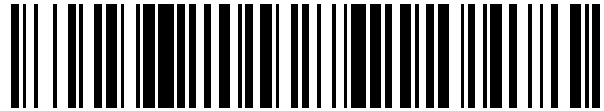


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 555 106**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.04.2011 E 11766613 (1)**

97 Fecha y número de publicación de la concesión europea: **02.09.2015 EP 2556171**

54 Título: **Ensayos biológicos codificados espacialmente**

30 Prioridad:

05.04.2011 US 201113080616
05.04.2010 US 321124 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.12.2015

73 Titular/es:

PROGNOSYS BIOSCIENCES, INC. (100.0%)
505 Coast Blvd. South
La Jolla, CA 92037, US

72 Inventor/es:

CHEE, MARK S.

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 555 106 T3

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

DESCRIPCIÓN

Ensayos biológicos codificados espacialmente

5 **Campo de la invención**

La presente invención se refiere a ensayos de moléculas biológicas, y más particularmente a ensayos para determinar las distribuciones espaciales de un gran número de moléculas biológicas en una muestra sólida de manera simultánea.

10 **Antecedentes de la invención**

En la presente discusión se describirán determinados artículos y métodos con la finalidad introductoria y de información de antecedentes. Nada contenido en el presente documento debe entenderse como una "admisión" de la técnica anterior. El solicitante se reserva el derecho de demostrar, en los casos donde sea adecuado, que los artículos y métodos citados en el presente documento no constituyen técnica anterior según las disposiciones legales aplicables.

El análisis exhaustivo de la expresión génica y el análisis de proteínas han sido herramientas útiles para entender los mecanismos de la biología. El uso de estas herramientas ha permitido la identificación de genes y proteínas implicadas en el desarrollo y en varias enfermedades, tales como el cáncer y enfermedades autoinmunes. Los métodos convencionales, tales como la hibridación *in situ* y otras detecciones multicanal de diferentes transcritos han revelado patrones espaciales de la expresión génica y han ayudado a arrojar luz acerca de la base molecular del desarrollo y las enfermedades. Otras tecnologías que han posibilitado el análisis cuantitativo de muchas secuencias de ARN por muestra incluyen micromatrices (véase Shi, et al., Nature Biotechnology, 24(9):1151-61 (2006); y Slonim y Yanai, Plos Computational Biology, 5(10):e1000543 (2009)); análisis en serie de la expresión génica (SAGE) (véase Velculescu, et al., Science, 270(5235):484-87 (1995)), implementación de alto rendimiento de la qPCR (véase Spurgeon, et al., Plos ONE, 3(2):e1662 (2008), y PCR *in situ* (véase Nuovo, Genome Res., 4:151-67 (1995)). Aunque estos métodos son útiles, sin embargo, no permiten la medida simultánea de la expresión de muchos genes o la presencia y/o actividad de múltiples proteínas en muchas localizaciones espaciales en una muestra. La microdissección de captura láser ha permitido el análisis de muchos genes en un pequeño número de localizaciones, pero es muy extensivo, trabajoso y no se escala bien. Determinados ensayos PCR en formato 2D conservan información espacial (véase Armani, et al., Lab on a Chip, 9(24): 3526-34 (2009)), pero estos métodos tienen una baja resolución espacial debido a que se basan en la transferencia física del tejido a los pocillos, lo que también evita el acceso al azar a muestras de tejido y altos niveles de multiplexación.

Actualmente no existen métodos prácticos para analizar a alta resolución los patrones de expresión espacial de grandes números de genes, proteínas, u otras moléculas biológicamente activas de manera simultánea. Por lo tanto hay una necesidad de mapas reproducibles, de alta resolución espacial de moléculas biológicas en tejidos. La presente invención aborda esta necesidad.

El documento WO 2010/019826 describe la detección y cuantificación de moléculas diana individuales en muestras biomoleculares usando sondas nanoindicadoras estables con una señal única detectable.

MIR KALIM et al., NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, R.U., vol. 37, n.º 1, 1 de enero de 2009 (01/01/2009), páginas e5-1, describen la secuenciación mediante ligación y escisión cíclica (CycliC) directamente sobre un molde capturado en una micromatriz.

Sumario de la invención

Se proporciona este sumario para introducir una selección de conceptos de manera simplificada que se describen adicionalmente después en la descripción detallada. Este sumario no tiene por objeto identificar características clave o esenciales de la materia objeto reivindicada, ni tampoco se prevé su uso para limitar el alcance de la materia objeto reivindicada. Otras características, detalles, utilidades y ventajas de la materia objeto reivindicada serán evidentes a partir de la siguiente descripción detallada escrita que incluye aquellos aspectos ilustrados en los dibujos adjuntos y definidos en las reivindicaciones adjuntas.

La divulgación se refiere a sistemas de ensayo que proporcionan mapas espaciales de alta resolución de la actividad biológica en tejidos. El sistema de ensayo comprende un ensayo capaz de altos niveles de multiplexación donde se proporcionan sondas codificadas para una muestra biológica en patrones espaciales definidos; instrumentación capaz de administrar de manera controlada el reactivo de acuerdo con los patrones espaciales; y un esquema de descodificación que proporciona una lectura que es de naturaleza digital. En pocas palabras, la presente invención proporciona la capacidad para observar muchas dianas biológicas en muchas localizaciones, proporcionando la resolución de la hibridación *in situ* con el análisis altamente paralelo de datos de la secuenciación.

Por lo tanto, la invención proporciona un método de acuerdo con la reivindicación 1 y un método de acuerdo con la reivindicación 2.

En aspectos particulares de la invención, las dianas biológicas comprenden ácidos nucleicos y las sondas codificadas son oligonucleótidos, y en algunos aspectos, hay dos sondas codificadas para cada una de las múltiples dianas de ácido nucleico. En algunos aspectos, las múltiples dianas biológicas comprenden proteínas, las regiones de sonda de las sondas codificantes son proteínas y los marcadores codificantes comprenden oligonucleótidos. En algunos aspectos, las múltiples dianas biológicas comprenden enzimas. En algunos aspectos las regiones de sonda de las sondas codificadas comprenden anticuerpos, aptámeros o moléculas pequeñas.

Algunos aspectos del sistema de ensayo comprenden además una etapa de amplificación entre la etapa de separación y la etapa de determinación. En algunos aspectos, la etapa de determinación se lleva a cabo mediante secuenciación de ácidos nucleicos, y en aspectos preferidos, la secuenciación es secuenciación de ácido nucleico digital de alto rendimiento.

En algunos aspectos de la invención, el producto de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra es mayor de 20, en algunos aspectos, el producto de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra es mayor de 50, en algunos aspectos, el producto de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra es mayor de 75, 100, 150, 500, 750, 1.000, 5.000, 10.000, 25.000, 50.000, 100.000, 500.000, o 1.000.000 o más. En otros aspectos, se determinan en paralelo las secuencias de al menos cincuenta mil sondas codificantes, en otros aspectos se determinan en paralelo las secuencias de al menos cien mil sondas codificantes, en algunos aspectos se determinan en paralelo las secuencias de al menos quinientas mil sondas codificantes, y en algunos aspectos se determinan en paralelo las secuencias de al menos un millón, diez millones, cien millones, mil millones, diez mil millones, cien mil millones o más sondas codificantes.

En algunos aspectos, el patrón espacial conocido se determina mediante características histológicas de la muestra. También en algunos aspectos, un equipo programado por un programa informático efectúa al menos dos etapas de la etapa de administración, la etapa de separación, la etapa de determinación y la etapa de asociación.

En algunos aspectos, las regiones de sonda de las sondas codificadas son proteínas y la etapa de separación se lleva a cabo mediante sondas codificadas que interactúan con las dianas biológicas que se estén capturando mediante un agente de captura por afinidad. En algunos aspectos, las regiones de sonda de las sondas codificantes son ácidos nucleicos y la etapa de separación se lleva a cabo mediante un lavado de la muestra.

La divulgación también se refiere a un sistema de ensayo para determinar patrones espaciales de abundancia o actividad o ambas de múltiples dianas de proteína en múltiples sitios en una muestra, donde el sistema de ensayo lleva a cabo las siguientes etapas: proporcionar una muestra fijada a un soporte; administrar sondas codificadas para las múltiples dianas proteicas a los múltiples sitios en la muestra con un patrón espacial conocido, donde cada sonda codificada comprende una región de sonda de proteína que puede interactuar con las dianas de proteína y un marcador de codificación que identifica una ubicación del sitio al que se administró la sonda y de la región de sonda de proteína de la sonda codificante de la que forma parte el marcador codificante; permitir que las sondas codificadas interactúen con las dianas de proteína; separar las sondas codificadas que interactúan con las dianas de proteína de sondas codificadas que no interactúan con las dianas de proteína; determinar la totalidad o parte de una secuencia de las sondas codificadas mediante secuenciación de alto rendimiento, y asociar la abundancia o actividad o ambas de las múltiples dianas de proteína a las localizaciones de los múltiples sitios en la muestra.

La divulgación también se refiere a un sistema de ensayo para determinar patrones espaciales de abundancia o actividad o ambas de múltiples dianas biológicas en múltiples sitios en una muestra, donde el sistema de ensayo lleva a cabo las siguientes etapas: proporcionar una muestra fijada a un soporte; administrar sondas codificadas para las múltiples dianas biológicas a los múltiples sitios en la muestra con un patrón espacial conocido, donde cada sonda codificada comprende una región de sonda que puede interactuar con las dianas biológicas y un marcador de codificación que identifica una localización del sitio al que se administró la sonda codificada e identifica a la diana biológica; permitir que las sondas codificadas interactúen con las dianas biológicas; determinar la totalidad o parte de una secuencia de las sondas codificadas, y asociar la abundancia o actividad o ambas de las múltiples dianas biológicas a las localizaciones de los sitios en la muestra.

La invención puede utilizar diversos mecanismos de detección, basándose en las moléculas que se van a detectar y en los reactivos necesarios para dicho sistema de detección. Los métodos ejemplares que pueden usarse con la invención se describen en más detalle a continuación.

Descripción de las figuras

La figura 1 proporciona una visión general simplificada del sistema de ensayo usando la presente.
 La figura 2 proporciona una visión general simplificada de una realización de los sistemas de ensayo usando la presente invención para detectar ácidos nucleicos.
 La figura 3 es una ilustración representativa de una realización del ensayo visto de manera general en la figura 2.
 La figura 4 ilustra un mecanismo general para una realización de un esquema de codificación combinatoria de los sistemas de ensayo usando la invención.

La figura 5 proporciona un ejemplo simplificado específico de la realización de un esquema de codificación combinatoria mostrado en la figura 4.

Definiciones

5 Los términos y expresiones usados en el presente documento pretenden tener el significado sencillo y convencional entendido por los expertos habituales en la materia. Las siguientes definiciones pretenden ayudar al lector en la comprensión de la presente invención, pero no pretenden variar o de otro modo limitar el significado de dichos términos, a menos que se indique específicamente.

10 El término "anticuerpo", tal como se usa en el presente documento, pretende hacer referencia a una inmunoglobulina o anticuerpo completo o a cualquier fragmento funcional de una molécula de inmunoglobulina que sea capaz de unirse específicamente a un antígeno (los anticuerpos y antígenos son "compañeros de unión", tal como se definen en el presente documento). Un "anticuerpo", tal como se usa en el presente documento, pretende incluir al anticuerpo completo así como a cualquier fragmento de anticuerpo capaz de unirse al antígeno o fragmento antigénico de interés. 15 Los ejemplos de dichos péptidos incluyen moléculas de anticuerpo completas, fragmentos de anticuerpo, tales como Fab, F(ab')₂, CDR, VL, VH, y cualquier otra porción de un anticuerpo que sea capaz de unirse específicamente a un antígeno. Los anticuerpos para los ensayos de la invención son inmunorreactivos o inmunoespecíficos para, Y Por lo tanto se unen específicamente y de manera selectiva a, proteínas ya sean detectadas (es decir, dianas biológicas) o 20 usadas para la detección (es decir, sondas) en los ensayos de la invención.

La expresión "agente de unión", tal como se usa en el presente documento, se refiere a cualquier agente que se una específicamente a una molécula biológica de interés.

25 "Complementaria" o "sustancialmente complementaria" se refiere a la hibridación o emparejamiento de bases o a la formación de un dúplex entre nucleótidos o ácidos nucleicos, tales como, por ejemplo, entre las dos hebras de una molécula de ADN bicatenario o entre un cebador oligonucleotídico y un sitio de unión a cebador en un ácido nucleico monocatenario. Los nucleótidos complementarios son, en general, A y T (o A y U), o C y G. Se dice que dos moléculas monocatenarias de ARN o ADN son sustancialmente complementarias cuando los nucleótidos de una hebra, 30 alineados y comparados de manera óptima con inserciones o deleciones de nucleótidos adecuadas, se emparejan con al menos aproximadamente el 80 % de la otra hebra, normalmente con al menos de aproximadamente el 90 % a aproximadamente el 95 %, e incluso con de aproximadamente un 98 % a aproximadamente un 100 %.

