

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 959 190**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.07.2016** **PCT/US2016/043487**

87 Fecha y número de publicación internacional: **02.02.2017** **WO17019481**

96 Fecha de presentación y número de la solicitud europea: **22.07.2016** **E 16831111 (6)**

97 Fecha y número de publicación de la concesión europea: **05.07.2023** **EP 3325669**

54 Título: **Composiciones y métodos de análisis de ARN**

30 Prioridad:

24.07.2015 US 201562196725 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
21.02.2024

73 Titular/es:

**THE JOHNS HOPKINS UNIVERSITY (100.0%)
Technology Ventures, 100 N. Charles Street 5th
Floor
Baltimore, MD 21201, US**

72 Inventor/es:

**LARMAN, HARRY, BENJAMIN y
CREDLE, JOEL**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 959 190 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

Composiciones y métodos de análisis de ARN

Solicitudes relacionadas

5 Esta solicitud es una Solicitud de Patente Internacional que reivindica el beneficio de prioridad según el Artículo 35 U.S.C. 119(e) de la Solicitud Provisional Núm.: 62/196.725, presentada el 24 de julio de 2015 y titulada "Compositions and Methods of RNA Analysis".

Antecedentes de la invención

10 La medicina de precisión se basa en la capacidad de investigadores y patólogos para caracterizar molecularmente especímenes de pacientes, incluidos tejidos resecados, p. ej., tejidos tumorales y materiales de biopsia. El desarrollo de técnicas avanzadas para el análisis de ADN, tales como la secuenciación de ADN de alto rendimiento (o secuenciación de ADN de próxima generación, NGS), ha contribuido en gran medida a la detección de lesiones genéticas asociadas con el cáncer. Sin embargo, la capacidad de caracterizar perfiles de expresión génica en muestras de tejido fijadas (p. ej., secciones de tejido fijadas con formalina e incluidas en parafina) se ha quedado atrás porque el procedimiento de fijación degrada la calidad del ARN en la muestra, lo que inhibe significativamente la capacidad de extraer información importante de la muestra.

15 El documento USRE38442E se refiere a un método de amplificación de ácidos nucleicos y a un método de amplificación de señales de hibridación (HSAM).

20 Dado que la información sobre los patrones de expresión génica en tales especímenes de tejido fijado es de gran valor, existe una necesidad insatisfecha de una plataforma que permita un análisis sólido del ARN dentro de especímenes de tejido fijado.

Compendio de la invención

La presente divulgación proporciona métodos para analizar la presencia y el nivel de expresión de ácidos nucleicos (p. ej., ARN) en una muestra (p. ej., una célula, un órgano, un tejido y similares).

La presente invención se refiere a un método que comprende:

25 aplicar una o más sondas multipartitas a una muestra fijada seleccionada del grupo que consiste en una célula, un órgano, un tejido y cualquier combinación de los mismos, en donde cada una de las una o más sondas multipartitas incluye al menos dos subsondas y en donde al menos dos subsondas se reasocian con una secuencia diana contigua en un ácido nucleico diana dentro de la muestra fijada:

30 reasociar al menos una de las una o más sondas multipartitas aplicadas con al menos un ARN diana dentro de la muestra fijada sin *extracción* de ARN;

ligar enzimáticamente al menos un extremo 3' a al menos un extremo 5-fosforilado de las al menos dos subsondas asociadas con la al menos una sonda multipartita reasociada dentro de la muestra fijada para crear un proxy de ácido nucleico diana que puede ser detectado;

liberar el proxy de ácido nucleico diana del ácido nucleico diana; y

35 amplificar el proxy de ácido nucleico diana.

40 En un aspecto, las subsondas comprenden un sitio de unión del cebador adjunto que facilita la amplificación posterior a la ligación del proxy de ácido nucleico diana. En algunos aspectos, cada subsonda comprende un oligonucleótido. En otros aspectos, las subsondas (que también se pueden denominar "subsondasceptoras") tienen una terminación 3' de al menos dos bases de ARN. En un aspecto, la subsonda es un oligonucleótido de ADN que tiene una terminación 3' de al menos dos bases de ARN.

Las al menos dos subsondas están ligadas con una enzima. La enzima puede ser una ligasa. En realizaciones particulares, la enzima puede ser una de las siguientes ligasas: una ARN ligasa 2 de T4 (Rnl2), una ADN ligasa de T4, una ADN ligasa del virus de Chlorella (ADN ligasa de PBCV-1), un derivado de Rnl2, un derivado de PBCV-1 o cualquier combinación de los mismos.

45 En algunos ejemplos, la muestra se puede fijar utilizando reactivos sin formalina, incluidos, por ejemplo, glutaraldehído, mercurial, agentes oxidantes, alcoholes y picratos. En otros ejemplos, la muestra puede incluir células fijadas en suspensión o un cultivo de tejido fijado. La muestra se puede obtener en una forma muy degradada antes de la fijación. En particular, la muestra puede ser una muestra de tejido fijada con formalina e incluida en parafina (FFPE).

50 En una realización, el proxy de ácido nucleico diana se puede liberar del ácido nucleico diana mediante una endonucleasa o recuperarse desnaturalizando el proxy de ácido nucleico diana que contiene al menos dos subsondas

ligadas. La endonucleasa puede ser ARNasaH, ARNasaA, ARNasa If o ARNasaHIII.

En una realización, el proxy de ácido nucleico diana amplificado se puede analizar utilizando una técnica tal como, por ejemplo, Secuenciación de Próxima Generación (NGS), Secuenciación Profunda, secuenciación basada en espectrometría de masas o análisis de longitud, secuencia de fragmentos de ADN o análisis de longitud mediante electroforesis en gel o electroforesis capilar y/o hibridación sobre sondas de detección inmovilizadas.

En una realización, el espécimen de tejido fijado se puede procesar en secciones que tienen un espesor de 1-100 micrómetros, p. ej., 1 micrómetro, 5 micrómetros, 10 micrómetros, 15 micrómetros, 20 micrómetros, 25 micrómetros, 30 micrómetros, 35 micrómetros, 40 micrómetros. 45 micrones, 50 micrones, 60 micrones, 65 micrones, 70 micrones, 75 micrones, 80 micrones, 85 micrones, 90 micrones, 95 micrones o 100 micrones.

En una realización, las al menos dos subsondas de la pluralidad de sondas multipartitas pueden tener una longitud de aproximadamente 10-200 nucleótidos, p. ej., una longitud de aproximadamente 10, aproximadamente 15, aproximadamente 20, aproximadamente 25, aproximadamente 30, aproximadamente 35, aproximadamente 40, aproximadamente 50, aproximadamente 75, aproximadamente 100, aproximadamente 125, aproximadamente 150, aproximadamente 175 o aproximadamente 200 nucleótidos. En otra realización, cada una de las al menos dos subsondas de la pluralidad de sondas multipartitas puede tener aproximadamente una longitud de 15-30 nucleótidos de longitud, p. ej., una longitud de aproximadamente 15, aproximadamente 16, aproximadamente 17, aproximadamente 18, aproximadamente 19, aproximadamente 20., aproximadamente 21, aproximadamente 22, aproximadamente 23, aproximadamente 24, aproximadamente 25, aproximadamente 26, aproximadamente 27, aproximadamente 28, aproximadamente 29 o aproximadamente 30 nucleótidos.

En una realización, el proxy de ácido nucleico se puede amplificar utilizando PCR. En una realización, la PCR incluye aproximadamente 20-50 ciclos, p. ej., aproximadamente 20, aproximadamente 25, aproximadamente 30, aproximadamente 35, aproximadamente 40, aproximadamente 45 o aproximadamente 50 ciclos.

En una realización relacionada, cada una de la pluralidad de sondas multipartitas incluye tres subsondas.

Opcionalmente, el método mencionado anteriormente se realiza secuencialmente en una muestra, es decir, el método se realiza dos veces, tres veces, cuatro veces, cinco veces, seis veces, siete veces, ocho veces, nueve veces o diez veces en la misma muestra. En algunos casos, el espécimen (es decir, la muestra) se archiva entre ensayos secuenciales.

Los métodos descritos en el presente documento también son útiles para diagnosticar infecciones, p. ej., infecciones virales, infecciones bacterianas o infecciones fúngicas. Por ejemplo, la muestra se obtiene de un sujeto sospechoso o en riesgo de desarrollar una infección viral, una infección bacteriana o una infección fúngica. Opcionalmente, el ácido nucleico diana se selecciona del grupo que consiste en un ácido nucleico viral, un ácido nucleico bacteriano y un ácido nucleico fúngico. En algunos casos, el método comprende adicionalmente liberar el proxy de ácido nucleico diana del ácido nucleico diana; amplificar el proxy de ácido nucleico diana; y secuenciar el proxy del ácido nucleico diana, identificando así un ácido nucleico viral, un ácido nucleico bacteriano o un ácido nucleico fúngico y diagnosticando una infección viral, una infección bacteriana o una infección fúngica, respectivamente, en el sujeto. Preferiblemente, después del diagnóstico, el sujeto se trata con un agente antifúngico, un agente antibacteriano o un agente antiviral.

Las infecciones fúngicas ilustrativas pueden incluir infecciones derivadas de los siguientes organismos: *Acremonium* sp., *Aspergillus clavatus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus glaucus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus ochraceus*, *Aspergillus terreus*, *Aspergillus unguis*, *Aspergillus ustus*, *Beauveria* sp., *Bipolaris* sp., *Blastoschizomyces* sp., *Blastomyces dermatitidis*, *Candida albicans*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lusitanae*, *Candida parapsilosis*, *Candida tropicalis*, *Chrysosporium* sp., *Cladosporium* sp., *Coccidioides immitis*, *Cryptococcus neoformans* var *gattii* serotipo B, *Cryptococcus neoformans* serotipo A, *Cryptococcus laurentii*, *Cryptococcus terreus*, *Curvularia* sp., *Fusarium* sp., *Filobasidium capsuligenum*, *Filobasidiella* (Cryptococcus) *neoformans* var *bacillispora* serotipo C, *Filobasidiella* (Cryptococcus) *neoformans* var *neoformans* serotipo D, *Filobasidium uniguttulatum*, *Geotrichum* sp., *Histoplasma capsulatum*, *Malbranchea* sp., *Mucor* sp., *Paecilomyces* sp., *Paracoccidioides brasiliensis*, especies de *Penicillium*, *Pneumocystis carinii*, *Pseudallescheria boydii*, *Rhizopus* sp., *Sporothrix schenckii*, *Scopulariopsis brevicaulis* sp., *Scopulariopsis brumptii*, *Saccharomyces cerevisiae*, y *Trichosporon beigelii*.

Las infecciones bacterianas ilustrativas pueden incluir infecciones derivadas de los siguientes organismos: *Bacillus anthracis*, *Bordetella pertussis*, *Borrelia burgdorferi*, *Brucella abortus*, *Brucella canis*, *Brucella melitensis*, *Brucella suis*, *Brucellosis*, *Campylobacter jejuni*, infección respiratoria por *Chlamydia pneumoniae*, *Chlamydia psittaci*, *Chlamydia trachomatis*, *Linfogranuloma venereum*, *Clostridium botulinum*, *Clostridium difficile*, *Clostridium perfringens*, *Clostridium tetani*, *Corynebacterium diphtheria*, *Enterococcus faecalis*, *Enterococcus faecium*, *Escherichia coli*, *Francisella tularensis*, *Haemophilus influenza*, *Helicobacter pylori*, *Legionella pneumophila*, *Leptospira interrogans*, *Listeria monocytogenes*, *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Rickettsia*, *Salmonella typhi*, *Salmonella typhimurium*, *Shigella sonnei*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus saprophyticus*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Treponema pallidum*, *Vibrio cólera*,

Yersinia pestis.

