incluir especies naturales o análogos funcionales de las mismas. Los análogos funcionales de ácidos nucleicos particularmente útiles son capaces de hibridar con un ácido nucleico de una manera específica de secuencia o pueden usarse como molde para la replicación de una secuencia de nucleótidos particular. Los ácidos nucleicos de origen natural generalmente tienen un esqueleto que contiene enlaces fosfodiéster. Una estructura análoga puede tener un enlace en el esqueleto alternativo que incluye cualquiera de una variedad de los conocidos en la técnica. Los ácidos nucleicos de origen natural generalmente tienen un azúcar desoxirribosa (p. ej., que se encuentra en el ácido desoxirribonucleico (ADN)) o un azúcar ribosa (p. ej., que se encuentra en el ácido ribonucleico (ARN)). Un ácido nucleico puede contener nucleótidos que tengan cualquiera de una variedad de análogos de estos restos de azúcar que se conocen en la técnica. Un ácido nucleico puede incluir nucleótidos nativos o no nativos. A este respecto, un ácido desoxirribonucleico nativo puede tener una o más bases seleccionadas del grupo que consiste

en adenina, timina, citosina o guanina y un ácido ribonucleico puede tener una o más bases seleccionadas del grupo que consiste en uracilo, adenina, citosina o guanina. Se conocen en la técnica bases no nativas útiles que pueden incluirse en un ácido nucleico o nucleótido.

Tal y como se usa en la presente memoria, el término "paso", cuando se usa en referencia a características de una matriz, pretende referirse al espacio de centro a centro para características adyacentes. Un patrón de características se puede caracterizar en términos de paso promedio. El patrón se puede ordenar de manera que el coeficiente de variación alrededor del paso medio sea pequeño o el patrón puede ser aleatorio, en cuyo caso el coeficiente de variación puede ser relativamente grande. En cualquier caso, el paso promedio puede ser, por ejemplo, al menos aproximadamente 10 nm, 0,1 µm, 0,5 µm, 1 µm, 5 µm, 10 µm, 100 µm o más. Alternativa o adicionalmente, el paso promedio puede ser, por ejemplo, como máximo aproximadamente 100 µm, 10 µm, 5 µm, 1 µm, 0,5 µm 0,1 µm o menos. Por supuesto, el paso promedio para un patrón particular de características puede estar entre uno de los valores más bajos y uno de los valores más altos seleccionados de los rangos anteriores.

Tal y como se usa en la presente memoria, los términos "proteína" o "aminoácido" pretenden ser consistentes con su uso en la técnica e incluir especies de origen natural o análogos funcionales de las mismas. Las proteínas de origen natural generalmente tienen un esqueleto que contiene enlaces peptídicos. Una estructura análoga puede tener un enlace en el esqueleto alternativo que incluye cualquiera de los conocidos en la técnica. Las proteínas de origen natural generalmente tienen aminoácidos nativos seleccionados de arginina, histidina, lisina, ácido aspártico, ácido glutámico, serina, treonina, asparagina, glutamina, cisteína, glicina, prolina, alanina, valina, isoleucina, leucina, metionina, fenilalanina, tirosina, triptófano. y modificaciones naturales de los mismos. Algunas modificaciones de origen natural incluyen fosforilación (p. ej., de serina, treonina, histidina, ácido aspártico y ácido glutámico), prenilación, isoprenilación, acilación, alquilación, glicosilación, biotinilación, ubiquitinación o similares. Una proteína puede incluir aminoácidos nativos que tengan restos no naturales en los mismos. Una proteína puede incluir aminoácidos nativos o no nativos.

Tal y como se usa en la presente memoria, el término "reaccionar", cuando se usa en referencia a un primer agente y un segundo agente, pretende referirse al acto de modificar la estructura química de uno o ambos agentes, crear uno o más enlaces covalentes entre los dos agentes, permitir que uno de los reactivos catalice una modificación de la estructura química del otro agente, o unir específicamente los dos agentes (p. ej., mediante interacciones no covalentes). Las reacciones ejemplares incluyen, pero no están limitadas a, reacciones químicas tales como reducción, oxidación, adición, eliminación, transposición, esterificación, amidación, eterificación, ciclación o sustitución; interacciones de unión en las que un primer agente se une a un segundo agente con afinidad específica; reacciones de disociación en las que dos o más agentes se separan entre sí; fluorescencia; luminiscencia; quimioluminiscencia; y reacciones biológicas, tales como replicación de ácido nucleico, amplificación de ácido nucleico, hibridación de ácido nucleico, ligación de ácido nucleico, fosforilación, catálisis enzimática, unión de receptor o unión de ligando.

Tal y como se usa en la presente memoria, el término "recombinante" pretende referirse a una construcción genética que no se produce de forma natural. Un ejemplo es el producto de combinar material genético de más de un origen. Las moléculas recombinantes ejemplares incluyen, pero no están limitadas a, ADN, ARN y proteína. Los orígenes de los que puede derivarse una molécula recombinante incluyen, por ejemplo, elementos genéticos similares de diferentes organismos, elementos genéticos diferentes del mismo organismo, elementos genéticos diferentes de diferentes organismos, elementos genéticos sintéticos o combinaciones de elementos genéticos sintéticos y naturales.

Tal y como se usa en la presente memoria, el término "agente de cribado" pretende significar un ítem que tiene una estructura o función que es selectiva para un primer agente candidato en comparación con un segundo agente candidato. En muchas realizaciones, la estructura o función del agente de cribado se conoce antes de su uso en un método o composición de la presente descripción. Los ítems ejemplares incluyen, pero no están limitados a, moléculas, células y componentes subcelulares. Las moléculas pueden ser opcionalmente moléculas biológicamente activas tales como proteínas, aminoácidos, ácidos nucleicos (p. ej., ADN o ARN), nucleótidos, polisacáridos, sacáridos, metabolitos, vitaminas, cofactores enzimáticos o similares. Las funciones selectivas ejemplares incluyen, pero no están limitadas a, activación de otro agente, inhibición de otro agente, modificación química de otro agente, degradación de otro agente, síntesis de otro agente; en donde el otro agente puede ser uno cualquiera o más de los ítems ejemplificados anteriormente como agentes candidatos. La estructura de un agente de cribado puede ser cualquier estructura conocida o sospechada para los ítems anteriores u otros ítems conocidos en la técnica.

Tal y como se usa en la presente memoria, la referencia a manipular "selectivamente" (o manipulación "selectiva" de) una primera cosa en comparación con la segunda cosa tiene la intención de significar que la manipulación tiene un mayor efecto sobre la primera cosa en comparación con el efecto sobre la segunda cosa. La manipulación no necesita tener ningún efecto sobre la segunda cosa. La manipulación puede tener un efecto en la primera cosa que sea al menos un 1 %, 10 %, 50 %, 90 % o 99 % mayor que el efecto en la segunda cosa. La manipulación puede tener un efecto en la primera cosa que sea al menos 2 veces, 5 veces, 10 veces, 100 veces, $1x10^3$ veces, $1x10^4$ veces o $1x10^6$ veces más alto que el efecto en la segunda cosa. La manipulación puede incluir, por ejemplo, modificar, poner en contacto, tratar, cambiar, escindir (p. ej., enlace químico), escindir fotoquímicamente (p. ej., un enlace químico), formar (p. ej., un enlace químico), formar fotoquímicamente (p. ej., un enlace químico), modificar covalentemente, modificar no covalentemente, destruir, fotoablación, eliminar, sintetizar, polimerizar, fotopolimerizar, amplificar (p. ej., un ácido nucleico), copiar (p. ej., un ácido nucleico), extender (p. ej., un ácido nucleico), ligar (p. ej., un ácido nucleico) u otra manipulación mostrada en la presente memoria o conocida de otro modo en la técnica.

Tal y como se usa en la presente memoria, el término "molécula pequeña" pretende significar un compuesto que tiene un peso molecular que es menor de aproximadamente 1.000 Dalton. En realizaciones particulares, una molécula pequeña no es polimérica. Sin embargo, una molécula pequeña puede, en otras realizaciones, ser un dímero o un trímero. También se entenderá que una molécula pequeña puede ser un monómero que puede incorporarse en un polímero. Las moléculas pequeñas particularmente útiles son compuestos orgánicos. Las moléculas pequeñas útiles pueden tener un peso molecular inferior a 900, 800, 600, 400, 200 o 100 Dalton.

Tal y como se usa en la presente memoria, el término "soporte sólido" se refiere a un sustrato rígido que es insoluble en un líquido acuoso. El sustrato puede ser no poroso o poroso. El sustrato puede opcionalmente ser capaz de captar un líquido (p. ej., debido a la porosidad), pero típicamente será lo suficientemente rígido como para que el sustrato no se hinche sustancialmente al captar el líquido y no se contraiga sustancialmente cuando el líquido se elimina por secado. Un soporte sólido no poroso es generalmente impermeable a líquidos o gases. Los soportes sólidos ejemplares incluyen, pero no están limitados a, vidrio y vidrio modificado o funcionalizado, plásticos (p. ej., acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™, olefinas cíclicas, poliimidas, etc.), nailon, cerámica, resinas, Zeonor, sílice o materiales a base de sílice que incluyen silicio y silicio modificado, carbono, metales, vidrios inorgánicos, haces de fibras ópticas y polímeros. Los soportes sólidos particularmente útiles para algunas realizaciones son componentes de una celda de flujo o están ubicados dentro de un aparato de celda de flujo. En la presente memoria se muestran con más detalle celdas de flujo ejemplares.

Tal y como se usa en la presente memoria, el término "sub-biblioteca" pretende significar una colección que tiene una representación que incluye ítems de una biblioteca o copias de ítems de la biblioteca. La representación en la sub-biblioteca puede ser completa o parcial en comparación con la representación en la biblioteca. Se puede obtener una sub-biblioteca separando al menos algunos de los ítems de una biblioteca o haciendo copias de al menos algunos de los ítems de una biblioteca.

Tal y como se usa en la presente memoria, el término "secuencia de etiqueta" pretende significar una serie de nucleótidos en un ácido nucleico que se puede usar para identificar o caracterizar un agente unido a (o asociado con) el ácido nucleico. La secuencia de etiqueta puede ser una secuencia de origen natural o una secuencia que no se produce de forma natural en el organismo del que se obtuvo el ácido nucleico. En realizaciones particulares, una o más secuencias de etiqueta que se usan con una muestra biológica no están presentes de forma natural en el genoma, transcriptoma u otros ácidos nucleicos de la muestra biológica. Por ejemplo, las secuencias de etiqueta pueden tener menos del 80 %, 70 %, 60 %, 50 % o 40 % de identidad de secuencia con las secuencias de ácido nucleico en una muestra biológica particular.

Tal y como se usa en la presente memoria, el término "secuencia universal" se refiere a una serie de nucleótidos que es común a dos o más moléculas de ácido nucleico incluso si las moléculas también tienen regiones de secuencia que difieren entre sí. Una secuencia universal que está presente en diferentes miembros de una colección de moléculas puede permitir la captura de múltiples ácidos nucleicos diferentes utilizando una población de ácidos nucleicos de captura universales que son complementarios a la secuencia universal. De manera similar, una secuencia universal presente en diferentes miembros de una colección de moléculas puede permitir la replicación o amplificación de múltiples ácidos nucleicos diferentes utilizando una población de cebadores universales que son complementarios a la secuencia universal. Por tanto, un ácido nucleico de captura universal o un cebador universal incluye una secuencia que puede hibridar específicamente con una secuencia universal.

Las realizaciones mostradas a continuación y enumeradas en las reivindicaciones pueden entenderse a la vista de las definiciones anteriores.

La presente descripción proporciona un método para cribar células, que comprende:

(a) proporcionar una pluralidad de células diferentes, en donde cada una de las células diferentes comprende una etiqueta de ácido nucleico que tiene una secuencia de etiqueta, estando unida la etiqueta de ácido nucleico a cada célula de la pluralidad de células mediante unión covalente a un lípido de la membrana plasmática, un ácido graso de la membrana plasmática, o una proteína en un lípido de la membrana plasmática de la célula, en donde cada secuencia de etiqueta comprende una secuencia única;

(b) poner en contacto una mezcla de las células diferentes con un soporte sólido para formar una matriz de células unida al soporte sólido;

(c) cribar la matriz de células en el soporte sólido con relación a al menos una característica óptica, en donde la reacción de cribado comprende detectar las células individuales que están unidas al soporte sólido;

(d) secuenciar las secuencias de etiqueta de etiquetas de ácido nucleico que están unidas al soporte sólido; y

(e) identificar al menos una célula en la matriz como una célula candidata basada en la característica óptica y la secuencia de etiqueta de la célula candidata. Las células para uso en la invención pueden incluir, por ejemplo, una célula de un organismo que incluye, pero no está limitada a, un mamífero tal como un roedor, ratón, rata, conejo, cobaya, ungulado, caballo, oveja, cerdo, cabra, vaca, gato, perro, primate (es decir, primate humano o no humano); una planta tal como *Arabidopsis thaliana,* maíz, sorgo, avena, trigo, arroz, canola o soja; un alga tal

como *Chlamydomonas reinhardtii;* un nematodo tal como *Caenorhabditis elegans;* un insecto tal como *Drosophila melanogaster,* mosquito, mosca de la fruta, abeja melífera o araña; un pez tal como el pez cebra; un reptil; un anfibio como una rana o *Xenopus laevis;* un *dictyostelium discoideum;* un hongo tal como *pneumocystis carinii, Takifugu rubripes,* levadura, *Saccharamoyces cerevisiae* o *Schizosaccharomyces pombe;* o un Plasmodium falciparum. Otros organismos incluyen un procariota tal como una bacteria, *Escherichia coli, estafilococos* o *mycoplasma pneumoniae;* una arquea; un virus tal como el virus de la hepatitis C o el virus de la inmunodeficiencia humana; o un viroide. Las células pueden derivarse de un cultivo o población homogénea de los organismos anteriores o, alternativamente, de una colección de varios organismos diferentes, por ejemplo, en un cultivo, comunidad o ecosistema mixto.

Las células que se usan en la presente memoria pueden ser de origen natural, por ejemplo, recogidos de una población nativa. Por ejemplo, las células pueden ser células genéticamente naturales aisladas de un organismo o comunidad multicelular. Alternativamente, las células candidatas pueden ser variantes modificadas por ingeniería de una célula natural. Por ejemplo, se pueden usar células, proteínas o ácidos nucleicos modificados por ingeniería genética en un método mostrado en la presente memoria. Las células o componentes celulares usados como agentes candidatos pueden derivarse de organismos unicelulares u organismos multicelulares. En realizaciones particulares, las células madre, las células inmunes o los componentes biológicos derivados de uno o más de estos tipos de células pueden usarse como agentes candidatos en la presente memoria.

Las etiquetas de ácido nucleico se pueden unir a agentes candidatos usando cualquiera de una variedad de químicas conocidas en la técnica apropiadas para su uso con ácidos nucleicos y el tipo particular de agente candidato en uso. En opciones particulares, las etiquetas de ácido nucleico se unirán covalentemente a agentes candidatos. Las químicas adecuadas para unir una etiqueta de ácido nucleico a un agente candidato (ya sea que el agente candidato un ácido nucleico u otra especie) incluyen, por ejemplo, ésteres de N-hidroxisuccinimida (ésteres de NHS), imidoésteres, hidrazinas, carbodiimidas, maleimidas, haloacetilos, disulfuros de piridinilo, diazirinas, química de clic (véanse, p. ej., las Pat. de EE. UU. Nos. 6.737.236; 7.427.678; 7.375.234; 7.763.736; o 8.129.542) o sulfhidrilos. Otras químicas útiles incluyen aquellas que se han empleado para unir ácidos nucleicos a perlas u otros soportes sólidos como se muestra en la Pat. de EE. UU. No. 7.259.258 o 7.504.499.

Las secuencias de etiqueta pueden tener una variedad de longitudes. Las secuencias más largas generalmente pueden acomodar un mayor número y variedad de etiquetas para una población. Generalmente, todas las sondas en una pluralidad tendrán la misma longitud en la etiqueta (aunque con diferentes secuencias), pero también es posible usar etiquetas de diferente longitud para diferentes sondas. Una secuencia etiqueta puede tener una longitud de al menos 2, 4, 6, 8, 10, 12, 15, 20 o más nucleótidos. Alternativamente o adicionalmente, la longitud de la secuencia de etiqueta puede ser como máximo de 20, 15, 12, 10, 8, 6, 4 o menos nucleótidos. Se exponen ejemplos de secuencias de etiqueta que se pueden usar se muestran, por ejemplo, en la Publ. de Solic. de Pat. de EE. UU. No. 2014/0342921 A1 y Pat. de EE. UU. No. 8.460.865.

El método de la presente invención incluye una etapa de puesta en contacto de una mezcla de células diferentes con un soporte sólido para formar una matriz de células unida al soporte sólido.

Puede usarse cualquiera de una variedad de soportes sólidos. Los soportes sólidos particularmente útiles son los que se utilizan para matrices de ácidos nucleicos. Los ejemplos incluyen vidrio, vidrio modificado, vidrio funcionalizado, vidrios inorgánicos, microesferas (p. ej., partículas inertes y/o magnéticas), plásticos, polisacáridos, nailon, nitrocelulosa, cerámica, resinas, sílice, materiales a base de sílice, carbono, metales, una fibra óptica o haces de fibras ópticas, polímeros y placas de múltiples pocillos (p. ej., de microtitulación). Los plásticos ejemplares incluyen acrílicos, poliestireno, copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos y Teflon™. Los materiales a base de sílice ejemplares incluyen silicio y diversas formas de silicio modificado.

En realizaciones particulares, un soporte sólido puede estar dentro o formar parte de un recipiente tal como un pocillo, tubo, canal, cubeta, placa de Petri, botella o similar. Un recipiente particularmente útil es una celda de flujo, por ejemplo, como se describe en WO 2014/142841 A1; Pub. de Solic. de Pat. de EE. UU. No. 2010/0111768 A1 y Pat. de EE. UU. No. 8,951,781 o Bentley et al., *Nature* 456: 53-59 (2008). Las celdas de flujo ejemplares son aquellas que están disponibles comercialmente en Illumina, Inc. (San Diego, CA) para su uso con una plataforma de secuenciación tal como una plataforma Genome Analyzer®, MiSeq®, NextSeq® o HiSeq®. Otro recipiente particularmente útil es un pocillo en una placa de múltiples pocillos o placa de microtitulación.

Opcionalmente, un soporte sólido puede incluir un recubrimiento de gel. La unión de ácidos nucleicos a un soporte sólido a través de un gel se ejemplifica mediante celdas de flujo disponibles comercialmente en Illumina Inc. (San Diego, CA) o descritas en Las Pub. de Solic. de Pat. de EE. UU. Nos. 2011/0059865 A1, 2014/0079923 A1 o 2015/0005447 A1; o Publ. PCT ºNo. WO 2008/093098. Los geles ejemplares que se pueden usar en los métodos y aparatos mostrados en la presente memoria incluyen, pero no están limitados a, aquellos que tienen una estructura coloidal, tales como agarosa; estructura de malla polimérica, tal como gelatina; o estructura de polímero reticulado, tal como poliacrilamida, SFA (véase, por ejemplo, la Pub. de Solic. de Pat. de EE. UU. No. 2011/0059865 A1) o PAZAM (véanse, por ejemplo, las Pub. de Solic. de Pat. de EE. UU. Nos. 2014/0079923 A1 o 2015/0005447 A1). Puede usarse un gel para unir agentes candidatos directamente (p. ej., mediante un enlace covalente entre el gel y los agentes candidatos) o mediante la hibridación del ácido nucleico unido al gel con ácidos nucleicos complementarios que ya

están unidos a los agentes candidatos.

En algunas realizaciones, un soporte sólido puede configurarse como una matriz de características a las que se pueden unir ácidos nucleicos y/o agentes candidatos. Las características pueden estar presentes en cualquiera de una variedad de formatos deseados. Por ejemplo, las características pueden ser pocillos, pozos, canales, crestas, regiones elevadas, clavijas, postes o similares. En algunas realizaciones, las características pueden contener perlas. Sin embargo, en realizaciones particulares, las características no necesitan contener una perla o partícula. Las características ejemplares incluyen pocillos que están presentes en sustratos utilizados para plataformas de secuenciación comerciales vendidas por 454 LifeSciences (una subsidiaria de Roche, Basilea Suiza) o Ion Torrent (una subsidiaria de Life Technologies, Carlsbad California). Otros sustratos que tienen pocillos incluyen, por ejemplo, fibras ópticas grabadas y otros sustratos descritos en las Pat de EE. UU. Nos. 6.266.459; 6.355.431; 6.770.441; 6.859.570; 6.210.891; 6.258.568; 6.274.320; Publ. de Solic. de Pat. de EE. UU. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; 2010/0282617 A1 o Publicación PCTNo. WO 2000/63437. En algunas realizaciones, los pocillos de un sustrato pueden incluir material de gel (con o sin perlas) como se muestra en la Publ. de Solic. de Pat. de EE. UU. No. 2014/0243224 A1.

Las características en un soporte sólido pueden ser características metálicas en una superficie no metálica, tal como vidrio, plástico u otros materiales ejemplificados anteriormente. Se proporcionan soportes sólidos ejemplares que tienen características metálicas y métodos para su fabricación en la Pat. de EE. UU. No. 8.895.249 o Pub. de Solic. de Pat. de EE. UU. No. 2014/0243224 A1.