La "hibridación" se refiere al proceso mediante el cual dos polinucleótidos monocatenarios se unen de manera no covalente para formar un polinucleótido bicatenario estable. El polinucleótido resultante (normalmente) bicatenario es un "híbrido" o un "dúplex". Las "condiciones de hibridación" incluirán normalmente concentraciones de sal de aproximadamente menos de 1 M, a menudo de menos de aproximadamente 500 mM y pueden ser menores de aproximadamente 200 mM. Un "tampón de hibridación" es una solución salina tamponada, tal como SSPE al 5 %, u otros tampones similares conocidos en la técnica. Las temperaturas de hibridación pueden ser tan bajas como 5 °C, 40 pero normalmente son mayores de 22 °C, y más normalmente mayores de aproximadamente 30 °C, y normalmente superiores a 37 °C. A menudo, las hibridaciones se llevan a cabo en condiciones rigurosas, es decir, condiciones en las que un cebador hibridará con su subsecuencia diana pero no hibridará con las otras secuencias no complementarias. Las condiciones rigurosas son dependientes de secuencia y son diferentes en circunstancias distintas. Por ejemplo, los fragmentos más largos pueden necesitar mayores temperaturas de hibridación para una 45 hibridación específica que los fragmentos cortos. Ya que otros factores pueden afectar a la rigurosidad de la hibridación, incluyendo la composición de la base y la longitud de las hebras complementarias, la presencia de disolventes orgánicos, y el alcance del emparejamiento erróneo de bases, la combinación de parámetros es más importante que la medición absoluta de cualquier otro parámetro de manera individual. En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5 °C más bajas que la T_f para la secuencia específica a una fuerza iónica y pH definidos. Las condiciones rigurosas ejemplares incluyen una concentración de sal de al menos 0,01 M a no más de 1 M de concentración de ión sodio (o de otra sal) a un pH de aproximadamente 7,0 a aproximadamente 8,3 y una temperatura de al menos 25 °C. Por ejemplo, condiciones de SSPE 5x (NaCl 750 mM, fosfato de sodio 50 mM, EDTA 5 mM a pH 7,4) y una temperatura de aproximadamente 30 °C son adecuadas para hibridaciones específicas de alelos, aunque una temperatura adecuada depende de la longitud y/o contenido de GC de la región 50 hibridada.

"Ligadura" significa formar un enlace o engarce covalente entre los extremos de dos o más ácidos nucleicos, por ejemplo, oligonucleótidos y/o polinucleótidos, en una reacción dirigida por molde. La naturaleza del enlace o engarce puede variar ampliamente y la ligadura puede llevarse a cabo enzimática o químicamente. Tal como se usa en el presente documento, las ligaduras se llevan a cabo normalmente de manera enzimática para formar un engarce 60 fosfodiéster entre un nucleótido terminal de carbono 5' de un nucleótido con un carbono 3' de otro nucleótido.

"Ácido nucleico", "oligonucleótido", "oligo" o los equivalentes gramaticales usados en el presente documento se refieren generalmente a al menos dos nucleótidos unidos covalentemente entre sí. Un ácido nucleico contendrá generalmente enlaces fosfodiéster, aunque en algunos casos pueden incluirse análogos de ácido nucleico que tengan 65 armazones alternativos, tales como de engarces de fosforamidita, fosforoditioato, o de metilfosforamidita; o armazones

y engarces de ácidos peptidonucleicos. Otros ácidos nucleicos análogos incluyen aquellos con estructuras bicíclicas, incluyendo ácidos nucleicos bloqueados, armazones positivos, armazones no iónicos y armazones sin ribosa. Pueden efectuarse modificaciones del armazón ribosa-fosfato para aumentar la estabilidad de las moléculas; por ejemplo, los híbridos APN:ADN pueden mostrar mayor estabilidad en algunos ambientes.

5 "Cebador" significa un oligonucleótido, ya sea natural o sintético, que es capaz, tras formar un dúplex con un molde polinucleotídico, de actuar como punto de inicio de la síntesis de ácido nucleico y de extenderse desde su extremo 3' a lo largo del molde de tal forma que se forma un dúplex extendido. La secuencia de nucleótidos añadida durante el proceso de extensión se determina mediante la secuencia del polinucleótido de molde. Normalmente se extienden los
10 cebadores mediante una ADN polimerasa.

El término "SNP" o la expresión "polimorfismo de un solo nucleótido" se refiere a una variación genética entre individuos; por ejemplo, una sola posición de base nitrogenada en el ADN de organismos que es variable. Los SNP se encuentran por todo el genoma; gran parte de la variación genética entre individuos se debe a variaciones en locus de
15 SNP, y esta variación genética da a menudo como resultado una variación fenotípica entre individuos. Los SNP para su uso en la presente invención y sus alelos respectivos pueden proceder de cualquier número de fuentes, tales como bases de datos públicas (U.C. Santa Cruz Human Genome Browser Gateway (<http://genome.ucsc.edu/cgi-bin/hgGateway>) o el sitio web dbSNP del NCBI (<http://www.ncbi.nlm.nih.gov/SNP/>), o pueden determinarse experimentalmente tal como se describe en la Patente de Estados Unidos n.º 6.969.589; y en la
20 Publicación de Estados Unidos n.º 2006/0188875 titulada "Human Genomic Polymorphisms". Aunque se describe el uso de SNP en algunas de las realizaciones presentadas en el presente documento, se entenderá que también pueden usarse otros marcadores genéticos bialélicos o multialélicos. Un marcador genético bialélico es uno que tiene dos formas polimórficas, o alelos. Tal como se ha mencionado anteriormente, para un marcador genético bialélico que está asociado con un rasgo, el alelo que es más abundante en la composición genética de un grupo de caso en
25 comparación con un grupo de control se denomina el "alelo asociado" y el otro alelo puede citarse como el "alelo no asociado". Por lo tanto, para cada polimorfismo bialélico que está asociado con un rasgo dado (por ejemplo, una enfermedad o una respuesta a fármaco), hay un alelo asociado correspondiente. Otros polimorfismos bialélicos que pueden usarse con los métodos presentados en el presente documento incluyen, pero sin limitación, cambios multinucleotídicos, inserciones, eliminaciones, y traslocaciones. Se apreciará además que las referencias a ADN en el
30 presente documento pueden incluir ADN genómico, ADN mitocondrial, ADN episómico, y/o derivados de ADN, tales como amplicones, transcritos de ARN, ADNc, análogos de ADN, etc. Los locus polimórficos que se exploran en un estudio de asociación pueden ser en un estado diploide o haploide y, de manera ideal, serán de sitios a lo largo del genoma.

35 Las expresiones "se une de manera selectiva", "unión selectiva" y similares, tal como se usan en el presente documento, cuando se refieren a un compañero de unión (por ejemplo, proteína, ácido nucleico, anticuerpo u otro agente de captura por afinidad, etc.), se refiere a una reacción de unión de dos o más compañeros de unión con alta afinidad y/o complementariedad para asegurar una hibridación selectiva en condiciones de ensayo determinadas. Normalmente, la unión específica será al menos tres veces la desviación estándar de la señal de fondo. Por lo tanto, en
40 condiciones determinadas, el compañero de unión se une a su molécula "diana" particular y no se une en una cantidad significativa a otras moléculas presentes en la muestra.

"Secuenciación", "determinación de secuencia" y similares significan la determinación de la información relacionada con la secuencia de bases nucleotídicas de un ácido nucleico. Dicha información puede incluir la identificación o
45 determinación de información de secuencia parcial así como total del ácido nucleico. La información de secuencia puede determinarse con diversos grados de fiabilidad o confianza estadística. En un aspecto, el término incluye la determinación de la identidad y orden de una diversidad de nucleótidos contiguos en un ácido nucleico. La "secuenciación digital de alto rendimiento" o la "secuenciación de última generación" significa determinación de secuencia usando métodos que determinan muchas (normalmente de miles a miles de millones) de secuencias de ácido nucleico de una manera intrínsecamente paralela, es decir, donde se preparan moldes de ADN para su
50 secuenciación no de uno en uno, sino en un proceso por lotes, y donde se leen muchas secuencias preferentemente en paralelo, o como alternativa usando un proceso en serie de ultra-alto rendimiento que en sí puede disponerse en paralelo. Dichos métodos incluyen, pero sin limitación, pirosecuenciación (por ejemplo, tal como se comercializa por 454 Life Sciences, Inc., Branford, CT); secuenciación mediante ligadura (por ejemplo, tal como se comercializa en la
55 tecnología SOLiD™, Life Technologies, Inc., Carlsbad, CA); secuenciación mediante síntesis usando nucleótidos modificados (tal como se comercializa en la tecnología TruSeq™ y HiSeq™ de Illumina, Inc., San Diego, CA, HeliScope™ de Helicos Biosciences Corporation, Cambridge, MA, y PacBio RS de Pacific Biosciences of California, Inc., Menlo Park, CA), secuenciación mediante tecnologías de detección de iones (Ion Torrent, Inc., South San Francisco, CA); secuenciación de nanobolas de ADN (Complete Genomics, Inc., Mountain View, CA); tecnologías de
60 secuenciación basadas en nanoporos (por ejemplo, tal como se han desarrollado por Oxford Nanopore Technologies, LTD, Oxford, R. U.), y métodos de secuenciación dispuestos en paralelo altamente similares.

El término "T_f" se usa en referencia a la "temperatura de fusión". La temperatura de fusión es la temperatura a la cual una población de moléculas de ácido nucleico bicatenario se disocia a la mitad en hebras individuales. Se conocen
65 bien en la técnica varias ecuaciones para calcular la T_f de ácidos nucleicos. Tal como se indica por las referencias convencionales, se calcula una estimación sencilla del valor de T_f mediante la ecuación $T_m = 81,5 + 0,41 (\% G+C)$,

cuando un ácido nucleico se encuentra en solución acuosa con NaCl 1 M (véase, por ejemplo, Anderson y Young, Quantitative Filter Hybridization, en Nucleic Acid Hybridization (1985)). Otras referencias (por ejemplo, Allawi y SantaLucia, Jr., Biochemistry, 36:10581-94 (1997)) incluyen métodos alternativos de imputación que tienen en cuenta características estructurales y ambientales así como de secuencia para el cálculo de la T_m.

5

Descripción detallada de la invención

La práctica de las técnicas descritas en el presente documento puede emplear, salvo que se indique lo contrario, técnicas y descripciones convencionales de química orgánica, tecnología de polímeros, biología molecular (incluyendo técnicas recombinantes), biología celular, bioquímica, y tecnología de secuenciación que se encuentran dentro de las capacidades de los expertos en la materia. Dichas técnicas convencionales incluyen síntesis de matrices poliméricas, hibridación y ligadura de polinucleótidos, y detección de la hibridación usando un marcador. Pueden obtenerse ilustraciones específicas de técnicas adecuadas por referencia a los ejemplos en el presente documento. Sin embargo, se pueden usar, por supuesto, otros procedimientos convencionales equivalentes. Dichas técnicas y descripciones convencionales pueden encontrarse en manuales de laboratorio convencionales tales como Green, et al., Eds., Genome Analysis: A Laboratory manual Series (Vol. I-IV) (1999); Weiner, Gabriel, Stephens, Eds., Generic Variation: A Laboratory Manual (2007); Dieffenbach, Dveksler, Eds., PGR Primer: A Laboratory Manual (2003); Bowtell y Sambrook, DNA Microarrays: A Molecular Cloning Manual (2003); Mount, Bioinformatics: Sequence and Genome Analysis (2004); Sambrook y Russell, Condensed Protocols from Molecular Cloning: A Laboratory Manual (2006); y Sambrook y Russell, Molecular Cloning: Laboratory Manual (2002) (todos de Cold Spring Harbour Laboratory Press); Stryer, Biochemistry (4^a Ed.) (1995) W.H. Freeman; Nueva York, N. Y.; Gait, "Oligonucleotide Synthesis: A Practical Approach" (2002) IRL Press, Londres; Nelson y Cox, Lehninger. Principles of Biochemistry (2000) 3^a Ed., W. H. Freeman Pub., Nueva York, N.Y.; y Berg, et al., Biochemistry (2002) 5^a Ed., W.H. Freeman Pub., Nueva York, N.Y.

10

15

20

25

Cabe destacar que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas del singular "un", "una" y "el" o "la" incluyen referencia en plural a menos que el contexto dicte claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "un ácido nucleico" se refiere a uno o más ácidos nucleicos, y la referencia a "el ensayo" incluye la referencia a etapas y métodos equivalentes conocidos para los expertos en la materia, y así sucesivamente.

30

A menos que se defina de otro modo, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado que el normalmente entendido por un experto habitual en la materia a la cual pertenece esta invención.

35

En los casos donde se proporciona un intervalo de valores, se entiende que cada valor intermedio entre el límite superior e inferior de ese intervalo y cualquier otro valor indicado o intermedio en dicho intervalo indicado está abarcado en la invención. Los límites superiores e inferiores de estos intervalos más pequeños pueden incluirse independientemente en los intervalos más pequeños, y también están abarcados en la invención, salvo cualquier límite excluido específicamente en el intervalo indicado. En los casos donde el intervalo indicado incluya uno o ambos límites, los intervalos que excluyan a cualquiera de estos límites incluidos están también incluidos en la invención.

40

En la siguiente descripción, se exponen numerosos detalles específicos para proporcionar una comprensión más exhaustiva de la presente invención. Sin embargo, será evidente para los expertos en la materia que la presente invención puede ponerse en práctica sin uno o más de estos detalles específicos. En otros casos, no se han descrito características y procedimientos bien conocidos para los expertos en la materia con el fin de no oscurecer la invención.

45

La invención en general (La invención se define en las reivindicaciones.)