Las infecciones virales ilustrativas pueden incluir infecciones derivadas de los siguientes organismos: *Adenovirus*, *virus de Influenza aviar*, virus tipo A de *Influenza*, virus tipo B de *Influenza*, sarampión, virus de *Parainfluenza*, *Virus sincitial respiratorio* (VRS), *Rinovirus*, *SARS-CoV* (Síndrome Respiratorio Agudo Severo), virus *Coxsackie*, *Enterovirus*, *Poliovirus*, *Rotavirus*, virus de la *Hepatitis B*, virus de la *Hepatitis C*, *Virus de la diarrea viral bovina* (sustituto), *Herpes simple 1*, *Herpes simple 2*, *Citomegalovirus humano*, *Virus de la varicela zóster*, *Virus de la inmunodeficiencia humana 1* (VIH-1), *Virus de la inmunodeficiencia humana 2* (VIH-2), *Virus de la inmunodeficiencia de simios* (SIV), *Virus de la inmunodeficiencia de simios similar al humano* (SHIV), *Virus del dengue*, *Hantavirus*, Virus de la fiebre hemorrágica, *Linfocitis*, Virus de la *coromeningitis*, Sustitutos del virus de la viruela (Viruela vacuna, Viruela del mono, Viruela del conejo), Virus *vaccinia*, Virus de la *encefalomielitis equina venezolana* (EEV), Virus del Nilo Occidental, Virus de la fiebre amarilla, Virus Zika.

En una realización, la muestra es una muestra de tejido fijada con formalina e incluida en parafina (FFPE).

En una realización, el proxy del ácido nucleico diana se somete a amplificación en fase sólida. En una realización, la amplificación por PCR en fase sólida puede ser una PCR Puente o una amplificación por círculo rodante (RCA). En una realización, la superficie diana comprende portaobjetos de vidrio recubiertos de polímero que contienen cebadores de PCR inmovilizados.

En una realización, la muestra de tejido comprende tejidos tumorales resecados o materiales de biopsia.

En otra realización, el método comprende adicionalmente estampar las sondas ligadas sobre una superficie de réplica recubierta con una pluralidad de cebadores de PCR inmovilizados. En realizaciones, este método puede denominarse "enfoque de estampado" o "método de estampado".

En un aspecto, el método de estampado descrito anteriormente puede utilizar una muestra de tejido fijada con formalina e incluida en parafina (FFPE), tejidos tumorales resecados o materiales de biopsia. En otras realizaciones, el proxy de ácido nucleico diana se somete a amplificación en fase sólida, que incluye, por ejemplo, PCR Puente o amplificación por círculo rodante (RCA), que incluye portaobjetos de microscopio de vidrio recubiertos de polímero que contienen cebadores de PCR Puente inmovilizados.

Definiciones

Por "agente" se entiende cualquier compuesto pequeño, anticuerpo, molécula de ácido nucleico o polipéptido, o fragmentos de los mismos.

Por "alteración" se entiende un cambio (aumento o disminución) en los niveles de expresión o actividad de un gen o polipéptido detectado mediante métodos conocidos en la técnica convencionales tales como los descritos en el presente documento. Como se emplea en el presente documento, una alteración incluye un cambio de 10% en los niveles de expresión, preferiblemente un cambio de 25%, más preferiblemente un cambio de 40% y lo más preferiblemente un cambio de 50% o más en los niveles de expresión.

La "PCR Puente" es un método de amplificación por PCR en fase sólida localizada, en donde los fragmentos de ADN se amplifican mediante cebadores anclados a una superficie (en lugar de en solución), formando así agrupamientos de ADN inmovilizados en la superficie. Si los agrupamientos están suficientemente separados, se pueden secuenciar clonalmente.

En esta divulgación, "comprende", "que comprende", "que contiene" y "que tiene" y similares pueden tener el significado que se les atribuye en la ley de patentes de EE. UU. y pueden significar "incluye", "que incluye" y similares; "que consiste esencialmente en" o "consiste esencialmente" también tiene el significado que se le atribuye en la ley de patentes de EE. UU. y el término es abierto, lo que permite la presencia de más de lo que se menciona, siempre y cuando se no se modifiquen las características básicas o novedosas de lo que se menciona por la presencia de más de lo que se menciona, pero excluye las realizaciones de la técnica anterior.

"Detectar" se refiere a identificar la presencia, ausencia o cantidad del ácido nucleico (p. ej., ARN) que se va a detectar.

Por "marca detectable" se entiende una composición que cuando se conecta a una molécula de interés hace que esta última sea detectable, por ejemplo, mediante medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos o químicos. Por ejemplo, las marcas útiles pueden incluir isótopos radiactivos, cuentas magnéticas, cuentas metálicas, partículas coloidales, colorantes fluorescentes, reactivos densos en electrones, enzimas (por ejemplo, como se emplean comúnmente en un ELISA), biotina, digoxigenina o haptenos.

Por "cantidad eficaz" se entiende la cantidad de un agente necesaria para mejorar los síntomas de una enfermedad con relación a un paciente no tratado. La cantidad eficaz del agentes o agentes activos para el tratamiento terapéutico de una enfermedad varía dependiendo de la forma de administración, la edad, el peso corporal y la salud general del sujeto. En última instancia, el médico o el veterinario a cargo decidirán la cantidad y el régimen de dosificación adecuados. Esta cantidad se denomina cantidad "eficaz".

Por "FFPE" se entiende tejido fijado con formalina e incluido en parafina. Las muestras FFPE se obtienen de tejidos (a menudo muestras sospechosas de tumores) que se fijan con formalina para preservar las características estructurales, espaciales y biomoleculares (p. ej., estructura citoesquelética y proteica) y después se incluyen en un tipo de cera de parafina para que el tejido pueda cortarse, típicamente 5-10 micras de espesor. La formalina entrecruza proteínas de forma irreversible a través de los grupos amino, preservando así la integridad estructural de las células para que se puedan teñir con colorantes o con inmunotinciones utilizadas para analizar anomalías en el tejido que indiquen condiciones celulares alteradas, p. ej., cáncer. Sin embargo, el efecto de estos fijadores de entrecruzamiento sobre los ácidos nucleicos de ARN y ADN dentro de la muestra es perjudicial para la sensibilidad y especificidad que se pueden lograr en los ensayos moleculares actuales, p. ej., ensayos moleculares que utilizan ADN o ARN derivado de muestras FFPE. Además, se pueden preparar muestras utilizando reactivos sin formalina, incluidos, por ejemplo, glutaraldehído, mercurial, agentes oxidantes, alcoholes y picratos.

Por "fragmento" se entiende una porción de un polipéptido o molécula de ácido nucleico. Esta porción contiene, preferiblemente, al menos 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% o 90% de la longitud total de la molécula de ácido nucleico o polipéptido de referencia. Un fragmento puede contener 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 nucleótidos o aminoácidos.

"Hibridación" significa enlaces de hidrógeno, que pueden ser enlaces de hidrógeno de Watson-Crick, Hoogsteen o Hoogsteen inverso, entre nucleobases complementarias. Por ejemplo, la adenina y la timina son nucleobases complementarias que se emparejan mediante la formación de enlaces de hidrógeno.

Por "infección" se entiende la invasión del cuerpo de un organismo por agentes causantes de enfermedades, su multiplicación y la reacción del anfitrión ante estos organismos y las toxinas que producen. La infección puede ser causada por cualquier microbio/microorganismo, incluidos, por ejemplo, bacterias, hongos y virus. Los microorganismos pueden incluir todas las especies bacterianas, arqueales y protozoarias. Este grupo también contiene algunas especies de hongos, algas y ciertos animales. En algunas realizaciones, los virus también pueden clasificarse como microorganismos.

Por "polinucleótido aislado" se entiende un ácido nucleico (p. ej., ARN, ADN, ADNc, etc.) que está libre de los genes que, en el genoma natural del organismo del que se obtiene la molécula de ácido nucleico, flanquean el gen. Por tanto, el término incluye, por ejemplo, un ADN recombinante que se incorpora a un vector; a un plásmido o virus que se replica de forma autónoma; o al ADN genómico de un procariota o eucariota; o que existe como una molécula separada (por ejemplo, un ADNc o un fragmento genómico o de ADNc producido por PCR o digestión con endonucleasa de restricción) independiente de otras secuencias. Además, el término incluye una molécula de ARN que se transcribe a partir de una molécula de ADN, así como un ADN recombinante que forma parte de un gen híbrido que codifica una secuencia polipeptídica adicional.

La "microdissección por captura láser" o "LCM" es un método para aislar células específicas de regiones microscópicas de tejidos, células u organismos. LCM es un método para obtener subpoblaciones de células de tejido bajo visualización microscópica directa. La tecnología LCM puede recolectar las células de interés directamente o aislar células específicas cortando las células no deseadas para obtener poblaciones de células enriquecidas histológicamente puras.

La hibridación con ligación in situ o "LISH" es una metodología descrita en el presente documento para la medición multiplexada de la expresión génica, adecuada para el análisis de especímenes de tejido fijados. Una realización de LISH utiliza la ARN Ligasa 2 de T4 (Rnl2) y las sondas de hibridación quiméricas de ADN-ARN, que cuando se reasocian, quedan ligadas, donde sirven como proxies fieles del nivel de expresión de una diana respectiva.

"LISH-stAmp" es un método para estampar y amplificar el producto de LISH en una superficie diana para preservar la información espacial.

Por "marcador" se entiende cualquier proteína o polinucleótido que tenga una alteración en el nivel de expresión o actividad que esté asociada con una enfermedad o trastorno.

Por "modular" se entiende alterar (aumentar o disminuir). Tales alteraciones se detectan mediante métodos convencionales conocidos en la técnica, como los descritos en el presente documento.

Por "neoplasia" se entiende una enfermedad o trastorno caracterizados por una proliferación excesiva o una apoptosis reducida. Los neoplasmas ilustrativos para los que se puede utilizar la invención incluyen, pero sin limitarse a, cáncer de páncreas, leucemias (p. ej., leucemia aguda, leucemia linfocítica aguda, leucemia mielocítica aguda, leucemia mieloblástica aguda, leucemia promielocítica aguda, leucemia mielomonocítica aguda, leucemia monocítica aguda, eritroleucemia aguda, leucemia crónica, leucemia mielocítica crónica, leucemia linfocítica crónica), policitemia vera, linfoma (enfermedad de Hodgkin, enfermedad no Hodgkin), macroglobulinemia de Waldenstrom, enfermedad de las cadenas pesadas y tumores sólidos tales como sarcomas y carcinomas (p. ej., fibrosarcoma, mixosarcoma), liposarcoma, condrosarcoma, sarcoma osteogénico, cordoma, angiosarcoma, endoteliosarcoma, linfangiosarcoma, linfangioendoteliosarcoma, sinovioma, mesotelioma, tumor de Ewing, leiomiomasarcoma, rabdomiosarcoma, carcinoma de colon, cáncer de mama, cáncer de ovario, cáncer de próstata, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma de glándulas sebáceas, carcinoma

papilar, adenocarcinomas papilares, cistadenocarcinoma, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma de conductos biliares, coriocarcinoma, seminoma, carcinoma embrionario, tumor de Wilm, cáncer de cuello uterino, cáncer de útero, cáncer testicular, carcinoma de pulmón, carcinoma de pulmón de células pequeñas, carcinoma de vejiga, carcinoma epitelial, glioma, glioblastoma multiforme, astrocitoma, meduloblastoma, craneofaringioma, ependimoma, pinealoma, hemangioblastoma, neuroma acústico, oligodendroglioma, schwannoma, meningioma, melanoma, neuroblastoma y retinoblastoma).

Por "NGS" se entiende Secuenciación de Próxima Generación. Las plataformas NGS realizan secuenciación masiva en paralelo, durante la cual se secuencian al unísono millones de fragmentos de ADN de una sola muestra. La tecnología de secuenciación masiva en paralelo facilita la secuenciación de alto rendimiento, lo que permite secuenciar un genoma completo en menos de un día. La creación de plataformas NGS ha hecho que la secuenciación sea accesible a más laboratorios, lo que ha aumentado rápidamente la cantidad de investigaciones y diagnósticos clínicos que se realizan con la secuenciación de ácidos nucleicos.