Las características pueden aparecer en un soporte sólido como una cuadrícula de manchas o parches. Las características pueden ubicarse en un patrón repetido o en un patrón irregular que no se repite. Los patrones repetidos particularmente útiles son patrones hexagonales, patrones rectilíneos, patrones de rejilla, patrones que tienen simetría reflectante, patrones que tienen simetría rotacional, o similares. Los patrones asimétricos también pueden ser útiles. El paso puede ser el mismo entre diferentes pares de características vecinas más cercanas o el paso puede variar entre diferentes pares de características vecinas más cercanas.

Las matrices de densidad alta se caracterizan por tener un paso promedio (para las características vecinas) de menos de aproximadamente 15 µm. Las matrices de densidad media tienen un paso promedio de aproximadamente 15 a 30 µm, mientras que las matrices de densidad baja tienen un paso promedio superior a 30 µm. Una matriz útil en la invención puede tener un paso promedio menor de 100 µm, 50 µm, 10 µm, 5 µm, 1 µm o 0,5 µm. Se pretende que los valores y rangos de paso promedio mostrados anteriormente o en cualquier otro lugar de la presente memoria sean aplicables a matrices ordenadas o matrices aleatorias.

En realizaciones particulares, las características en un soporte sólido pueden tener cada una un área que es mayor de aproximadamente 100 nm.$^2$, 250 nm$^2$, 500 nm$^2$, 1 µm$^2$, 2,5 µm$^2$, 5 µm$^2$, 10 µm$^2$, 100 µm$^2$ o 500 µm$^2$ o más. Alternativamente o adicionalmente, las características pueden tener cada una un área menor de aproximadamente 1 mm$^2$, 500 µm$^2$, 100 µm$^2$, 25 µm$^2$, 10 µm$^2$, 5 µm$^2$, 1 µm$^2$, 500 nm$^2$, o 100 nm$^2$ o menos. Los rangos anteriores pueden describir el área aparente de una perla u otra partícula sobre un soporte sólido cuando se ve o se forma una imagen desde arriba.

En realizaciones particulares, un soporte sólido puede incluir una colección de perlas u otras partículas. Los ejemplos de matrices que tienen perlas ubicadas en una superficie incluyen aquellas en las que las perlas están ubicadas en pocillos tales como una matriz BeadChip™ (Illumina Inc., San Diego CA), sustratos utilizados en plataformas de secuenciación de 454 LifeSciences (una subsidiaria de Roche, Basilea Suiza) o sustratos utilizados en plataformas de secuenciación de Ion Torrent (una subsidiaria de Life Technologies, Carlsbad California). Otros soportes sólidos que tienen perlas ubicadas en una superficie se describen en las Pat. de EE. UU. Nos. 6.266.459; 6.355.431; 6.770.441; 6.859.570; 6.210.891; 6.258.568; o 6.274.320; Publ. de Solic. de Pat. de EE. UU. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; o 2010/0282617 A1 o Publicación PCTNo. WO 00/63437. Varias de las referencias anteriores describen métodos para unir ácidos nucleicos a perlas antes de cargar las perlas en o sobre un soporte sólido. Como tal, la colección de perlas puede incluir diferentes perlas, teniendo cada una un ácido nucleico único unido. Sin embargo, se entenderá que se puede hacer que las perlas incluyan cebadores universales, y las perlas se pueden cargar luego en una matriz, formando de ese modo matrices universales para su uso en un método mostrado en la presente memoria. Los agentes candidatos se pueden unir a las perlas antes o después de que las perlas se carguen en un soporte sólido. Como se mostró anteriormente en la presente memoria, los soportes sólidos que se usan típicamente para matrices de perlas se pueden usar sin perlas. Por ejemplo, los ácidos nucleicos (tales como sondas o cebadores) o agentes candidatos pueden unirse directamente a los pocillos o al material de gel en los pocillos. Por tanto, las referencias anteriores son ilustrativas de materiales, composiciones o aparatos que pueden modificarse para su uso en los métodos y composiciones mostrados en la presente memoria.

Por consiguiente, un soporte sólido utilizado en un método mostrado en la presente memoria puede incluir una matriz de perlas, en el que diferentes agentes candidatos o diferentes ácidos nucleicos se unen a diferentes perlas de la matriz. En esta realización, cada perla se puede unir a un agente candidato o ácido nucleico diferente y las perlas se pueden distribuir aleatoriamente sobre el soporte sólido con el fin de unir eficazmente los diferentes ácidos nucleicos al soporte sólido. Opcionalmente, el soporte sólido puede incluir pocillos que tengan dimensiones que no admitan más de una sola perla o un único agente candidato. En tal configuración, las perlas se pueden unir a los pocillos debido a

las fuerzas resultantes del ajuste de las perlas en los pocillos. También es posible utilizar adhesivos o químicas de fijación para sujetar las perlas en los pocillos.

Un soporte sólido puede incluir, o puede prepararse mediante los métodos mostrados en la presente memoria para unir, una pluralidad de ácidos nucleicos o agentes candidatos. Por ejemplo, un soporte sólido puede incluir al menos 10, 100, $1 \times 10^3$, $1 \times 10^4$, $1 \times 10^5$, $1 \times 10^6$, $1 \times 10^7$, $1 \times 10^8$, $1 \times 10^9$ o más ácidos nucleicos o agentes candidatos diferentes. Alternativamente o adicionalmente, un soporte sólido puede incluir como máximo $1 \times 10^9$, $1 \times 10^8$, $1 \times 10^7$, $1 \times 10^6$, $1 \times 10^5$, $1 \times 10^4$, $1 \times 10^3$, 100, 10 o menos ácidos nucleicos o agentes candidatos diferentes. Se entenderá que cada uno de los diferentes ácidos nucleicos o agentes candidatos puede estar presente en varias copias, por ejemplo, cuando los componentes de ácido nucleico de los agentes candidatos se han amplificado para formar un grupo. Por tanto, los rangos anteriores pueden describir el número de diferentes agentes candidatos o grupos de ácidos nucleicos en un soporte sólido. También se entenderá que los rangos anteriores pueden describir el número de etiquetas diferentes u otros elementos de secuencia mostrados en la presente memoria como únicos para ácidos nucleicos o agentes candidatos particulares. Alternativamente, o adicionalmente, los rangos pueden describir el número de ácidos nucleicos extendidos o agentes candidatos modificados creados sobre un soporte sólido usando un método mostrado en la presente memoria.

Las características pueden estar presentes en un soporte sólido antes de poner en contacto el soporte sólido con ácidos nucleicos o agentes candidatos. Por ejemplo, en realizaciones en las que los ácidos nucleicos o los agentes candidatos se unen a un soporte mediante hibridación con cebadores, los cebadores pueden unirse a las características, mientras que las áreas intersticiales fuera de las características carecen sustancialmente de cualquiera de los cebadores. Los ácidos nucleicos o los agentes candidatos pueden capturarse en las características preformadas en un soporte sólido y, opcionalmente, amplificarse en el soporte sólido, usando métodos mostrados en la Pat. de EE. UU. No. 8.895.249, Pat. de EE. UU. No. 8.778.849 o Pub. de Solic. de Pat. de EE. UU. No. 2014/0243224 A1.

En algunas realizaciones, las características se forman durante o después de la unión de etiquetas de ácido nucleico y/o agentes candidatos a un soporte sólido. Por ejemplo, un soporte sólido puede tener un césped de cebadores o puede carecer de otra forma de características. En este caso, se puede formar una característica en virtud de la unión de un ácido nucleico o agente candidato sobre el soporte sólido. Opcionalmente, un ácido nucleico capturado se puede amplificar en el soporte sólido de manera que el grupo resultante se convierta en una característica. Aunque la unión se ejemplifica anteriormente como captura entre un cebador y una porción complementaria de otro ácido nucleico, se entenderá que pueden estar presentes restos de captura distintos de los cebadores en características preformadas o como césped. Otros restos de captura ejemplares incluyen, pero no están limitados a, restos químicos capaces de reaccionar con un ácido nucleico o agente candidato para crear un enlace covalente o receptores capaces de unirse de forma no covalente a un ligando en un ácido nucleico o agente candidato.

En realizaciones particulares, los cebadores de ácido nucleico se unirán a un soporte sólido. Una biblioteca de agentes candidatos, que tiene etiquetas de ácido nucleico unidas, puede unirse al soporte sólido mediante la hibridación de las etiquetas de ácido nucleico con los cebadores de ácido nucleico unidos. Por ejemplo, las etiquetas de ácido nucleico pueden incluir una secuencia de unión de cebador universal, los cebadores de ácido nucleico pueden incluir una secuencia de cebador universal y los agentes candidatos pueden unirse al soporte sólido mediante la hibridación de la secuencia de unión del cebador universal a la secuencia de cebador universal. Como alternativa a los cebadores universales, el soporte sólido puede incluir cebadores específicos de diana que hibridan con secuencias de etiqueta específicas.

Se puede formar una matriz de agentes candidatos únicos por característica mediante la unión de agentes candidatos a un soporte sólido. Por consiguiente, una o más características de un soporte sólido pueden incluir cada una un único agente candidato (p. ej., una única molécula, una única célula u otro ítem único). Las características pueden configurarse, en algunas realizaciones, para acomodar no más de un único agente candidato de un tipo particular. Sin embargo, ya sea que la característica pueda acomodar o no a más de un agente candidato, la característica no obstante puede incluir no más de un único agente candidato, no más de una única etiqueta de ácido nucleico, o no más de tanto el agente candidato único como la etiqueta de ácido nucleico único. Alternativamente, una característica individual puede incluir una pluralidad de agentes candidatos y/o etiquetas de ácido nucleico. En realizaciones particulares, el conjunto puede producirse mediante amplificación a partir de un único molde de ácido nucleico para producir amplicones, por ejemplo, como un grupo unido a cada característica.

Un método mostrado en la presente memoria puede usar cualquiera de una variedad de técnicas de amplificación. Las técnicas ejemplares que pueden usarse incluyen, pero no están limitadas a, reacción en cadena de la polimerasa (PCR), amplificación por círculo rodante (RCA), amplificación de desplazamiento múltiple (MDA) o amplificación de cebado aleatorio (RPA). En algunas realizaciones, la amplificación se puede realizar en disolución. Preferiblemente, una técnica de amplificación utilizada en un método de la presente descripción se llevará a cabo en fase sólida. En particular, una o más especies de cebadores (p. ej., cebadores universales para uno o más sitios de unión de cebadores universales presentes en una etiqueta de ácido nucleico) que están unidos a un soporte sólido y que se hibridan con una etiqueta de ácido nucleico pueden extenderse en una técnica de amplificación. Tomando una realización de PCR en fase sólida como ejemplo, uno o ambos cebadores usados para la amplificación pueden unirse a un soporte sólido (p. ej., mediante un gel). Los formatos que utilizan dos especies de cebadores unidos a un soporte sólido a menudo se denominan amplificación en puente porque los amplicones bicatenarios forman una estructura

similar a un puente entre los dos cebadores unidos a un soporte sólido que flanquean la secuencia molde que se ha copiado. Los reactivos y las condiciones ejemplares que se pueden usar para la amplificación en puente se describen, por ejemplo, en las Pat. de EE. UU. Nos. 5.641.658, 7.115.400 o 8.895.249; o Publ. de Pat. de EE. UU. Nos. 2002/0055100 A1, 2004/0096853 A1, 2004/0002090 A1, 2007/0128624 A1 o 2008/0009420 A1. La amplificación por PCR en fase sólida también se puede llevar a cabo con uno de los cebadores de amplificación unido a un soporte sólido y el segundo cebador en disolución. Un formato ejemplar que usa una combinación de un cebador unido a un soporte sólido y un cebador soluble es el formato usado en la PCR en emulsión como se describe, por ejemplo, en Dressman et al., *Proc. Natl. Acad. Sci. USA* 100: 8817-8822 (2003), WO 2005/010145 o Publ. de Solic. de Pat. de EE. UU. Nos. 2005/0130173 A1 o 2005/0064460 A1. La PCR en emulsión es ilustrativa del formato y se entenderá que, para los propósitos de los métodos mostrados en la presente memoria, el uso de una emulsión es opcional y, de hecho, para varias realizaciones no se usa una emulsión.

Las técnicas de RCA pueden modificarse para su uso en un método de la presente descripción. Los componentes ejemplares que se pueden usar en una reacción de RCA y los principios mediante los cuales la RCA produce amplicones se describen, por ejemplo, en Lizardi et al., *Nat. Genet.* 19: 225-232 (1998) y Publ. de Solic. de Pat. de EE. UU. No. 2007/0099208 A1. Los cebadores utilizados para RCA pueden estar en disolución o unidos a un soporte sólido. Los cebadores pueden ser uno o más de los cebadores universales descritos en la presente memoria.

Las técnicas de MDA pueden modificarse para su uso en un método de la presente descripción. Algunos principios básicos y condiciones útiles para MDA se describen, por ejemplo, en Dean et al., *Proc Natl. Acad. Sci. USA* 99: 5261-66 (2002); Lage et al., *Genome Research* 13: 294-307 (2003); Walker et al., *Molecular Methods for Virus Detection*, Academic Press, Inc., 1995; Walker et al., *Nucl. Acids Res.* 20: 1691-96 (1992)); US 5.455.166; US 5.130.238; y US 6.214.587. Los cebadores utilizados para MDA pueden estar en disolución o unidos a un soporte sólido en un sitio de amplificación. Nuevamente, los cebadores pueden ser uno o más de los cebadores universales descritos en la presente memoria.

En realizaciones particulares, puede usarse una combinación de las técnicas de amplificación ejemplificadas anteriormente. Por ejemplo, se pueden usar RCA y MDA en una combinación en la que se usa RCA para generar un amplicón concatemérico en disolución (p. ej., usando cebadores en fase de disolución). A continuación, el amplicón se puede usar como molde para MDA usando cebadores que están unidos a un soporte sólido (p. ej., cebadores universales). En este ejemplo, los amplicones producidos después de las etapas combinadas de RCA y MDA se unirán al soporte sólido.

El método de la presente invención puede incluir una etapa de puesta en contacto de una matriz de células candidatas con un agente o estímulo de cribado. Uno o más agentes candidatos en la matriz pueden opcionalmente reaccionar con el agente de cribado o responden al estímulo. Como resultado, el uno o más agentes candidatos se pueden clasificar como aciertos.

En algunas realizaciones, un agente de cribado reaccionará con uno o más agentes candidatos en una matriz uniéndose a la una o más células candidatas o bloqueando la unión entre la célula candidata y un analito que tenga afinidad por el agente candidato. La matriz se puede detectar para identificar las características a las que se une el agente de cribado. Por ejemplo, el agente de cribado puede producir una señal que es detectable, por ejemplo, debido a la presencia de un marcador en el agente de cribado.

En algunas realizaciones, un agente de cribado modificará una o más características en una matriz donde reside una célula candidata. Por ejemplo, un agente de cribado puede reaccionar con un agente candidato modificando químicamente el agente candidato, el agente de cribado o ambos. Se puede detectar una matriz para identificar características en las que se han realizado modificaciones. En particular, la modificación puede producir una señal que sea detectable, por ejemplo, debido a que se une un marcador a una característica donde reside un agente candidato. Las modificaciones ejemplares incluyen la adición de un resto marcador a una característica donde reside un agente candidato; la adición de un resto de afinidad a una característica donde reside un agente candidato, en donde el resto de afinidad se une a un marcador; la eliminación de un apantallador fluorescente en una característica donde reside un agente candidato, en donde surge una señal luminiscente debido a la eliminación del apantallador; la adición de un resto donante o aceptor de transferencia de energía de resonancia de Forster (FRET) a una característica donde reside un agente candidato, en el que la adición altera la longitud de onda de luminiscencia aparente emitida por la característica; la eliminación de un resto marcador de una característica donde reside un agente candidato; la eliminación de un resto de afinidad de una característica donde reside un agente candidato, en el que el resto de afinidad se une a un marcador; la eliminación de un resto donante o aceptor de transferencia de energía de resonancia de Forster (FRET) de una característica donde reside un agente candidato, en donde la eliminación altera la longitud de onda de luminiscencia aparente emitida por la característica, o similar.

En algunas realizaciones, un agente de cribado reaccionará con uno o más agentes candidatos en una matriz para producir un analito. Se puede detectar una matriz para identificar las características en las que se ha producido un analito. Por ejemplo, el analito puede producir una señal que es detectable. Los analitos ejemplares que se pueden producir incluyen un marcador luminiscente, un ligando para un receptor detectable, un sustrato usado por una enzima para producir un producto detectable, o similares.

En algunas realizaciones, no es necesario que un agente de cribado esté en contacto directo con una biblioteca de agentes candidatos. Más bien, un agente de captura puede ponerse en contacto con el agente de cribado para rendir un producto de captura y luego el producto de captura puede ponerse en contacto con una colección de agentes candidatos. En otra realización, un agente de captura puede ponerse en contacto con una colección de agentes candidatos para rendir un producto de captura y luego el producto de captura puede ponerse en contacto con un agente de cribado. Como se ejemplifica por las realizaciones que usan un producto de captura, no es necesario que un agente de cribado entre en contacto directo con un agente candidato para que el agente de cribado sea útil para indicar una estructura o función deseada o sospechada para un agente candidato.

Un método establecido en el presente documento puede incluir además una etapa de detectar una matriz de agentes candidatos durante o después de poner en contacto la matriz con un agente de cribado, determinando así que al menos un agente candidato en la matriz reacciona con el agente de cribado. El soporte sólido al que se une la matriz puede estar en cualquiera de una variedad de estados mostrados en la presente memoria. Por ejemplo, el soporte sólido puede incluir agentes candidatos junto con etiquetas de ácido nucleico que están unidas al mismo. Alternativamente, el soporte sólido puede no incluir etiquetas de ácido nucleico, sino que se encuentra en un estado que sigue a la eliminación de las etiquetas de ácido nucleico de los agentes candidatos. Los ácidos nucleicos (p. ej., etiquetas y/o agentes candidatos basados en ácidos nucleicos) pueden estar en forma de molécula única o en forma de conjunto durante una etapa de detección. En otras realizaciones más, el soporte sólido puede no incluir agentes candidatos, sino que puede encontrarse en un estado que sigue a la eliminación de los agentes candidatos. Por consiguiente, la detección puede ocurrir en cualquiera de una variedad de puntos en un método mostrado en la presente memoria.

Cualquiera de una variedad de señales se puede detectar en una etapa de cribado mostrada en la presente memoria que incluye, por ejemplo, una señal óptica tal como absorbancia de radiación, emisión de luminiscencia, duración de la luminiscencia, polarización de luminiscencia o similares; dispersión de Rayleigh y/o Mie; propiedades magnéticas; propiedades eléctricas; carga; masa; radiactividad o similares. Los marcadores ejemplares que se pueden detectar en un método mostrado en la presente memoria incluyen, sin limitación, un fluoróforo, luminóforo, cromóforo, nanopartícula (p. ej., nanotubos de oro, plata, carbono), átomo pesado, isótopo radiactivo, marcador de masa, marcador de carga, marcador de spin, receptor, ligando o similares.

Las realizaciones particulares utilizan técnicas de imagenología. Se puede obtener una imagen usando dispositivos de detección conocidos en la técnica. Los ejemplos incluyen microscopios configurados para imagenología óptica, campo brillante, campo oscuro, contraste de fase, fluorescencia, reflexión, interferencia o confocal. En realizaciones particulares, se puede usar un microscopio de fluorescencia (p. ej., un microscopio fluorescente confocal) para detectar un agente de cribado (u otro analito) que sea fluorescente, por ejemplo, en virtud de un marcador fluorescente. También se pueden obtener imágenes de muestras fluorescentes utilizando un dispositivo de secuenciación de ácido nucleico que tenga ópticas para la detección fluorescente, tal como un dispositivo de plataforma Genome Analyzer®, MiSeq®, NextSeq® o HiSeq® comercializado por Illumina, Inc. (San Diego, CA); o una plataforma de secuenciación SOLiD™ comercializada por Life Technologies (Carlsbad, CA). Otras ópticas de imagenología que se pueden utilizar incluyen las que se encuentran en los dispositivos de detección descritos en Bentley et al., *Nature* 456: 53-59 (2008), Publ. PCT Nos. WO 91/06678, WO 04/018497 o WO 07/123744; Pat. de EE. UU. Nos. 7.057.026, 7.329.492, 7.211.414, 7.315.019 o 7.405.281 y Publ. de Solic. de Pat. de EE. UU. No. 2008/0108082.

Se puede obtener una imagen de un soporte sólido con la resolución deseada, por ejemplo, para distinguir los agentes candidatos en una matriz sobre el soporte sólido. Por consiguiente, la resolución puede ser suficiente como para distinguir características de una matriz que están separadas por al menos 0,5 µm, 1 µm, 5 µm, 10 µm, 50 µm, 100 µm, 500 µm, 1 mm o más. Alternativamente, o adicionalmente, la resolución se puede establecer para distinguir características de una matriz que están separadas como máximo por 1 mm, 500 µm, 100 µm, 50 µm, 10 µm, 5 µm, 1 µm, 0,5 µm o menos.