La invención se refiere a ensayos multiplexados codificados espacialmente que comprenden 1) un ensayo capaz de altos niveles de multiplexación con un esquema de codificación espacial eficaz; 2) instrumentación capaz de suministrar reactivos de acuerdo con un patrón espacial; y 3) descodificación determinada mediante una lectura que es de naturaleza digital. Los sistemas de ensayo detectan la presencia o ausencia y la cantidad relativa de una diana biológica o actividad biológica indicativa de una diana biológica, así como la localización de la diana biológica o la actividad en una muestra biológica, por ejemplo, una sección de tejido u otra estructura biológica dispuesta sobre un soporte, tal como un portaobjetos de microscopía o una placa de cultivo.

50

55

El sistema de ensayo proporciona además instrumentación con capacidad para administrar reactivos con un patrón espacialmente definido. Esta instrumentación, junto con programas informáticos, reactivos y protocolos, proporciona un componente clave del sistema de ensayo altamente innovador, permitiendo medir numerosas dianas o actividades biológicas en un ambiente espacial significativo, incluyendo la expresión génica y la localización del péptido. Un esquema de codificación usado en estos sistemas de ensayo permiten determinar la localización o actividad de dianas biológicas (o la ausencia de las mismas en las muestras biológicas después de que se hayan retirado los productos del ensayo multiplexado de la muestra biológica y se hayan agrupado para su análisis). La descodificación del esquema de codificación puede efectuarse mediante, por ejemplo, secuenciación de última generación, que proporciona fácilmente de millones a miles de millones de puntos de datos a bajo coste. Los resultados del ensayo, tales como la cantidad o actividad de dianas biológicas, pueden volver a mapearse a una localización específica en la muestra

60

65

biológica. Los sistemas de ensayo abren una nueva ventana analítica hacia los complejos patrones espaciales de la función y regulación celular en muestras biológicas.

5 En la figura 1 se proporciona una vista general simplificada del sistema 100 de ensayo usando la presente invención. En la etapa 110, se proporciona una muestra biológica fijada a un soporte. La muestra biológica contiene dianas biológicas de interés. Las dianas biológicas pueden incluir cualquier molécula de interés, tal como ácidos nucleicos (incluyendo, por ejemplo, transcritos de ARN, secuencias de ADN genómico, ADNc, amplicones, u otras secuencias de ácido nucleico) y proteínas, enzimas y similares. En la etapa 120, se suministran sondas codificadas a la muestra biológica de acuerdo con un patrón espacial. Las sondas codificadas comprenden sondas que pueden interactuar con

10 dianas biológicas de interés, y marcadores de codificación, que identifican las posiciones en la muestra de las dianas biológicas que se están ensayando, y por lo tanto pueden usarse para vincular los resultados del ensayo de nuevo a las localizaciones en la muestra. Los marcadores de codificación en la mayoría de realizaciones son oligonucleótidos. Sin embargo, los marcadores de codificación también pueden ser marcadores de masa, marcadores fluorescentes, u otros restos.

15 En algunas realizaciones, las porciones de sonda y marcador de codificación de la sonda codificada se acoplan previamente antes de administrarse a la muestra biológica. Por ejemplo, en el caso donde las sondas codificadas son oligonucleótidos, tanto la sonda como la secuencia de marcador de codificación pueden sintetizarse como un solo oligonucleótido. Como alternativa, la sonda y las porciones de marcador de codificación de las sondas de codificación

20 pueden sintetizarse u obtenerse por separado y combinarse antes de la administración a la muestra biológica (por ejemplo, pueden sintetizarse dos oligonucleótidos separados y acoplarse mediante, por ejemplo, ligadura; o pueden prepararse un anticuerpo y un oligonucleótido por separado y combinarse antes de la administración a la muestra biológica). Asimismo, tal como se describe en las figuras 2-5, las sondas y los marcadores de codificación (en oligonucleótidos codificantes) se sintetizan por separado, y se administran a la muestra biológica en diferentes etapas

25 (por ejemplo, en primer lugar las sondas y posteriormente los marcadores de codificación, o *vice versa*) del ensayo.

30 En la etapa 130, se deja que las sondas codificadas reaccionen o interactúen con las dianas biológicas, es decir, se proporcionan condiciones para permitir, por ejemplo, que los oligonucleótidos hibriden con dianas de ácido nucleico, que las dianas catalicen reacciones con dianas de proteína, que los anticuerpos se unan a epítopos, etc. En el caso donde las dianas biológicas son ácidos nucleicos, las sondas codificadas son normalmente oligonucleótidos e hibridan con los ácidos nucleicos diana. En caso de que las dianas biológicas sean proteínas, las sondas codificadas son normalmente aptámeros, moléculas pequeñas, o proteínas conjugadas a oligonucleótidos que interactúan con

35 proteínas diana uniéndose a ellas o reaccionando con ellas (es decir, una de las proteínas es un sustrato para la otra). Los oligonucleótidos codificantes pueden acoplarse a las sondas (proteínas) mediante conjugación, reticulación química o fotónica mediante grupos adecuados y similares.

Una vez han interactuado las sondas codificadas con las dianas biológicas, las sondas codificadas que interactuaron con las dianas biológicas tienen que separarse de las sondas codificadas que no interactuaron con las dianas biológicas en la etapa 140. En el caso donde las dianas biológicas sean ácidos nucleicos y las sondas codificadas sean

40 oligonucleótidos, la separación puede lograrse mediante, por ejemplo, el lavado de las sondas codificadas no hibridadas de la muestra. De manera similar, para otros ensayos que se basan en la unión por afinidad, incluyendo aquellos que usan sondas de aptámero, molécula pequeña, y de proteína, pueden usarse etapas de lavado para eliminar moléculas de unión de baja afinidad. En el caso donde la sonda se transforme mediante interacción con la diana, por ejemplo, en el caso de un péptido, por ejemplo, mediante escisión por una proteasa o fosforilación mediante

45 una cinasa, es conveniente recoger todas las sondas codificadas, tanto las sondas codificadas que interactuaron con las dianas biológicas y que se transformaron como las sondas codificadas que no se transformaron. Después de la recogida o agrupamiento, puede usarse un anticuerpo u otro agente de captura por afinidad para capturar sondas que se transformaron mediante la adición de un resto (por ejemplo, un grupo fosfato). En los casos donde las sondas se han transformado mediante escisión, pueden separarse las sondas transformadas, por ejemplo, mediante la captura

50 de las sondas no transformadas mediante un marcador que se retira de las sondas transformadas durante la transformación (por ejemplo, mediante escisión), o añadiendo un nuevo marcador en el sitio de escisión.

Una vez que las sondas codificadas que han reaccionado (transformadas) o interactuado se separan de las sondas codificadas que no han reaccionado o interactuado, se determina la secuencia de las sondas codificadas que han

55 reaccionado y/o interactuado mediante, preferentemente, secuenciación. La secuencia de las sondas codificadas permite el mapeo de los resultados de ensayo de nuevo a localizaciones en la muestra biológica.

La figura 2 proporciona una visión general simplificada de un sistema de ensayo usando la presente invención que realiza una implementación eficaz de un esquema de codificación combinatoria para la codificación de la información

60 espacial. Para los fines de esta visión general, las sondas son oligonucleótidos, pero tal como se explica en otras partes, también pueden usarse otros tipos de sondas. En la etapa 210, se proporciona una muestra biológica fijada a un soporte, por ejemplo, una muestra de tejido u otra estructura biológica. En la etapa 220 se suministran una o más sondas oligonucleotídicas a la muestra biológica, donde las sondas oligonucleotídicas son capaces de hibridar con dianas biológicas en la muestra biológica. En la etapa 230, se deja que las sondas oligonucleotídicas interactúen con

65 (hibriden con) dianas de ácido nucleico; es decir, se proporcionan condiciones adecuadas donde las sondas oligonucleotídicas pueden hibridar con los ácidos nucleicos diana.

En la etapa 240, se retiran las sondas oligonucleotídicas que no hibridan con ácidos nucleicos diana, y de este modo se separan de las sondas oligonucleotídicas que hibridaron con ácidos nucleicos diana. En esta realización, la separación puede efectuarse mediante, por ejemplo, lavado de la muestra y eliminación de las sondas oligonucleotídicas no hibridadas. A continuación, en la etapa 250, los oligonucleótidos codificantes (los agentes codificantes) se administran a la muestra biológica de acuerdo con un patrón espacial seleccionado, donde los oligonucleótidos codificantes comprenden marcadores de codificación que se usan para codificar la localización de las dianas biológicas en la muestra biológica. Nótese que a diferencia del sistema de ensayo en la figura 1, en este caso las sondas y los agentes codificantes (oligonucleótidos codificantes) se administran en etapas separadas. En la etapa 260, los oligonucleótidos codificantes se acoplan a las sondas oligonucleotídicas para crear sondas codificadas. En este caso donde las sondas son oligonucleótidos, los oligonucleótidos codificantes pueden acoplarse a las sondas oligonucleotídicas mediante, por ejemplo, ligadura. Como alternativa, puede transferirse la información en los oligonucleótidos codificantes usando una ADN polimerasa para extender una sonda oligonucleotídica que actúa como cebador, y de este modo copiar e incorporar la secuencia de los oligonucleótidos codificantes.

En la etapa 270, se determina la secuencia de los marcadores de codificación en las sondas codificadas así como la secuencia o una parte de la secuencia de la sonda en sí, y en la etapa 280, se vuelven a mapear los ácidos nucleicos en la muestra biológica. En algunas realizaciones, la abundancia de secuencias revela la cantidad relativa de dianas biológicas en la localización. Aunque esta realización muestra las etapas individuales en un orden concreto, para explicar mejor la invención, puede variarse el orden de las etapas. Por ejemplo, pueden combinarse las etapas 220 y 250, de tal forma que se administra una mezcla de las sondas y oligonucleótidos de acuerdo con un patrón espacial seleccionado. La etapa de acoplamiento 260 puede llevarse a cabo entonces inmediatamente después de las etapas combinadas 220 y 250, o de manera concomitante con estas. En este caso, la etapa 240 puede tener lugar después de la etapa 260. Por lo tanto puede apreciarse que los dos resultados clave de esta serie de etapas, es decir, la codificación específica de localización de las moléculas de sonda y la separación de moléculas de sonda basándose en su capacidad para interactuar con las moléculas diana correspondientes, puede lograrse con cierta flexibilidad en la implementación de las etapas concretas. De manera similar, hay una flexibilidad considerable en el diseño del esquema de codificación. Tal como se describe más adelante, los ensayos de la invención son particularmente susceptibles a métodos combinatorios.

Por lo tanto, la presente invención proporciona una capacidad para observar muchas dianas biológicas diferentes en muchas localizaciones, proporcionando la resolución de la hibridación *in situ* con el análisis altamente paralelo de datos de la secuenciación. En algunas realizaciones, la suma de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra biológica es mayor de 20, en otras realizaciones, la suma de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra biológica es mayor de 50, en otras realizaciones, la suma de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra biológica es mayor de 100, mayor de 500, 1.000, 10.000, 25.000, 100.000, 500.000, 1.000.000. Se apreciará que, debido a la dimensión de codificación espacial de la invención, pueden contemplarse números incluso más grandes. Por ejemplo, el ensayo de 10.000 dianas por localización x 10.000 localizaciones podría generar 10^8 ensayos diferentes, y pueden contemplarse fácilmente números incluso mayores que estos, particularmente si se usan localizaciones espaciales con resolución en el orden de las células individuales. Además, en realizaciones donde se emplea secuenciación digital de alto rendimiento, se determinan en paralelo normalmente las secuencias de al menos 1.000 sondas codificantes. Más normalmente, usando una lectura digital, es deseable obtener múltiples lecturas de secuencia para cada ensayo (definidas mediante una sonda y un código de localización espacial). Es deseable obtener una media de al menos 3 copias por ensayo, y más normalmente de al menos 10 o al menos 30 copias por ensayo, dependiendo del diseño del experimento y de las necesidades del ensayo. Para una lectura cuantitativa con un intervalo dinámico adecuado, puede ser deseable obtener al menos 1.000 lecturas por ensayo. Por lo tanto, si se llevan a cabo 1.000.000 de ensayos, el número de lecturas de secuencia puede ser de 1.000 millones o más. Con la secuenciación digital de alto rendimiento, y permitiendo redundancia, se determinan las secuencias de al menos 10.000 sondas codificantes en paralelo o se determinan la secuencia de al menos 100.000, 500.000, 1.000.000, 10.000.000, 100.000.000, 1.000.000.000 o más sondas codificantes en paralelo.

Ensayos

La parte de ensayo de los sistemas de ensayo comprende las siguientes etapas generales: administrar las sondas y agentes codificantes donde los agentes codificantes, (en algunas realizaciones acoplados previamente a las sondas) se administran a la muestra de acuerdo con un patrón espacial conocido, permitiendo que las sondas interactúen o reaccionen con dianas biológicas en la muestra, y, si no se han acoplado previamente las sondas y los agentes codificantes, acoplando los agentes codificantes a las sondas.