Los polinucleótidos que tienen "identidad sustancial" con una secuencia endógena típicamente son capaces de hibridar con al menos una hebra de una molécula de ácido nucleico de hebra sencilla o doble. Por "hibridar" se entiende emparejar para formar una molécula de doble hebra entre secuencias de polinucleótidos complementarias (p. ej., un gen descrito en el presente documento), o porciones de las mismas, en diversas condiciones de rigurosidad. (Véase, p. ej., Wahl, G. M. y S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

Por ejemplo, la concentración de sal rigurosa normalmente será inferior a aproximadamente 750 mM de NaCl y 75 mM de citrato trisódico, preferiblemente menos de aproximadamente 500 mM de NaCl y 50 mM de citrato trisódico, y más preferiblemente menos de aproximadamente 250 mM de NaCl y 25 mM de citrato trisódico. La hibridación de baja rigurosidad se puede obtener en ausencia de disolvente orgánico, p. ej., formamida, mientras que la hibridación de alta rigurosidad se puede obtener en presencia de al menos aproximadamente 35% de formamida, y más preferiblemente al menos aproximadamente 50% de formamida. Las condiciones de temperatura rigurosas normalmente incluirán temperaturas de al menos aproximadamente 30°C, más preferiblemente de al menos aproximadamente 37°C y lo más preferiblemente de al menos aproximadamente 42°C. Los expertos en la técnica conocen bien la variación de parámetros adicionales, tales como el tiempo de hibridación, la concentración de detergente, p. ej., dodecilsulfato de sodio (SDS), y la inclusión o exclusión del ADN portador. Se logran varios niveles de rigurosidad combinando estas diversas condiciones según sea necesario. En una realización preferida, la hibridación se producirá a 30°C en NaCl 750 mM, citrato trisódico 75 mM y SDS al 1%. En una realización más preferida, la hibridación se producirá a 37°C en NaCl 500 mM, citrato trisódico 50 mM, SDS al 1%, formamida al 35% y 100 µg/ml de ADN de esperma de salmón desnaturalizado (ADNss). En una realización más preferida, la hibridación se producirá a 42°C en NaCl 250 mM, citrato trisódico 25 mM, SDS al 1%, formamida al 50% y 200 µg/ml de ADNss. Las variaciones útiles de estas condiciones resultarán fácilmente evidentes para los expertos en la técnica.

Para la mayoría de las aplicaciones, las etapas de lavado que siguen a la hibridación también variarán en rigurosidad. Las condiciones de rigurosidad del lavado se pueden definir por la concentración de sal y por la temperatura. Como se indicó anteriormente, la rigurosidad del lavado se puede aumentar disminuyendo la concentración de sal o aumentando la temperatura. Por ejemplo, la concentración de sal rigurosa para las etapas de lavado será preferiblemente inferior a aproximadamente 30 mM de NaCl y 3 mM de citrato trisódico, y lo más preferiblemente inferior a aproximadamente 15 mM de NaCl y 1,5 mM de citrato trisódico. Las condiciones de temperatura rigurosas para las etapas de lavado normalmente incluirán una temperatura de al menos aproximadamente 25°C, más preferiblemente de al menos aproximadamente 42°C, e incluso más preferiblemente de al menos aproximadamente 68°C. En una realización preferida, las etapas de lavado se producirán a 25°C en NaCl 30 mM, citrato trisódico 3 mM y SDS al 0,1%. En una realización más preferida, las etapas de lavado se producirán a 42°C en NaCl 15 mM, citrato trisódico 1,5 mM y SDS al 0,1%. En una realización más preferida, las etapas de lavado se producirán a 68°C en NaCl 15 mM, citrato trisódico 1,5 mM y SDS al 0,1%. Variaciones adicionales sobre estas condiciones serán fácilmente evidentes para los expertos en la técnica. Los mecanismos de hibridación son bien conocidos por los expertos en la técnica y son descritos, por ejemplo, por Benton y Davis (*Science* 196:180, 1977); Grunstein y Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, Nueva York, 2001); Berger y Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, Nueva York); y Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Nueva York.

"Secuenciación" o cualquier equivalente gramatical como se emplean en el presente documento se pueden referir a un método utilizado para secuenciar el proxy de ácido nucleico diana amplificado. La técnica de secuenciación puede incluir, por ejemplo, Secuenciación de Próxima Generación (NGS), Secuenciación Profunda, análisis de secuencia o longitud basado en espectrometría de masas, o análisis de secuencia o longitud de fragmentos de ADN mediante electroforesis en gel o electroforesis capilar. Se pueden utilizar técnicas de secuenciación compatibles, incluida la secuenciación en tiempo real de una sola molécula (Pacific Biosciences), Semiconductor de iones (secuenciación Ion Torrent), pirosecuenciación (454), secuenciación por síntesis (Illumina), secuenciación por ligación (secuenciación SOLiD), terminación de cadena (Secuenciación de Sanger), secuenciación de ADN Nanopore (Oxford Nanosciences Technologies), secuenciación de molécula única Heliscope (Heliscope Inc.), secuenciación con espectrometría de masas, secuenciación de nanobolas de ADN, secuenciación por hibridación y secuenciación de ADN con corrientes por efecto túnel.

Por "sustancialmente idéntico" se entiende un polipéptido o molécula de ácido nucleico que presenta al menos 50% de identidad con una secuencia de aminoácidos de referencia (por ejemplo, una cualquiera de las secuencias de aminoácidos descritas en el presente documento) o secuencia de ácido nucleico (por ejemplo, una cualquiera de las secuencias de ácidos nucleicos descritas en el presente documento). Preferiblemente, tal secuencia es al menos 60%, más preferiblemente 80% o 85%, y más preferiblemente 90%, 95% o incluso 99% idéntica a nivel de aminoácidos o ácido nucleico a la secuencia utilizada para la comparación.

La identidad de secuencia se mide típicamente utilizando soporte lógico de análisis de secuencia (por ejemplo, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, programas BLAST, BESTFIT, GAP o PILEUP/PRETTYBOX). Tal soporte lógico empareja secuencias idénticas o similares asignando grados de homología a diversas sustituciones, delecciones y/u otras modificaciones. Las sustituciones conservativas incluyen típicamente sustituciones dentro de los siguientes grupos: glicina, alanina; valina, isoleucina, leucina; ácido aspártico, ácido glutámico, asparragina, glutamina; serina, treonina; lisina, arginina; y fenilalanina, tirosina. En un enfoque ilustrativo para determinar el grado de identidad, se puede utilizar un programa BLAST, con una puntuación de probabilidad entre e^{-3} y e^{-100} que indica una secuencia estrechamente relacionada.

"Conjunto de cebadores" significa un conjunto de oligonucleótidos que se puede utilizar, por ejemplo, para PCR. Un conjunto de cebadores consistiría en al menos 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 30, 40, 50, 60, 80, 100, 200, 250, 300, 400, 500, 600 o más cebadores.

Se entiende que los rangos proporcionados en el presente documento son una abreviatura de todos los valores dentro del rango. Por ejemplo, se entiende que un rango de 1 a 50 incluye cualquier número, combinación de números o subrango del grupo que consiste en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 o 50, así como todos los valores decimales intermedios entre los números enteros antes mencionados, tales como por ejemplo 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8 y 1,9. Con respecto a los subrangos, se contemplan específicamente los "subrangos anidados" que se extienden desde cualquier punto extremo del rango. Por ejemplo, un subrango anidado de un rango ilustrativo de 1 a 50 puede comprender 1 a 10, 1 a 20, 1 a 30 y 1 a 40 en una dirección, o 50 a 40, 50 a 30, 50 a 20, y 50 a 10 en la otra dirección.

Por "reduce" se entiende una alteración negativa de al menos 10%, 25%, 50%, 75% o 100%.

Por "referencia" se entiende una condición convencional o de control.

Una "secuencia de referencia" es una secuencia definida utilizada como base para la comparación de secuencias o una comparación de la expresión génica. Una secuencia de referencia puede ser un subconjunto o la totalidad de una secuencia especificada; por ejemplo, un segmento de un ADNc o secuencia genética de longitud completa, o el ADNc o la secuencia genética completos. Para polipéptidos, la longitud de la secuencia polipeptídica de referencia será generalmente de al menos aproximadamente 16 aminoácidos, preferiblemente al menos aproximadamente 20 aminoácidos, más preferiblemente al menos aproximadamente 25 aminoácidos, e incluso más preferiblemente aproximadamente 35 aminoácidos, aproximadamente 50 aminoácidos, o aproximadamente 100 aminoácidos. Para los ácidos nucleicos, la longitud de la secuencia de ácido nucleico de referencia será generalmente de al menos aproximadamente 40 nucleótidos, preferiblemente al menos aproximadamente 60 nucleótidos, más preferiblemente al menos aproximadamente 75 nucleótidos, e incluso más preferiblemente aproximadamente 100 nucleótidos o aproximadamente 300 o aproximadamente 500 nucleótidos o cualquier número entero alrededor o entre ellos.

Como se emplea en el presente documento, "obtener" como "obtener un agente" incluye sintetizar, comprar o adquirir de otro modo el agente.

Por "sujeto" se entiende un mamífero, incluido, pero sin limitarse a, un mamífero humano o no humano, tal como un bóvido, équido, cánido, óvido o férido.

Como se emplea en el presente documento, el término "subsonda" se puede referir a cualquiera de las dos o más sondas que se unen a la secuencia diana contigua sin dejar ningún nucleótido intermedio no unido. En algunas realizaciones, la sonda multipartita descrita en el presente documento puede incluir al menos dos "subsondas". En otra realización, cada una de las al menos dos subsondas de la pluralidad de sondas multipartitas puede tener aproximadamente 10-50 nucleótidos de longitud. Una vez ligadas las sondas, la sonda multipartita ligada (alternativamente, la "subsonda ligada") se puede liberar del ARN. En algunas realizaciones, la subsonda puede contener un sitio de unión del cebador adjunto (p. ej., adaptadores) para facilitar la amplificación posterior del proxy de ácido nucleico diana. En otras realizaciones, al menos una de las dos o más subsondas se puede denominar "subsonda aceptora" que tienen una terminación 3' de al menos dos bases de ARN.

Como se emplea en el presente documento, los sitios de "unión de cebadores adjuntos" pueden referirse a sitios de unión dentro de la sonda o subsondas multipartitas descritas en el presente documento que facilitan la amplificación del proxy de ácido nucleico diana. Los "sitios de unión de cebadores adjuntos" también pueden denominarse "adaptadores".

Como se emplean en el presente documento, los términos "tratar", "tratando", "tratamiento" y similares se refieren a reducir o mejorar un trastorno y/o síntomas asociados con el mismo. Se apreciará que, aunque no se excluye, tratar un trastorno o afección no requiere que el trastorno, afección o síntomas asociados con los mismos sean eliminados por completo.

- 5 Como se emplean en el presente documento, los términos "prevenir", "previniendo", "prevención", "tratamiento profiláctico" y similares se refieren a reducir la probabilidad de desarrollar un trastorno o afección en un sujeto, que no tiene, pero está en riesgo de, o es susceptible de, desarrollar un trastorno o afección.

- 10 A menos que se indique específicamente o sea obvio por el contexto, como se emplea en el presente documento, se entiende que el término "o" es inclusivo. A menos que se indique específicamente o sea obvio por el contexto, como se emplean en el presente documento, se entiende que los términos "un", "uno", "una", "el" y "la" son singulares o plurales.

- 15 A menos que se indique específicamente o sea obvio por el contexto, como se emplea en el presente documento, se entiende que el término "aproximadamente" está dentro de un rango de tolerancia normal en la técnica, por ejemplo, dentro de 2 desviaciones típicas de la media. Aproximadamente puede entenderse como dentro de 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0,5%, 0,1%, 0,05% o 0,01% del valor establecido. A menos que se desprendo lo contrario del contexto, todos los valores numéricos proporcionados en el presente documento están modificados por el término aproximadamente.

- 20 La mención de una lista de grupos químicos en cualquier definición de una variable en el presente documento incluye definiciones de esa variable como cualquier grupo único o combinación de grupos enumerados. La mención de una realización para una variable o aspecto en el presente documento incluye esa realización como cualquier realización única o combinada con cualquier otra realización o parte de la misma.

Cualquiera de las composiciones o métodos proporcionados en el presente documento se puede combinar con una o más de cualquiera de las otras composiciones y métodos proporcionados en el presente documento.

- 25 A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que entiende comúnmente un experto en la técnica a la que pertenece esta invención. Aunque en la práctica o prueba de la presente invención se pueden utilizar métodos y materiales similares o equivalentes a los descritos en el presente documento, a continuación, se describen métodos y materiales adecuados. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluidas las definiciones. Además, los materiales, métodos y ejemplos son únicamente ilustrativos y no se pretende que sean limitantes.