Un método de la presente descripción puede incluir una etapa de secuenciación de una etiqueta de ácido nucleico para determinar la secuencia de la etiqueta que está unida a los agentes candidatos en un soporte sólido (p. ej., matriz). En muchas realizaciones, los agentes candidatos se ubican aleatoriamente en una matriz y la reacción de secuenciación proporciona información para localizar cada uno de los diferentes agentes candidatos. La secuenciación se puede llevar a cabo en el soporte sólido al que se unen un agente candidato y su etiqueta de ácido nucleico. Por tanto, no es necesario eliminar una etiqueta de ácido nucleico del soporte sólido para determinar su secuencia. En algunos casos, los agentes candidatos se pueden eliminar opcionalmente de un soporte sólido antes de realizar una reacción de secuenciación en la etiqueta de ácido nucleico.

Un método de la presente descripción puede emplear detección basada en el tiempo o mediciones cinéticas de una matriz de agentes candidatos. Por tanto, la detección de una matriz puede incluir la adquisición de señales en varios puntos de tiempo para una o más de las características individuales de la matriz. Cualquiera de una variedad de las señales y marcadores mostrada en la presente memoria puede detectarse en un análisis cinético o basado en el tiempo.

Un método de la presente divulgación puede incluir una etapa de secuenciación de una etiqueta de ácido nucleico para determinar la secuencia de la etiqueta que se une a un agente candidato en una matriz. En algunas realizaciones, el agente candidato es un ácido nucleico, el agente candidato está codificado por un ácido nucleico al que está unido (p. ej., un agente candidato que es una proteína unida a un ARNm o ADNc) o el agente candidato contiene un ácido

nucleico (p. ej., una célula que contiene ácidos nucleicos). Las técnicas de secuenciación, tales como las técnicas de secuenciación por síntesis (SBS), son un método particularmente útil.

La SBS se puede llevar a cabo de la siguiente manera. Para iniciar un primer ciclo de SBS, uno o más nucleótidos marcados, ADN polimerasa, cebadores de SBS, etc., se pueden poner en contacto con una o más características en un soporte sólido (p. ej., característica(s) donde los ácidos nucleicos están unidos al soporte sólido). Pueden detectarse aquellas características en las que la extensión del cebador SBS cause la incorporación de un nucleótido marcado. Opcionalmente, los nucleótidos pueden incluir un resto de terminación reversible que termina la extensión adicional del cebador una vez que se ha añadido un nucleótido al cebador SBS. Por ejemplo, un análogo de nucleótido que tiene un resto terminador reversible puede añadirse a un cebador de manera que la extensión posterior no pueda ocurrir hasta que se administre un agente de desbloqueo para eliminar el resto. Por tanto, para las realizaciones que utilizan terminación reversible, se puede administrar un reactivo de desbloqueo al soporte sólido (antes o después de que se produzca la detección). Pueden llevarse a cabo lavados entre las distintas etapas de administración. El ciclo puede repetirse entonces n veces para extender el cebador en n nucleótidos, detectando así una secuencia de longitud n. Los procedimientos de SBS, sistemas fluídicos y plataformas de detección ejemplares que pueden adaptarse fácilmente para su uso con una composición, aparato o método de la presente descripción se describen, por ejemplo, en Bentley et al., *Nature* 456: 53-59 (2008), Publ. PCT Nos. WO 91/06678, WO 04/018497 o WO 07/123744; Pat. de EE. UU. Nos. 7.057.026, 7.329.492, 7.211.414, 7.315.019 o 7.405.281 y Publ. de Solic. de Pat. de EE. UU. No. 2008/0108082.

Se pueden utilizar otros procedimientos de secuenciación que utilizan reacciones cíclicas, tales como la pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) a medida que se incorporan nucleótidos particulares en una cadena de ácido nucleico naciente (Ronaghi, et al., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1), 3-11 (2001); Ronaghi et al. *Science* 281(5375), 363 (1998); o Pat. de EE. UU. Nos. 6.210.891, 6.258.568 o 6.274.320). En la pirosecuenciación, el PPi liberado puede detectarse convirtiéndolo inmediatamente en trifosfato de adenosina (ATP) por la ATP sulfurilasa, y el nivel de ATP generado puede detectarse mediante fotones producidos por luciferasa. Por tanto, la reacción de secuenciación se puede monitorizar mediante un sistema de detección de luminiscencia. Las fuentes de radiación de excitación utilizadas para los sistemas de detección basados en fluorescencia no son necesarias para los procedimientos de pirosecuenciación. Los sistemas fluídicos, detectores y procedimientos útiles que se pueden usar para la aplicación de pirosecuenciación a aparatos, composiciones o métodos de la presente descripción se describen, por ejemplo, en Publ. de Solic. de Pat. PCT No. WO2012/058096, Publ. de Solic. de Pat. EE. UU. No. 2005/0191698 A1 o Pat. de EE. UU. Nos. 7.595.883 o 7.244.559.

Las reacciones de secuenciación por ligación también son útiles, incluyendo, por ejemplo, las descritas en Shendure et al. *Science* 309: 1728-1732 (2005); o Pat. de EE. UU. Nos. 5.599.675 o 5.750.341. Algunas realizaciones pueden incluir procedimientos de secuenciación por hibridación como se describe, por ejemplo, en Bains et al., *Journal of Theoretical Biology* 135(3), 303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251(4995), 767-773 (1995); o Publ. de Solic. de Pat. PCT No. WO 1989/10977. Tanto en los procedimientos de secuenciación por ligación como de secuenciación por hibridación, los ácidos nucleicos que están presentes en los sitios de una matriz se someten a ciclos repetidos de administración y detección de oligonucleótidos. Las composiciones, aparatos o métodos mostrados en la presente memoria o en las referencias citadas en la presente memoria pueden adaptarse fácilmente para procedimientos de secuenciación por ligación o secuenciación por hibridación. Típicamente, los oligonucleótidos están marcados con fluorescencia y pueden detectarse usando detectores de fluorescencia similares a los descritos con respecto a los procedimientos de SBS en la presente memoria o en las referencias citadas en la presente memoria.

Algunas realizaciones de secuenciación pueden utilizar métodos que implican la monitorización en tiempo real de la actividad de la ADN polimerasa. Por ejemplo, las incorporaciones de nucleótidos pueden detectarse mediante interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa que lleva fluoróforo y nucleótidos marcados con γ-fosfato, o con guías de onda en modo cero (ZMW). Las técnicas y los reactivos para la secuenciación basada en FRET se describen, por ejemplo, en Levene et al. *Science* 299, 682-686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026-1028 (2008); Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176-1181 (2008).

Algunas realizaciones de secuenciación incluyen la detección de un protón liberado tras la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede utilizar un detector eléctrico y técnicas asociadas que están disponibles comercialmente en Ion Torrent (Guilford, CT, una subsidiaria de Thermo Fisher) o métodos y sistemas de secuenciación descritos en las Publ. de Solic. de Pat. de EE. UU. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; o US 2010/0282617 A1.

Un método de la presente descripción puede incluir además una etapa de eliminación de uno o más agentes candidatos de un soporte sólido. En realizaciones particulares, los agentes candidatos pueden eliminarse de una característica de una matriz antes de secuenciar una etiqueta de ácido nucleico en la característica. Esto puede ser beneficioso cuando el agente candidato interfiere con la técnica de secuenciación o cuando se desea proteger al agente candidato de los reactivos utilizados en la técnica de secuenciación. En algunas realizaciones, uno o más agentes candidatos que se han identificado como aciertos en un cribado pueden eliminarse selectivamente de una matriz.

En algunas realizaciones, se puede eliminar un agente candidato escindiendo un enlace que lo une a un soporte sólido. Por ejemplo, un agente candidato puede estar unido a un soporte sólido mediante una etiqueta de ácido

nucleico a la que está unido. En los casos en los que el ácido nucleico etiquetado se une al soporte mediante hibridación con un ácido nucleico complementario, la eliminación se puede lograr desnaturalizando el complejo híbrido. Alternativamente, la etiqueta de ácido nucleico puede incluir un sitio de escisión ubicado entre el punto de unión al agente candidato y la secuencia de la etiqueta. Como tal, el agente candidato puede separarse de la secuencia de la etiqueta mediante una reacción de escisión dirigida al sitio de escisión. En algunos métodos, se puede introducir un sitio de escisión en un ácido nucleico durante una etapa de amplificación o modificación. Por ejemplo, se puede introducir un sitio de escisión en un cebador extendido durante la etapa de extensión.

Los sitios de escisión ejemplares incluyen, pero no están limitados a, restos que son susceptibles a un proceso químico, enzimático o físico que da como resultado la rotura de enlaces. Por ejemplo, la ubicación puede ser una secuencia de nucleótidos reconocida por una endonucleasa. Las endonucleasas adecuadas y sus secuencias de reconocimiento son bien conocidas en la técnica y en muchos casos están incluso disponibles comercialmente (p. ej., en New England Biolabs, Beverley MA; ThermoFisher, Waltham, MA o Sigma Aldrich, St. Louis MO). Una endonucleasa particularmente útil romperá un enlace en una cadena de ácido nucleico en un sitio que es está alejado en 3' de su sitio de unión en el ácido nucleico, ejemplos de los cuales incluyen endonucleasas de restricción de Tipo II o Tipo II.

Un resto fotolábil proporciona un sitio de escisión útil. Los restos fotolábiles ejemplares incluyen, pero no están limitados a, éster de (1-(4,5-dimetoxi-2-nitrofenil)etilo) (es decir, DMNPE) y éster de (1-(2-nitrofenil)etilo) (es decir, NPE). Véase, *Meth. Enzymol.* 291: 307-347 (1998). Otros restos fotolábiles y los métodos para su síntesis y uso se describen en WO 91/06678 o Publ. de Solic. de Pat. de EE. UU. No. 20100092957 A1. Puede usarse ventajosamente un resto fotolábil para la escisión dirigida del agente candidato de una característica individual. La radiación resuelta espacialmente (p. ej., de un láser, una fuente de luz enfocada y/o un filtro/máscara espacial) se puede usar para abordar una característica individual en la que un agente candidato particular está unido a través de un resto fotolábil. Como resultado, ese agente candidato en particular se puede eliminar de forma selectiva (en comparación con los agentes candidatos en otras características en la misma matriz).

En algunas realizaciones, un sitio de escisión es un sitio abásico o un nucleótido que tiene una base que es susceptible de ser eliminada para crear un sitio abásico. Los ejemplos de nucleótidos que son susceptibles de ser eliminados para formar un sitio abásico incluyen uracilo y 8-oxo-guanina. En una realización particular, se usa el reactivo USER™ disponible en New England Biolabs. Otros ejemplos de sitios de escisión y métodos que pueden usarse para escindir ácidos nucleicos se muestran, por ejemplo, en la Pat. de EE. UU. No. 7.960.120.

También se describe en la presente memoria un método para producir una matriz de proteínas. El método puede incluir las etapas de (a) proporcionar una biblioteca de moléculas de ADNc que están unidas a un soporte sólido; (b) amplificar las moléculas de ADNc en el soporte sólido para formar grupos, en el que cada grupo incluye múltiples copias de una molécula de ADNc particular de la biblioteca; (c) transcribir las múltiples copias en los grupos para producir múltiples moléculas de ARNm unidas a cada uno de los grupos; y (d) traducir las moléculas de ARNm en los grupos para producir múltiples proteínas unidas a cada uno de los grupos.

También se describe en la presente memoria un método para producir una matriz de proteínas que incluye las etapas de (a) proporcionar una biblioteca de moléculas de ARNm, en donde las moléculas de ARNm individuales en la biblioteca incluyen una secuencia diana y una secuencia de etiqueta, (b) derivar una primera sub-biblioteca de la biblioteca, la primera sub-biblioteca que incluye ácidos nucleicos que tienen las secuencias de etiqueta o complementos de las mismas, en la que los ácidos nucleicos están unidos a características individuales en un soporte sólido, (c) derivar una segunda sub-biblioteca de la biblioteca, la segunda sub-biblioteca que incluye ácidos nucleicos que tienen las secuencias diana y las secuencias de etiqueta o complementos de las mismas; (d) poner en contacto la segunda sub-biblioteca con la primera sub-biblioteca, uniendo así los ácidos nucleicos de la segunda sub-biblioteca al soporte sólido mediante la hibridación de las secuencias de etiqueta y los complementos de las mismas; y (e) traducir las secuencias diana en el soporte sólido para producir una matriz de proteínas unidas a las características individuales.

En el presente documento también se describe una biblioteca de moléculas de ADN genómico (ADNg), ARN mensajero (ARNm) o ADN de copia (ADNc) puede derivarse de uno o más organismos que expresan una o más proteínas de interés. Por ejemplo, la biblioteca puede obtenerse de un organismo del que se sospecha que expresa una proteína deseada o una proteína con una actividad deseada. Puede usarse cualquier organismo conocido en la técnica, tales como los mostrados en la presente memoria. El organismo puede tener un interés industrial, terapéutico, de diagnóstico o pronóstico. Por ejemplo, puede obtenerse una biblioteca de un organismo humano y cribarse con fines de pronóstico o diagnóstico. Alternativamente, se puede sospechar que el organismo expresa un gen que produce un efecto terapéutico o que tiene una aplicación industrial. Una biblioteca puede obtenerse de un organismo no humano o que de otro modo no tiene origen humano.

En el presente documento también se describe una una biblioteca de moléculas de ADNg, ARNm o ADNc que incluye variantes de un solo gen que se expresan en una biblioteca de organismos recombinantes. Dichas variantes se pueden producir usando técnicas conocidas de ingeniería de proteínas que incluyen, pero no están limitadas a, las mostradas en Sambrook et al., *Molecular Cloning: A Laboratory Manual, Tercera Ed.,* Cold Spring Harbor Laboratory, Nueva York (2001) y en Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1999). Por ejemplo, las variantes pueden producirse mediante mutagénesis aleatoria de todo o parte del gen. También se pueden utilizar técnicas de mutagénesis semialeatoria.

Los genes que se expresan (p. ej., como ARNm, ADNc o proteína) en un método mostrado en la presente memoria se pueden seleccionar del grupo que consiste en anticuerpos, enzimas, receptores, quinasas, fosfatasas, polimerasas, proteasas, esterasas, receptores de hormonas nucleares, enzimas modificadoras de histonas y otros ejemplificados en la presente memoria. Los genes pueden ser recombinantes, por ejemplo, expresarse en un organismo que es diferente de la fuente natural del gen o tener modificaciones que no son naturales para el gen. En realizaciones particulares, se puede aislar un ADNg, ARNm, ADNc o biblioteca de proteínas de una bacteria, levadura, insecto, mamífero u otro sistema celular que se usa comúnmente para clonación o expresión.

Se puede construir una biblioteca de diferentes ácidos nucleicos (p. ej., ADNg, ARNm, ADNc) de manera que cada miembro individual de la biblioteca incluya una secuencia codificante de proteína y una secuencia de etiqueta. La secuencia de etiqueta se puede asignar a cada secuencia codificante de proteínas de forma aleatoria o sistemática (es decir, cuando una etiqueta conocida está asociada *a priori* con una variante conocida). Como ejemplo, las etiquetas se pueden asignar aleatoriamente incorporando nucleótidos aleatorios en una o más posiciones en la región de la etiqueta cuando se sintetiza la construcción de ácido nucleico que también codificará la proteína. En otro ejemplo, una población de diferentes etiquetas de ácido nucleico se puede ligar aleatoriamente a una población de diferentes ácidos nucleicos que codifican proteínas de modo que, de promedio, cada ácido nucleico que codifica proteínas se ligará a una única secuencia de etiqueta. Como ejemplo de etiquetado sistemático, una etiqueta específica se puede sintetizar en una construcción de ácido nucleico de una manera que se correlacione *a priori* con una mutación o variante en la secuencia que codifica la proteína que está en la construcción.

En el presente documento también se describe un método que incluye una etapa de derivación de una primera y una segunda sub-biblioteca a partir de una biblioteca. Opcionalmente, las sub-bibliotecas pueden derivarse eliminando fracciones de una biblioteca o dividiendo físicamente la biblioteca. Generalmente, una biblioteca será un fluido que incluye múltiples copias de cada agente candidato. Puede usarse cualquiera de una variedad de agentes candidatos tales como ADNg, ARNm, ADNc, macrociclos, péptidos cíclicos, moléculas fusionadas (p. ej., fusión de ácido nucleico-proteína), construcciones presentadas (p. ej., péptidos en fagos) o proteína o similares. Como tales, se espera que las sub-bibliotecas que se obtienen fraccionando o dividiendo el fluido tengan aproximadamente el mismo contenido. En algunos casos, se puede llevar a cabo una etapa de amplificación en una biblioteca de ácidos nucleicos (p. ej., biblioteca de ADN o ARN) con el fin de crear las múltiples copias.

Se puede unir una biblioteca de ADNg, ARNm, ADNc o de proteínas a un soporte sólido usando métodos mostrado en otra parte en la presente memoria. En una opción particular, se usa una biblioteca de moléculas de ARNm, en donde cada una de las moléculas de ARNm incluye una secuencia que codifica un agente candidato junto con una secuencia de etiqueta. Una primera sub-biblioteca se deriva mediante un método que incluye poner en contacto moléculas de ARNm de la biblioteca con el soporte sólido para unir las moléculas de ARNm al soporte sólido. Por ejemplo, el soporte sólido puede tener cebadores de ácido nucleico unidos a un soporte sólido y las moléculas de ARNm pueden unirse al soporte sólido mediante hibridación con los cebadores de ácido nucleico. A modo de ejemplo más específico, las moléculas de ARNm pueden incluir una secuencia de unión de cebador universal, los cebadores de ácido nucleico y una secuencia de cebador universal, y las moléculas de ARNm pueden unirse al soporte sólido mediante la hibridación de la secuencia de unión del cebador universal a la secuencia del cebador universal. Las moléculas de ARNm unidas pueden copiarse o amplificarse entonces sobre el soporte sólido para producir complementos de las secuencias de etiquetas. Después, una segunda sub-biblioteca que se deriva de la misma biblioteca que la primera sub-biblioteca puede ponerse en contacto con los amplicones (o copias) de la primera sub-biblioteca, uniendo así los ácidos nucleicos de la segunda sub-biblioteca al soporte sólido. a través de la hibridación de las secuencias de etiqueta y los complementos de las mismas. De esta manera, el soporte sólido se ha modificado para incluir una biblioteca de moléculas de ARNm que está disponible para una reacción de cribado o para traducción para producir una biblioteca de proteínas expresadas por las moléculas de ARNm de la segunda sub-biblioteca.

En un método descrito en la presente memoria, una de la primera sub-biblioteca de ácidos nucleicos o segunda sub-biblioteca de ácidos nucleicos se puede copiar o amplificar en disolución. Como tal, la sub-biblioteca amplificada o copiada incluirá secuencias de etiqueta complementarias que se pueden hibridar con secuencias de etiqueta en la otra sub-biblioteca. Cualquiera de las sub-bibliotecas se puede unir al soporte sólido y la otra sub-biblioteca se puede hibridar para crear un soporte sólido que tenga una biblioteca de moléculas de ácido nucleico que esté disponible para una reacción de cribado, reacción de transcripción para producir ARNm y/o reacción de traducción para producir una biblioteca de proteínas.

Para los métodos descritos en la presente memoria que utilizan una biblioteca de ARNm, se puede derivar una primera sub-biblioteca mediante un método que incluye la transcripción inversa de algunas o todas las moléculas de ARNm individuales en la biblioteca para producir ADNc. La transcripción inversa puede ocurrir en disolución o sobre un soporte sólido. Las moléculas de ADNc resultantes pueden amplificarse entonces en disolución o sobre el soporte sólido. Por tanto, pueden producirse moléculas de ADNc que tengan complementos de las secuencias de etiqueta que estaban en la biblioteca. En algunos casos, las moléculas de ADNc no pueden incluir más que una porción de las moléculas de ARNm, por ejemplo, una porción que expresa un dominio particular o un subconjunto de dominios de una proteína multidominio.

También se describe en el presente documento un método que puede incluir una etapa de transcripción de ADNc a ARNm. La transcripción se puede llevar a cabo utilizando mezclas conocidas tales como las mostradas Sambrook et

al., *Molecular Cloning: A Laboratory Manual, Tercera Ed.*, Cold Spring Harbor Laboratory, Nueva York (2001) y en Ansubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1999) o los que están disponibles comercialmente. Opcionalmente, la transcripción se puede llevar a cabo en moléculas de ADNc unidas a un soporte. Los productos de ARNm de la reacción de transcripción pueden unirse a las características donde se producen. Por ejemplo, la unión se puede lograr usando las técnicas mostradas en el Ejemplo II, en la presente memoria. También puede ser útil unir proteínas a las moléculas de ADN que las codifican (p. ej., mediante unión directa o colocalización en gotitas de emulsión). Las proteínas también se pueden unir en las proximidades de las moléculas de ADN que las codifican (p. ej., una matriz de proteínas programables de ácido nucleico (NAPPA)).

También se describe en el presente documento un método que puede incluir una etapa de traducción de ARNm en proteína. La traducción se puede llevar a cabo utilizando mezclas conocidas como las mostradas en Sambrook et al., *Molecular Cloning: A Laboratory Manual, Tercera Ed.*, Cold Spring Harbor Laboratory, Nueva York (2001) y en Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1999) o los que están disponibles comercialmente. Opcionalmente, la traducción se puede llevar a cabo en moléculas de ARNm unidas a un soporte. Los productos proteicos de la reacción de traducción se pueden unir a las características donde se producen. Por ejemplo, la unión se puede lograr usando las técnicas mostradas en el Ejemplo II, en la presente memoria.