Las muestras incluyen virtualmente cualquier muestra biológica o muestras que pueden fijarse a un soporte o proporcionarse esencialmente de un modo bidimensional, donde es importante la capacidad para anclar una diana biológica ensayada o una actividad de nuevo a la localización dentro de la muestra biológica. Las muestras biológicas ejemplares incluyen secciones de tejido (por ejemplo, incluyendo secciones completas de animales y biopsias de tejido), poblaciones de células en portaobjetos o placas de cultivo, y similares. Los sistemas de ensayo son particularmente ventajosos en tanto que son compatibles con numerosos tipos de muestra biológica, incluyendo

muestras frescas, tales como secciones de tejido primario, y muestras conservadas (incluyendo, pero sin limitación, muestras congeladas y muestras fijadas en paraformalina, incluidas en parafina (FFPE). Un aspecto importante de los sistemas de ensayo es que las muestras biológicas se inmovilizan sobre una superficie de sustrato que tiene áreas discretas, medibles de manera independiente.

5 Las dianas biológicas que se van a detectar pueden ser cualquier molécula biológica incluyendo, pero sin limitación, proteínas, ácidos nucleicos, lípidos, hidratos de carbono, iones, o complejos multicomponente que contienen cualquiera de los anteriores. Los ejemplos de dianas subcelulares incluyen orgánulos, por ejemplo, mitocondrias, aparato de Golgi, retículo endoplasmático, cloroplastos, vesículas de endocitosis, vesículas de exocitosis, vacuolas, lisosomas, etc.

15 En algunas realizaciones particulares, el sistema de ensayo se usa para analizar ácidos nucleicos, por ejemplo, mediante genotipado, cuantificación del número de copias de ADN o de transcritos de ARN, localización de transcritos concretos dentro de las muestras, y similares. La figura 3 ilustra un esquema general para un ensayo ejemplar para, por ejemplo, detectar polimorfismos de nucleótidos individuales (SNP) que pueden usarse con el sistema de ensayo. En la figura 3, se proporcionan dos sondas oligonucleotídicas. Cada sonda oligonucleotídica comprende una región específica de diana (situada a cada lado del SNP que va a analizarse) observada en 305 y 307, y las regiones de ligadura, observadas en 301 y 303. Se deja que las sondas oligonucleotídicas hibriden con un ácido nucleico diana (no mostrado) en la muestra biológica. En la etapa 302, se extiende una de las sondas oligonucleotídicas para que incorpore la secuencia del SNP y se liga a la otra sonda para formar una sonda extendida que comprende la región 309 de ácido nucleico diana y las regiones 301 y 303 de ligadura.

25 Dos agentes codificantes, comprendiendo cada uno un marcador de codificación (observado en 315 y 317), una región de ligadura (observada en 311 y 313) y una región de cebador (observada en 319 y 321) se combinan con y se ligan a la sonda extendida en la etapa 304 para formar un oligonucleótido específico de diana codificado. De nuevo, a diferencia de la figura 1, las sondas y los agentes codificantes se administran en etapas separadas. Hacer eso permite el uso de las realizaciones combinatorias descritas más adelante. En realizaciones preferidas, los oligonucleótidos codificantes en un par de oligonucleótidos codificantes se ligan específicamente a un lado de la secuencia diana o de la otra (es decir, en 5' o 3' de la secuencia diana). Asimismo, normalmente, las regiones de ligadura y de cebador de los oligonucleótidos y sondas codificantes son universales; es decir, el conjunto de regiones de ligadura y cebador usadas para construir las sondas y oligonucleótidos codificantes es constante, y solo difieren las regiones específicas de diana de las sondas y los marcadores de codificación de los oligonucleótidos codificantes. Sin embargo, de nuevo en realizaciones alternativas, las regiones de ligadura y de cebador no son universales y difieren entre sondas y agentes codificantes.

35 Después de la ligadura, se eluyen las sondas codificantes, se agrupan, y, opcionalmente, se añaden adaptadores de secuenciación a las sondas codificantes mediante PCR. En realizaciones alternativas, pueden ligarse cebadores de secuenciación a los oligonucleótidos codificantes, o pueden incluirse secuencias de cebador de secuenciación como parte del oligonucleótido codificante. Tal como se observa en la figura 3, cada adaptador de secuenciación comprende la región 319 o 321 de cebador, compatible con las regiones 319 y 321 de cebador en las sondas codificadas. La construcción final que comprende el primer adaptador 327, la primera región de cebador 319, el primer marcador de codificación 315, las regiones 311 y 301 de ligadura, la región diana 309, las regiones 313 y 303 de ligadura, el segundo marcador de codificación 317, la segunda región 325 de cebador y el segundo adaptador 329 está ahora preparado para la entrada en un proceso de secuenciación digital de alto rendimiento.

45 En la figura 3 se ejemplifica una combinación de reacciones de extensión y de ligadura, pero debe apreciarse que pueden usarse una diversidad de reacciones para acoplar los oligonucleótidos codificantes a los oligonucleótidos específicos de diana, incluyendo únicamente ligadura (por ejemplo, para oligonucleótidos que hibridan con porciones contiguas de la secuencia de ácido nucleico diana). Como alternativa, puede emplearse un ensayo que utilice un oligonucleótido adicional, tal como en el ensayo GOLDENGATE® (véase Fan, et al., Cold Spring Symp. Quant. Biol., 68:69-78 (2003); (Illumina, Inc., San Diego, CA)).

55 En otras realizaciones, también puede usarse el sistema de ensayo para analizar péptidos o proteínas, la presencia de anticuerpos, actividades enzimáticas o de otras proteínas, modificaciones postraduccionales, formas activas y no activas de péptidos, así como isoformas de péptidos en una muestra biológica. Por consiguiente, las sondas pueden comprender una región activa de una enzima, un dominio de unión de una inmunoglobulina, dominios de proteínas definidos, proteínas completas, péptidos sintéticos, péptidos con mutaciones introducidas, aptámeros y similares.

60 En determinados aspectos, las sondas son sustratos para enzimas o proenzimas, por ejemplo, cinasas, fosfatasa, zimógenos, proteasas, o fragmentos de las mismas. En determinados aspectos, las sondas son sustratos de fosforilación usados para detectar proteínas implicadas en una o más rutas de transducción de señales, por ejemplo, una cinasa o una fosfatasa. En otra estrategia específica, las sondas son sustratos de proteasas específicos que se asocian únicamente con proteasas individuales o clases de proteasas. En otras estrategias, las sondas son diferentes formas procesadas, isoformas y/o dominios de una enzima. Las sondas basadas en proteínas están normalmente conjugadas o de otro modo ligadas a agentes codificantes de oligonucleótidos. Los agentes codificantes de oligonucleótidos en este caso también podrían incluir un componente de secuencia nucleotídica que permita la

identificación de la sonda de proteína.

En determinados aspectos, la presente invención proporciona ensayos para evaluar diferencias en la cantidad y/o actividad de dianas biológicas entre diferentes localizaciones en una muestra y/o entre muestras. El método incluye determinar una diversidad de resultados codificados de la muestra biológica y evaluar las diferencias en cantidad de las dianas biológicas en cada localización en la muestra biológica.

Realizaciones combinatorias

Para maximizar la eficacia de la codificación, puede usarse una estrategia combinatoria que usa pares de marcadores de codificación en los oligonucleótidos codificantes. Al desacoplar la información específica de diana y los marcadores de codificación, se reduce dramáticamente el número de oligonucleótidos necesario, con la consiguiente disminución de los costes.

La figura 4 ilustra un mecanismo general para una realización de un esquema de codificación combinatorio de la invención, donde se ensayan ácidos nucleicos en una sección de tejido representativa (mostrada en 416). La figura 4 muestra en A dos construcciones de oligonucleótido específico de diana/codificante 420 y 422 (por ejemplo, formado entre las etapas 302 y 304 de la figura 3) unidas específicamente a un ácido nucleico diana 402 de interés. La primera sonda codificada 420 comprende el marcador 408 de codificación asociado con, por ejemplo, un sitio de cebado universal para la amplificación de los productos de ensayo o un adaptador para permitir la identificación de los identificadores de codificación usando tecnologías de secuenciación 404. La segunda sonda codificada 422 comprende el marcador 406 de codificación asociado con, por ejemplo, un sitio de cebado universal para la amplificación de los productos de ensayo o un adaptador para permitir la identificación de los identificadores de codificación usando tecnologías de secuenciación 410.

La figura 4 muestra en B el patrón espacial que puede usarse para veinte marcadores de codificación diferentes, a1 a a10 (marcador 406 de codificación en la sonda codificada 420) y b1 a b10 (marcador 408 de codificación en la sonda codificada 422). El marcador codificante a1, por ejemplo, se deposita sobre la muestra biológica en diez áreas o puntos discretos (mostrados como la primera línea horizontal de puntos en 412). El marcador de codificación a2 se deposita sobre la muestra biológica en diez puntos sobre la segunda línea horizontal en 412. El marcador de codificación a3 se deposita sobre la muestra biológica en diez puntos sobre la tercera línea horizontal en 412, y así sucesivamente. Mientras que los marcadores "a" se depositan en diez filas horizontales, los marcadores "b" se depositan en tres filas verticales, tal como se muestra en 414. Por ejemplo, el marcador de codificación b1 se deposita sobre la muestra biológica en diez puntos discretos en la primera fila vertical de 414, el marcador de codificación b2 se deposita sobre la muestra biológica en diez puntos discretos en la segunda fila vertical de 414, y así sucesivamente. El uso de dicha configuración permite que veinte marcadores de codificación definan de manera única 100 localizaciones diferentes sobre la muestra biológica.

La figura 4 muestra en c una sección de tejido 416 representativa coincidente con la cuadrícula 418 del marcador de codificación. Las flechas muestran como los marcadores de codificación "a" y los marcadores de codificación "b" se depositan sobre la cuadrícula 418 que es coincidente con la sección de tejido 416. Si, una vez secuenciados, los marcadores de codificación a1 y b4, por ejemplo, se asocian con una secuencia de ácido nucleico diana, entonces esta secuencia de ácido nucleico diana (es decir, diana biológica) estaba presente en la sección de tejido en la localización a1, b4.

La figura 5 proporciona un ejemplo específico simplificado del esquema de codificación de los sistemas de ensayo usando la invención. La figura 5 muestra los oligonucleótidos codificantes 510, que comprenden a1, a2, a3, a4 y b1, b3, b3 y b4. Los oligonucleótidos específicos de diana (TSO) (sondas) 1 y 2 se muestran en 520. Se muestra un esquema de depósito o de dispersión en 530. Al igual que la cuadrícula ejemplificada en la figura 4, los oligonucleótidos codificantes a1 a a4 se depositan en puntos en un patrón (en este caso, en un patrón vertical), y los oligonucleótidos codificantes b1 a b4 se depositan en puntos en un patrón (en este caso, un patrón horizontal). La cuadrícula, aunque se muestra como un cuadrado con puntos, es realmente un patrón de deposición sobre una muestra biológica (no mostrada), tal como la sección de tejido 416 mostrada en la figura 4.

Los oligonucleótidos específicos de diana se administran a la muestra biológica, donde los oligonucleótidos específicos de diana hibridan con ácidos nucleicos diana en la muestra biológica si están presente los ácidos nucleicos diana. Entonces se eliminan los oligonucleótidos específicos de diana no hibridados, por ejemplo, mediante lavado. Los oligonucleótidos codificantes se administran entonces a la muestra biológica de acuerdo con el patrón espacial mostrado en 530. Se ligan los oligonucleótidos codificantes (o, por ejemplo, se extienden y ligan) a cualquier oligonucleótido específico de diana que hibridó con el ácido nucleico diana en la muestra biológica, y después las construcciones ligadas se eluyen de la muestra biológica, se agrupan, y se añaden adaptadores de secuenciación mediante, por ejemplo, PCR o ligadura, si no se habían incluido las secuencias previamente en los oligonucleótidos codificantes. Las construcciones ligadas se secuencian mediante, por ejemplo, secuenciación de alto rendimiento o de "última generación".

El grupo de secuencias resultante se muestra en 540. Se obtuvo una lectura de secuencia para el oligonucleótido específico de diana 1 solo en a4b1, a4b2, a1b3, a2b3, a3b3, a4b3 y a4b4 (posiciones mostradas con líneas

horizontales). Se obtuvo una lectura de secuencia para el oligonucleótido específico de diana 2 solo en a1b1 (posición mostrada con líneas verticales). Se obtuvo una lectura de secuencia para ambos oligonucleótidos específicos de diana 1 y 2 en las posiciones a2b1, a3b1, a1b2, a2b2, y a3b2 (posiciones mostradas con líneas cruzadas). No se obtuvo lectura de secuencia para los oligonucleótidos específicos de diana en a1b4, a2b4 o a3b4 (posiciones mostradas sin sombreado). Por lo tanto, en la muestra biológica en la que tuvo lugar el ensayo se detectó el primer ácido nucleico diana en una gran porción del lado izquierdo y en la parte inferior de la muestra biológica, el segundo ácido nucleico diana se detectó solo en la parte superior izquierda de la muestra biológica, y no se detectó ningún ácido nucleico en la parte superior derecha de la muestra biológica. La expresión diferencial de los dos ácidos nucleicos diana ahora puede volver a mapearse en la muestra biológica y en las estructuras biológicas o tipos celulares en estas localizaciones en la muestra biológica.

Además de la información de localización, puede obtenerse información relativa a la abundancia de los marcadores codificados. Por ejemplo, se observa que hay diez veces más secuencias a4T1b1 en el conjunto de datos en comparación con las secuencias a4T1b2, esto podría indicar que la secuencia de ácido nucleico diana 1 es diez veces más abundante en la localización a4T1b1 que en la localización a4T1b2.