30 **Breve descripción de los dibujos**

- 35 La Figura 1A-Figura 1C representa un electroferograma, un gráfico de barras y un gráfico respectivamente. La Figura 1A muestra un esquema que muestra un electroferograma capilar después de la ligación enzimática. Las bases de ARN están coloreadas en gris, las bases de ADN están coloreadas en negro y FAM es la marca fluorescente. La ADN ligasa de T4 no produce un producto de ligación mensurable (arriba). Rnl2 (ARN ligasa 2 de T4) liga la sonda química de ADN-ARN establecida en presencia de la cadena molde de ARN (centro), pero no en su ausencia (abajo). La Figura 1B representa un gráfico de barras que muestra que se evaluaron siete conjuntos únicos de sondas RASL (secuenciación con reasociación, selección y ligación de ARN) dirigidas al ARNm de GAPDH (gliceraldehído 3-fosfato deshidrogenasa) en el ARN celular. Se muestra la multiplicidad de diferencia en la eficiencia de ligación de Rnl2 frente a ADN ligasa de T4 para cada conjunto de sondas. La Figura 40 1C representa un gráfico que muestra la señal de un ensayo RASL basado en Rnl2 diseñado para medir una secuencia de ARN incorporada en un fondo constante de ARN celular. Nivel de fondo (BKGD) del ensayo qPCR con SYBR.

- 45 La Figura 2A-Figura 2C representa un diagrama, un diagrama de procedimiento y una imagen, respectivamente. La Figura 2A muestra un diagrama de procedimiento que representa la secuenciación mediante hibridación con ligación in situ (LISH-seq). Una reserva de conjuntos de sondas específicas de genes se hibrida con su ARN diana en el tejido fijado. Después de eliminar el exceso de sonda, Rnl2 liga las sondas reasociadas de manera adyacente. A continuación, el producto de la ligación se puede recuperar, amplificar y analizar la secuencia. La Figura 2B representa una viñeta que muestra experimentos iniciales que comparan la cuantificación del transcrito mediante RT-qPCR, RASL y LISH utilizando especímenes de tejido congelados y FFPE (fijados con formalina e incluidos en parafina) como se describe. La Figura 2C es una imagen que representa la utilidad de LISH-seq 50 para el análisis de tejidos microdisecados.

- 55 La Figura 3A-Figura 3B muestra un diagrama de molde y una imagen de molde/réplica, respectivamente. La Figura 3A muestra un diagrama de molde que representa el estampado de ADN en la superficie diana. Se muestra un molde para micromatrices de ADNss que se utiliza para guiar el suministro de ADNc biotinilado sobre una superficie de réplica recubierta con película de hidrogel con estreptavidina incorporada. La Figura 3B es una imagen que representa el ADNc marcado con Cy5/biotina que se transfirió a una superficie de réplica. (i) muestra el molde antes de la transferencia, (ii) muestra después de la transferencia. Se obtienen imágenes de la superficie de la réplica (iii) antes de la hibridación y (iv) después de la hibridación con una secuencia complementaria de

Cy3. Las manchas miden ~100 micrómetros.

La Figura 4 representa un análisis con sonda LISH-stAmp de una muestra fijada con formalina e incluida en parafina (FFPE) según una realización ilustrativa de la divulgación. La imagen superior representa el procedimiento LISH-stAmp. La hibridación y ligación de la sonda LISH se realizaría en la sección FFPE. La imagen central representa el estampado del producto de ligación sobre la superficie diana de la PCR puente (cada punto representa una única molécula de producto de ligación). La imagen inferior representa la PCR puente, que amplificaría la señal LISH in situ para su posterior análisis utilizando paneles de sondas de hibridación.

La Figura 5A-Figura 5D representa un diagrama, un gráfico de barras, un gel y un gráfico de barras, respectivamente. La Figura 5A representa un diagrama que describe un enfoque para medir de manera eficaz muchas secuencias de ARN en muestras FFPE sin purificación de ARN ni transcripción inversa. La Figura 5B es un gráfico de barras que indica que las secuencias de ADN que contienen dirribonucleótidos reasociadas con ADN fueron escindidas por la ARNasa H2, mientras que las secuencias reasociadas con ARN no. La Figura 5C es un gel que indica que el ARN FFPE podría servir como molde adecuado para la ligación *in situ* de la sonda mediada por Rnl2. La Figura 5D es un gráfico de barras que indica que el análisis de PCR de punto final del producto de ligación recuperado de la ARNasa H mediante secuenciación de ADN de próxima generación reveló que los pares de sondas correctos estaban efectivamente ligados *in situ*, mientras que los conjuntos de sondas con emparejamientos erróneos no se ligaron a niveles significativos.

Las Figuras 6A-6C representan imágenes de biopsia, un gel y un gráfico de barras, que indican respectivamente que LISH se puede utilizar para diagnosticar infección utilizando muestras clínicas FFPE conservadas. La Figura 6A representa biopsias de tejido cerebral infectado con *Exserohilum rostratum* (izquierda) y *Aspergillus fumigatus* (derecha). La Figura 6B representa un gel que indica que las sondas detectoras de *Exserohilum rostratum* informaron de una fuerte señal en la diana del paciente índice, pero no de biopsias de tejidos no infectados o secciones que contenían infecciones no relacionadas, tal como las de *Aspergillus fumigatus* estrechamente relacionadas. La Figura 6C es un gráfico de barras que indica que las sondas detectoras de *Exserohilum rostratum* informaron de una fuerte señal en la diana del paciente índice, pero no de biopsias de tejidos no infectados o secciones que contenían infecciones no relacionadas, tales como las de *Aspergillus fumigatus* estrechamente relacionadas.

Las Figuras 7A-7C representan un diagrama, un gel y un gel, que indican respectivamente que el ARN FFPE podría servir como molde adecuado para la ligación de sondas mediada por Rnl2 in situ. La Figura 7A representa un diagrama que indica los posibles eventos de ligación de subsondas "deseadas" y "no deseadas" que podrían ocurrir entre dos conjuntos de sondas diferentes. La Figura 7B representa un gel que indica el análisis de PCR de punto final de una reacción de ligación forzada que muestra que es posible, en condiciones inadecuadas, producir y detectar fácilmente productos de ligación fuera de la diana. La Figura 7C representa un gel que indica el requisito absoluto de ligación mediada por Rnl2 para lograr un producto de ligación detectable entre dos sondas.

La Figura 8 muestra un diagrama que indica la microdissección por captura láser (LCM) para recuperar el producto LISH. Se evaluó el producto de ligación del conjunto de sondas GAPDH obtenido a partir de tamaños de muestra de LCM de 200 μm^2 , 2.000 μm^2 y 20.000 μm^2 .

Descripción detallada de la invención

La presente divulgación proporciona métodos que son útiles para realizar análisis de expresión génica multiplexada en especímenes de tejido fijados tal como, por ejemplo, fijados con formalina e incluidos en parafina (FFPE). En particular, la presente invención proporciona un método que incluye

aplicar una o más sondas multipartitas a una muestra fijada seleccionada del grupo que consiste en una célula, un órgano, un tejido y cualquier combinación de los mismos, en donde cada una de las una o más sondas multipartitas incluye al menos dos subsondas y en donde la al menos dos subsondas se reasocian con una secuencia diana contigua en un ácido nucleico diana dentro de la muestra fijada:

reasociar al menos una de las una o más sondas multipartitas aplicadas con al menos un ARN diana dentro de la muestra fijada sin extracción de ARN;

ligar enzimáticamente al menos un extremo 3' a al menos un extremo 5-fosforilado de las al menos dos subsondas asociadas con al menos una sonda multipartita reasociada dentro de la muestra fijada para crear un proxy de ácido nucleico diana que puede ser detectado;

liberar el proxy de ácido nucleico diana del ácido nucleico diana; y

amplificar el proxy de ácido nucleico diana. Por ejemplo, las técnicas del presente documento proporcionan que la hibridación con ligación in situ ("LISH") puede aprovechar el desarrollo de una nueva química de ligación basada en Rnl2, que permite la conversión de alta eficiencia de secuencias de ARN entrecruzadas directamente

en paneles de secuencias de ADN amplificables por PCR, específicas de genes, sin depender de la extracción de ARN o la transcripción inversa. En otro aspecto, las técnicas del presente documento proporcionan que la combinación de LISH con análisis de secuenciación de ADN (NGS) de próxima generación ("LISH-seq") puede permitir la medición cuantitativa de paneles de ARNm en especímenes de tejido FFPE con una sensibilidad y precisión sin precedentes, desbloqueando potencialmente vastos archivos de datos de especímenes fijados de pacientes para nuevos análisis. LISH-seq puede superar las limitaciones anteriores asociadas con el análisis de expresión de genes multiplexados de tejidos microdisecados y las suspensiones de células fijadas limitantes. La combinación de LISH con un protocolo único de amplificación PCR puente y estampado de ADN ("LISH-stAmp") puede permitir la resolución espacial del análisis de expresión génica multiplexada en una sección de histología FFPE o una muestra de citología fijada con resolución de una sola célula. La presente divulgación se basa, al menos en parte, en el descubrimiento de que sondas multipartitas (p. ej., sondas que incluyen dos o más subsondas) se pueden reasociar con una diana de ácido nucleico (p. ej., ARN) en una muestra FFPE con un alto grado de especificidad y después ligar (p. ej., mediante RNA Ligasa 2 de T4) para crear una molde para la amplificación, sin requerir extracción de ARN ni transcripción inversa. Las técnicas del presente documento hacen uso del entrecruzamiento covalente de ARN en tejidos fijados como una característica beneficiosa a explotar. Es importante destacar que las técnicas descritas en el presente documento (p. ej., LISH) son resistentes a la degradación y la formación de aductos, y dependen completamente de reactivos y equipos fácilmente disponibles.

Oncología de precisión

La oncología de precisión se basa en la capacidad de investigadores y patólogos para caracterizar molecularmente los tejidos tumorales resecados y los materiales de biopsia. El desarrollo de técnicas avanzadas para el análisis de ADN, tales como la secuenciación de ADN de alto rendimiento o la secuenciación de ADN de próxima generación (NGS), ha contribuido enormemente a la capacidad de detectar lesiones genéticas asociadas con el cáncer. Sin embargo, la capacidad de caracterizar perfiles de expresión génica en muestras fijadas de pacientes se ha quedado atrás, debido en gran parte a la mala calidad del ARN de tales muestras.

Diagnóstico clínico de patógenos

El desarrollo de ensayos moleculares que puedan identificar rápidamente patógenos humanos sigue siendo un objetivo importante de la comunidad de diagnóstico. Habitualmente se encuentran disponibles muestras de tejido de personas sospechosas de tener una infección fúngica, bacteriana o viral aún no identificada. Sin embargo, el grado de degradación y procesamiento de biomoléculas en los que incurren estas muestras durante los procedimientos de fijación de rutina las hace inadecuadas para su uso en la mayoría de los ensayos moleculares disponibles actualmente. El desarrollo de ensayos alternativos que sean refractarios al procedimiento de fijación perjudicial, al mismo tiempo que mantengan la sensibilidad y sean altamente múltiples mejoraría en gran medida los resultados asociados con el diagnóstico y el tratamiento oportuno de infecciones patógenas.

La preservación clínica de tejido suele emplear un procedimiento de fijación con formalina e inclusión en parafina (FFPE), y las muestras FFPE son los especímenes de pacientes más comunes examinadas por los patólogos. Su relativa estabilidad proporciona un sustrato ideal para la tinción y permite un archivo económico de los tejidos del paciente. Desafortunadamente, sin embargo, este método para preservar la arquitectura del tejido degrada y modifica covalentemente el ARN del tejido (p. ej., convirtiéndolo en una forma altamente degradada), lo que dificulta su análisis utilizando enfoques tradicionales basados en la purificación del ARN y la transcripción inversa. La medición del ARN de células cultivadas y tejidos no fijados típicamente comienza con la purificación del ARN, seguida de la transcripción inversa (RT) y la reacción en cadena de la polimerasa (PCR). Sin embargo, la RT-PCR es ineficiente y poco confiable cuando se realiza sobre ARN de especímenes fijados químicamente. Kashofer, K., et al. *PLoS One*. 8(7) (2013)), Abrahamsen, HN., et al. *J Mol Diagn*. 5(1):34-41. (2013), Stephen. A., et al. *J Biomol Tech*. 15(3): 155-166. (2004)). FFPE es el método más ampliamente utilizado para preservar muestras de tejido clínico, debido al mantenimiento de la arquitectura del tejido, estabilización de biomoléculas y compatibilidad con una amplia variedad de tinciones, incluidas las inmunotinciones. La fijación con formalina daña el ARN mediante la adición de aductos tales como grupos hidroximetilo y mediante el entrecruzamiento del ARN consigo mismo y con otras biomoléculas mediante la formación de puentes de metileno. Además, la fragmentación del ARN típicamente ocurre antes y durante la preservación del tejido, y después continúa a un ritmo menor durante el almacenamiento (Evers, DL et al., *J Mol Diagn*. 13(3):282-8. (2011), Masuda, N., *Nucleic Acids Res*. 27(22):4436-43. (1999)). Dado que la RT-PCR requiere tramos relativamente largos de ARN mínimamente modificado, las técnicas tradicionales adolecen de una sensibilidad reducida y sesgos de medición impredecibles.