Los métodos mostrados en la presente memoria para producir una matriz de proteínas se pueden utilizar en un método para cribar proteínas. El método puede incluir las etapas de (i) producir la matriz de proteínas; (ii) poner en contacto la matriz de proteínas con un agente de cribado, en el que una o más proteínas de la matriz reaccionan con el agente de cribado (o estímulo); y (iii) detectar la matriz de proteínas durante o después del contacto con el agente de cribado (o estímulo), determinando de ese modo que al menos una proteína de la matriz reacciona con el agente de cribado (o estímulo).

En opciones particulares, el agente de cribado reacciona con la una o más proteínas uniéndose a la una o más proteínas o bloqueando la unión entre la una proteína y un analito que tiene afinidad por la proteína. La detección de la matriz puede incluir la detección del agente de cribado que está unido a la una o más proteínas. Por ejemplo, el agente de cribado puede ser luminiscente y la detección se puede llevar a cabo detectando luminiscencia en la matriz.

En algunas opciones, el agente de cribado reacciona con la una o más proteínas modificando químicamente la una o más proteínas. En este caso, la detección de la matriz se puede lograr detectando la una o más proteínas modificadas. Por ejemplo, la una o más proteínas modificadas pueden ser luminiscentes y la detección se puede lograr detectando luminiscencia en la matriz.

En algunas opciones, el agente de cribado reacciona con la una o más proteínas produciendo un producto analito. En este caso, la detección de la matriz incluye la detección del producto analito. Por ejemplo, el producto analito puede ser luminiscente y la detección se puede llevar a cabo detectando la luminiscencia en la matriz.

Como se ha mostrado anteriormente en la presente memoria, la detección de la matriz puede incluir la adquisición de señales en varios puntos de tiempo para una o más de las características individuales de la matriz.

Las secuencias de etiqueta que forman parte de las moléculas de ADNg, ADNc o ARNm pueden secuenciarse usando métodos mostrados en la presente memoria o conocidos en la técnica. Por ejemplo, se puede llevar a cabo una reacción de secuenciación en el soporte sólido donde se lleva a cabo, se llevó a cabo o se llevará a cabo un cribado. Por tanto, un método mostrado en la presente memoria puede incluir una etapa de secuenciación de secuencias de etiqueta, o complementos de las mismas, en el soporte sólido, determinando así las ubicaciones de las secuencias de etiqueta, o complementos de las mismas, en las características individuales del soporte sólido. Dichos métodos pueden incluir además una etapa de identificación de al menos una proteína en la matriz que reacciona con un agente de cribado basado en la secuencia de etiqueta que está unida a la al menos una proteína.

También se describe en la presente memoria una matriz que incluye (a) un soporte sólido; (b) una biblioteca de diferentes moléculas de ADNc unidas al soporte sólido, en la que cada molécula de ADNc diferente está unida a una característica individual del soporte sólido, y en la que cada característica incluye múltiples copias de una molécula de ADNc particular; (c) moléculas de ARNm unidas a las moléculas de ADNc, en las que cada una de las moléculas de ADNc es complementaria a la respectiva molécula de ARNm unida; y (d) moléculas de proteína unidas a las moléculas de ARNm, donde cada una de las moléculas de proteína está codificada por la respectiva molécula de ARNm unida.

También se describe en la presente memoria una matriz que incluye (a) una biblioteca de moléculas de ARNm, en la que las moléculas de ARNm individuales en la biblioteca comprenden una secuencia diana y una secuencia de etiqueta, (b) un soporte sólido que comprende ácidos nucleicos que tienen complementos de las secuencias de etiqueta, donde los ácidos nucleicos están unidos a características individuales en un soporte sólido, en el que las secuencias de etiqueta de las moléculas de ARNm individuales se hibridan con las respectivas secuencias de etiqueta complementarias en las características individuales del soporte sólido, y en las que se unen proteínas derivadas de la traducción de las moléculas de ARNm a las respectivas moléculas de ARNm.

Una matriz de la presente descripción puede incluir componentes resultantes de una o más etapas de un método mostrado en la presente memoria. Por ejemplo, los ácidos nucleicos, tales como especies de ADNg, ADNc o ARNm, se pueden unir a un soporte sólido de acuerdo con los métodos mostrados en la presente memoria.

De manera similar, las proteínas se pueden unir a un soporte sólido de acuerdo con los métodos mostrados en la presente memoria. En opciones particulares, el ADNc se puede unir a un soporte sólido, el ARNm codificado se puede unir al ADNc y la proteína codificada se puede unir al ARNm, como se muestra en los Ejemplos II y III. Como se muestra en esos ejemplos, la unión mediante ADNc es opcional y, en cambio, el ARNm puede unirse al soporte sólido (p. ej., mediante hibridación con una etiqueta complementaria o mediante otros métodos de unión), y la proteína codificada se puede unir al ARNm.

Opcionalmente, se puede unir una molécula de ARNm a su molécula de ADNc codificante mediante un complejo formado con una ARN polimerasa. La unión puede estar mediada por reticulación covalente entre el complejo de ADNc, ARNm y ARN polimerasa. De manera similar, una molécula de proteína puede unirse a su ARNm codificante a través de un complejo formado con un ribosoma. La unión puede estar mediada por reticulación covalente entre el complejo de ARNm, la proteína y el ribosoma.

Cualquiera de una variedad de agentes de cribado, marcadores o productos de una etapa de cribado mostrada en la presente memoria puede estar presente en una matriz de la presente descripción. Por ejemplo, las características donde se localizan las proteínas pueden marcarse luminiscentemente, opcionalmente, mediante la unión no covalente específica de una molécula luminiscente (tal como un agente de cribado) a las proteínas o mediante un resto luminiscente unido covalentemente a las proteínas. No todas las características estarán necesariamente marcadas. Más bien, un subconjunto de características que contienen proteínas que tienen una actividad deseada o seleccionada puede marcarse selectivamente mientras que otras características no están marcadas. Por ejemplo, se pueden marcar menos del 50 %, 25 %, 10 %, 5 % o 1 % de las características, dependiendo del contenido de la biblioteca y la naturaleza del cribado.

Cualquiera de una variedad de agentes candidatos y/o etiquetas de ácido nucleico mostrada en la presente memoria puede estar presente en una matriz de la presente descripción. El número de especies diferentes en una matriz, la densidad de características que contienen las especies en una matriz o el número de cada especie unida a una característica particular puede estar en un rango mostrado en la presente memoria con respecto a los métodos para preparar y usar las matrices.

En esta divulgación se proporciona un método de cribado de células. El método puede incluir las etapas de (a) proporcionar una pluralidad de células diferentes, en donde cada una de las células diferentes comprende una etiqueta de ácido nucleico que tiene una secuencia de etiqueta, estando unida la secuencia de etiqueta de ácido nucleico a cada célula de la pluralidad de células mediante unión covalente a un lípido de membrana plasmática, ácido graso de membrana plasmática o proteína en un lípido de membraba plasmática de la célula, en donde cada secuencia de etiqueta comprende una secuencia única; (b) poner en contacto una mezcla de las diferentes células con un soporte sólido para formar una matriz de células unidas al soporte sólido; (c) cribar la matriz de células en el soporte sólido para al menos una característica óptica, en donde la reacción de cribado comprende detectar las células individuales que están unidas al soporte sólido; (d) secuenciar las secuencias de etiqueta de las etiquetas de ácido nucleico que están unidas al soporte sólido; e (e) identificar al menos una célula en la matriz como una célula candidata basándose en la característica óptica y la secuencia de etiqueta de la célula candidata.

En realizaciones particulares, las células se utilizan como agentes candidatos en un método mostrado en la presente memoria. Las células pueden ser células naturales aisladas de un organismo multicelular, células naturales que comprenden organismos unicelulares, células modificadas genéticamente de un organismo multicelular u organismos unicelulares modificados genéticamente. Las células utilizadas en un método o composición mostrado en la presente memoria pueden obtenerse de una fuente nativa o pueden obtenerse de un cultivo *ex vivo*.

Las células que se pueden copiar, cultivar o expandir *ex vivo* son particularmente útiles. Por ejemplo, puede ser útil hacer una o más copias de una célula candidata antes o después de una etapa de cribado. Por tanto, una célula que se identifica como un acierto en un cribado (o un clon de la célula) puede aislarse de un material preformado o del soporte sólido utilizado en el cribado. La célula aislada se puede utilizar o manipular adicionalmente, por ejemplo, para caracterizar más completamente la célula o para emplear la célula en un procedimiento terapéutico.

Para las realizaciones que emplean células modificadas genéticamente, la modificación genética puede ser cualquiera de una variedad conocida en la técnica que incluye, por ejemplo, aquellas que dan como resultado la expresión de una proteína recombinante no nativa, la expresión de una proteína recombinante mutante, la deleción de la totalidad o parte de la secuencia codificante de una proteína de origen natural, la inhibición de la expresión de una proteína de origen natural, el aumento de la expresión de una proteína de origen natural, la producción de un analito no nativo o la inhibición de la producción de un analito nativo. Por ejemplo, las secuencias codificantes de uno o más genes en una biblioteca de células candidatas pueden contener mutaciones puntuales, deleciones (p. ej., eliminación de la secuencia codificante de la proteína completa, un dominio u otra porción de la proteína) o inserciones (p. ej., quimeras).

Las células candidatas que se encuentran en una biblioteca pueden incluir etiquetas de ácido nucleico. En el Ejemplo IV y/o a continuación se muestran varios métodos ejemplares para etiquetar células.

En algunas realizaciones, las etiquetas de ácido nucleico pueden unirse covalentemente a la superficie de las células. Generalmente, se emplean métodos para etiquetar cada célula en una biblioteca con solo una única secuencia de

etiqueta de ácido nucleico (aunque pueden estar presentes múltiples moléculas de ácido nucleico que tienen cada una, una copia de la misma secuencia de etiqueta en o sobre cada célula). Por ejemplo, las células pueden aislarse físicamente de modo que cada célula pueda reaccionar individualmente con una molécula de etiqueta de ácido nucleico. Los métodos para separar físicamente las células incluyen, por ejemplo, la separación de cada célula en un recipiente individual, pocillo en una microplaca, característica en una matriz, perla, gota fluídica en un dispositivo accionador de gotas, gota de líquido en una emulsión o vesícula.

Los métodos particularmente útiles para crear gotitas a las que se pueden suministrar etiquetas de ácido nucleico incluyen, por ejemplo, los comercializados por RainDance Technologies (Billerica, MA) o descritos en las Pat. de EE. UU. Nos. 9.017.623 o 8.857.462. Los métodos adicionales para crear y añadir etiquetas a las gotitas están comercializados por 10X Genomics o se describen en las Pub. de Solic. de Pat. de EE. UU. Nos. 2014/0155295 A1; 2014/0206554 A1; 2014/0227684 A1 o 2014/0378322 A1. Estos métodos se pueden modificar de manera que las células (u otros agentes candidatos mostrados en la presente memoria) se carguen en gotitas individuales y las gotitas cargadas interactúen con fluidos que contienen etiquetas de ácido nucleico de manera que las gotitas individuales terminen con una especie de ácido nucleico de etiqueta única y una especie de agente candidato único. Por supuesto, pueden estar presentes múltiples copias de la especie de ácido nucleico o múltiples copias de la especie de agente candidato en una gotita individual. En algunas realizaciones, las etiquetas de ácido nucleico se unen a perlas que se administran a las gotitas individuales. Las etiquetas de ácido nucleico pueden unirse a la superficie de las células usando las químicas de unión mostradas en la presente memoria, por ejemplo, en el Ejemplo IV.

Un dispositivo de manipulación de gotitas particularmente útil que se puede usar para separar células (u otros agentes candidatos), por ejemplo, para etiquetar las células, es un accionador de gotitas como se describe, por ejemplo, en la Pat. de EE. UU. No. 8.637.242, Pat. de EE. UU. No. 6.911.132, titulado "Apparatus for Manipulating Droplets by Electrowetting-Based Techniques", expedida el 28 de junio de 2005; Pamula et al., Pub. de Patente de EE. UU. No. 20060194331, titulada "Apparatuses and Methods for Manipulating Droplets on a Printed Circuit Board", publicado el 31 de agosto de 2006; Pollack et al., Pub. de Patente Internacional. No. WO/2007/120241, titulada "Droplet-Based Biochemistry", publicado el 25 de octubre de 2007; Shenderov, Patente de EE. UU. No. 6.773.566, titulado "Electrostatic Actuators for Microfluidics and Methods for Using Same", expedida el 10 de agosto de 2004; Shenderov, Patente de EE. UU. No. 6.565.727, titulada "Actuators for Microfluidics Without Moving Parts", expedida el 20 de mayo de 2003; Kim et al., Pub. de Patente de EE. UU. No. 20030205632, titulada "Electrowetting-driven Micropumping", publicada el 6 de noviembre de 2003; Kim et al., Pub. de Patente de EE. UU. No. 20060164490, titulada "Method and Apparatus for Promoting the Complete Transfer of Liquid Drops from a Nozzle", publicada el 27 de julio de 2006; Kim et al., Pub. de Patente de EE. UU. No. 20070023292, titulada "Small Object Moving on Printed Circuit Board", publicada el 1 de febrero de 2007; Shah et al., Pub. de Patente de EE. UU. No. 20090283407, titulada "Method for Using Magnetic Particles in Droplet Microfluidics", publicada el 19 de noviembre de 2009; Kim et al., Pub. de Patente de EE. UU. No. 20100096266, titulada "Method and Apparatus for Real-time Feedback Control of Electrical Manipulation of Droplets on Chip", publicada el 22 de abril de 2010; Velev, Patente de EE. UU. No. 7.547.380, titulada "Droplet Transportation Devices and Methods Having a Fluid Surface", expedida el 16 de junio de 2009; Sterling et al., Patente de EE. UU. No. 7.163.612, titulada "Method, Apparatus and Article for Microfluidic Control via Electrowetting, for Chemical, Biochemical and Biological Assays and the Like", expedida el 16 de enero de 2007; Becker et al., Patente de EE. UU. No. 7.641.779, titulada "Method and Apparatus for Programmable Fluidic Processing", expedida el 5 de enero de 2010; Becker et al., Patente de EE. UU. No. 6.977.033, titulada "Method and Apparatus for Programmable Fluidic Processing", expedida el 20 de diciembre de 2005; Decre et al., Patente de EE. UU. No. 7.328.979, titulada "System for Manipulation of a Body of Fluid", expedida el 12 de febrero de 2008; Yamakawa et al., Pub. de Patente de EE. UU. No. 20060039823, titulada "Chemical Analysis Apparatus", publicada el 23 de febrero de 2006; Wu, Pub. de Patente de EE. UU. No. 20110048951, titulada "Digital Microfluidics Based Apparatus for Heat-exchanging Chemical Processes", publicada el 3 de marzo de 2011; Fouillet et al., Pub. de Patente de EE. UU. No. 20090192044, titulada "Electrode Addressing Method", publicada el 30 de julio de 2009; Fouillet et al., Patente de EE. UU. No. 7.052.244, titulada "Device for Displacement of Small Liquid Volumes Along a Micro-catenary Line by Electrostatic Forces", expedida el 30 de mayo de 2006; Marchand et al., Pub. de Patente de EE. UU. No. 20080124252, titulada "Droplet Microreactor", publicada el 29 de mayo de 2008; Adachi et al., Pub. de Patente de EE. UU. No. 20090321262, titulada "Liquid Transfer Device", publicada el 31 de diciembre de 2009; Roux et al., Pub. de Patente de EE. UU. No. 20050179746, titulada "Device for Controlling the Displacement of a Drop Between Two or Several Solid Substrates", publicada el 18 de agosto de 2005; y Dhindsa et al., "Virtual Electrowetting Channels: Electronic Liquid Transport with Continuous Channel Functionality", Lab Chip, 10: 832-836 (2010).

En realizaciones particulares, las células de una biblioteca pueden modificarse genéticamente para incluir una etiqueta de ácido nucleico. Por ejemplo, las células individuales en una población pueden portar un plásmido que codifica una secuencia de etiqueta o el genoma de las células individuales puede modificarse para incluir una secuencia de etiqueta. En algunos casos, la secuencia de etiqueta está codificada en una construcción de ácido nucleico que también incluye una variante genética que se va a cribar en un método mostrado en la presente memoria.

Puede construirse una biblioteca de diferentes células de modo que cada célula individual adquiera una secuencia de etiqueta asignada aleatoriamente o cada célula pueda modificarse para incluir una secuencia de etiqueta conocida. Por ejemplo, el etiquetado aleatorio se puede llevar a cabo incorporando nucleótidos aleatorios en una o más posiciones en una región de etiqueta cuando se sintetiza una construcción de ácido nucleico que también codificará una variante genética que se va a cribar. En otro ejemplo, se puede ligar aleatoriamente una población de diferentes

etiquetas de ácido nucleico a una población de diferentes ácidos nucleicos codificantes variantes de manera que, de promedio, cada ácido nucleico variante se ligará a una secuencia de etiqueta única. Después, las construcciones se pueden añadir a las células en una biblioteca. De manera similar, una población de diferentes etiquetas de ácido nucleico se puede unir covalentemente a una población de diferentes células de manera que, de promedio, cada célula se una a una secuencia de etiqueta única.

En otras realizaciones, se puede construir una biblioteca de diferentes células de modo que cada célula individual adquiera una etiqueta conocida que está asociada *a priori* con una célula conocida. Como ejemplo de un etiquetado *a priori*, una etiqueta específica se puede sintetizar en una construcción de ácido nucleico de una manera que se correlacione con una mutación o variante en esa construcción. En algunas realizaciones, las células se separan físicamente entre sí y los ácidos nucleicos con secuencias de etiqueta conocidas se ponen en contacto con las células para formar una unión covalente de un solo tipo de etiqueta con una sola célula.

En una realización particular, se puede unir una etiqueta de ácido nucleico a una perla y la perla se puede unir a una célula. Opcionalmente, la perla se puede unir a un anticuerpo que tenga afinidad de unión específica a la célula. Sin embargo, se pueden usar otras modalidades de unión para unir perlas a células que incluyen, por ejemplo, las ejemplificadas en la presente memoria en el contexto de la unión de ácidos nucleicos a agentes candidatos o soportes sólidos.

En algunas realizaciones, una etiqueta de ácido nucleico puede unirse a una célula mediante la unión covalente de la etiqueta de ácido nucleico a un lípido o ácido graso de la membrana plasmática de la célula. Alternativamente, se puede unir una etiqueta de ácido nucleico a una célula mediante unión covalente a una proteína en un lípido de la membrana plasmática de la célula. Como alternativa a la unión covalente, una etiqueta de ácido nucleico puede incluir un receptor que se une a un ligando en la superficie celular o una etiqueta de ácido nucleico puede incluir un ligando que se une a un receptor en la superficie celular. En el Ejemplo IV a continuación se muestran métodos de unión ejemplares.

Se puede unir una biblioteca de células a una matriz usando uno o más de los métodos mostrados en la presente memoria para unir otros tipos de agentes candidatos a un soporte sólido. Por ejemplo, un soporte sólido puede incluir cebadores de ácido nucleico y las células pueden unirse al soporte sólido mediante la hibridación de etiquetas de ácido nucleico con los cebadores de ácido nucleico. En algunos casos, las etiquetas de ácido nucleico pueden incluir una secuencia de unión de cebador universal, los cebadores de ácido nucleico pueden incluir una secuencia de cebador universal y los agentes candidatos pueden unirse al soporte sólido mediante la hibridación de la secuencia de unión del cebador universal a la secuencia de cebador universal.

Un método de la presente descripción puede incluir una etapa de extensión de cebadores unidos a un soporte sólido a los que se hibridan ácidos nucleicos, tales como etiquetas de ácido nucleico. Los cebadores extendidos resultantes incluirán secuencias de etiqueta y otras secuencias de los ácidos nucleicos (aunque en forma complementaria). Los cebadores extendidos son, por tanto, versiones etiquetadas espacialmente de los ácidos nucleicos de los agentes candidatos. Se entenderá que los elementos de la secuencia, distintos de las secuencias de etiqueta, que están presentes en el ácido nucleico también pueden incluirse en los cebadores extendidos. Dichos elementos incluyen, por ejemplo, sitios de unión de cebadores, sitios de escisión, otras secuencias de etiqueta (p. ej., etiquetas de identificación de muestras), secuencias de captura, sitios de reconocimiento para proteínas de unión a ácidos nucleicos o enzimas de ácidos nucleicos, o similares.

La extensión de cebadores puede llevarse a cabo usando métodos ejemplificados en la presente memoria o conocidos de otro modo en la técnica para la amplificación de ácidos nucleicos o secuenciación de ácidos nucleicos. En realizaciones particulares, pueden añadirse uno o más nucleótidos al extremo 3' de un cebador, por ejemplo, mediante catálisis por polimerasa (p. ej., ADN polimerasa, ARN polimerasa o transcriptasa inversa). Se pueden utilizar métodos químicos o enzimáticos para añadir uno o más nucleótidos al extremo 3' o 5' de un cebador. Pueden añadirse uno o más oligonucleótidos al extremo 3' o 5' de un cebador, por ejemplo, mediante métodos químicos o enzimáticos (p. ej., catálisis por ligasa). Un cebador puede extenderse de una manera dirigida por molde, por lo que el producto de la extensión es complementario a un ácido nucleico molde que se hibrida con el cebador. En algunas realizaciones, un cebador de ADN se extiende mediante una transcriptasa inversa utilizando un molde de ARN, produciendo así un ADNc. Por tanto, una sonda extendida preparada en un método mostrado en la presente memoria puede ser una molécula de ADN transcrita de forma inversa. Los métodos ejemplares para extender ácidos nucleicos se muestran en la Publ. de Solic. de Pat. de EE. UU. No. US 2005/0037393 A1 o Pat. de EE. UU. No. 8.288.103 o 8.486.625.