En el caso del análisis de nucleótidos tal como se muestra en la figura 3, al ligar los marcadores de codificación directamente a oligonucleótidos específicos de diana, solo se necesitan 2n oligonucleótidos específicos de diana para las dianas. Por ejemplo, al usar la estrategia combinatoria indicada en la figura 2, el ensayo de 100 dianas diferentes en 10.000 localizaciones espaciales requeriría de 2 x 100 oligonucleótidos específicos de diana y de 2 x 100 oligonucleótidos codificantes: El recuento total de oligonucleótidos de ensayo sería únicamente de 400 (200 específicos de diana y 200 codificantes), sin contar los cebadores universales. Por el contrario, si no se desacoplan los oligonucleótidos codificantes de los oligonucleótidos específicos de diana, se necesitarían (n x X códigos de posición) + (n x Y códigos de posición), o en el ejemplo anterior, 20.000 oligonucleótidos, sin contar las secuencias de cebador universales. Además, aunque las realizaciones mostradas en las figuras 2-5 ilustran un esquema combinatorio usando dos agentes de codificación (marcadores de codificación), pueden usarse tres, cuatro o más agentes de codificación y marcadores de codificación, y unirse a la sonda o entre sí mediante diversos medios y con diversas combinaciones de etapas.

Debido al aspecto de codificación espacial de este sistema de ensayo, puede generarse una gran cantidad de información con incluso un número modesto de ensayos. Por ejemplo, cinco o más dianas biológicas ensayadas en cinco o más posiciones en la muestra generan 25 o más combinaciones. Al usar secuenciación digital como lectura, el número óptimo de lecturas de secuencia por combinación depende de la sensibilidad e intervalo dinámico necesario, y puede ajustarse. Por ejemplo, si para cada combinación se muestrean de media 100 lecturas, el total para 25 combinaciones es de 25.000 lecturas. Si se ensayan 1.000 dianas en 1.000 localizaciones con una media de profundidad de muestreo de 1.000, se requieren entonces 10^9 lecturas. Estos números, aunque son grandes, se encuentran dentro de la capacidad de los métodos de secuenciación digitales intrínsecamente paralelos, que pueden generar conjuntos de datos de miles de millones o de billones de lecturas en un espacio de tiempo razonable y con un coste por lectura muy bajo. Por lo tanto, al variar el número de posiciones interrogadas o de dianas biológicas ensayadas, o ambas, y usando secuenciación digital, pueden obtenerse grandes cantidades de información. En aspectos específicos, se interrogan múltiples localizaciones para dos o más moléculas biológicas.

Sistemas de administración de reactivo

El sistema de administración de reactivos incluye instrumentación que permite la administración de reactivos a partes discretas de la muestra biológica, manteniendo la integridad de los patrones espaciales del esquema de codificación. Los sistemas de administración de reactivo de los sistemas de ensayo comprenden medios de obtención de imágenes opcionales; equipamiento de administración de reactivo y programas informáticos de control. La administración del reactivo puede lograrse de una cantidad de formas diferentes. Cabe destacar que la administración puede ser a muchas muestras biológicas a la vez. En el presente documento se ha ejemplificado una sola sección de tejido; sin embargo, pueden fijarse y analizarse simultáneamente múltiples muestras biológicas. Por ejemplo, pueden analizarse secciones en serie de una muestra de tejido en paralelo y combinarse los datos para construir un mapa en 3D.

La instrumentación que permite la formación de patrones espaciales de reactivos en la muestra biológica es una parte integral del sistema de ensayo. Las tecnologías para formular y administrar ambas moléculas biológicas (por ejemplo, oligonucleótidos o anticuerpos) y reactivos químicos (por ejemplo, moléculas pequeñas o dNTP) se conocen en la técnica, y el uso de estos sistemas instrumentales se conoce por los expertos en la materia y son fácilmente adaptables a los sistemas de ensayo. Un ejemplo de un sistema de administración de reactivo adecuado es el manipulador de líquidos acústico Labcyte™ Echo, que puede usarse para administrar gotas a escala de nanolitro que contienen moléculas biológicas con alta precisión y reproducibilidad. Un experto en la materia podría incorporar este dispositivo de administración de reactivo al sistema general, usando programas informáticos para especificar las localizaciones sobre las que deben administrarse los reactivos.

Otros instrumentos que pueden usarse para la deposición de agentes y/o identificadores de codificación sobre muestras biológicas incluyen, pero sin limitación, deposición por chorro de tinta; deposición mecánica mediante un punzón, bolígrafo o capilar; impresión de micro contacto; métodos fotoquímicos o fotolitográficos; y similares. Para

varias aplicaciones, puede preferirse segmentar o secuestrar determinadas áreas de las muestras biológicas en una o más áreas de ensayo para diferentes distribuciones de reactivo y/o determinaciones de diana biológica. Las zonas de ensayo pueden separarse físicamente usando barreras o canales.

5 En un aspecto ejemplar, el sistema de administración de reactivo puede ser un sistema basado en flujo. Los sistemas basados en flujo para la administración de reactivo en la presente invención pueden incluir instrumentación, tal como una o más bombas, válvulas, depósitos de fluido, canales, y/o celdas de almacenamiento de reactivo. Los sistemas de administración de reactivo están configurados para mover fluido para que entre en contacto con una sección discreta de la muestra biológica. El movimiento de los reactivos puede estar dirigido por una bomba dispuesta, por ejemplo, 10 aguas abajo de los reactivos fluidos. La bomba puede dirigir cada reactivo fluido hacia (y atravesar) el compartimento de reacción. Como alternativa, los reactivos pueden dirigirse hacia el fluido mediante gravedad. Las publicaciones de Estados Unidos n.º 20070166725 y 20050239192 divulgan determinadas herramientas de manejo de fluidos de propósito general que pueden usarse con los sistemas de ensayo, permitiendo la manipulación precisa de los materiales de vidrio, líquidos y sólidos para lograr manipulaciones analíticas muy complejas con un equipo 15 relativamente sencillo.

En un ejemplo más específico, pueden unirse una o más celdas de flujo a la muestra biológica fijada al sustrato desde arriba. La celda de flujo puede incluir tubos de entrada y salida conectados a esta y se usa opcionalmente una bomba externa para administrar reactivos a la celda de flujo y a través de la muestra biológica. Las celdas de flujo se 20 configuran para administrar reactivos únicamente a determinadas partes de la muestra biológica, restringiendo la cantidad y tipo de reactivo administrado a cualquier sección específica de la muestra biológica.

En otro aspecto, puede integrarse un sistema de microfluidos al sustrato sobre el que se dispone la muestra biológica o unirse externamente sobre el sustrato. Los pasadizos de microfluidos para mantener y portar el fluido pueden formarse sobre y/o por encima del sustrato plano mediante una capa de fluidos apoyada sobre el sustrato. Los reactivos fluidos pueden seleccionarse y administrarse según la apertura y cierre de válvulas dispuestas entre depósitos de reactivo. 25

Las bombas incluyen generalmente cualquier mecanismo para mover fluidos y/o reactivos dispuestos en un fluido. En algunos ejemplos, puede configurarse la bomba para que mueva fluidos y/o reactivos a través de los pasadizos con volúmenes pequeños (es decir, estructuras microfluidas). La bomba puede operar de manera mecánica ejerciendo una presión positiva o negativa en el fluido y/o en una estructura que porta el fluido, eléctricamente mediante la aplicación adecuada de campos eléctricos, o ambas, entre otros medios. Las bombas mecánicas ejemplares pueden incluir bombas de jeringa, bombas peristálticas, bombas rotatorias, gas a presión, pipeteadores, etc. Las bombas mecánicas 30 pueden estar micromecanizadas, moldeadas, etc. Las bombas eléctricas ejemplares pueden incluir electrodos y pueden funcionar mediante electroforesis, electroosmosis, electrocapilaridad, dielectroforesis (incluyendo formas de viaje de onda de las mismas), y/o similares. 35

Las válvulas incluyen generalmente cualquier mecanismo para regular el paso de fluidos a través de un canal. Las válvulas pueden incluir, por ejemplo, miembros deformables que puedan deformarse de manera selectiva para cerrar parcial o completamente un canal, una proyección móvil que pueda extenderse de manera selectiva dentro de un canal para bloquear total o parcialmente un canal, una estructura de electrocapilaridad, y/o similares. 40

Puede acoplarse una junta abierta a la parte superior de la muestra biológica y la muestra y los reactivos pueden inyectarse en la junta. Los materiales de junta adecuados incluyen, pero sin limitación, neopreno, nitrilo, y silicona. Como alternativa, puede formarse una cámara de reacción estanca al agua mediante una junta dispuesta en sándwich entre la muestra biológica o el sustrato y un material químicamente inerte resistente al agua tal como, pero sin limitación, aluminio negro anodizado, termoplásticos (por ejemplo, poliestireno, policarbonato, etc), vidrio, etc. 45

En una realización opcional, el sistema de ensayo comprende medios para la obtención de imágenes para determinar características y la organización de la muestra biológica de interés. Las imágenes obtenidas, por ejemplo, pueden usarse para diseñar el patrón de deposición de los reactivos. Los medios para la obtención de imágenes son opcionales, ya que un individuo puede en su lugar observar la muestra biológica usando, por ejemplo, un microscopio, analizar la organización de la muestra biológica, y especificar un patrón espacial para la administración de los reactivos de ensayo. En caso de incluirse, el sistema de administración puede comprender una disposición de microcircuitos que incluya un dispositivo de imagen, tal como un dispositivo para la obtención de imágenes basado en CCD o en IGFET (por ejemplo, basado en CMOS) y un pulverizador ultrasónico para la administración de reactivos, tal como se describe en la publicación de Estados Unidos n.º 20090197326. Asimismo, cabe destacar que aunque las figuras 4 y 5 ilustran el uso de una configuración de cuadrícula x,y, pueden usarse otras configuraciones, tales como, por ejemplo, 50 siguiendo la topología de una muestra de tejido; usando como dianas determinados grupos de células, capas de células y/o tipos celulares en un tejido, y similares. 55 60

En otra alternativa más, el sistema de administración de reactivo controla la administración de los reactivos a patrones específicos en la superficie de una muestra biológica usando técnicas de semiconductor, tales como enmascaramiento y pulverización. Las áreas específicas de una muestra biológica pueden protegerse de la exposición a los reactivos mediante el uso de una máscara para proteger áreas específicas de la exposición. Los reactivos pueden introducirse 65

en la muestra biológica usando técnicas convencionales, tales como pulverización o flujo de fluido. El uso de una administración con máscara da como resultado un esquema de administración patronado sobre la superficie del sustrato.

5 En una estrategia preferida, la instrumentación de administración de reactivo se basa en la tecnología de impresión de chorro de tinta. Hay una diversidad de diferentes mecanismos de impresión por chorro de tinta (por ejemplo, térmica, piezoeléctrica) y se ha demostrado su compatibilidad con formulaciones de tinta acuosas y orgánicas. Pueden usarse conjuntos de boquillas activadas para administrar múltiples reactivos a la vez, y pueden lograrse resoluciones muy altas.

10 Para dirigirse a sitios específicos de interés, puede usarse una imagen informativa de la muestra biológica que se va a ensayar como ayuda en los métodos de administración de reactivos y el esquema de codificación asociado. Las regiones de muestra de la muestra biológica pueden identificarse usando procesamiento de imagen (por ejemplo, imágenes de tipos celulares diferenciadas mediante inmunohistoquímica u otras químicas de tinción) integradas con otras características del sistema de ensayo. En algunos aspectos, se usan programas informáticos para traducir automáticamente la información de la imagen en un patrón de administración de reactivo. Por lo tanto, un mecanismo para registrar y alinear de manera muy precisa la muestra biológica para la administración de reactivo es un componente importante de los sistemas de ensayo que usa la invención. Los mecanismos, tales como el uso de marcadores fiduciales sobre portaobjetos y/u otros sistemas de posicionamiento físico muy preciso pueden adaptarse para este fin.

La invención se usa preferentemente con un conjunto completo de programas informáticos diseñados a medida para el sistema de ensayo. De manera opcional, se usan programas informáticos de diseño de oligonucleótidos para diseñar los nucleótidos codificantes (y en realizaciones en las que se ensayan ácidos nucleicos, los oligonucleótidos específicos de diana) para ejecutar el ensayo específico, y pueden integrarse como parte del sistema. También opcionalmente, pueden integrarse algoritmos y programas informáticos para la administración de reactivos y el análisis de datos (es decir, análisis de secuencia) para determinar los resultados del ensayo. El análisis integrado de datos es particularmente útil, ya que el tipo de conjunto de datos que se genera puede ser masivo a consecuencia de la escala. Los algoritmos y herramientas informáticas que están especialmente diseñados para el análisis de los datos asociados espacialmente generados por los sistemas de ensayo, incluyendo programas informáticos de análisis de patrones y herramientas de visualización, potencian el valor de los datos generados por los sistemas de ensayo.

En determinados aspectos, el sistema de ensayo comprende procesos para producir y llevar a cabo el control de calidad de los reactivos, por ejemplo, la integridad y fidelidad de la secuencia de los grupos de oligonucleótidos. En particular, los reactivos se formulan según factores tales como la volatilidad, estabilidad a temperaturas clave, y la compatibilidad química para la compatibilidad con la instrumentación de administración de reactivo y pueden analizarse mediante instrumentación integrada en el sistema de ensayo.