Sin embargo, a pesar de la importancia de los especímenes FFPE, pocas técnicas permiten un análisis de expresión de ARNm robusto y altamente multiplexado de las muestras fijadas de los pacientes, debido principalmente a los desafíos asociados con la extracción de ARN y la transcripción inversa en tales muestras. Los enfoques existentes utilizan la hibridación de sondas sin amplificación de la diana (p. ej., el ensayo Prosigna de NanoString) o la transcripción inversa (RT) dirigida y el análisis de productos de PCR (p. ej., el ensayo qPCR Oncotype DX de Genomic Health; el ensayo de matriz de cuentas citométricas de transcritos de fusión Signature LTx 12-plex de Assuragen). Mientras que el enfoque sin amplificación está limitado por la necesidad de instrumental y reactivos costosos y patentados, las técnicas dependientes de RT están limitadas por la baja calidad del ARN (p. ej., debido a la

fragmentación, el entrecruzamiento, la formación de aductos covalentes, etc.), lo que reduce la sensibilidad y limita el diseño del ensayo. El análisis transcriptómico imparcial utilizando NGS es una alternativa emergente, pero también está limitado por su requisito de RT, el alto coste de la secuenciación imparcial, los desafíos asociados con la cuantificación computacional del transcrito y la dificultad de interpretación de los datos. Además, cada uno de estos ensayos depende de la extracción del molde de ARN, lo que se asocia con degradación adicional, pérdida de material y pérdida de resolución espacial.

Patología molecular

La patología molecular se centra en el estudio y diagnóstico de enfermedades mediante el examen de moléculas dentro de órganos, tejidos o fluidos corporales. La patología molecular comparte algunos aspectos de la práctica con la patología anatómica y clínica, la biología molecular, la bioquímica, la proteómica y la genética. Una consideración clave es que es posible un diagnóstico más preciso cuando el diagnóstico se basa tanto en los cambios morfológicos de un tejido como en los cambios moleculares dentro de las células del tejido. La patología molecular se utiliza comúnmente en el diagnóstico de cáncer y enfermedades infecciosas. Las técnicas ilustrativas incluyen reacción en cadena de la polimerasa cuantitativa (qPCR), PCR múltiple, micromatrices de ADN, hibridación *in situ*, secuenciación de ADN, ensayos de tejido por inmunofluorescencia basados en anticuerpos, obtención de perfiles moleculares de patógenos y análisis de genes bacterianos para determinar la resistencia a antimicrobianos.

La aparición de la patología molecular ha dado lugar a una nueva era en la medicina de precisión. Las tecnologías para genotipificar y/o fenotipificar tejidos tumorales, por ejemplo, se utilizan ahora de forma rutinaria para proporcionar diagnósticos precisos e intervenciones terapéuticas óptimas. Las mediciones de expresión génica multiplexadas tienen el potencial de proporcionar información fenotípica rica, pero hasta la fecha no se han realizado ampliamente en especímenes de tejido fijados debido a las limitaciones asociadas con la calidad del ARN en tales especímenes.

Hibridación con ligación *in situ* (FISH)

Las técnicas del presente documento proporcionan una plataforma poderosa para la medición multiplexada de la expresión génica, que es ideal para el análisis de especímenes de tejido fijados y que supera las limitaciones técnicas críticas asociadas con el análisis de ARN FFPE. En una realización, LISH utiliza la ARN ligasa de T4 (Rnl2) para unir eficientemente conjuntos de sondas de ARN-ADN quiméricas específicas de secuencia en secuencias diana de ARN fijadas *in situ*. El tratamiento posterior del tejido con ARNasa H libera el producto de ligación con molde de ARN, al tiempo que destruye el producto de ligación con molde de ADN de fondo. Como se observa en la Figura 5A, etapa I, las sondas no tienen preferencia por la reasociación con dianas de ARNm o ADN que tienen una secuencia complementaria. Por lo tanto, durante el tratamiento con ARNasa H (etapa III), existen dos reservas distintas, sonda ligada:molde de ARN y sonda ligada:molde de ADN. La ARNasa H digiere específicamente el ARN de los dúplex de ARN:ADN. Por lo tanto, la ARNasa H reconocerá y digerirá moldes de ARN en dúplex de sonda ligada:molde de ARN y, a la inversa, digerirá y destruirá el componente de ARN (dirrabo interno) de las sondas ligadas en dúplex de sonda ligada:molde de ADN.

En un aspecto, las técnicas del presente documento proporcionan sondas de ácido nucleico multipartitas que se aplican a un espécimen de tejido fijado (p. ej., FFPE). Cada una de las sondas multipartitas incluye dos o más subsondas configuradas para reasociarse con una secuencia diana contigua en un ácido nucleico diana (p. ej., un ARN), de manera que cada una de las dos o más sondas se une a la secuencia diana contigua sin dejar ningún nucleótido no unido en la secuencia diana contigua entre cualquiera de las dos o más sondas que son adyacentes. Por ejemplo, si una sonda multipartita incluye dos sondas (p. ej., una sonda bipartita), una vez que las dos sondas se unen a la secuencia diana contigua, el extremo 3' de una de las sondas será inmediatamente adyacente al extremo 5' de la otra sonda de manera que las dos sondas estén unidas en una configuración de extremo a extremo.

Una vez que la sonda multipartita se une a la secuencia diana contigua, se puede utilizar una ligasa (p. ej., ARN Ligasa 2 de T4 (Rnl2)) para ligar las sondas multipartitas adyacentes unidas en contrapartes de ADNc amplificables, sin requerir extracción de ARN ni depender de transcripción inversa. Las técnicas del presente documento (p. ej., LISH) pueden ser susceptibles de un método de análisis de ARNm altamente multiplexado. Por ejemplo, se contempla dentro del alcance de la invención que un espécimen fijado típico pueda analizarse con 10-10.000 sondas multipartitas simultáneamente. Las técnicas del presente documento (p. ej., LISH) amplían el trabajo previo sobre la química de ligación de sondas de ADN con molde de ARN (véase, p. ej., Larman, HB et al., *Nucleic Acids Research* 42, 9146-9157 (2014)), para convertir moléculas de ARN entrecruzadas en secuencias de ADN amplificables por PCR. Ventajosamente, las técnicas descritas en el presente documento son resistentes a la degradación (p. ej., una muestra altamente degradada) y la formación de aductos, y dependen completamente de reactivos y equipos fácilmente disponibles.

Diseño y validación de sondas

Las sondas multipartitas se pueden diseñar con un canal de diseño de conjuntos de sondas similar a Primer-BLAST e implementarse utilizando Primer 3, BLASTN, Melting, pandas y la biblioteca convencional Python. Las sondas se diseñaron para reasociarse lo más cerca posible de la cola poli(A) del ARNm diana. Se implementaron configuraciones personalizadas de Primer3 para diseñar hasta 20 sondas separadas de 36 nucleótidos antisentido para el transcrito

diana. Las sondas diseñadas por Primer3 se extendieron 4 pares de bases en la dirección 5' de la sonda (hacia la cola poli(A)). A continuación, cada secuencia de 40 nucleótidos se dividió por la mitad y se anexaron adaptadores comunes (AD1 para sondas aceptoras y RCAD2 para sondas donadoras). Después convocó Primer3 para calcular las propiedades de cada adaptador oligo plus de sonda. Se utilizaron umbrales derivados empíricamente para los cálculos de Primer3 para filtrar la sonda candidata. Se eliminaron las sondas restantes con una Tm no deseada a menos de 10°C de la temperatura de fusión deseada prevista. A continuación, se empleó un esquema de clasificación no paramétrico que utiliza la distancia a la cola de poli(A) y la penalización del Primer 3 (basado en la sonda original de diseño Primer3 de 36 nucleótidos) para seleccionar los dos conjuntos de sondas que mejor predicen que se reasocian al menos a una distancia de 10 nucleótidos entre sí para cada transcrito diana. Las bases oligo aceptoras 3'-terminal y 3'-penúltimas se cambiaron a sus contrapartes de ARN.

Los parámetros críticos del diseño de la sonda incluyen la longitud de la sonda y la temperatura de fusión de las secuencias de reasociación. Para una secuencia diana determinada, el aumento de longitud de la sonda aumenta la fuerza de la interacción de unión específica, pero también puede aumentar las ligaciones inapropiadas mediante unión no específica a secuencias distintas de las deseadas y/o disminuir la concentración eficaz de la sonda en la reacción.

Se creó una biblioteca de sondas donadoras y aceptoras para explorar el impacto de la longitud de la secuencia de reasociación del transcrito, que oscila entre 12 y 22 nucleótidos. Las posiciones de unión se mantuvieron constantes para eliminar variables potencialmente confusas, tales como el sesgo de la secuencia de nucleótidos. Tanto para las sondas donadoras como para las aceptoras, el rendimiento de la ligación deseada dependía de la longitud de la sonda. Debido a que el coste de la sonda aumenta con la longitud y debido a que se observaron mejoras decrecientes en la cuantificación relativa después de ~20 nucleótidos para la mayoría de las sondas, se desarrolló un algoritmo de diseño de sonda para identificar secuencias de 20 nucleótidos adyacentes en los transcritos diana. Sin embargo, se contempla dentro del alcance de la divulgación que la longitud de las sondas dentro de una sonda multipartita puede oscilar entre 10 y 200 nucleótidos de longitud.

Se desarrolló una estrategia de señuelo de sonda para reducir el muestreo de transcritos expresados a niveles muy altos para optimizar la eficiencia del análisis de secuencia. Cada conjunto de sondas está flanqueado por una secuencia de unión del cebador común, de modo que los productos de ligación se pueden amplificar por separado mediante cebadores que contienen un código de barras de ADN corto específico para cada pocillo. Las sondas señuelo carecen de las secuencias de unión del cebador, por lo que forman productos de ligación no amplificables a los niveles deseados. De este modo, los productos de PCR de múltiples muestras se pueden agrupar para secuenciarlos y las lecturas individuales posteriormente se pueden desconvolucionar mediante sus correspondientes códigos de barras. El número finito de lecturas de secuenciación de ADN obtenidas da como resultado un sobremuestreo de transcritos muy abundantes. La estrategia de señuelo de sonda supera este sesgo de muestreo.

Sonda multipartita

Las sondas multipartitas presentan una forma de reunir y localizar dos o más secuencias de ácidos nucleicos. Una sonda multipartita puede incluir una secuencia de unión a ácido nucleico diana capaz de hibridarse con la secuencia genética diana y una secuencia de sonda de unión a ácido nucleico adicional adyacente a la secuencia de unión a ácido nucleico diana. Por ejemplo, las sondas multipartitas pueden incluir cada una dos o más sondas de ácido nucleico configuradas para reasociarse con una secuencia diana contigua en un ácido nucleico diana (p. ej., un ARN), de manera que cada una de las dos o más sondas se une a la secuencia diana contigua sin dejar ningún nucleótido libre en la secuencia diana contigua entre cualquiera de las dos o más sondas que son adyacentes (p. ej., una subsonda). Se puede hibridar una sonda multipartita con la muestra genética que incluye la secuencia genética diana y ligar para incluir una molécula de amplificación detectable. La molécula de amplificación detectable puede incluir una secuencia de sonda de captura que es una secuencia de ácido nucleico capaz de hibridarse con la secuencia de la sonda de unión a ácido nucleico para la sonda bipartita. Cada sonda dentro de la sonda multipartita puede tener una longitud de entre 10 y 200 bases. Por ejemplo, se entiende que 10 a 200 incluye cualquier número, combinación de números o subrango de números, así como "subrangos anidados" que se extienden desde cualquier punto final del rango. Por ejemplo, un subrango anidado de un rango ilustrativo de 10 a 200 puede comprender de 10 a 20, de 10 a 30, de 10 a 40, de 10 a 50, de 10 a 60, de 10 a 70, de 10 a 80, de 10 a 90, 10 a 100, 10 a 110, 10 a 120, 10 a 130, 10 a 140, 10 a 150, 10 a 160, 10 a 170, 10 a 180, 10 a 190 y 10 a 200.