Todo o parte de un ácido nucleico que se hibrida con un cebador se puede copiar por extensión. Por ejemplo, una sonda extendida puede incluir al menos 1, 2, 5, 10, 25, 50, 100, 200, 500, 1.000 o más nucleótidos que se copian a partir de un ácido nucleico. La longitud del producto de extensión se puede controlar, por ejemplo, usando nucleótidos terminados de manera reversible en la reacción de extensión y ejecutando un número limitado de ciclos de extensión. Los ciclos se pueden ejecutar como se ejemplifica para las técnicas de SBS y no es necesario el uso de nucleótidos marcados. Por consiguiente, un cebador extendido producido en un método mostrado en la presente memoria puede incluir no más de 1.000, 500, 200, 100, 50, 25, 10, 5, 2 o 1 nucleótidos que se copian a partir de un ácido nucleico. Por supuesto, las sondas extendidas pueden tener cualquier longitud dentro o fuera de los rangos mostrados anteriormente.

Se puede cribar una biblioteca de células que está unida a una matriz usando los métodos mostrados en la presente

memoria con respecto al cribado de otros agentes candidatos. En realizaciones particulares, el cribado de la matriz de células puede incluir una etapa de tratamiento de las células con un agente de cribado. Como resultado, el agente de cribado puede unirse al menos a una célula candidata en la matriz. Opcionalmente, el agente de cribado es luminiscente y la reacción de cribado se llevará a cabo detectando la luminiscencia de la al menos una célula candidata.

En algunas realizaciones, un agente de cribado que se administra a una matriz modificará al menos una célula candidata en la matriz. Por ejemplo, el agente de cribado puede estimular al menos a una célula candidata en la matriz. Alternativamente, el agente de cribado puede inhibir o incluso matar al menos a una célula candidata en la matriz.

Un agente de cribado que se pone en contacto con una matriz de células puede aumentar o disminuir la luminiscencia de al menos una célula candidata y la reacción de cribado puede incluir etapas para detectar la luminiscencia de la al menos una célula candidata.

Como se muestra en otra parte en la presente memoria, la etapa de detección puede incluir una medición cinética o basada en el tiempo. Por ejemplo, la detección de una matriz de células puede incluir una etapa de adquisición de señales en varios puntos de tiempo para una o más de las características individuales de la matriz.

Las etiquetas de ácido nucleico que se unen a las células o están presentes en las células pueden secuenciarse mediante un método mostrado en la presente memoria. Los ácidos nucleicos que están unidos a las células se pueden secuenciar en las células. Alternativamente, los ácidos nucleicos se pueden transferir a un soporte sólido, por ejemplo, mediante métodos de extensión de cebadores tales como los ejemplificados en el Ejemplo IV, y las regiones de los cebadores extendidos que contienen la secuencia de etiqueta (o complemento de la misma) se pueden secuenciar en el soporte sólido. Los métodos mostrados en otra parte en la presente memoria para secuenciar ácidos nucleicos unidos a un soporte sólido pueden usarse para etiquetas de ácido nucleico que están unidas a células o a características próximas a las células de las que se derivan las etiquetas de ácido nucleico.

En algunos casos, puede ser deseable eliminar las células del soporte sólido dejando las etiquetas de ácido nucleico (o copias de las etiquetas) unidas al soporte sólido. Las secuencias de etiqueta se pueden secuenciar luego en ausencia de las células.

Para realizaciones en las que están presentes etiquetas de ácido nucleico en una célula, las células pueden lisarse en la superficie de la matriz, lo que conduce a la liberación del contenido celular y la captura localizada de las etiquetas de ácido nucleico. Las etiquetas de ácido nucleico pueden incluir regiones de secuencia universal que son complementarias a los cebadores en el soporte sólido. Por tanto, las etiquetas de ácido nucleico se pueden capturar en el soporte sólido en un formato que sea susceptible de amplificación y/o secuenciación. Los ácidos nucleicos que se capturan en el soporte sólido pueden opcionalmente amplificarse. Los ácidos nucleicos que se capturan en el soporte sólido o los amplicones producidos a partir de los mismos pueden secuenciarse en el soporte sólido usando los métodos mostrados en la presente memoria.

Un método de la presente descripción puede incluir una etapa de retirada de una o más células de la matriz. Por ejemplo, una o más células que se identifican como aciertos en una etapa de cribado se pueden retirar de forma selectiva. Opcionalmente, se pueden usar condiciones que retengan la viabilidad de las células. Como tales, la o las células retiradas se pueden cultivar, copiar o expandir. La o las células se pueden retirar usando técnicas y reactivos mostrados en otra parte en la presente memoria para otros agentes candidatos. Por ejemplo, el uso de enlazadores fotolábiles y haces de luz filtrados espacialmente puede ser particularmente útil para aislar una célula particular de otras en la matriz.

En algunas realizaciones, no es necesario retirar las células viables de la matriz después de una etapa de cribado. Más bien, la secuencia de etiqueta puede haber sido asociada *a priori* con una célula conocida de modo que la secuenciación de la etiqueta sea suficiente para identificar la célula sin la necesidad adicional de aislar o caracterizar las células. El conocimiento *a priori* de la etiqueta presente en una característica particular puede correlacionarse con las señales observadas de la característica en una etapa de cribado para identificar la célula.

Los siguientes ejemplos pretenden ilustrar, pero no limitar la presente invención.

Ejemplo I (no reivindicado)

Cribado de moléculas pequeñas

Las plataformas de secuenciación de última generación, tales como las comercializadas por Illumina (San Diego, CA), proporcionan una base para construir una plataforma de cribado integrada de alto rendimiento para el descubrimiento de fármacos. Al modificar los ácidos nucleicos con un grupo funcional que permite la unión a las moléculas diana, es posible utilizar una celda de flujo de secuenciación como sustrato para construir matrices para el cribado de compuestos de alto rendimiento. Se puede implementar una variedad de ensayos de cribado utilizando la misma plataforma con cambios en las dianas unidas. Un método ejemplar para cribar moléculas pequeñas en una plataforma de secuenciación se esquematiza en la Fig. 1A a la Fig. 1F, y se describe a continuación.

La Fig. 1A muestra una primera etapa en la que se sintetizan 50.000 moléculas de ácido nucleico diferentes, cada una

con una longitud de 29 nucleótidos, en un instrumento de síntesis basado en cinta. La cinta incluye sitios direccionables individualmente en los que se sintetizan especies de ácido nucleico individuales, respectivamente. Las moléculas de ácido nucleico también incluyen un grupo funcional (FG) tal como un grupo sulfhidrilo, amina o N-hidroxisuccinimida en el extremo 5'. La región 5' de los ácidos nucleicos incluye 10 nucleótidos que codifican un cebador universal (que es el mismo para todos los 50.000 ácidos nucleicos) y la región 3' codifica una de 50.000 secuencias de etiqueta diferentes (también llamadas secuencias de "código"). Un ejemplo de instrumento de síntesis de ADN basado en cinta se describe en la Publ. de Solic. de Pat. de EE. UU. No. 2011/0178285 A1.

Una segunda etapa se muestra en la Fig. 1B, donde un ácido nucleico híbrido monomolde se liga a los ácidos nucleicos en cada pocillo de la cinta de síntesis para crear un cebador protuberante universal. Como se muestra en el diagrama, la ligación se produce después de la hibridación de la secuencia del cebador universal de los diferentes ácidos nucleicos con una protuberancia sobre el híbrido monomolde. Opcionalmente, los diferentes ácidos nucleicos se pueden retirar de la cinta (p. ej., para purificar o modificar los ácidos nucleicos) antes del evento de ligación. Otras configuraciones opcionales incluyen tener el grupo funcional unido al híbrido monomolde en lugar de los diferentes ácidos nucleicos. Alternativamente, o adicionalmente, el híbrido monomolde puede tener un resto de unión de superficie tal como biotina, azida o grupo alquino.

Como se muestra en la Fig. 1C, se lleva a cabo una tercera etapa para añadir robóticamente un agente candidato específico (por ejemplo, "compuesto") a cada uno de los pocillos individuales en la cinta sintetizadora. Cada uno de los agentes candidatos incluye un grupo reactivo que reacciona con el grupo funcional de los diferentes ácidos nucleicos. Como resultado, cada uno de los agentes candidatos se une a una etiqueta de ácido nucleico particular en la cinta sintetizadora, creando así agentes candidatos codificados.

La etapa 4, como se muestra en la Fig. 1D, se lleva a cabo para inmovilizar los agentes candidatos codificados en sitios distinguibles en la superficie de una celda de flujo. La inmovilización puede ocurrir mediante la hibridación de las porciones de monomolde de los ácidos nucleicos con cebadores universales complementarios que se encuentran en la superficie de la celda de flujo. Los cebadores se pueden unir a la superficie de la celda de flujo mediante gel PAZAM u otro hidrogel como se describe, por ejemplo, en la Pub. de Solic. de Pat. de EE. UU. No. 2011/0059865 A1 o Pat. de EE. UU. No. 9.012.022.

En la etapa 5, como se muestra en la Fig. 1E, un agente de cribado (también llamado una "molécula diana") se pone en contacto con los agentes candidatos codificados inmovilizados que están en la superficie de la celda de flujo. El agente de cribado incluye un marcador fluorescente que se puede detectar utilizando un dispositivo óptico tal como un instrumento de secuenciación Illumina. Pueden obtenerse imágenes, por ejemplo, en tiempo real para determinar la cinética de unión para el agente de cribado en cada uno de los sitios distinguibles. En el ejemplo de la Fig. 1E, el agente candidato unido al código N es un "acierto" según se determina por un perfil de cinética de unión deseado.

Como se muestra en la Fig. 1F, la secuencia para el código N se puede determinar basándose en un protocolo de secuenciación llevado a cabo en la celda de flujo. La secuencia del código N puede correlacionarse con la ubicación en la cinta sintetizadora donde se sintetizó y, a su vez, la identidad del "acierto" se puede determinar en función del conocimiento de qué agente candidato se administró en la ubicación en la cinta sintetizadora donde se sintetizó el código N.

Ejemplo II (no reivindicado)

Cribado de proteínas

Los métodos de amplificación en fase sólida permiten una presentación altamente multiplexada de ácidos nucleicos. Un método de amplificación en fase sólida particularmente útil es la amplificación en puente (también llamada formación de grupos) que se puede llevar a cabo como se describe, por ejemplo, en la Pat. de EE. UU. No. 5.641.658; Publ. de Solic. de Pat. de EE. UU. No. 2002/0055100; Pat. de EE. UU. No. 7.115.400; Publ. de Solic. de Pat. de EE. UU. No. 2004/0096853; Publ. de Solic. de Pat. de EE. UU. No. 2004/0002090; Publ. de Solic. de Pat. de EE. UU. No. 2007/0128624; y Publ. de Solic. de Pat. de EE. UU. No. 2008/0009420.

Este ejemplo describe el uso de los grupos de ácidos nucleicos que resultan de la amplificación en fase sólida para formar una matriz de proteínas en una celda de flujo. Algunas ventajas clave de esta técnica son (a) proporciona un gran número de características que contienen proteínas que tienen un tamaño relativamente pequeño, pero que se pueden distinguir ópticamente en una superficie de manera que se puede cribar una gran variedad de proteínas diferentes para determinar su actividad; y (b) las características contienen los ácidos nucleicos que codifican la proteína, siendo los ácidos nucleicos susceptibles de secuenciación por lo que se puede identificar la proteína en cada característica de la celda de flujo.

Como se muestra en el diagrama de la Fig. 2, se construye una biblioteca de ADNc para que tenga adaptadores con sitios de unión de cebadores P5 y P7 junto con sitios de transcripción y traducción. La biblioteca se une a una celda de flujo y los miembros de la biblioteca se amplifican en la superficie de la celda de flujo usando amplificación en puente para formar grupos. Para la descripción de las secuencias de unión del cebador P5 y P7, los métodos para hacer bibliotecas de ácidos nucleicos con adaptadores y los métodos para hacer grupos, véanse, por ejemplo, las referencias citadas anteriormente con respecto a las técnicas de amplificación en puente, Bentley et al., *Nature* 456: 53-59 (2008),

Pub. de Solic. de Pat. de EE. UU. No. 2011/0059865 A1 y Pat. de EE. UU. No. 7,741,463.

Los grupos de ADNc resultantes se transcriben en ARNm usando ARN polimerasa. El ARNm se mantiene cerca del grupo de ADNc a partir del cual se transcribe por (a) no tener un sitio de terminación de la transcripción al final de la construcción de ADNc de manera que la ARN polimerasa se detiene y el transcrito de ARNm permanece unido al ADNc, (b) el uso de sitios de pausa de transcripción conocidos (p. ej., del operón Trp de *E. coli*) que de manera similar detendrá la ARN polimerasa, (c) el empleo de métodos químicos para pausar la transcripción (p. ej., biciclomicina), o (d) tener una secuencia complementaria en el grupo de ADNc con el que hibridará el ARN transcrito.

El ARNm, que se localiza en el grupo de ADNc a partir del que se transcribió, se traduce entonces para producir proteína. Esto se verá facilitado por un codón de inicio en el ARNm (que puede introducirse junto con la secuencia de inicio de la transcripción en el adaptador en el extremo 5' de la construcción de la biblioteca de ADNc). Además, el ARNm no tendrá un codón de parada en el extremo 3'. Como se muestra en el diagrama de la Fig.2, la falta de un codón de parada da lugar a un bloqueo del ribosoma al final del ARNm, de modo que la proteína permanece unida al ARNm a través del ribosoma (el ARNm, a su vez, se une al grupo de ADNc por una ARN polimerasa bloqueada). Alternativamente, los ribosomas se pueden bloquear por medios químicos (por ejemplo, cloranfenicol/puromicina). La transcripción y traducción acopladas se pueden realizar usando lisado de reticulocitos de conejo, S30 bacteriano o sistemas de extracto de germen de trigo fácilmente disponibles.

En un enfoque alternativo, las proteínas se pueden generar en disolución y luego unirlas a una celda de flujo u otra superficie. Aunque este enfoque no incluye la amplificación en la superficie, puede ser beneficioso cuando las técnicas basadas en disolución están más fácilmente disponibles o son aplicables a la proteína que se está evaluando. La matriz resultante de proteínas se criba con un agente de cribado u otro estímulo para identificar los aciertos. El ADNc de los grupos se secuencia para identificar los aciertos de proteínas.

Ejemplo III (no reivindicado)

Cribado de proteínas usando etiquetas

Las plataformas de secuenciación de última generación, tales como las comercializadas por Illumina (San Diego, CA) proporcionan una base para construir una plataforma de cribado integrada de alto rendimiento para la evolución de proteínas. En la Fig. 3A y la Fig. 3B se muestra un diagrama de un método ejemplar.

Como primera etapa, se crea una biblioteca de vectores que tiene mutaciones aleatorias o semi-aleatorias en una proteína diana de interés y etiquetas aleatorias (también llamadas "códigos"). Los vectores se construyen para expresar un ARNm que incluye la secuencia codificante de la proteína y la etiqueta. El número de etiquetas aleatorias diferentes supera el número de variantes de proteínas en la biblioteca, de modo que cada ARNm variante tendrá una alta probabilidad de incluir una etiqueta única. Por ejemplo, una secuencia aleatoria de 10 nucleótidos proporcionará $10^6$ etiquetas diferentes y una secuencia aleatoria de 20 nucleótidos proporcionará $10^{12}$ etiquetas diferentes.

Alternativamente a la introducción aleatoria de mutaciones y al uso de etiquetas aleatorias, la primera etapa se puede realizar de manera que se realicen mutaciones predefinidas y se asocien con etiquetas conocidas.

En una segunda etapa, se cultiva una biblioteca de bacterias que se han transformado con los vectores para producir ARNm a partir de los vectores respectivos. La biblioteca se lisa para liberar una mezcla de transcritos de ARNm de los diversos vectores de la biblioteca (es decir, una biblioteca de ARNm).

En una tercera etapa, la biblioteca de ARNm se divide para crear dos sub-bibliotecas. La biblioteca incluye múltiples copias de cada miembro, de modo que la división de la biblioteca da como resultado que los mismos miembros estén presentes en ambas sub-bibliotecas. De nuevo, se une una secuencia de etiqueta única a cada secuencia de variante de proteína.

En una cuarta etapa, como se muestra en la Fig. 3B, la primera sub-biblioteca se une a una celda de flujo, se amplifica en la celda de flujo, y se secuencian las secuencias codificantes y las etiquetas respectivas de los amplicones resultantes. De esta forma, se ubica el sitio donde está unido cada miembro. A continuación, los miembros de la segunda sub-biblioteca se ponen en contacto con la celda de flujo para permitir que los miembros de la segunda sub-biblioteca se hibriden con los amplicones generados a partir de los miembros idénticos de la primera sub-biblioteca mediante la complementariedad de las secuencias de etiqueta. De esta manera, el sitio donde se encuentra cada miembro de la segunda sub-biblioteca se puede inferir en función de los resultados de la secuenciación y la hibridación esperada entre las etiquetas y sus complementos. Los miembros de la segunda sub-biblioteca, que se han capturado en la celda de flujo, se traducen entonces para expresar las proteínas diana en la celda de flujo. Cada una de las proteínas se expresa de manera que permanece unida al ARN a partir del que se tradujo. Por ejemplo, la unión se puede lograr usando las técnicas mostradas en el Ejemplo II. A continuación, las proteínas se criban con un agente de cribado u otro estímulo que permite identificar una función deseada con un alto rendimiento. El resultado es una señal de actividad de la proteína que se localiza en un sitio en la celda de flujo que se correlaciona con la secuencia de ARN que codifica la proteína.

Esta metodología puede permitir un alto rendimiento de 30 millones de proteínas por carril de celda de flujo o más.

Este proceso de cribado evita los desafíos asociados con el cribado basado en emulsión. Por ejemplo, se pueden obtener datos cuantitativos y basados en cinética en cada característica de la celda de flujo utilizando un instrumento de secuenciación Illumina para permitir un criterio de cribado más refinado y selectivo que el disponible utilizando técnicas tradicionales basadas en emulsión.

Ejemplo IV

Cribado celular

Este ejemplo describe una metodología para el cribado de células que proporciona las ventajas de alto rendimiento de la clasificación de células activadas por fluorescencia (FACS) con la capacidad de seguir procesos dinámicos en tiempo real, una ventaja típicamente asociada con la microscopía de fluorescencia, al tiempo que se añaden las ventajas de la secuenciación de última generación. para identificar células individuales en la población. Como resultado, los métodos mostrados en la presente memoria permiten la recuperación y caracterización de células que muestran un comportamiento deseado. Específicamente, este ejemplo adapta una plataforma de secuenciación de última generación de Illumina Inc. (San Diego, CA) para monitorizar por fluorescencia, el comportamiento fenotípico (p. ej., medido por la expresión de indicadores fluorescentes) de células individuales en respuesta a estímulos específicos con un rendimiento ultra alto, seguido de la secuenciación de etiquetas que se muestran en la superficie de cada célula individual para su posterior identificación y recuperación. Esta plataforma puede proporcionar un método estándar para el cribado de alto rendimiento para la ingeniería celular.

Las células tienen capacidades terapéuticas potenciales que son diferentes y, en muchos casos, una mejora, de las de las moléculas pequeñas y las proteínas. Las células pueden moverse activamente a ubicaciones específicas del cuerpo, pueden detectar múltiples estímulos externos y pueden integrar múltiples fuentes de información para responder con resultados precisos. Además, las respuestas basadas en células pueden tener patrones dinámicos complejos que se pueden controlar mediante funciones preprogramadas o mediante la adición de factores externos. Actualmente, se están explorando terapias basadas en células para el tratamiento de una amplia gama de enfermedades, desde el uso de microorganismos diseñados por ingeniería para tratar la enfermedad de Crohn hasta el uso de células inmunes humanas diseñadas por ingeniería derivadas de pacientes para tratar el cáncer. Es importante destacar que se pueden diseñar por ingeniería microorganismos o células humanas para que sean terapéuticamente útiles.

Flujo de trabajo para la ingeniería celular

La ingeniería celular (ya sea en células microbianas o humanas) es un proceso complejo con muchas variables a considerar. Por lo tanto, en lugar de intentar diseños únicos, es beneficioso diseñar bibliotecas de células candidatas que luego se pueden seleccionar o cribar. Específicamente, la ingeniería celular puede incluir las siguientes etapas: (1) diseño de circuitos genéticos que pueden incluir uno o más sensores para señales externas, una red de integración de información y una respuesta dependiente de la señal (la respuesta a menudo implica expresión génica); (2) modificación de una o más dianas genéticas para los componentes mencionados anteriormente a través de una biblioteca de células candidatas (p. ej., usando CRISPR-Cas9); y (3) cribado de alto rendimiento para identificar aciertos, entre las numerosas células candidatas diferentes, que confieren comportamientos celulares deseables.