Secuenciación

40 Pueden usarse numerosos métodos para identificar los marcadores de codificación y secuencias de sonda en las sondas codificadas de los sistemas de ensayo. Los marcadores de codificación pueden detectarse usando técnicas, tales como espectrometría de masa (por ejemplo, Maldi-Tof, CL-EM/EM), obtención de imágenes por resonancia magnética nuclear, o preferentemente, secuenciación de ácidos nucleicos. Los ejemplos de técnicas para descodificar los marcadores de descodificación pueden encontrarse, por ejemplo, en la publicación de Estados Unidos n.º 20080220434. Por ejemplo, los marcadores de codificación pueden ser marcadores de masa de oligonucleótidos (OMT o massTag). Dichos marcadores se describen, por ejemplo, en la publicación de Estados Unidos n.º 20090305237. En otra alternativa más, las sondas codificadas pueden amplificarse e hibridarse a una micromatriz. Esto podría requerir que se lleven a cabo reacciones de amplificación separadas, en las que cada amplificación sea específica para un código espacial o subconjunto de códigos particular, efectuada usando cebadores específicos de código. Cada amplificación también podría incorporar un marcador resoluble diferente (por ejemplo, un fluoróforo). Después de la hibridación, pueden determinarse las cantidades relativas de una diana particular mapeando en diferentes localizaciones espaciales en la muestra por medio de las abundancias relativas de los marcadores resolubles.

55 En un aspecto particularmente preferido, los marcadores de codificación resultantes de acuerdo con el sistema de ensayo son sustratos para secuenciación de alto rendimiento de última generación, y se usan métodos de secuenciación de última generación altamente paralelos para confirmar la secuencia de los marcadores de codificación, por ejemplo, con la tecnología SOLiD™ (Life Technologies, Inc.) o Genome Analyzer (Illumina, Inc.). Dichos métodos de secuenciación de última generación pueden llevarse a cabo, por ejemplo, usando un método de secuenciación en un pase o usando secuenciación de extremo emparejado. Los métodos de secuenciación de última generación incluyen, pero sin limitación, métodos basados en hibridación, tales como los divulgados en, por ejemplo, Drmanac, patentes de Estados Unidos n.º 6.864.052; 6.309.824; y 6.401.267; y Drmanac et al, publicación de Estados Unidos 2005/0191656; métodos de secuenciación mediante síntesis, por ejemplo, las Patentes de Estados Unidos n.º 6.210.891; 6.828.100; 6.969.488; 6.897.023; 6.833.246; 6.911.345; 6.787.308; 7.297.518; 7.462.449 y 7.501.245; las solicitudes de publicación de Estados Unidos n.º 20110059436; 20040106110; 20030064398; y 20030022207;

Ronaghi, et al, Science, 281: 363-365 (1998); y Li, et al, Proc. Natl. Acad. Sci., 100: 414-419 (2003); métodos basados en ligadura, por ejemplo, las Patentes de Estados Unidos n.º 5.912.148 y 6.130.073; y las solicitudes de patente de Estados Unidos n.º 20100105052, 20070207482 y 20090018024; secuenciación nanopore, por ejemplo, las solicitudes de patente de Estados Unidos n.º 20070036511; 20080032301; 20080128627; 20090082212; y Soni y Meller, Clin Chem 53: 1996-2001 (2007)), así como otros métodos, por ejemplo, las solicitudes de patente de Estados Unidos n.º 20110033854; 20090264299; 20090155781; y 20090005252; también véanse, McKernan, et al., Genome Res., 19:1527-41 (2009) y Bentley, et al., Nature 456:53-59 (2008).

Aplicaciones del sistema de ensayo

Será evidente para un experto en la materia tras la lectura de la presente divulgación que hay numerosas áreas importantes de investigación biológica, diagnósticos, y desarrollo de fármacos que se beneficiarían de un sistema de ensayo multiplexado de alto rendimiento que pueda medir de manera simultánea la cantidad y localización espacial de una diana biológica en una muestra biológica. Por ejemplo, la combinación de la agilidad para estimar la abundancia relativa de diferentes transcritos de ARN con la capacidad para reconstruir una imagen de patrones espaciales de abundancia a lo largo de muchas localizaciones, que pueden ser tan pequeñas como o incluso más pequeñas que las células individuales, en un tejido posibilita muchas áreas diferentes de investigación básica. Lo siguiente son usos ejemplares y en ningún modo se pretende que sean limitativos en cuanto a su alcance.

En un ejemplo, se determinan patrones en 3 dimensiones de la expresión génica analizando una serie de secciones de tejido, de una manera análoga a la reconstrucción de imágenes en un barrido TC. Dicho método puede usarse para medir cambios en la expresión génica en una patología, por ejemplo, en tejido canceroso y/o un tejido tras lesionarse, inflamarse o infectarse. Con los sistemas de ensayo que usan la invención, se obtiene información más detallada de la expresión génica y de la localización de proteínas en tejidos complejos, dando lugar a nuevas perspectivas acerca de la función y la regulación en estados tanto normales como patológicos, y proporciona que puedan probarse nuevas hipótesis. Por ejemplo, un sistema de ensayo que use la invención puede posibilitar algunas de las perspectivas obtenidas con muchos estudios individuales y programas más grandes, tales como ENCODE (Birney, et al., Nature, 447:799-816 (2007)) y modENCODE para su integración a nivel de tejido. Los sistemas de ensayo también asisten en los esfuerzos computacionales para modelar redes interconectadas de expresión génica en el campo de la biología de sistemas.

Los sistemas de ensayo también proporcionan una nueva estrategia para el análisis de la variación somática, por ejemplo, mutaciones somáticas en el cáncer, variabilidad en respuesta a organismos infecciosos. Por ejemplo, los tumores son normalmente altamente heterogéneos, conteniendo células cancerosas así como células genéticamente normales en un ambiente local anormal. Las células cancerosas sufren mutación y selección, y en este proceso no es infrecuente que se desarrollen clones. La identificación de mutaciones somáticas relativamente raras en el contexto de tumores puede permitir el estudio de mutaciones clave en la selección de variantes clonales. Los patrones transcripcionales asociados con la angiogénesis, inflamación, u otros procesos relacionados con el cáncer en células tanto cancerosas como genéticamente normales pueden analizarse para obtener una visión de la biología del cáncer y ayudar en el desarrollo de nuevos agentes terapéuticos para el tratamiento de cánceres. En otro ejemplo, los individuos tienen una susceptibilidad variable a organismos infecciosos, y pueden usarse los sistemas de ensayo que usan la invención para estudiar la interacción entre microbios y tejidos o los diferentes tipos celulares dentro del tejido.

De manera importante, además de proporcionar información asociada espacialmente, la invención permite un gran aumento en la sensibilidad de la detección de mutaciones raras, ya que puede aumentarse dramáticamente la relación de señal a ruido al ensayarse únicamente una pequeña localización en una reacción dada. En un ensayo típico para mutaciones raras en una muestra mixta, la muestra se trata en bruto, es decir, se extraen los ácidos nucleicos de muchas células en un solo grupo. Por lo tanto, si una mutación está presente en una de cada 10.000 células, tiene que detectarse contra un fondo de ADN normal de ~10.000 células. Por el contrario, pueden analizarse muchas células con los sistemas de ensayo que usan la invención, pero se identificarían las células individuales o grupos de células mediante el sistema de codificación espacial. Por lo tanto, en los sistemas de ensayo que usan la presente invención, el fondo se reduce en órdenes de magnitud, aumentando enormemente la sensibilidad. Además, puede observarse la organización espacial de células mutantes, lo que puede ser particularmente importante para detectar mutaciones clave en secciones de tejido de un cáncer. Los análisis histológicos moleculares ya están proporcionando perspectivas en la biología del cáncer y pueden tener un potencial para su uso en diagnósticos. La tecnología de la invención es prometedora para aumentar la potencia de dichas estrategias.

Ejemplos

Los siguientes ejemplos se han incluido para proporcionar a las personas normalmente expertas en la materia una divulgación y descripción completas de cómo preparar y usar la presente invención, y no pretenden limitar el alcance de lo que el inventor considera como su invención, ni tampoco pretenden representar o implicar que los experimentos a continuación sean todos o los únicos experimentos llevados a cabo. Las presentes realizaciones, por lo tanto, deben considerarse en todos los aspectos ilustrativas y no restrictivas.

Se han realizado esfuerzos para asegurar la fiabilidad con respecto a los números utilizados (por ejemplo, cantidades,

temperatura, etc.) pero deben tenerse en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es un peso molecular promedio, la temperatura está en grados centígrados, y la presión es o está próxima a la atmosférica.

5 EJEMPLO 1: Prueba de concepto inicial del esquema de codificación

Como prueba inicial de concepto, se desarrolla un sistema modelo usando una micromatriz para demostrar un ensayo funcional monoplexado. El diseño básico valida el concepto del ensayo, y establece un ensayo funcional antes de abordar asuntos relacionados con el análisis de una muestra biológica más complicada. Se usa secuenciación convencional como lectura para esta prueba de concepto.

Se usa una micromatriz como aproximación a una sección de tejido. Las secuencias diana de la micromatriz están completamente especificadas, de tal forma que se conoce la composición de las dianas y pueden variarse de manera sistemática. Se acoplan moldes de oligonucleótidos sintéticos a un portaobjetos de vidrio mediante una modificación de amino 5'. Cada portaobjetos tiene una única secuencia de molde de oligonucleótido, y los ensayos que se llevan a cabo pueden emplear ligadura o extensión seguida de ligadura ya que esto puede ser útil para determinar determinados polimorfismos.

Una vez que se ha completado la parte *in situ* del ensayo, se eluyen los productos de reacción y se analizan mediante qPCR para determinar la presencia o ausencia de un producto y estimar el rendimiento, y mediante secuenciación convencional para determinar la estructura de los productos de ensayo. Los ensayos monoplexados que se ensayan incluyen controles positivos y negativos adecuados, y una variante de un solo nucleótido (SNV) para comprobar la capacidad para distinguir variantes de una sola base.

25 EJEMPLO 2: Escalabilidad

La complejidad del sistema de ensayo se aumenta para establecer la escalabilidad del ensayo para su uso en estudios de alto rendimiento. La escalabilidad de la codificación espacial y de los sistemas de ensayo se demuestra llevando a cabo un ensayo de 24 canales x 24 sitios usando un sistema modelo de micromatriz.

La cantidad de diana biológica, en este caso una secuencia diana de ADN, en cada localización de ensayo se varía sistemáticamente sobre el sustrato de la micromatriz. Por ejemplo, en una micromatriz con un tamaño de punto de 50 micrómetros (de centro a centro), un área de 1 mm² contiene ~400 puntos. La región alrededor de cada sitio está ocupada opcionalmente por una región que está desprovista de estos puntos para permitir la resolubilidad individual de las secuencias diana. Como alternativa, pueden agruparse los puntos con dos o más puntos directamente adyacentes rodeados por o adyacentes a una región que está desprovista de secuencias diana.

Para demostrar que la codificación espacial es precisa, los sitios comprenden diferentes composiciones de diana para demostrar que la lectura del ensayo coincide con la composición esperada de cada sitio. Con 24 secuencias diana, se efectúa un patrón digital simple, teniendo cada sitio un conjunto diferente de 12 dianas presentes y 12 dianas ausentes, para producir un código binario (0 = ausente, 1 = presente). Entonces se determina la lectura del ensayo para demostrar que las regiones detectadas coinciden con la señal esperada después de la descodificación espacial. En este ejemplo en particular, el código espacial es lo suficientemente grande (2²⁴), de tal forma que incluso unos pocos errores no darían como resultado que se mezclasen diferentes códigos. Además, este diseño permite la identificación de errores y permite una estimación no solo de la precisión de la codificación espacial, sino también de la precisión mostrando la presencia o ausencia de secuencias diana.

Se evalúa la capacidad para detectar diferencias cuantitativas generando curvas de respuesta a la dosis para cada uno de los 24 ensayos que se llevan a cabo en cada sitio de un ensayo de 24 sitios. Esto permite la estimación del límite de detección, del intervalo dinámico, y de la potencia para detectar un cambio múltiple dado a lo largo del intervalo.

En un aspecto, se usa una disposición en cuadro latino para representar dianas individuales en diferentes proporciones variando el número de características para cada diana. En otras palabras, con múltiples puntos en un sitio, el número de puntos ubicados en cada una de las 24 secuencias diana pueden variarse y cada uno de los 24 sitios puede tener una composición distinta. Una micromatriz de 2,54 x 7,62 cm es lo suficientemente grande como para permitir múltiples replicados. Este conjunto mayor de 24 secuencias necesitará desconvolución, y esto se logra usando técnicas de alto rendimiento, tales como tecnologías de secuenciación de última generación (por ejemplo, la tecnología SOLiD™ (Life Technologies, Inc., Carlsbad, CA) o Genome Analyzer (Illumina, Inc., San Diego, CA)). El uso del ensayo de 24 canales demuestra tanto la precisión de la codificación y descodificación espacial, como la respuesta cuantitativa del sistema de ensayo.