Multiplexación

El análisis de expresión génica múltiplex proporciona una medición directa y cuantitativa de múltiples secuencias de ácidos nucleicos simultáneamente utilizando un sistema de detección. Los ensayos multiplex utilizan una estrategia en donde se amplifica y cuantifica más de una diana a partir de una única alícuota de muestra. En la PCR múltiple, se consulta una alícuota de muestra con múltiples sondas que contienen colorantes fluorescentes en una única reacción de PCR. Esto aumenta la cantidad de información que se puede extraer de la muestra. Con la técnica LISH descrita en el presente documento, se pueden analizar de cientos a miles de sondas al mismo tiempo.

Ligasa

Rnl2 (ARNdh Ligasa) es una ligasa de ARNdh dependiente de ATP que sella eficazmente mellas 3'-OH/5'PO₄ en ARN

dúplex. Este proceso se produce mediante la adenilación de la ligasa, la transferencia de AMP al 5'PO₄ en la hebra donadora, y el ataque por la hebra aceptora 3'-OH en la hebra donadora 5'-adenilada, dando como resultado la formación del enlace fosfodiéster covalente. Rnl2 tolera la sustitución completa de su sustrato de ARN dúplex con desoxirribonucleótidos, siempre que el extremo 3' de las cadenasceptoras termine en un dirribonucleótido. La técnica emplea sondas donadoras completamente desoxirribonucleotídicas con extremos 5'-PO₄, que se someten a una ligación dependiente de molde altamente eficaz a sondasceptoras híbridas de desoxirribonucleótido-3'-dirribonucleótido. En resumen, Rnl2 puede ligar sondas de ARN o sondas híbridas de ARN-ADN en donde una sonda tiene las dos bases 3' como ARN cuando se reasocia sobre moldes de ARN o ADN. Rnl2 no puede ligar de manera eficaz sondas completamente de ADN reasociadas sobre moldes de ADN o sondas completamente de ADN sobre moldes de ARN.

La ligasa puede introducirse para unir conjuntos de sondas donadoras y aceptoras reasociadas de manera adyacente. La ligación enzimática une covalentemente las sondas, que después sirven como molde para la amplificación de señales basada en PCR. En condiciones típicas, todos los componentes de la reacción de ligación están en exceso sobre el ARNm diana, lo que garantiza la proporcionalidad directa entre las moléculas molde y los eventos de ligación. Posteriormente, los productos de ligación se pueden amplificar y marcar con códigos de barras durante una PCR.

Se contempla dentro del alcance de la divulgación que se pueden utilizar otros tipos de ligasas según las técnicas del presente documento, incluidas, pero sin limitarse a, ADN ligasa de T4 y ADN ligasa de PBCV-1 (ADN ligasa del virus de Chlorella).

ARNasaH (Ribonucleasa H)

Una vez ligadas las sondas, la sonda multipartita ligada (p. ej., las subsondas) se puede liberar del molde de ARN para que se pueda recuperar y utilizar como fuente de amplificación. Según las técnicas del presente documento, se puede utilizar ARNasaH para liberar la sonda multipartita ligada (p. ej., las subsondas). La ARNasaH pertenece a una familia de endonucleasas no específicas de secuencia que catalizan la escisión del ARN mediante un mecanismo hidrolítico. Los miembros de la familia ARNasaH se pueden encontrar en casi todos los organismos, desde bacterias hasta arqueas y eucariotas. Debido a que la ARNasaH degrada específicamente solo el ARN en los híbridos ARN:ADN, se utiliza comúnmente en biología molecular para destruir el molde de ARN después de la síntesis de ADNc de la primera hebra mediante transcripción inversa, así como en procedimientos tales como los ensayos de protección de nucleasas. La ARNasa H también se puede utilizar para degradar hebras de ARN específicas cuando se hibrida el oligonucleótido de ADNc, tal como la eliminación de la cola de poliadenina del ARNm hibridado con oligo (dT), o la destrucción de un ARN no codificante seleccionado dentro o fuera de la célula viva. La RNasaH hidroliza específicamente los enlaces fosfodiéster del ARN, que se hibrida con el ADN. Esta enzima no digiere el ADN de hebra sencilla o doble. Para terminar la reacción, a menudo se añade un quelante, tal como EDTA, para secuestrar los iones metálicos en la mezcla de reacción, o la enzima se puede destruir térmicamente.

Después de la ligación de sondas reasociadas de manera adyacente, el producto de ligación (producto LISH) se puede recuperar mediante la incubación de la muestra con ARNasaH, que destruye el componente de ARN de las hélices híbridas de ARN/ADN y libera las subsondas ligadas en solución. A continuación, el producto se puede amplificar y, finalmente, analizar la secuencia. Los métodos alternativos para recuperar los conjuntos de sondas ligadas incluyen, pero sin limitarse a, otras enzimas ARNasa (tales como ARNasa A, ARNasa I_F, ARNasa HII, por ejemplo), tratamiento térmico para fundir las hélices híbridas o extracción mecánica de tejido (p. ej., mediante microdissección por captura láser o raspado) antes de la amplificación por PCR.

Las técnicas contenidas en el presente documento tienen el potencial de ampliar en gran medida la caja de herramientas moleculares disponible para los investigadores y patólogos que analizan muestras de tejido fijado. Los enfoques actuales para el análisis de la expresión génica altamente multiplexada de tejidos FFPE requieren la extracción de ARN dañado y de calidad relativamente baja. Por el contrario, los análisis basados en LISH utilizan moléculas de ARN entrecruzadas para obtener secuencias amplificables por PCR y, por lo tanto, no requieren extracción de ARN ni transcripción inversa, lo que proporciona varias ventajas distintas relacionadas con la sensibilidad del ensayo y la simplicidad del flujo de trabajo. Es importante destacar que el sistema utiliza instrumental y reactivos ampliamente disponibles, lo que lo hace accesible de inmediato. El ensayo LISH-seq propuesto en esta solicitud se adoptará cada vez más a medida que disminuya el coste de la secuenciación de ADN de alto rendimiento y aumente su disponibilidad.

Para aplicaciones en donde la resolución espacial del análisis de la expresión génica es importante, LISH-stAmp tiene el potencial de superar drásticamente los ensayos alternativos tanto en términos de sensibilidad como de número de transcritos que se pueden analizar simultáneamente. Al ampliar significativamente las capacidades para medir la expresión génica en tejidos fijados, la plataforma LISH mejorará en gran medida la capacidad de caracterizar fenotipos tumorales y comprender las complejas interacciones con las células del sistema inmunológico. Finalmente, debido a que la plataforma LISH es ideal para muestras de tejido fijadas, desbloqueará los vastos archivos de muestras de pacientes para nuevos análisis de ARN de alta dimensión que de otro modo no serían factibles.

Ejemplos

Ejemplo 1: Desarrollo de secuenciación mediante hibridación con ligación in situ (LISH-seq)

Se determina un protocolo óptimo para producir eficazmente un producto de ligación de sonda en secciones FFPE. Se emplean conjuntos de sondas de genes constitutivos para explorar el impacto de diversos parámetros del protocolo sobre la señal (ligación de la sonda deseada) y el ruido (ligación de la sonda no deseada) del ensayo. El rendimiento de LISH se evalúa mediante comparación directa con RT-qPCR. La secuenciación de ADN de alto rendimiento de productos LISH mixtos ("LISH-seq") demuestra el grado de multiplexación que proporciona el sistema (p. ej., de cientos a miles de transcritos por sección de tejido). LISH-seq. analiza una colección de especímenes almacenados durante períodos de tiempo cada vez mayores. Este conjunto de datos revela hasta qué punto LISH-seq desbloquea los vastos archivos de especímenes de tumores de pacientes para análisis de ARNm altamente multiplexados. Se presenta una estrategia que utiliza LISH-seq para superar las limitaciones anteriores asociadas con el análisis de ARN extraído de tejidos microdisecados y pequeñas poblaciones de células en suspensión fijadas.

Ensayo de secuenciación por reasociación, selección y ligación de ARN (RASL-seq)

En 2012, el grupo de Fu de la UCSD introdujo la "secuenciación por Reasociación, Selección y Ligación de ARN" (RASL-seq) y demostró su capacidad para realizar el seguimiento de cientos de transcritos en miles de muestras a un coste menor y un rendimiento mayor que el que es posible actualmente con enfoques alternativos (véanse, p. ej., Li, H., et al. Proc Natl Acad Sci U S A 109, 4609-4614 (2012) y Zhou, Z., et al. Molecular cell 47, 422-433 (2012)).

Esta hazaña técnica se logra mediante la utilización de una reserva de conjuntos de sondas de ADN específicas de genes, cuyos pares se reasocian de manera adyacente en transcritos diana. Después de la ligación del conjunto de sondas enzimáticas con moldes de transcritos, se realiza la amplificación por PCR del producto de ligación. Los cebadores de PCR tienen típicamente códigos de barras con secuencias de ADN específicas de la muestra, lo que permite la posterior combinación de productos de PCR de una gran cantidad de muestras para el análisis NGS simultáneo, lo que reduce drásticamente el coste por muestra. Una característica clave del ensayo RASL es que la ligación de la sonda se modela directamente mediante secuencias cortas de ARN, evitando así el requisito de transcripción inversa (RT).

Los protocolos RASL-seq publicados anteriormente se basaban en la ADN ligasa de T4 para realizar la etapa de ligación crítica. Sin embargo, se ha observado que esta ligación de hebras de ADN con moldes de ARN daba como resultado un rendimiento extremadamente bajo (véase, p. ej., la Tabla 1 en Bullard, DR. & Bowater, RP The Biochemical Journal 398, 135-144 (2006) Figura 1A Larman, HB et al., Nucleic acids research 42, 9146-9157 (2014)) y, por tanto, una sensibilidad del ensayo deficiente.

Para abordar este problema, como se describe en el presente documento, se desarrolla una química de ligación que utiliza la ARN ligasa 2 de T4 (Rnl2), una enzima que repara naturalmente las roturas en el ARN de doble hebra. Es importante destacar que Rnl2 tolera la sustitución de su sustrato de ARNd_h mellado por ADN_h mellado, siempre que el extremo 3' de la hebra aceptora tenga dos bases de ARN (Figura 1A) (Véase, p. ej., Nandakumar, J. y Shuman, S. Molecular cell 16, 211221 (2004)). Como se describe en detalle a continuación, Rnl2 liga conjuntos de sondas químicas de ADN-ARN sobre un molde de ARN con un rendimiento cuantitativo, que como promedio es 400 veces mayor que el de la ADN ligasa de T4 (Figura 1B). LISH también podría combinarse con etapas adicionales para mapear los niveles de expresión de ARN en especímenes fijados. Esto permitiría combinar la información espacial relativa a la distribución de la diana con los niveles de expresión de la diana, una característica que no es posible con los ensayos moleculares basados en RASL o RT. A pesar de contener el elemento dirribonucleotídico, el producto de ligación se amplifica por PCR mediante las múltiples ADN polimerasas probadas. Es importante destacar que el ensayo RASL es capaz de medir un transcrito incorporado en un rango dinámico lineal de más de seis órdenes de magnitud (Figura 1C).