Una técnica de cribado robusta y de alto rendimiento puede proporcionar ventajas significativas para la ingeniería celular, ya que la cantidad de células que deben ensayarse es grande (a menudo se ensayan miles de diseños diferentes) y una gama amplia y variada de comportamientos celulares se manifestará debido a las modificaciones introducidas en el proceso de ingeniería. Un cribado deseable también será capaz de seguir dinámicas temporales complejas, mientras logra un rendimiento muy alto. El cribado mostrado en la presente memoria proporciona estas ventajas.

El método de cribado celular tiene las siguientes dos etapas:

Etapa 1: Monitorización fenotípica. Se carga una biblioteca de células vivas diseñadas por ingeniería con indicadores de fluorescencia adecuados en una celda de flujo para una plataforma de secuenciación (p. ej., plataforma Illumina MiSeq®, NextSeq®, HiSeq® o Genome Analyzer®), como se muestra en la Fig. 4. Las células individuales se distribuirán aleatoriamente y se unirán a la superficie de la celda de flujo (utilizando químicas de unión específicas dependiendo de las células particulares que se van a analizar). La fluorescencia de cada célula individual se registra a intervalos de tiempo definidos, escaneando la superficie de la celda de flujo como se hace actualmente para las celdas de flujo de Illumina durante la secuenciación. Obsérvese que las células individuales tendrán tamaños promedio comparables a (o más grandes que) los grupos de secuenciación actuales, por lo que la resolución de la plataforma de secuenciación se adapta bien al cribado. Además, las células pueden portar miles de moléculas indicadoras fluorescentes, de modo que las intensidades de señal serán similares (o mayores) que las detectadas actualmente por las plataformas de secuenciación. En tiempos específicos, las células se exponen a agentes de cribado (u otros estímulos) y se detecta la respuesta de las células (p. ej., como se refleja en cambios en la expresión de genes indicadores o cambios en la intensidad de la fluorescencia de cualquier otro tipo de indicador celular - tal como la concentración de calcio en el citoplasma, extinción de fluoróforos por cambios en la localización subcelular, etc.). Al monitorizar los cambios en los indicadores fluorescentes a lo largo del tiempo, será posible determinar la dinámica temporal de los cambios fenotípicos en respuesta a diversos estímulos (o combinaciones de estímulos).

Etapa 2: Identificación de una sola célula mediante la decodificación de etiquetas individuales. A cada célula de la población se le asignará una etiqueta. A continuación, se describen tres técnicas diferentes para etiquetar células. En una primera etapa, las etiquetas se unen a células individuales. Independientemente del tipo de etiqueta (p. ej., perla, ácido nucleico unido a la membrana, etc.), las células se clasifican primero en recipientes individuales (p. ej., micropocillos) de modo que una célula se encuentre en cada recipiente. Después, se proporciona una etiqueta única a la célula en cada recipiente, como se muestra en la Fig.5.

2.1 Código de barras de célula única con perlas. Cada célula de la población (p. ej., ~10 µm de diámetro para las células T) se etiquetan con 1 a 10 perlas. Las perlas pueden tener un diámetro de aproximadamente 1-2 µm. Las perlas útiles incluyen, por ejemplo, las que se utilizan en la tecnología BeadArray de Illumina o se describen en las Pat. de EE. UU. Nos. 7.622.294 o 8.741.630. Cada célula se etiquetará con perlas que codifican una sola etiqueta, aislando previamente las células individuales en recipientes individuales y añadiendo a cada recipiente perlas que comparten la misma etiqueta única (es decir, habrá una correspondencia conocida entre cada recipiente y cada etiqueta). El etiquetado celular podría lograrse de varias formas, por ejemplo, uniendo covalentemente las perlas con anticuerpos que reconocen epítopos específicos presentes de forma natural en la superficie de las células diana. Obsérvese que, de esta manera, aunque cada célula etiquetada con perlas se distribuirá aleatoriamente en la superficie de la celda de flujo, cada célula se colocalizará espacialmente con su o sus propias perlas (Fig. 6). Por tanto, después de la Etapa 1, la identidad de cada célula se revelará decodificando la o las perlas en la celda de flujo y asignando etiquetas identificadas a la célula más cercana a la o las perlas correspondientes. Las etiquetas de las perlas se pueden decodificar usando una técnica de secuenciación por síntesis (véase, por ejemplo, Bentley et al., *Nature* 456: 53-59 (2008), Publ. de PCT Nos. WO 91/06678, WO 04/018497 o WO 07/123744; Pat de EE. UU. Nos. 7.057.026, 7.329.492, 7.211.414, 7.315.019 o 7.405.281 y Publ. de Solic. de Pat. de EE. UU. No. 2008/0108082 A1) o una técnica de decodificación mostrada en la Pat. de EE. UU. No. 8.460.865 o en Gunderson et al., *Genome Research* 14: 870-877 (2004).

2.2 Identificación de una sola célula con etiquetas de secuencia. Una alternativa o adición al etiquetado de perlas es el uso de identificadores de secuencia únicos que se muestran directamente en la superficie de cada celda. Para eso, cada célula de la población se modificará con etiquetas de ADN monocatenario individuales que se muestran en la superficie antes de la carga celular en la celda de flujo (p. ej., antes de la Etapa 1). La química de unión de la etiqueta de ácido nucleico dependerá del tipo de célula particular. Por ejemplo, las células humanas pueden etiquetarse con moléculas de ADN monocatenario unidas covalentemente a los lípidos de la membrana plasmática, según lo informado por Sleden et al., *J Am Chem Soc* 134: 765-8 (2012). Como otro ejemplo, las células de levadura pueden etiquetarse con moléculas de ADN monocatenario unidas covalentemente a proteínas HaloTag mostradas en los complejos de pared celular Aga1-Aga2.

Cada etiqueta individual se compone de al menos tres regiones. La región 1 está conectada a un resto de unión (p. ej., lípido o proteína, dependiendo del tipo de célula). La región 2 consiste en un número definido de bases que depende de la plexidad requerida (usando los 4 nucleótidos es posible generar 4 combinaciones para una secuencia de longitud "n"; p. ej. una secuencia de 4 bases de longitud abarcará etiquetas para la identificación de 256 células individuales diferentes, 5 darán como resultado 1.024 etiquetas diferentes, etc.). Por tanto, con tan solo una secuencia de etiqueta de 9 bases, será posible identificar individualmente más de 250.000 células diferentes. La región 3 consiste en una secuencia de ADN complementaria a un cebador de secuenciación. Obsérvese que cada célula se etiquetará con ácidos nucleicos que codifican una sola etiqueta, aislando previamente las células individuales en los recipientes y añadiendo a cada recipiente ácidos nucleicos que comparten la misma secuencia de etiqueta única (como antes, habrá una correspondencia conocida entre cada recipiente y cada etiqueta).

En una etapa separada (realizada antes o después de que se hayan obtenido las mediciones fenotípicas), se determinará la identidad de cada célula individual, por ejemplo, de una de las dos formas que se describen a continuación.

2.2.1 Secuenciación in situ de la etiqueta particular unida a cada célula. Cada célula mostrará varias copias de una etiqueta en particular. Por tanto, la señal es fácilmente detectable en un instrumento de secuenciación (véase la Fig. 7). Además, la longitud de lectura para identificar cada etiqueta es relativamente corta (p. ej., 9 bases son suficientes para identificar >250.000 etiquetas) dentro de las longitudes de lectura de cientos de nucleótidos típicas de los instrumentos de secuenciación disponibles comercialmente.

La secuenciación de etiquetas in situ se puede lograr de varias formas. Una posibilidad es secuenciar directamente en la superficie de células vivas. Otra posibilidad es fijar las células antes de la secuenciación (en el caso en el que los nucleótidos que se escapan de las células vivas interfieran con la reacción de secuenciación). Una tercera posibilidad es separar y capturar las etiquetas de ácido nucleico de cada célula. Para ello, la superficie de la celda de flujo se modifica con una sonda de captura (p. ej., un cebador de secuenciación) que contiene una región complementaria a una región común a todos los ácidos nucleicos unidos a las células, lo que permite la unión inicial de las células a la superficie y facilita la captura de los ácidos nucleicos desprendidos antes de la secuenciación. Opcionalmente, las células se pueden retirar después de la captura superficial de las etiquetas de ácido nucleico.

2.2.2. Secuenciación in situ de copias de etiquetas unidas a la celda de flujo. Como se muestra en la Fig. 8, las etiquetas de ácido nucleico que se muestran en la superficie celular se hibridan con los cebadores unidos a la superficie de la celda de flujo (con secuencias complementarias parciales). Las etiquetas de ácido nucleico de la superficie

celular se separan de las células mediante digestión con una enzima de restricción específica, y luego las células se eliminan por lavado de la celda de flujo. Las etiquetas de ácido nucleico unidas a la superficie de la celda de flujo se extenderán, creando de manera efectiva copias de las etiquetas de la célula original, que luego se secuenciarán. En muchos casos, no será necesario ampliar las copias en puente, ya que es probable que una célula deje una "mancha" de aproximadamente 1 µm de diámetro compuesta por miles de ácidos nucleicos. Sin embargo, se puede llevar a cabo la amplificación en puente si se desea (en cuyo caso se pueden añadir sitios de unión de cebadores apropiados a las etiquetas de ácido nucleico y se puede unir un segundo tipo de cebador a la superficie de la celda de flujo).

En algunos casos, se puede unir una etiqueta a una célula individual utilizando un ácido nucleico anclado a la membrana, como se muestra en la Fig. 9. El ácido nucleico anclado a la membrana incluirá una región enlazadora, un sitio de restricción (RE), un sitio de unión del cebador de secuenciación (SBS3), secuencia de etiquetas (código de barras) y secuencia de captura. Las células pueden luego ser capturadas en la superficie por ácidos nucleicos unidos a la superficie con una secuencia complementaria a la secuencia de captura, como se describe en Sleden et al., *J Am Chem Soc* 134: 765-8 (2012) o como se muestra en la Fig.10.

Después del registro de las señales de fluorescencia indicativas del comportamiento fenotípico celular en una etapa de cribado (p. ej., midiendo la fluorescencia de un informador durante un período de tiempo y en condiciones de medio específicas), las etiquetas de ácido nucleico se pueden copiar en la superficie de la celda de flujo y luego las células pueden eliminarse por lavado. Las copias en la superficie preservan efectivamente la información espacial con respecto a qué células se ubicaron en la superficie de la celda de flujo, como se muestra en la Fig.11.

Al registrar la ubicación espacial de cada célula en la superficie, será posible asignar identidades de etiqueta a cada una de las series de tiempo de fluorescencia registradas, conectando así de manera efectiva las respuestas fenotípicas con la identidad celular. Estas identidades se pueden usar para identificar clones individuales en recipientes individuales (donde las células se etiquetaron inicialmente, como se muestra en la Fig.12), o para recuperar células seleccionadas de mezclas de células sedimentando con ácidos nucleicos que son complementarios a las etiquetas presentes en la superficie de las células seleccionadas.

El término "que comprende" está destinado en la presente memoria a tener extremos abiertos, que incluye no solo los elementos enumerados, sino que además abarca cualquier elemento adicional.

Aunque la invención se ha descrito con referencia a los ejemplos proporcionados anteriormente, debe entenderse que se pueden realizar varias modificaciones sin apartarse de la invención. Por consiguiente, la invención está limitada únicamente por las reivindicaciones.

**REIVINDICACIONES**

1. Un método de cribado de células, que comprende

(a) proporcionar una pluralidad de células diferentes, en donde cada una de las células diferentes comprende una etiqueta de ácido nucleico que tiene una secuencia de etiqueta, estando unida la etiqueta de ácido nucleico a cada célula de la pluralidad de células mediante unión covalente a un lípido de la membrana plasmática, un ácido graso de la membrana plasmática, o una proteína en un lípido de la membrana plasmática de la célula, en donde cada secuencia de etiqueta comprende una secuencia única;

(b) poner en contacto una mezcla de las células diferentes con un soporte sólido para formar una matriz de células unida al soporte sólido;

(c) cribar la matriz de células en el soporte sólido con relación a al menos una característica óptica, en donde la reacción de cribado comprende detectar las células individuales que están unidas al soporte sólido;

(d) secuenciar las secuencias de etiqueta de etiquetas de ácido nucleico que están unidas al soporte sólido; y

(e) identificar al menos una célula en la matriz como una célula candidata basada en la característica óptica y la secuencia de etiqueta de la célula candidata.

2. El método de la reivindicación 1, en donde la etapa (a) comprende separar las células diferentes en recipientes separados y añadir un ácido nucleico que tiene una secuencia de etiqueta a cada uno de los recipientes, proporcionando de ese modo una pluralidad de células diferentes, en donde cada una de las células diferentes está en un recipiente separado y comprende una etiqueta de ácido nucleico que tiene una secuencia de etiqueta distintiva.

3. El método de la reivindicación 1, en donde el ácido nucleico que tiene una secuencia de etiqueta está unido a una perla y la perla se une a una célula de la pluralidad de células diferentes, opcionalmente en donde la perla comprende un anticuerpo que tiene afinidad de unión específica por la célula.

4. El método de la reivindicación 1, en donde el ácido nucleico que tiene una secuencia de etiqueta está unido a una célula de la pluralidad de células diferentes mediante unión covalente a un lípido o ácido graso de la membrana plasmática.

5. El método de la reivindicación 1, en donde el ácido nucleico que tiene una secuencia de etiqueta está unido a una célula de la pluralidad de células diferentes mediante unión covalente a una proteína en un lípido de la membrana plasmática.

6. El método de la reivindicación 1, en donde el soporte sólido comprende cebadores de ácido nucleico y las células se unen al soporte sólido mediante hibridación de las etiquetas de ácido nucleico a los cebadores de ácido nucleico, opcionalmente en donde las etiquetas de ácido nucleico comprenden una secuencia de unión a cebador universal, los cebadores de ácido nucleico comprenden una secuencia de cebador universal, y los agentes candidatos se unen al soporte sólido mediante hibridación de la secuencia de unión del cebador universal a la secuencia de cebador universal.

7. El método de la reivindicación 1, en donde la detección de la matriz comprende adquirir señales en distintos puntos de tiempo para una o más de las características individuales en la matriz.

8. El método de la reivindicación 1, que además comprende amplificar las etiquetas de ácido nucleico para producir amplicones de las etiquetas de ácido nucleico en el soporte sólido, opcionalmente en donde la secuenciación de la etiqueta de ácido nucleico comprende secuenciar los amplicones de las etiquetas de ácido nucleico.

9. El método de la reivindicación 1, en donde el soporte sólido está localizado dentro de una celda de flujo.

10. El método de la reivindicación 9, que además comprende eliminar la al menos una célula candidata del soporte sólido y opcionalmente que además comprende cultivar la al menos una célula candidata después de eliminar el soporte sólido, replicando de ese modo la al menos una célula.

11. El método de la reivindicación 1, en donde el cribado de la matriz de células comprende tratar las células con un agente de cribado.

12. El método de la reivindicación 11, en donde el agente de cribado se une a la al menos una célula candidata.

13. El método de la reivindicación 11, en donde el agente de cribado modifica o estimula la al menos una célula candidata, opcionalmente en donde el agente de cribado aumenta o disminuye la luminiscencia de la al menos una célula candidata y en donde la reacción de cribado comprende detectar la luminiscencia de la al menos una célula candidata.

14. El método de la reivindicación 1, que además comprende amplificar una etiqueta de ácido nucleico que tiene una secuencia de etiqueta para unir copias múltiples de la etiqueta de ácido nucleico en cada una de las células diferentes.

15. El método de la reivindicación 1, en donde (d) comprende:

(i) eliminar las células individuales del soporte sólido en condiciones en donde las etiquetas de ácido nucleico están unidas al soporte sólido y después secuenciar las secuencias de etiquetas de ácido nucleico que están unidas al soporte sólido; o

(ii) copiar las etiquetas de ácido nucleico para producir copias de etiquetas de ácido nucleico que están unidas al soporte sólido y después secuenciar las secuenciar las secuencias de etiqueta de las copias de etiqueta de ácido nucleico que están unidas al soporte sólido; o

(iii) copiar las etiquetas de ácido nucleico para producir copias de etiquetas de ácido nucleico que están unidas al soporte sólido y después eliminar las células individuales del Soporte sólido en condiciones en donde las copias de etiqueta de ácido nucleico se unen al soporte sólido y después secuenciar las secuencias de etiqueta de etiquetas de ácido nucleico que están unidas al soporte sólido.

Etapa 1: Sintetizar 500.000 oligos 29-mer + grupo de unión (-SH, amina; NHS).

1. Sintetizador = 50.000 oligos/día 500.000 oligos en 10 días

2. Código de barras = 19-mer; "Cebador" = 10-mer

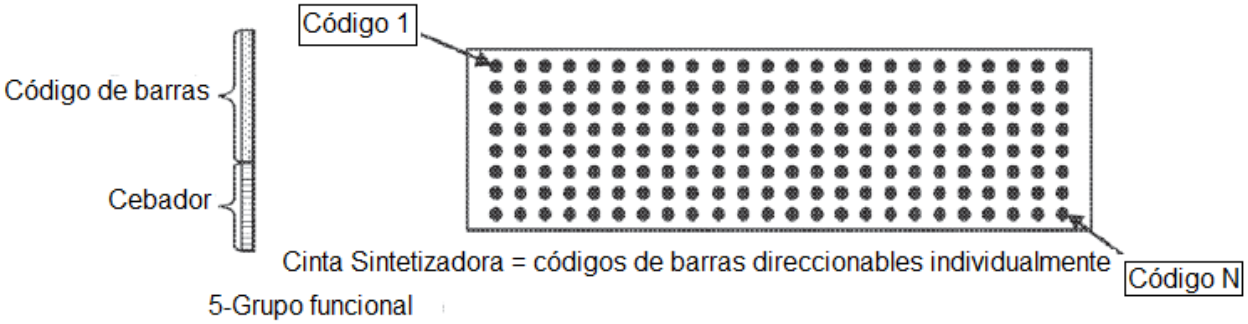



Fig. 1A

Etapa 2: Ligar el cebador protuberante con secuencia monomolde en pocillos individuales

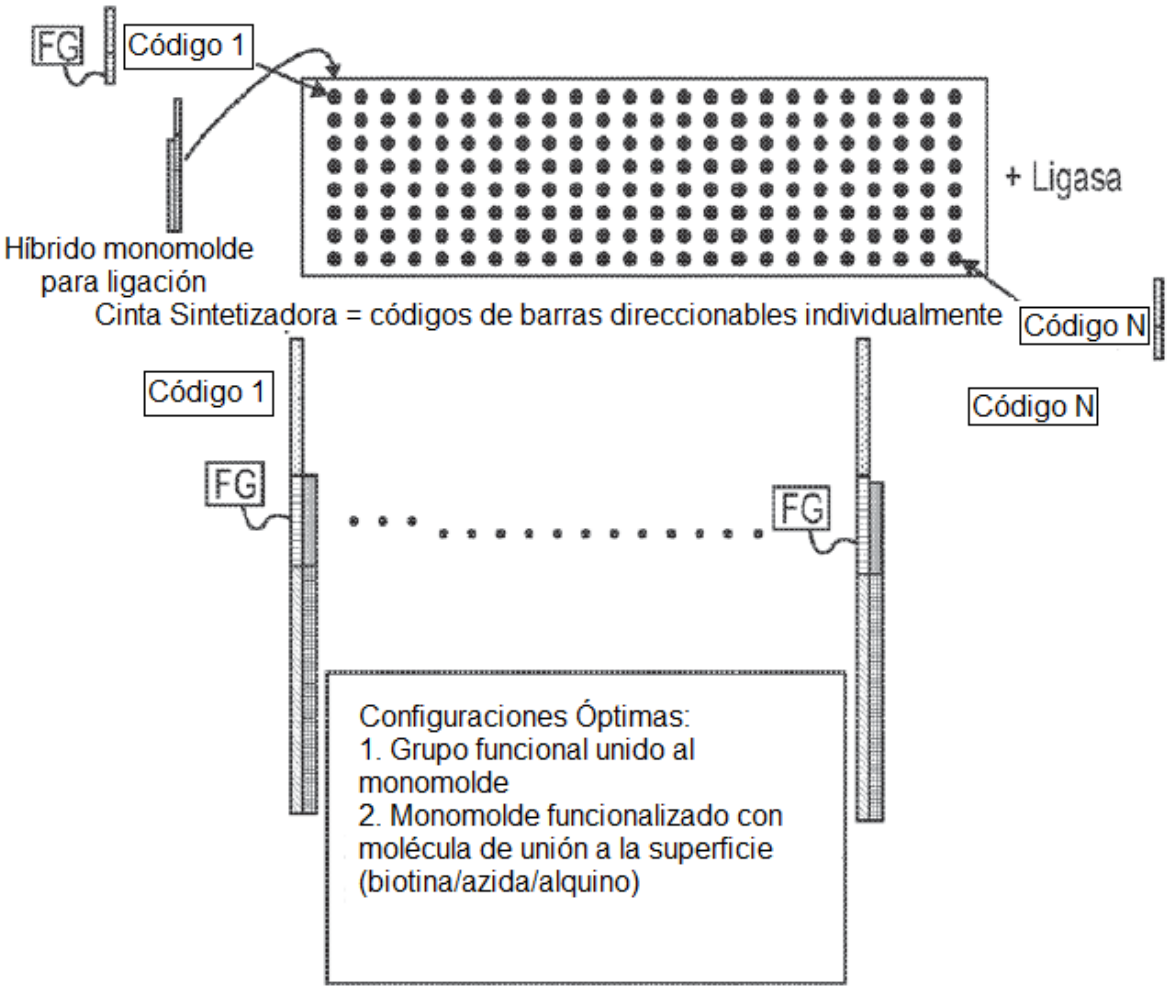



Fig. 1B

Etapa 3: Añadir robóticamente compuestos con grupo reactivo (RG) a pocillos individuales con códigos funcionalizados