EJEMPLO 3: Adaptación del ensayo a muestras conservadas

Se ensaya ADN genómico como prueba de concepto para ensayar ARN, ya que proporciona una forma de establecer una señal de referencia de una sola copia. Una vez que se ha desarrollado un ensayo funcional para muestras de

- 5 FFPE, se adapta a un ensayo de ARN. Para este fin, se ensayan concentraciones de oligonucleótido de ensayo para asegurar la compatibilidad con una alta multiplexación. Asumiendo un diámetro celular de 10 micrómetros, y una administración de una gota de reactivo de 10 micrómetros de diámetro a una célula individual, el volumen de la gota será de ~500 μ l y puede contener $\sim 3 \times 10^{11}$ moléculas a una concentración de 1 μ M. Para ensayar 1.000 secuencias diana en 10.000 células, se necesitarán ~2.000 oligonucleótidos diana en una gota. Por lo tanto, cada gota podría contener ~160 millones de copias de cada oligo de ensayo, un gran exceso frente a los pocos miles de secuencias en una célula.
- 10 La manipulación de números absolutos pequeños de moléculas de producto generados a partir de muestras muy pequeñas o comprometidas está potenciada para contrarrestar el problema de la baja eficacia de recuperación; es decir, la eficacia de la elución y se evitan las pérdidas resultantes de la absorción de las moléculas en las superficies. Una estrategia para abordar el último problema es incluir un material transportador, tal como glucógeno o ácidos nucleicos transportadores.
- 15 **EJEMPLO 4: Adaptación del ensayo a una muestra biológica.**
- 20 Se inmoviliza un molde de ARN de control sobre un soporte sólido para crear un sistema artificial. El ensayo se lleva a cabo usando ADN ligasa de T4, que puede reparar daños en los híbridos de ADN/ARN. Los ensayos se llevan a cabo en portaobjetos emparejados, o en diferentes secciones del mismo portaobjetos, donde se ensaya en un caso ADNg y en el otro se ensaya ARN. Cuando se ensaya ADNg, puede tratarse previamente al portaobjetos con RNasa, y cuando se ensaya ARN el portaobjetos se trata previamente con DNasa. Los resultados del ensayo se confirman extrayendo ADNd o ARN y cuantificando las cantidades relativas mediante qPCR o TR-qPCR, respectivamente.
- 25 Para hacer que los ensayos de ARN de sección de tejido sean lo más informativos posibles, se usa información preexistente acerca de los niveles de expresión en tejidos biológicos para dirigirse a los transcritos a lo largo de un intervalo de abundancias en el diseño del ensayo. Los transcritos tanto de alta abundancia, así como algunos transcritos de media y baja abundancia, se usan como diana para permitir una evaluación inicial de las características de rendimiento cuantitativo del ensayo.
- 30 Lo anterior solo ilustra los principios de la invención. Además, todos los ejemplos y el lenguaje condicional citado en el presente documento pretenden ayudar al lector a comprender los principios de la invención y los conceptos aportados por los inventores para desarrollar la técnica, y deben entenderse como pertenecientes sin limitación a dichos ejemplos y condiciones citados específicamente. Además, todas las afirmaciones en el presente documento que citan principios, aspectos, y realizaciones de la invención así como ejemplos específicos de la misma pretenden abarcar
- 35 equivalentes tanto estructurales como funcionales de los mismos. Además, se pretende que dichos equivalentes incluyan tanto equivalentes conocidos en la actualidad como equivalentes desarrollados en el futuro, es decir, cualquier elemento desarrollado que cumpla la misma función, independientemente de la estructura. El alcance de la presente invención, por lo tanto, no pretende estar limitado a las realizaciones ejemplares mostradas y descritas en el presente documento.
- 40

REIVINDICACIONES

1. Un método para determinar patrones espaciales de abundancia o de actividad o de ambos de múltiples dianas de ácido nucleico en múltiples sitios en una muestra, comprendiendo las siguientes etapas:
 - 5 proporcionar una muestra fijada a un soporte; administrar sondas oligonucleotídicas para múltiples dianas de ácido nucleico a los múltiples sitios en la muestra;
 - permitir que las sondas oligonucleotídicas hibriden con las dianas de ácido nucleico;
 - lavar las sondas oligonucleotídicas no hibridadas de la muestra;
 - 10 administrar agentes de codificación a localizaciones de los múltiples sitios en la muestra de acuerdo con un patrón espacial conocido, en donde al menos dos agentes codificantes se administran a cada uno de los múltiples sitios y la combinación de los agentes de codificación administrados a cada sitio es diferente;
 - acoplar los agentes de codificación a las sondas oligonucleotídicas hibridadas a las dianas de ácido nucleico para formar sondas codificadas;
 - 15 determinar la totalidad o parte de una secuencia de las sondas codificadas; y
 - asociar la abundancia o la actividad o ambas de las múltiples dianas biológicas a las localizaciones de múltiples sitios en la muestra.

2. Un método para determinar patrones espaciales de abundancia o de actividad o de ambos de múltiples dianas biológicas en múltiples sitios en una muestra, comprendiendo las siguientes etapas:
 - 20 proporcionar una muestra fijada a un soporte;
 - administrar sondas codificadas para múltiples dianas biológicas con un patrón espacial conocido a los múltiples sitios en la muestra, en de cada sonda codificada comprende una región de sonda capaz de interactuar con una
 - 25 diana biológica y un marcador de codificación que comprende un oligonucleótido que identifica la localización del sitio al que se administra la sonda codificada;
 - permitir que las sondas codificadas interactúen con las dianas biológicas;
 - determinar una secuencia de las sondas codificadas, en donde la secuencia comprende la secuencia oligonucleotídica del marcador de codificación, identificando de este modo la localización del sitio al que se
 - 30 administra la sonda codificada; y
 - asociar la abundancia o la actividad o ambas de las múltiples dianas biológicas a las localizaciones de los sitios en la muestra.

3. El método de la reivindicación 2, que además comprende después de la etapa en que se deja reaccionar y antes de la etapa de determinación una etapa de separación de las sondas codificadas que interactúan con las dianas biológicas de las sondas codificadas que no interactúan con las dianas biológicas.

4. El método de la reivindicación 3, en el que las dianas biológicas son ácidos nucleicos y las sondas codificadas son sondas oligonucleotídicas.

- 40 5. El método de las reivindicaciones 1 o 4, en el que hay dos sondas oligonucleotídicas para cada una de las dianas biológicas de ácido nucleico.

6. El método de la reivindicación 3, en el que las múltiples dianas biológicas son proteínas, las regiones de sonda de las sondas codificantes son proteínas y los marcadores codificantes comprenden oligonucleótidos.

7. El método de la reivindicación 3, en el que las múltiples dianas biológicas comprenden enzimas.

8. El método de la reivindicación 3, en el que las regiones de sonda de las sondas codificadas comprenden anticuerpos, aptámeros, moléculas pequeñas, sustratos enzimáticos, sustratos enzimáticos putativos o agentes de
- 50 captura por afinidad.

9. El método de una cualquiera de las reivindicaciones 1, 3 y 6, que además comprende una etapa de amplificación entre la etapa de separación o la etapa de acoplamiento y la etapa de determinación.

- 55 10. El método de una cualquiera de las reivindicaciones 1-3, en el que la etapa de determinación se lleva a cabo mediante secuenciación de ácido nucleicos o secuenciación de alto rendimiento.

11. El método de la reivindicación 1, en el que la etapa de acoplamiento se lleva a cabo mediante ligadura o mediante
- 60 extensión seguida de ligadura.

12. El método de una cualquiera de las reivindicaciones 1-3, en el que el producto de las múltiples dianas biológicas que se están ensayando y los múltiples sitios en la muestra es mayor de 1.000.000.

- 65 13. El método de una cualquiera de las reivindicaciones 1-3 y 6, en el que se determinan las secuencias de al menos un millón de sondas codificantes en paralelo.

14. El método de una cualquiera de las reivindicaciones 1-3, en el que el patrón espacial conocido se determina mediante características histológicas de la muestra.

5 15. El método de una cualquiera de las reivindicaciones 1-3, en el que un hardware programado por un programa informático efectúa al menos dos etapas de la etapa de administración, la etapa de separación, la etapa de determinación y la etapa de asociación.