Optimización y caracterización de LISH

Los xenoinjertos tumorales derivados de líneas celulares humanas se procesan en secciones seriadas FFPE de 4-40 µm de espesor y se utilizan para el análisis de transcritos de genes constitutivos individuales mediante Hibridación in situ por Ligación (LISH, Figura 2A). El mismo protocolo empleado por el laboratorio de patología clínica JHH (Johns Hopkins Hospital) para hibridar sondas de ADN por RNAscope (véase, p. ej., Wang, F., et al. The Journal of molecular diagnostic: JMD 14, 22-29 (2012)) utilizado actualmente para detectar el ARN del VPH in situ. Los conjuntos de sondas RNAscope son muy similares a los conjuntos de sondas LISH, lo que respalda este enfoque. La ligación mediada por Rnl2 sigue un protocolo de tinción de anticuerpos modificado. Después de la extracción del producto de ligación, se utiliza qPCR con SYBR green para la cuantificación de los productos de ligación. Con un protocolo funcional básico en mano, los efectos evaluados incluyen:

espesor de la sección,

fusión previa a la hibridación de la estructura secundaria del ARN,

concentración del conjunto de sondas LISH,

temperatura y duración de la hibridación del conjunto de sondas, rigurosidad del lavado,
 concentración de Rnl2 ligasa,
 temperatura y duración de la etapa de ligación, y
 método de recuperación de conjuntos de sondas LISH ligadas

- 5 La "señal" del ensayo se evalúa mediante el análisis qPCR con SYBR green de conjuntos de sondas emparejadas correctamente, mientras que el "ruido" del ensayo se evalúa mediante el análisis de conjuntos de sondas emparejadas incorrectamente. La razón señal-ruido del ensayo se maximiza para una serie de conjuntos de sondas de genes constitutivos, ya que los parámetros del protocolo varían como se describe anteriormente. Se desarrolla un protocolo LISH que es compatible con los análisis posteriores tradicionales, tales como inmunohistoquímica.
- 10 El producto de ligación de LISH se puede recuperar eficazmente mediante el uso de ARNasaH (véase, por ejemplo, la etapa 3 en la Figura 5A), que destruye el componente de ARN de las hélices híbridas de ARN/ADN, para liberar conjuntos de sondas ligadas en solución. Las zonas específicas de interés en un portaobjetos se aíslan mediante raspado mecánico o el uso de una pluma de barrera hidrófoba. Después de una breve incubación con solución de ARNasaH, se añade una mezcla maestra de PCR directamente al eluyente y se inicia el termociclado de la PCR. Para
- 15 el análisis de paneles de sondas LISH utilizando NGS, se utilizan conjuntos de cebadores con código de barras de ADN para la amplificación por PCR para permitir la multiplexación de muestras.

Establecimiento de la sensibilidad de LISH

Se realizan los dos experimentos siguientes: I. Uso de LISH-seq para desbloquear los archivos de especímenes de tejido fijados y II. Se realizan análisis LISH-seq de tejidos microdisecados para establecer la sensibilidad de LISH. Se

20 utilizan secciones en serie archivadas de tejido del bazo fijado en un ensayo LISH o se utiliza una sección en serie de tamaño equivalente para la extracción del ARN total. Se procesa una única sección de tejido para extraer ARN como referencia; la segunda sección en serie se procesa para su uso en un ensayo LISH para comparar la sensibilidad entre los dos métodos. El ARN derivado de secciones FFPE se somete a un análisis convencional de RT-qPCR de un solo gen y a un análisis de ensayo RASL de un solo gen de varios transcritos de genes constitutivos (Figura 2B). El ARN

25 obtenido de tejido de bazo humano congelado no fijado se utiliza para determinar la señal máxima posible a partir de cantidades equivalentes de ARNm inicial cuando se utiliza en reacciones RT-qPCR o RASL. El uso de los mismos conjuntos de sondas RASL en el ensayo LISH proporciona una estimación cuantitativa de la fracción de moléculas de ARN en el tejido fijado que se convierten en una señal amplificable mediante el protocolo LISH. Además, una comparación directa entre el análisis RT-qPCR y LISH de ARN fijado puede demostrar la sensibilidad superior de los

30 ensayos basados en LISH.

I. Uso de LISH-seq para desbloquear los archivos de especímenes de tejido fijados

Debido a que los productos de ligación de LISH se amplifican mediante PCR utilizando un único par de cebadores, es fácilmente compatible con la secuenciación NGS (LISH-seq). Además, los patrones relativos de expresión génica se

35 conservan bien incluso cuando la cantidad total de molde de ARN fijado se pierde con el tiempo debido a la degradación. Por lo tanto, el rendimiento del análisis LISH-seq se evalúa en función del tiempo que las secciones FFPE se almacenan a temperatura ambiente.

Para la primera serie de experimentos, en experimentos LISH se utilizan secciones de tumores humanos FFPE previamente caracterizadas con un panel de genes utilizando RT-qPCR para comparar perfiles de expresión génica entre los diferentes ensayos. El análisis del ARNm del gen constitutivo y los paneles de genes previamente

40 caracterizados ayudarán a establecer un marco importante para interpretar los datos de LISH-seq a partir de especímenes de tejido históricos y revelará hasta qué punto LISH-seq desbloquea los archivos de especímenes de tejido fijados para una nuevos análisis de transcritos altamente multiplexados.

II. Análisis LISH-seq de tejidos microdisecados y células en suspensión fijadas.

Los análisis de expresión génica de tejidos fijados microdisecados por captura láser (LCM) son valiosos, pero siguen siendo especialmente desafiantes debido a la cantidad minúscula y la baja calidad del ARN extraíble (véanse, por

45 ejemplo, Pagedar, NA, et al. Brain Research 1091, 289-299 (2006) y Kaneko, T. et al., Microscopy research and technique 72, 908-912 (2009)). La realización de LISH en una sección FFPE intacta, antes de LCM, supera esta limitación. Se desarrolla y evalúa un protocolo para amplificar por PCR el producto de ligación directamente a partir de material LCM utilizando las mismas secciones de FFPE descritas anteriormente.

De manera análoga a la LCM, es extremadamente valioso medir un panel de transcritos dentro de una pequeña población de células en suspensión después de la clasificación FACS (p. ej., células T CD4+/IFN γ + infiltrantes de tumores marcadas con tetrámero de péptido MHC). La fijación celular se utiliza típicamente para teñir citocinas intracelulares y factores de transcripción, lo que dificulta mucho la extracción de ARN de alta calidad a partir de células

50 clasificadas. Se desarrolla un protocolo LISH para células en suspensión fijadas y permeabilizadas, que también es compatible con análisis posteriores, tales como la tinción de citocinas intracelulares. A continuación, las células

55

clasificadas (o posiblemente células individuales) se someten directamente a una reacción de PCR (sin purificación de ácido nucleico) para un análisis LISH-seq de alta dimensión.

Estrategias alternativas

El protocolo para la reasociación de conjuntos de sondas LISH en secciones de tejido FFPE utiliza condiciones optimizadas (temperaturas, tiempos) y tampones para garantizar la máxima rigurosidad y sensibilidad. Si no se logra una ligación eficaz de las sondas LISH reasociadas in situ utilizando Rnl2, se evalúa la actividad de la ADN Ligasa del virus de Chlorella, también conocida como ADN Ligasa PBCV-1, que ha sido comercializada recientemente por NEB como ligasa "SplintR" (Véase, p. ej., Lohman, GJ, et al. Nucleic Acids Res 42, 1831-1844 (2014)). En el pasado, la ligasa PBCV-1 y Rnl2 se compararon directamente y se descubrió que Rnl2 funciona de manera más óptima (véase, p. ej., Larman, HB et al., Nucleic acids research 42, 9146-9157 (2014)), pero se espera que la ligasa PBCV-1 funcione suficientemente bien en aplicaciones LISH.

Además, la ligación de sondas no específicas es una reacción indeseable y puede suprimirse fuertemente incluyendo sondas "señuelo" que compitan con las sondas funcionales por la ligación no específica. Las sondas señuelo están diseñadas para dirigirse a secuencias irrelevantes y carecen de los adaptadores de unión del cebador de PCR presentes en las sondas LISH funcionales. Cuando se añaden sondas señuelo a la reacción en exceso, la fracción no deseada del producto de ligación amplificado puede reducirse drásticamente (Véase, p. ej., Larman, HB et al., Nucleic acids research 42, 9146-9157 (2014)).

Ejemplo 2: Mantenimiento de la resolución espacial del producto LISH amplificado: LISH-stAmp

El valor del análisis de expresión génica multiplexada aumenta enormemente si se preserva la distribución espacial de los niveles de transcritos. Tal tecnología es particularmente útil para caracterizar las complejas interacciones entre los tejidos tumorales y las células del sistema inmunológico. Una plataforma que permite el análisis altamente multiplexado y resuelto espacialmente de la expresión génica en secciones de tejido FFPE es transformadora tanto para los investigadores como para los patólogos del cáncer.

La utilidad de LISH se amplía a un método para estampar y Amplificar el producto LISH en una superficie diana ("LISH-stAmp") para preservar la información espacial. Este esfuerzo se basa en el estampado de moléculas de ADN para replicar micromatrices de ADN de alta densidad. Para amplificar localmente el producto LISH estampado, se utiliza PCR puente, que se ha descrito en la bibliografía y es la base de la tecnología de secuenciación de ADN de alto rendimiento de Illumina. Después de definir el protocolo LISH-stAmp, se desarrollan las técnicas para analizar el producto de amplificación por PCR puente. Estos estudios utilizan secciones de xenoinjertos tumorales derivados de líneas celulares humanas y extensiones de líneas celulares en suspensión fijadas. Los niveles de expresión de 5 a 10 genes se miden simultáneamente hibridando el producto de la PCR puente con oligonucleótidos fluorescentes multicoloreados. Se utilizan ciclos secuenciales de hibridación, obtención de imágenes y eliminación de oligonucleótidos para analizar una cantidad sin precedentes de transcritos de un solo portaobjetos FFPE.

Las herramientas para caracterizar la arquitectura espacial de los microambientes tisulares son invaluable para comprender la fisiopatología del cáncer, particularmente en el contexto de interacciones complejas entre células inmunitarias y tumorales.

Nanoestampado supramolecular (SuNS):

El nanoestampado supramolecular (SuNS) es un método desarrollado por Francesco Stellacci y otros para transferir información química con patrones reversibles (p. ej., ADN) de una superficie a otra (Figura 3A) (Véanse, p. ej., Yu, A.A., et al. Nano letters 5, 1061-1064 (2005), Akbulut, O., et al. Nano letters 7, 3493-3498 (2007) y Kim, J. y Crooks, R.M. Analytical chemistry 79, 7267-7274 (2007)). La metodología SuNS se amplió para 'replicar' enzimáticamente micromatrices de ADN del tamaño de portaobjetos de microscopio (Cuppoletti, A. y Larman, H.B. (Véase p. ej., Patente: PCT/US2008/076717, 2009)). Los experimentos emplean la siguiente estrategia:

- (i) síntesis enzimática de una hebra de ADNc marcada con biotina en una micromatriz de molde de ADNss,
- (ii) contacto de la micromatriz de ADNdh con una superficie de captura recubierta con película de hidrogel con estreptavidina incorporada,
- (iii) liberación térmica del ADNc biotinilado en el hidrogel donde se une a la estreptavidina, y
- (iv) separación de los portaobjetos y preparación del molde para el siguiente ciclo de replicación. Los estudios iniciales demostraron tanto el fiel mantenimiento de la resolución de la superficie de la réplica como su funcionalidad (Figura 3B).

En el presente documento, el estampado de ADN se adapta para resolver espacialmente el análisis de expresión del panel de genes en secciones FFPE.

Diseño experimental

Los estudios iniciales utilizan la experiencia en el estampado de ADN a partir de una micromatriz molde sobre una superficie de réplica recubierta con una película de hidrogel (véase, p. ej., Cuppoletti, A. y Larman, H.B. Patente: PCT/US2008/076717, 2009) y Larman, H.B. y Stellacci, F. (Véase p. ej., Patente: PCT/US2008/076723, 2010)). Sin embargo, en las técnicas presentadas en el presente documento, se realiza el estampado de ADN (producto de ligación LISH) de un "hidrogel" (p. ej., una sección de tejido FFPE) sobre una superficie de réplica recubierta con cebadores de PCR inmovilizados covalentemente.

PCR "puente"

La PCR puente es un método de amplificación por PCR en fase sólida localizada, en donde los fragmentos de ADN se amplifican mediante cebadores anclados a una superficie (en lugar de en solución), formando así agrupamientos de ADN inmovilizados en la superficie (véase, p. ej., Adams, C.P. & Kron, S.J. (Patente: US 08/285,385, 1997)). Si los agrupamientos están suficientemente separados, se pueden secuenciar clonalmente (véanse, p. ej., Mitra, RD et al., Analítica biochemistry 320, 55-65 (2003) y Mitra, RD et al., Proc Natl Acad Sci U S A 100, 5926-5931 (2003)), que es la base de la tecnología de secuenciación de ADN de alto rendimiento de Illumina (véanse, p. ej., Shendure, J., et al. Science 309, 1728-1732 (2005) y Porreca, GJ, et al. Current protocols in molecular biology / editado por Frederick M. Ausubel. [et al.] Capítulo 7, Unidad 7 8 (2006)). Sin embargo, si los agrupamientos se superponen total o parcialmente, la contribución relativa de cada secuencia se determina mediante hibridación con secuencias de ADNc marcadas con fluorescencia, que es la base de la tecnología LISH-stAmp (Figura 4).