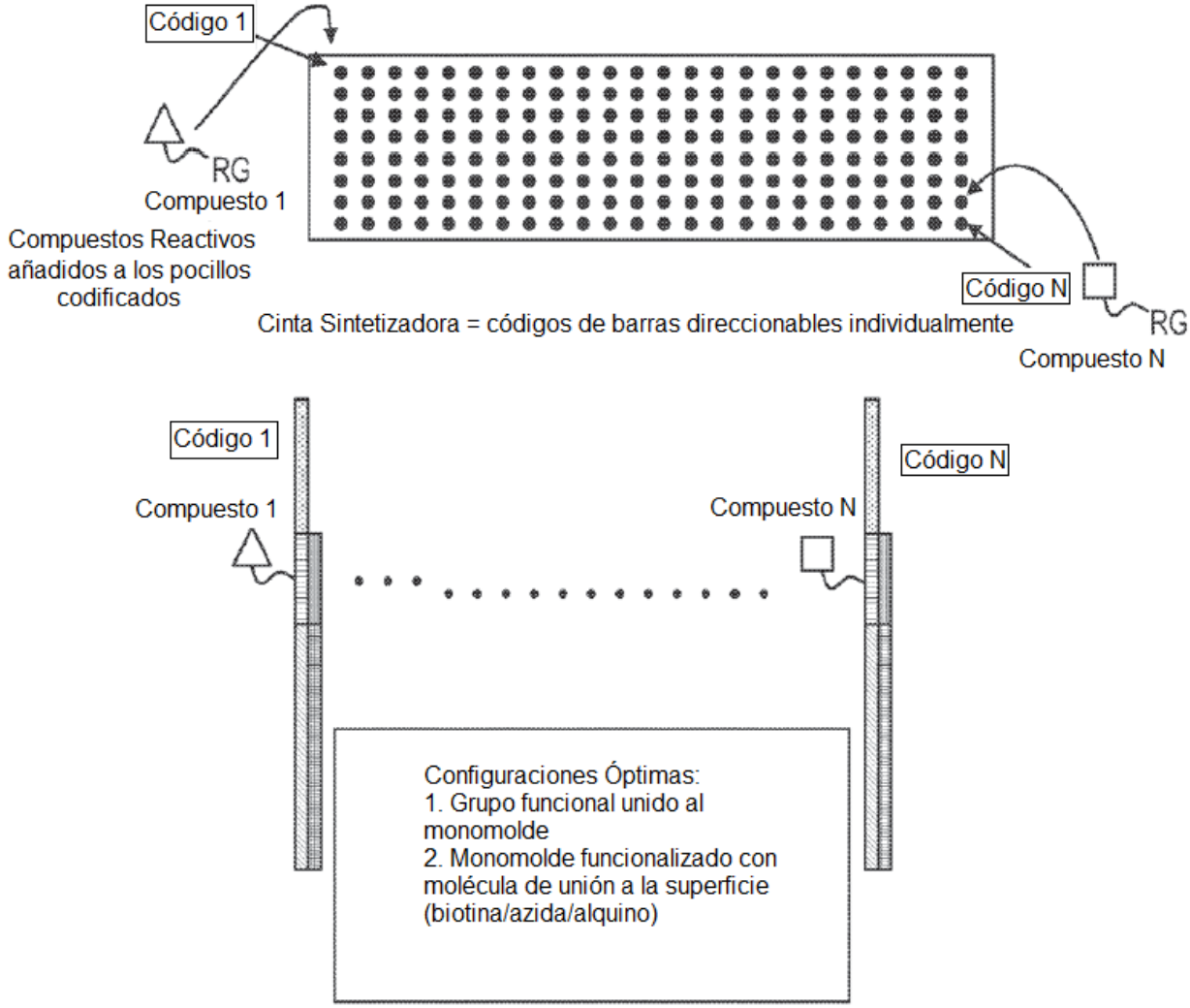



Fig. 1C

Etapa 4: Inmovilizar compuestos mediante hibridación o grupo funcional superficial a superficie de molécula única (o superficie de celda de flujo Pazam)

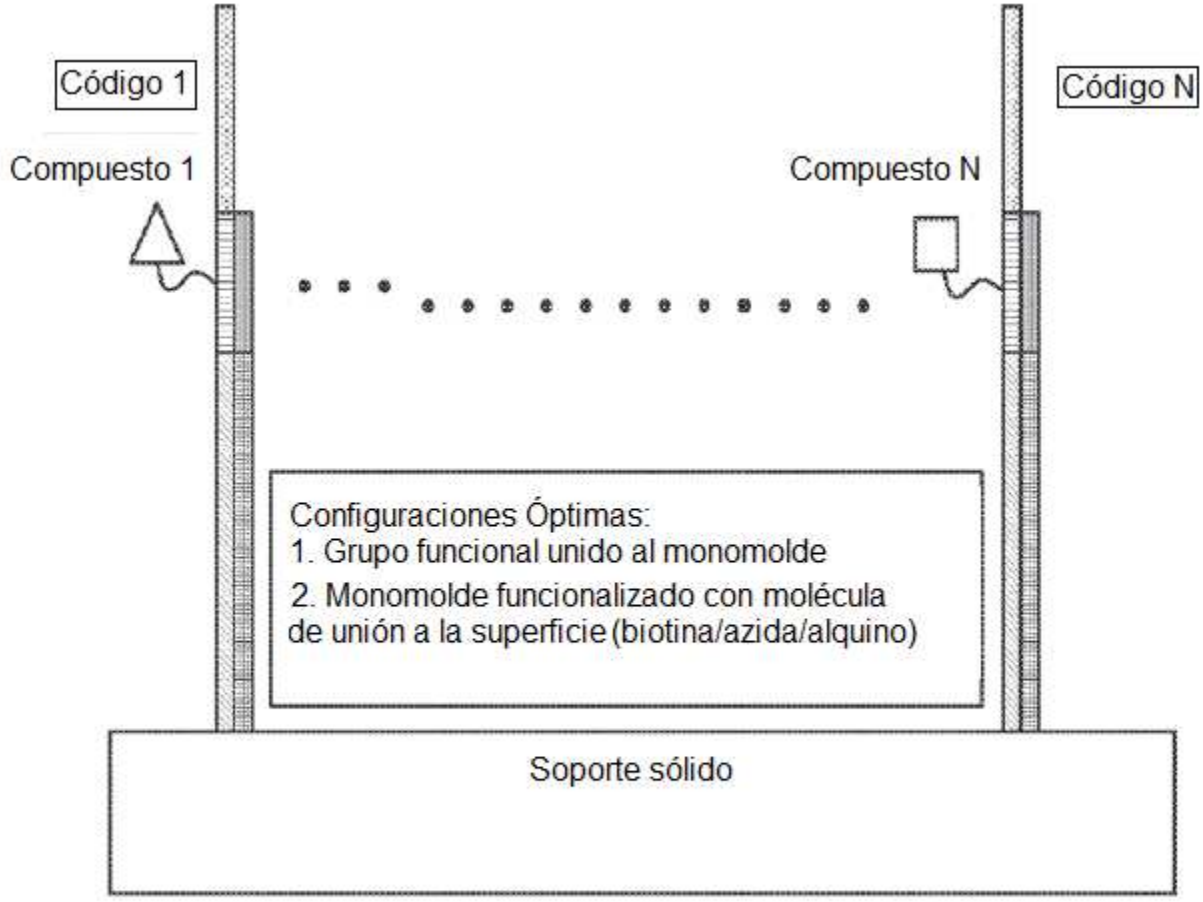



Fig. 1D

Etapa 5: Flujo en moléculas diana y registro de imágenes en tiempo real de la cinética de unión a compuestos.

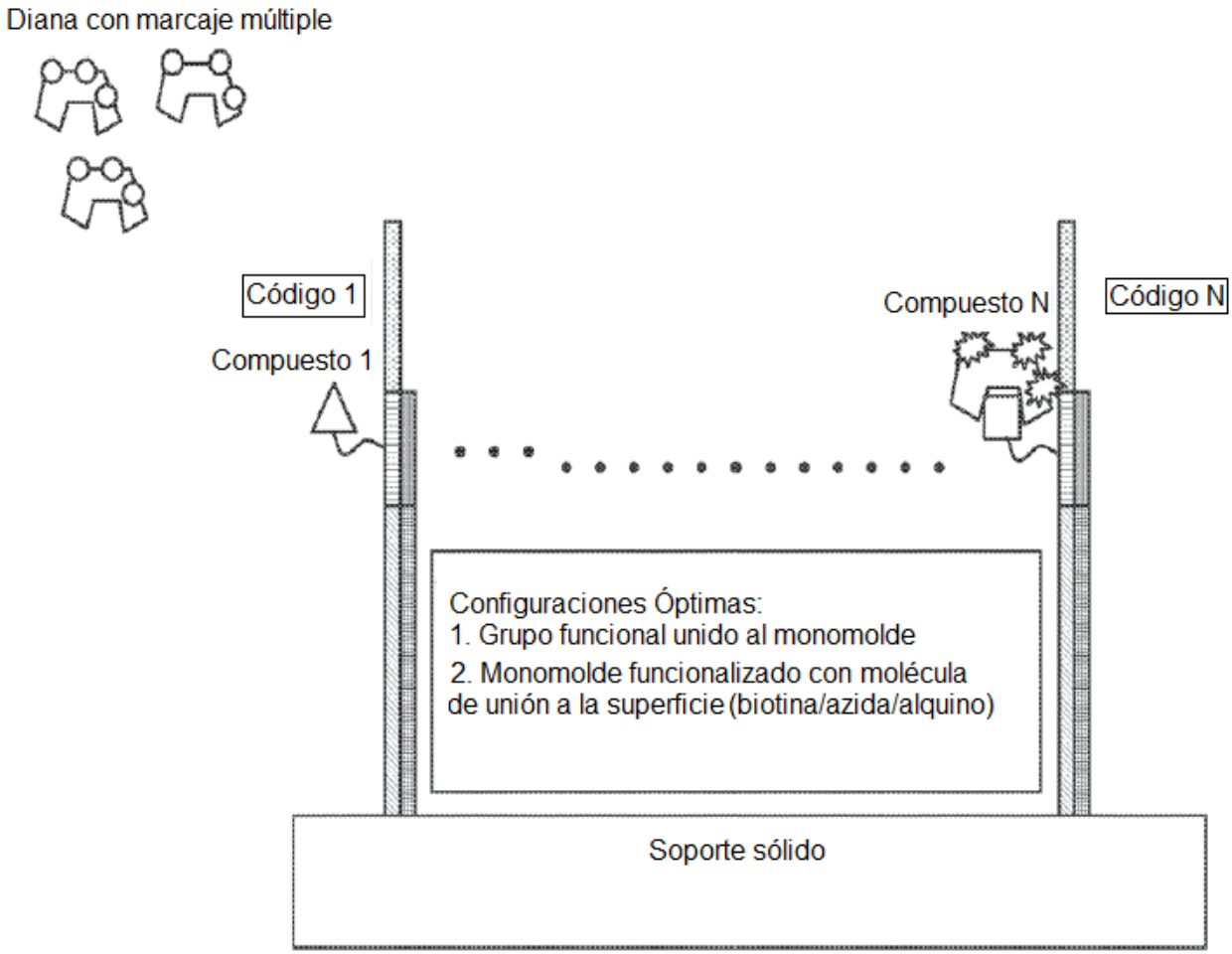



Fig. 1E

Etapa 6: Moléculas codificadas por secuencia para identificar "pistas" o "aciertos".

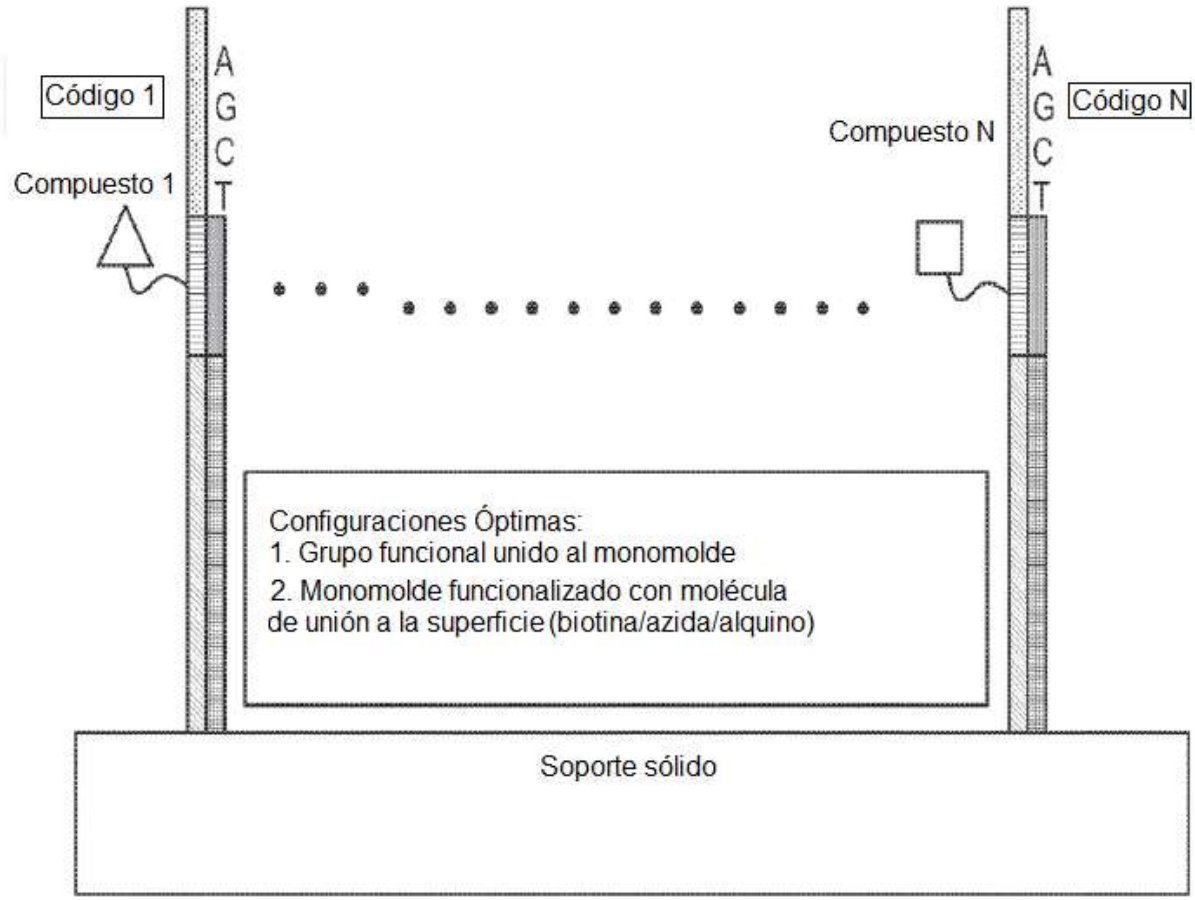



Fig. 1F

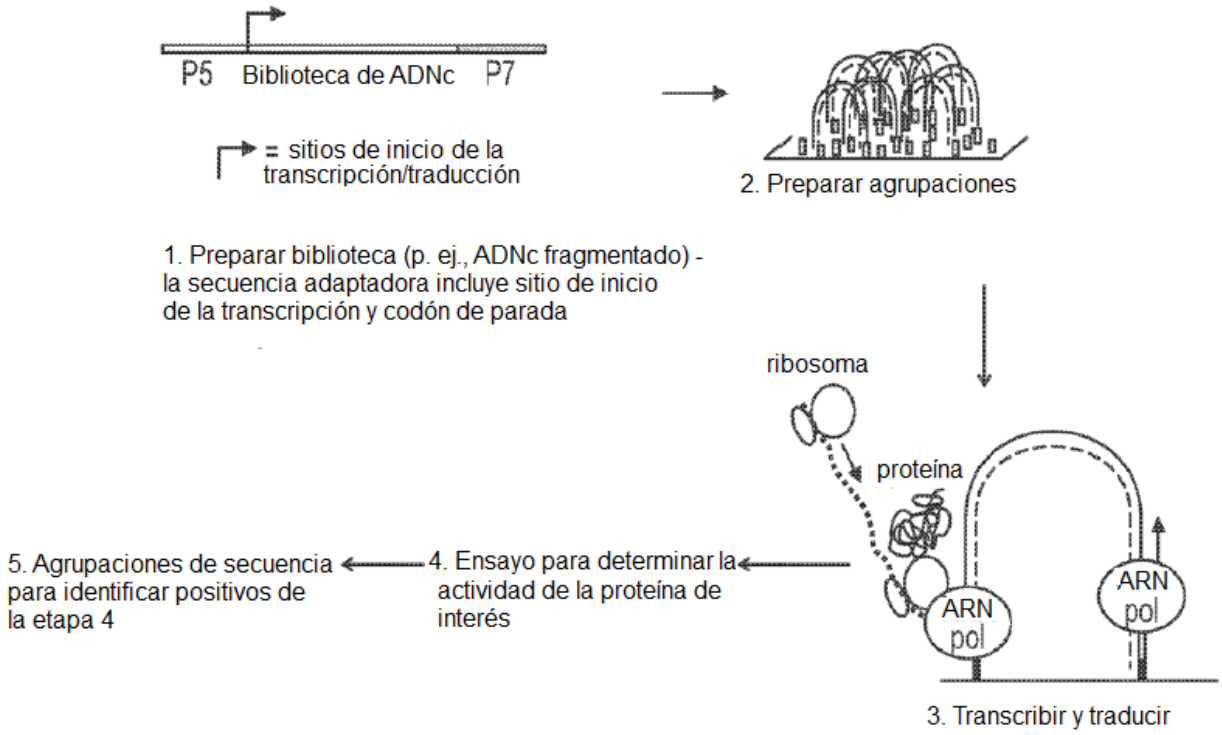
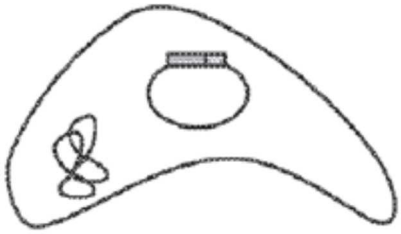
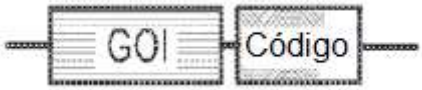

P5 Biblioteca de ADNc P7
= sitios de inicio de la transcripción/traducción
2. Preparar agrupaciones
1. Preparar biblioteca (p. ej., ADNc fragmentado) - la secuencia adaptadora incluye sitio de inicio de la transcripción y codón de parada
ribosoma
proteína
ARN pol
ARN pol
3. Transcribir y traducir
4. Ensayo para determinar la actividad de la proteína de interés
5. Agrupaciones de secuencia para identificar positivos de la etapa 4

Fig. 2

1. Crear una biblioteca bacteriana con transcritos de ARNm + índice (índice: 10 pb = $10^6$ códigos; 20 pb = $10^{12}$ códigos)



2. Lisar las bacterias - recoger los transcritos de ARNm




3. 1/2 muestra → La secuenciación determina los índices para mutaciones.