100

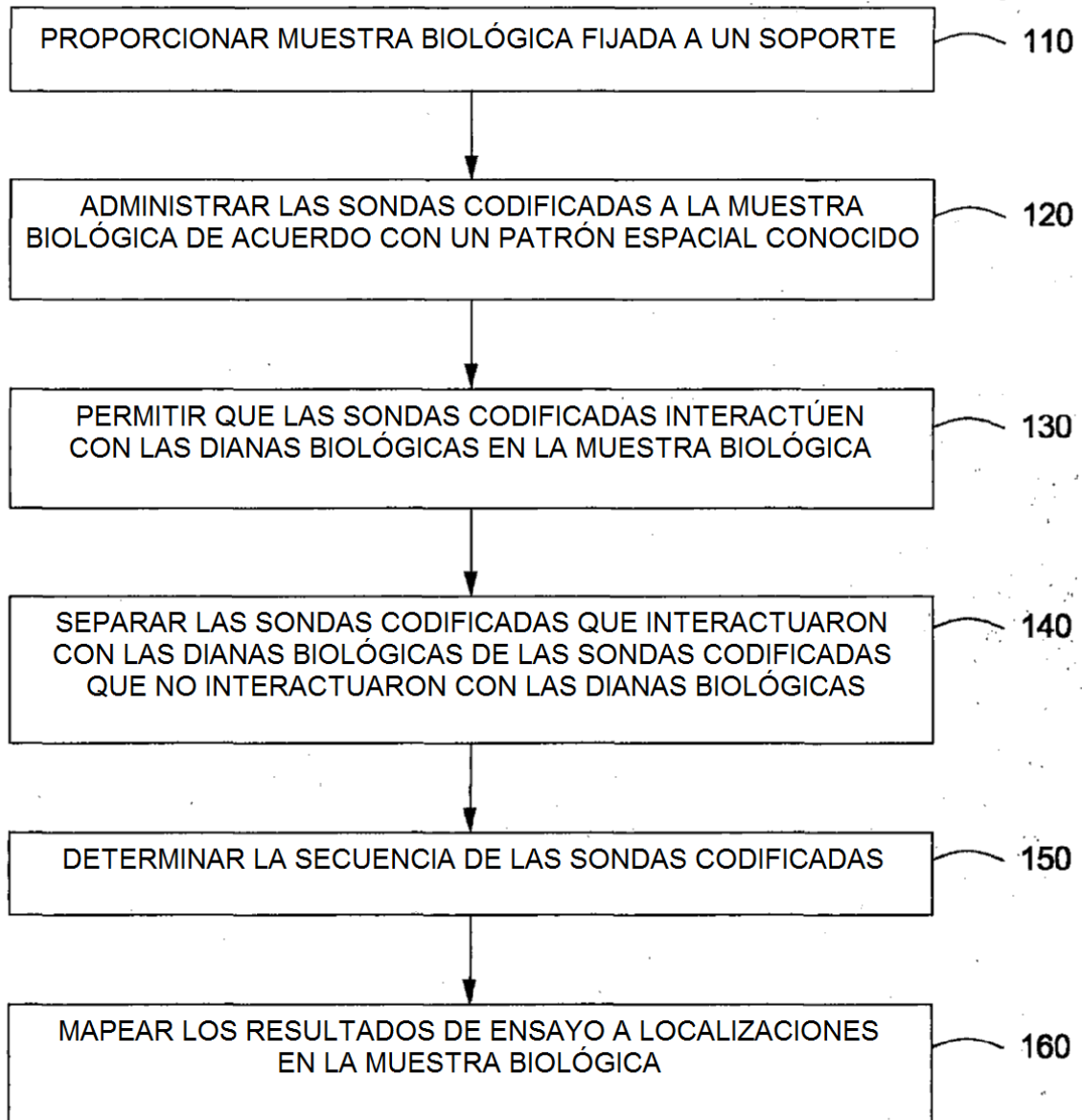


FIG. 1

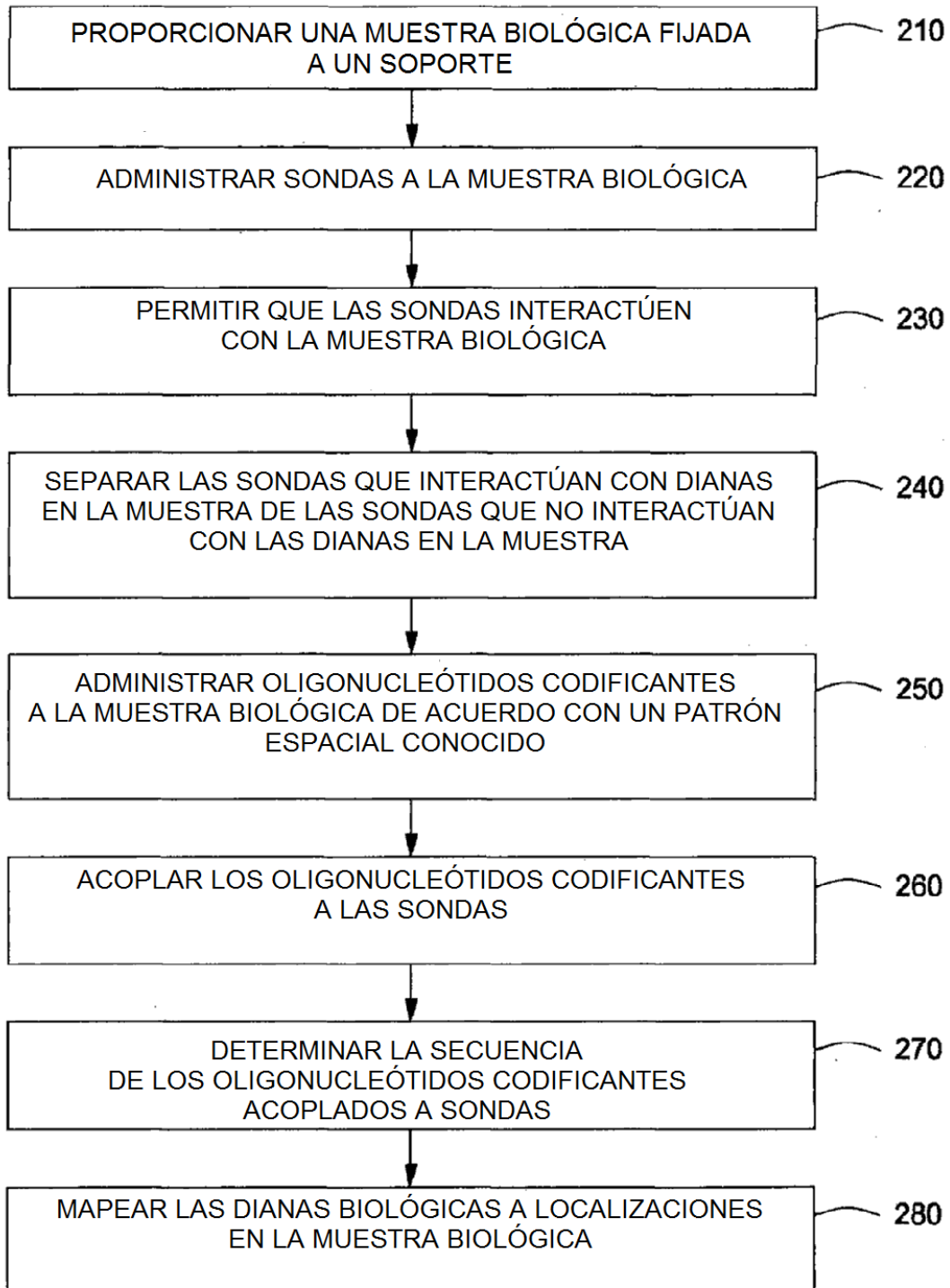
200

FIG. 2

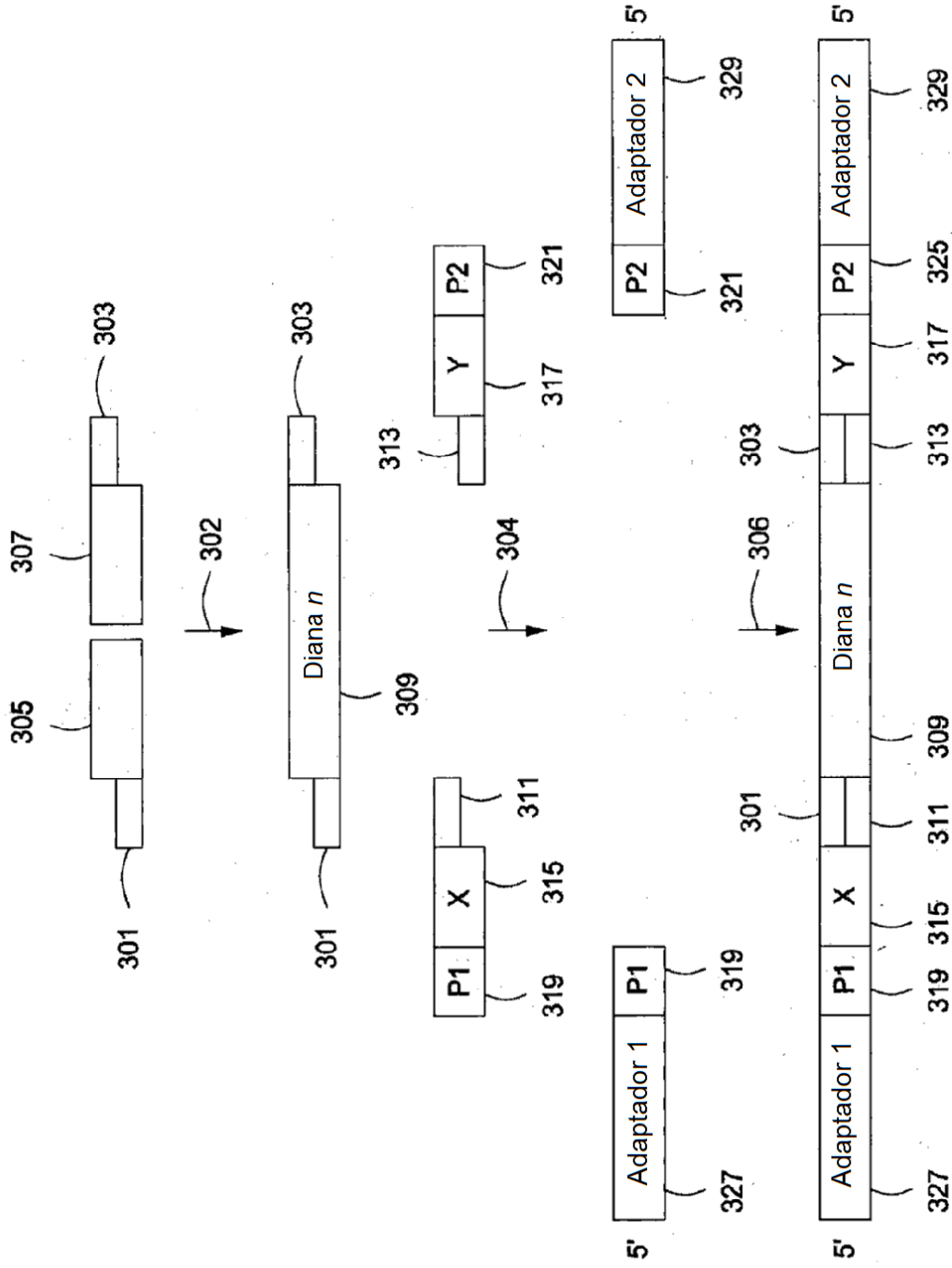


FIG. 3

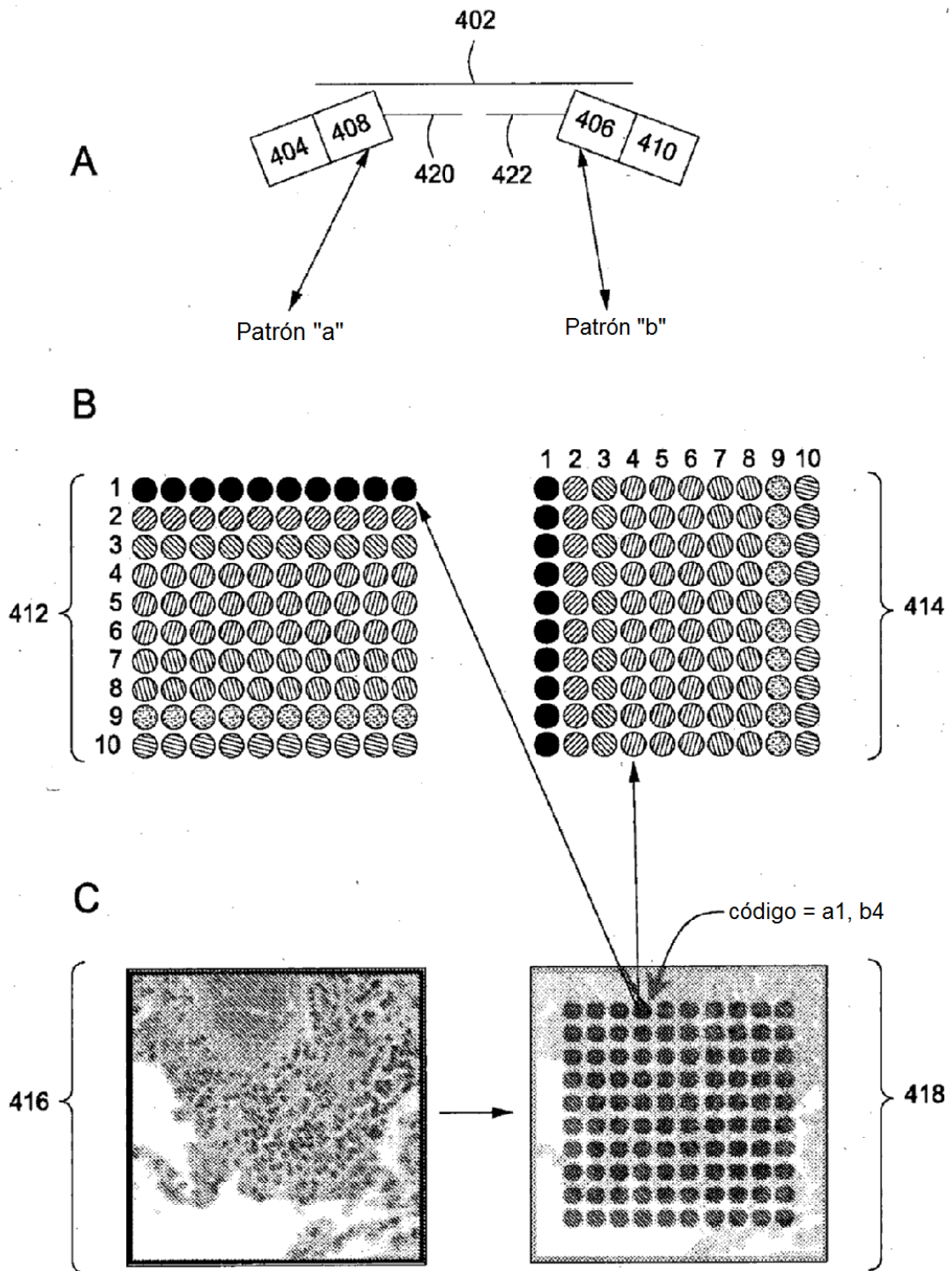


FIG. 4

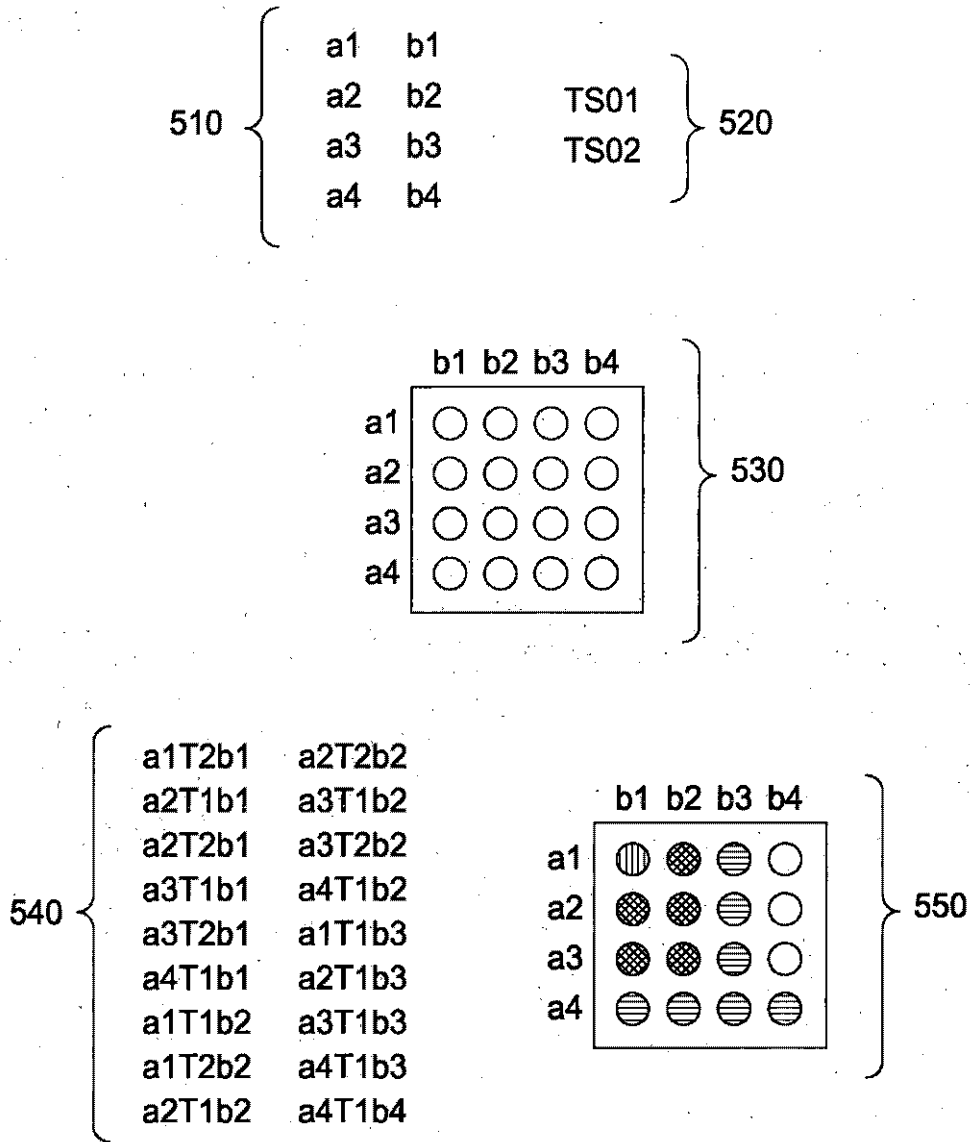


FIG. 5