Mutación $Y_1$ = Código $X_1$

Mutación $Y_2$ = Código $X_2$

Mutación $Y_N$ = Código $X_N$

4. 1/2 muestra → en traducción celda de flujo/ensayo

Fig. 3A

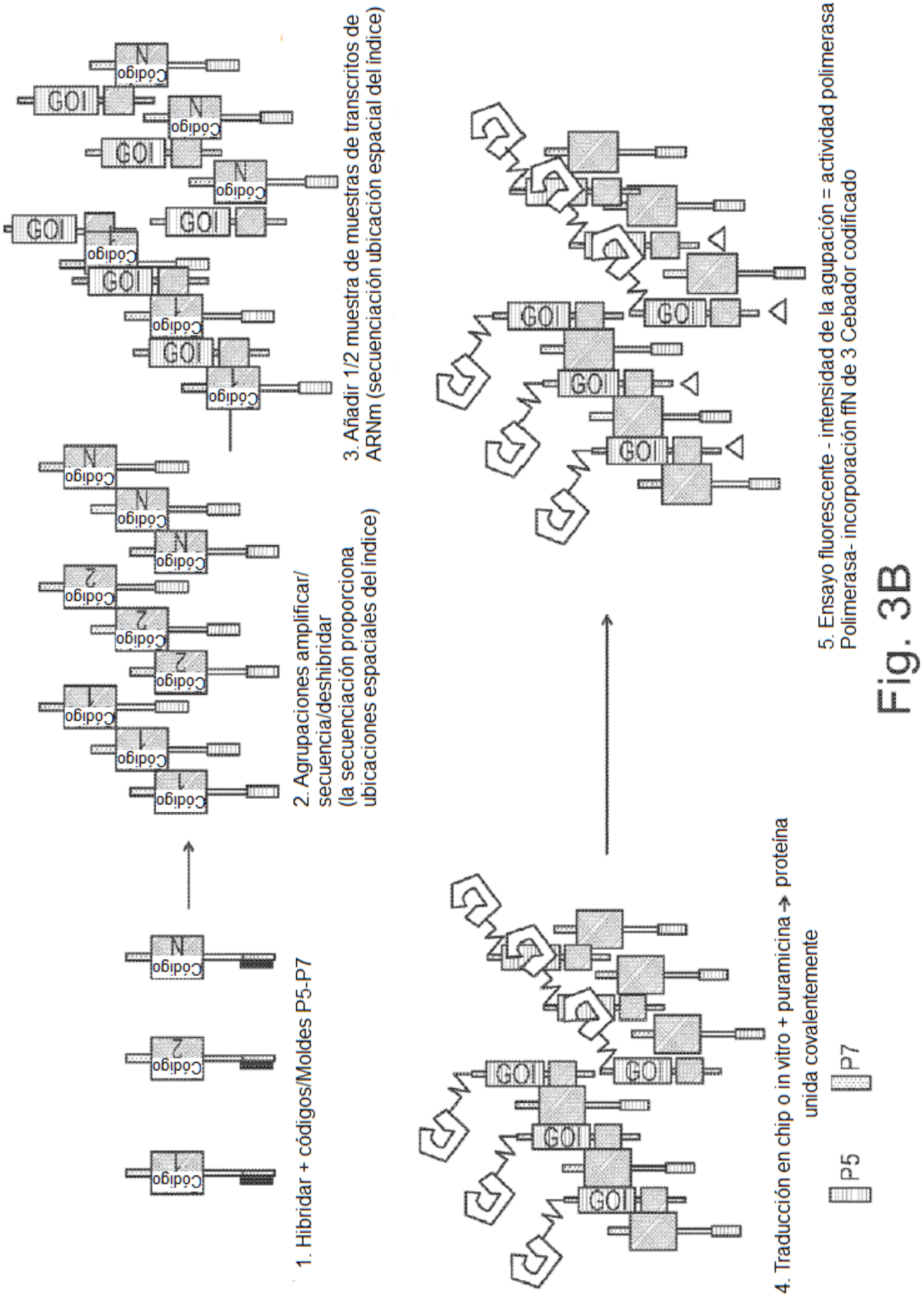

1. Hibridar + códigos/Moldes P5-P7
2. Agrupaciones amplificar/ secuencia/deshibridar (la secuenciación proporciona ubicaciones espaciales del índice)
3. Añadir 1/2 muestra de muestras de transcritos de ARNm (secuenciación ubicación espacial del índice)
4. Traducción en chip o in vitro + puramicina → proteína unida covalentemente
5. Ensayo fluorescente - intensidad de la agupación = actividad polimerasa Polimerasa- incorporación ffN de 3 Cebador codificado
P5
P7
GOI
Código
N
1
2

Fig. 3B

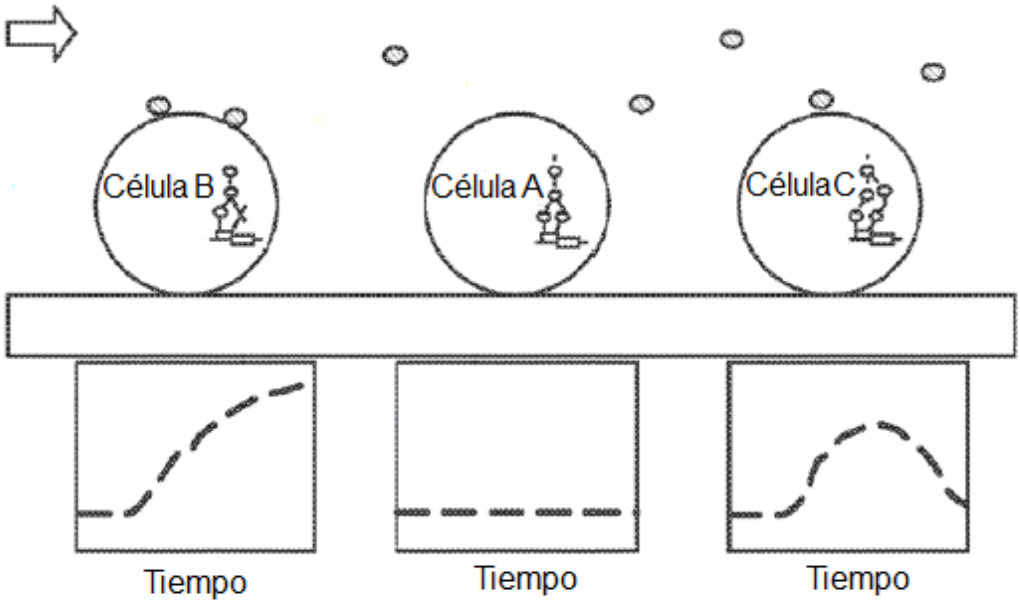

Célula B
Célula A
Célula C
Tiempo
Tiempo
Tiempo

Fig. 4

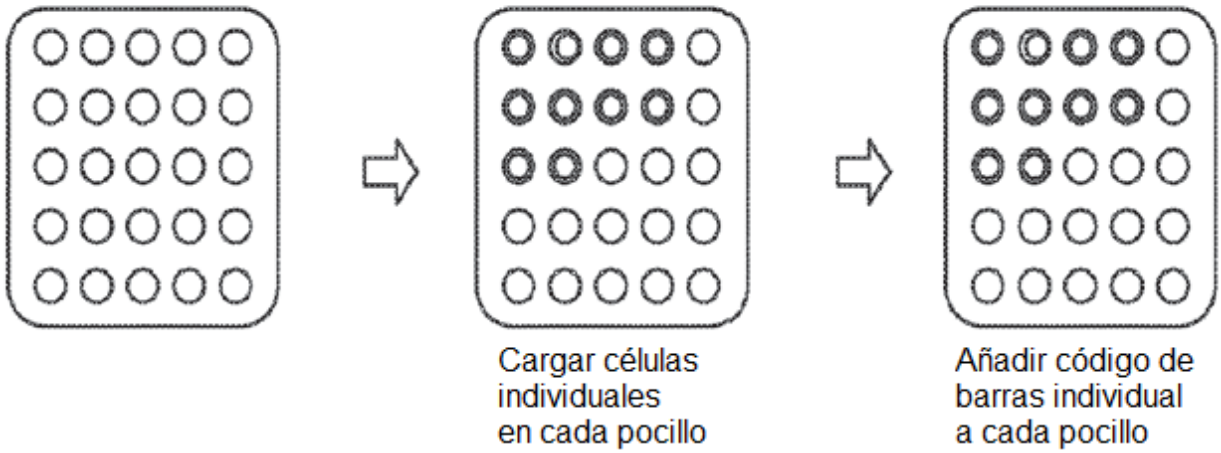

Cargar células individuales en cada pocillo
Añadir código de barras individual a cada pocillo

Fig. 5

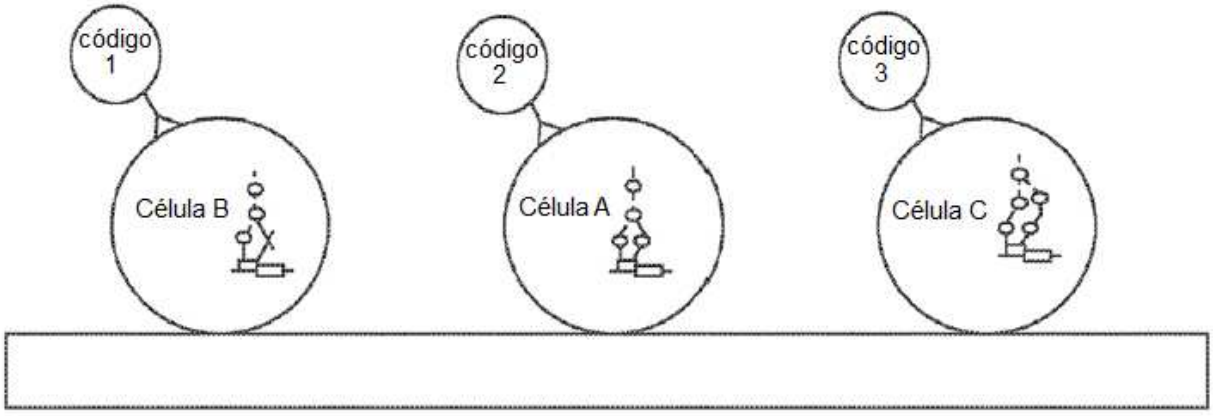

código 1
código 2
código 3
Célula B
Célula A
Célula C

Fig. 6

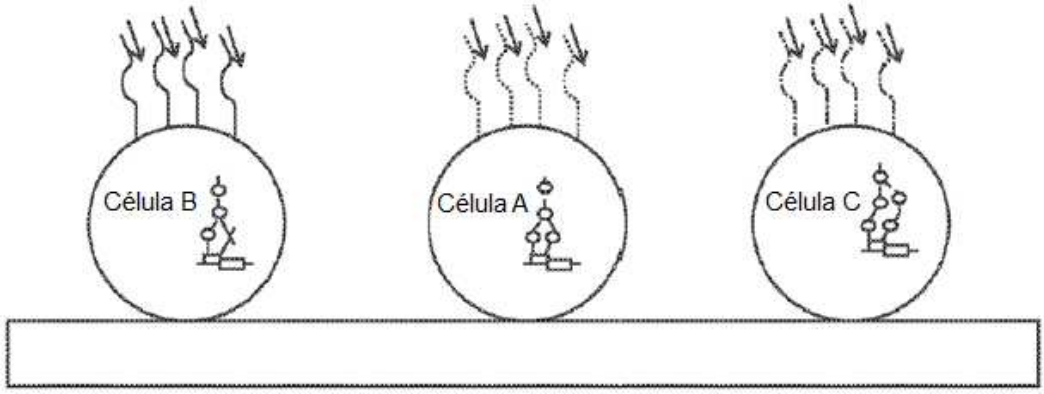

Célula B
Célula A
Célula C

Fig. 7

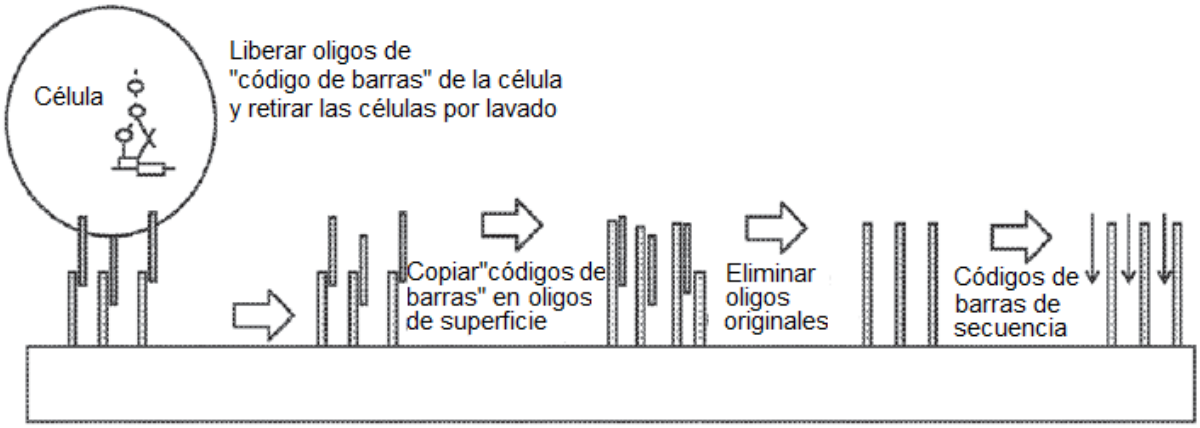

Célula
Liberar oligos de
"código de barras" de la célula
y retirar las células por lavado
Copiar"códigos de
barras" en oligos
de superficie
Eliminar
oligos
originales
Códigos de
barras de
secuencia

Fig. 8

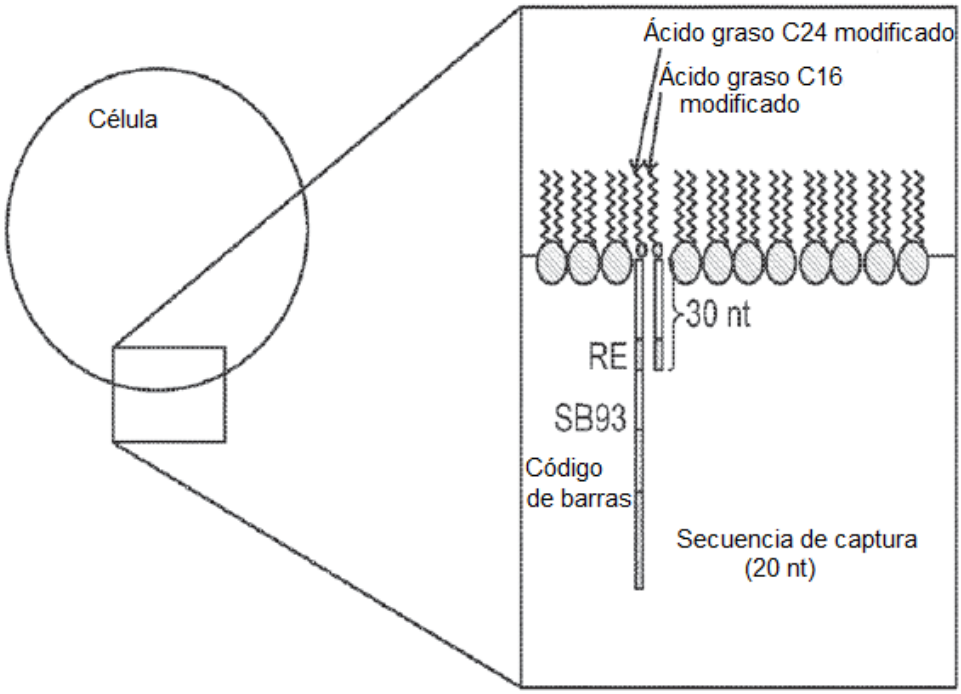

Célula
Ácido graso C24 modificado
Ácido graso C16
modificado
30 nt
RE
SB93
Código
de barras
Secuencia de captura
(20 nt)

Fig. 9

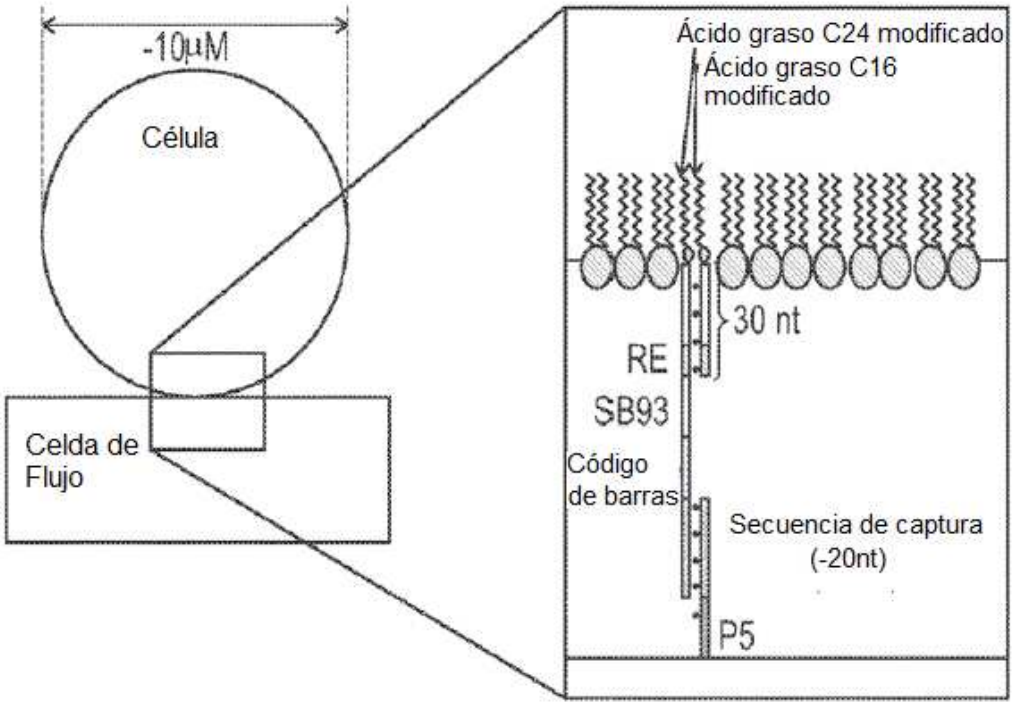

-10µM
Célula
Celda de Flujo
Ácido graso C24 modificado
Ácido graso C16 modificado
30 nt
RE
SB93
Código de barras
Secuencia de captura (-20nt)
P5

Fig. 10

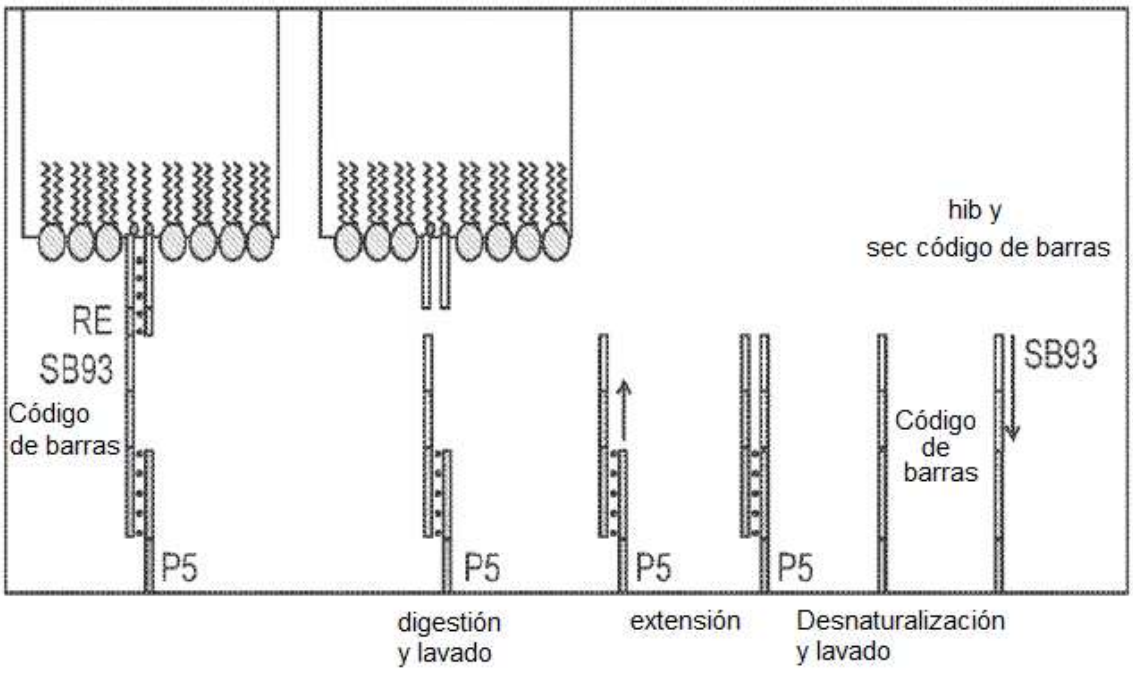

RE
SB93
Código de barras
P5
P5
P5
P5
hib y sec código de barras
SB93
Código de barras
digestión y lavado
extensión
Desnaturalización y lavado

Fig. 11

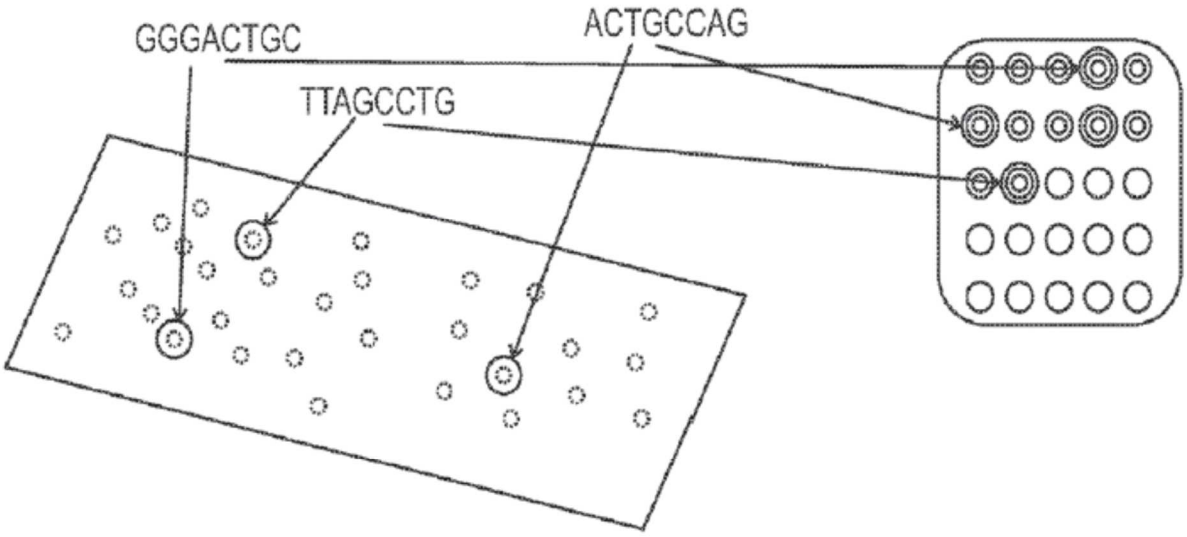



Fig. 12