La primera serie de experimentos establece la superficie diana de la PCR del puente. Se utilizan portaobjetos de vidrio recubiertos de epoxisilano (Schott), que son ideales para la inmovilización covalente de cebadores oligonucleotídicos terminados en amino 5' (IDT). Para evaluar la calidad (densidad y uniformidad de la superficie) de las superficies diana de la PCR del puente, se utilizan sondas de ADN de unión a cebadores fluorescentes.

Utilizando superficies diana, se optimiza el protocolo de PCR puente para el producto LISH, comenzando con los métodos publicados. Se aplica a mano una solución diluida de sonda LISH preligada directamente sobre la superficie del portaobjetos para imitar el estampado. Después de la PCR puente, el portaobjetos se analiza utilizando una sonda de hibridación complementaria para medir la amplificación.

Después de que la PCR puente satisfactoria utilizando el producto de ligación aplicado sea exitosa, se inicia la experimentación con muestras de tejido fijadas como fuente del producto de ligación de la sonda LISH. Primero se establece el método más simple para lograr un contacto conforme de área grande y confiable entre el portaobjetos de tejido y el portaobjetos diana. Se desarrolló un sistema de prensa con temperatura controlada para estampar ADN (véase, p. ej., Larman, H.B. y Stellacci, F. (Patente: PCT/US2008/076723, 2010)), y se exploró su utilidad para LISH-stAmp. También se pueden lograr resultados similares utilizando un simple sellador a vacío de alimentos. Una vez en contacto, los portaobjetos se calientan para permitir la liberación térmica del producto de ligación LISH de la sección FFPE y después se enfrían a 45°C durante el tiempo suficiente para permitir la captura mediante los cebadores de PCR puente en la superficie diana. Los portaobjetos se separan y el portaobjetos diana se somete a amplificación por PCR puente. Se realiza un análisis de rendimiento versus número de termociclos. Para desarrollar la metodología se utilizan secciones FFPE de xenoinjertos tumorales derivados de líneas celulares humanas.

Estudios de LISH-stAmp

Los experimentos LISH-stAmp utilizan un panel de 5 a 10 conjuntos de sondas, que se analizan simultáneamente utilizando oligos de ADNc marcados con colorantes fluorescentes que se pueden discernir en el microscopio invertido Zeiss Axiovert 200 (con el módulo confocal LSM510-Meta, siete líneas láser y cuatro conjuntos de filtros) que se encuentra en el centro de microscopía Johns Hopkins. Los estudios se centran principalmente en dos tipos de tejidos: (i) citoespinos de líneas celulares humanas y (ii) xenoinjertos tumorales derivados de líneas celulares humanas. Las preparaciones de citoespinos permiten el análisis de la resolución espacial de LISH-stAmp. Es importante destacar que los productos de ligación de sondas individuales se aplican a mano en portaobjetos de PCR puente separados para que sirvan como controles de compensación de fluorescencia para los análisis LISH-stAmp multiplexados. Se emplea un protocolo convencional de "extracción por arrastre" de micromatrices de ADN para permitir hibridaciones secuenciales de conjuntos de sondas, lo que permite analizar un número mucho mayor de sondas LISH de las que de otro modo se podrían obtener imágenes simultáneamente.

Estrategias alternativas

El análisis de niveles de expresión génica muy diferentes en portaobjetos LISH-stAmp puede resultar problemático. En parte, esto se debe a la competencia por los cebadores de PCR puente entre productos LISH de genes de alta expresión y productos LISH de genes de baja expresión. Anteriormente se abordó un desafío análogo en el desarrollo de RASL-seq mediante el uso de sondas RASL señuelo específicas de genes (véase, p. ej., Larman, HB et al., Nucleic acids research 42, 9146-9157 (2014)). Por lo tanto, se espera que el efecto de los conjuntos de sondas de genes constitutivos parcialmente señuelo aumente la disponibilidad de cebadores de PCR puente al producto LISH de transcritos de menor abundancia. En señales fluorescentes extremadamente desequilibradas, la señal de transcritos muy abundantes se amortigua diluyendo las sondas de hibridación marcadas en un fondo de sondas de hibridación

no marcadas.

Ejemplo 3: Medición de secuencias de ARN en FFPE

Un informe reciente ha demostrado que la química de ligación de sondas con molde de ARN es sensible, específica y adecuada para análisis multiplexados masivamente (Larman, HB et al., *Nucleic Acids Research* 42, 9146-9157 (2014)). El enfoque utiliza ARN Ligasa 2 de T4 (Rnl2), que liga conjuntos de sondas de oligonucleótidos de secuencia específica reasociados de forma adyacente en una "tablilla" de ARN. En este contexto, Rnl2 requiere que el oligo aceptor esté terminado en 3' con dos ribonucleótidos (aquí denominado "sonda 3'-dirribo"). En algunos ejemplos, el oligo aceptor puede tener una terminación 3' de al menos dos bases de ARN. La sonda donadora fosforilada (denominada aquí "sonda 5'-fosfo") puede ser completamente ADN (FIG. 5A). El conjunto de sondas 3'-dirribo y 5'-fosfo ligadas es, por tanto, una molécula de ADN que contiene una secuencia de dirribonucleótido interna.

No se conoce ninguna ADN ligasa que muestre preferencia por las tabllas de ARN frente a las de ADN, e incluso Rnl2 liga conjuntos de sondas quiméricas con aproximadamente la misma eficacia tanto en tabllas de ARN como de ADN. Por lo tanto, la cuantificación precisa de las secuencias de ARN se verá confundida por la señal de fondo de la ligación entablillada de ADN, a menos que el producto de la ligación del ADN se elimine específicamente. La ARNasa H es una nucleasa que digiere específicamente el componente de ARN de las hélices híbridas de ARN-ADN. Por lo tanto, la ARNasa H cumplió un doble propósito simultáneamente (i) destruyendo el producto de ligación que contiene dirribonucleótido entablillado en ADN (FIG. 5A (i)) y (ii) liberando el producto de ligación que contiene dirribonucleótido entablillado en ARN a partir de ARN fijado en una solución para su análisis (Figura 5A (ii)). De hecho, las secuencias de ADN que contienen dirribonucleótidos reasociadas con ADN fueron escindidas por la ARNasa H2, mientras que las secuencias reasociadas con ARN no (FIG. 5B). La ARNasa H1 requiere al menos cuatro bases contiguas de ARN para la digestión, por lo que esta enzima no destruyó las secuencias de ADN que contienen dirribonucleótidos reasociadas con ADN, como se esperaba. Por lo tanto, las ARNasa H2 y H1 se pueden utilizar para recuperar conjuntos de sondas ligadas, ocasionando o no la destrucción simultánea del producto de ligación entablillado en ADN, respectivamente. Este nuevo sistema de medición de secuencias de ARN en tejido fijado se denomina "Hibridación *in situ* por ligación" ("LISH").

El ARN FFPE sirvió como molde para la ligación *in situ* de sondas mediada por Rnl2

Se utilizaron dos conjuntos de sondas dirigidas a los genes constitutivos GAPDH y RPL19 para analizar una sección de 10 micrómetros de bazo humano. El análisis de PCR de punto final del producto de ligación recuperado con la ARNasa H reveló que, de hecho, se ligaron los pares de sondas correctos *in situ*, mientras que los conjuntos de sondas con emparejamientos erróneos no se ligaron a niveles detectables (FIG. 5C y 5D).

Para cuantificar la cantidad de productos de ligación específicos formados durante una reacción múltiple, se realizaron ensayos de PCR cuantitativa para cada uno de los cuatro posibles productos de ligación (dos "deseados" y dos "no deseados"). A continuación, estos ensayos se utilizaron para determinar la cantidad de señal (ligación de la sonda deseada) con respecto a ruido (ligación de la sonda no deseada), a la vez que se optimizaron los parámetros del protocolo LISH.

Evaluación de archivos clínicos de especímenes FFPE

Existen vastos archivos clínicos de especímenes FFPE en todo el mundo; por lo tanto, una técnica que permita recuperar información transcripcional de tejidos archivados sería de extrema utilidad, especialmente para investigar especímenes limitantes o de alto valor. Mientras que se espera que la magnitud del producto de ligación recuperado disminuya modestamente en función del tiempo, el perfil relativo de expresión génica debe mantenerse fielmente mediante el análisis LISH.

Ensayos LISH independientes realizados secuencialmente en la misma sección de tejido

El uso de ARNasa H para recuperar el producto de ligación de LISH debería preservar las moléculas de ARN que no sirvieron como moldes para la ligación de sondas. Como resultado, las secuencias de ARN no dirigidas de una sección previamente analizada con LISH deberían permanecer disponibles para un segundo análisis LISH no solapante. Esta es una característica útil de LISH para investigaciones de especímenes limitantes o de alto valor. Se pueden realizar ensayos LISH independientes de forma secuencial en la misma sección de tejido, de modo que los resultados de un ensayo LISH se puedan utilizar para guiar la selección de sondas informativas que se utilizarán en un segundo ensayo LISH de la misma sección de tejido FFPE.

Compatibilidad del ensayo LISH con una variedad de tinciones analíticas

Se utiliza una rica diversidad de tinciones analíticas para interrogar especímenes FFPE. Muchos de estos, incluidos la hematoxilina y eosina (H y E) y las inmunotinciones, serán útiles para emplearlos junto con los análisis LISH. Las sondas LISH ligadas son relativamente estables y, por lo tanto, se espera que permanezcan intactas *in situ* durante un protocolo de tinción posterior. Después de la tinción, el producto LISH se puede recuperar utilizando ARNasa H, microdissección por captura láser (LCM) u otros medios. Por lo tanto, el análisis LISH se puede combinar con el análisis de las mismas secciones de tejido utilizando una amplia variedad de tinciones comunes.

Tamaño del tejido microdisecado empleado mediante el ensayo LISH

Los productos LISH multiplexados son de tamaño uniforme y, por lo tanto, deben amplificarse con un sesgo de PCR mínimo. Esta calidad de LISH debería facilitar el análisis multiplexado de tejidos fijados extremadamente pequeños, tales como los obtenidos mediante microdisección por captura láser (LCM). Se determinó la cantidad más pequeña de tejido microdisecado que aún proporcionaría una medición sólida de la expresión génica. Se obtuvieron fragmentos de tejido de la sección del bazo de diez micrómetros cuyo tamaño variaba de 20.000 micrómetros cuadrados a 200 micrómetros cuadrados (el equivalente aproximado de ~300 a 3 áreas de esplenocitos, respectivamente). La cuantificación de los productos de ligación de la sonda constitutiva reveló una señal fiable, que era proporcional al tamaño del fragmento de LCM (FIG. 8). Por lo tanto, LISH funciona favorablemente en términos de la cantidad de tejido necesaria para obtener una medición multiplexada sólida de la expresión génica. Una ventaja adicional es que toda la sección se hibridó y ligó uniformemente antes de la microdisección de las áreas de tejido deseadas. Este enfoque minimizó la variación técnica y evitó los costes y las pérdidas de material asociados con la extracción de ARN de fragmentos individuales.

Ejemplo 4: Ensayos LISH para diagnóstico de infección

Una técnica de alta sensibilidad para detectar una gran cantidad de secuencias de ARN diferentes en especímenes FFPE sería extremadamente útil para el diagnóstico de infección. En micología clínica, por ejemplo, las tinciones químicas a menudo revelan la presencia de un organismo fúngico, pero la definición posterior de la especie puede resultar muy difícil. Este fue el caso en 2012, cuando más tarde se atribuyó un misterioso brote de meningitis fúngica a viales contaminados de metilprednisolona.

Se obtuvo una sección de biopsia cerebral extremadamente pequeña de un paciente atendido en el Hospital Johns Hopkins. Los ensayos LISH multiplex se realizaron utilizando un panel de conjuntos de sondas diseñados para distinguir entre tres especies de hongos estrechamente relacionadas. Las sondas detectoras de *Exserohilum rostratum* informaron de una fuerte señal deseada del paciente índice que había sido infectado mediante una inyección de metilprednisolona contaminada con *E. rostratum*, pero no de biopsias de tejidos no infectados o secciones que contenían infecciones no relacionadas, tales como las *Aspergillus fumigatus* íntimamente relacionado (Figuras 6A-6C). Por lo tanto, se utilizó LISH para detectar infecciones fúngicas utilizando especímenes clínicos conservados. De esta manera se podrían utilizar reservas de sondas muy complejas para ensayos de diagnóstico completos. En este contexto, se podría realizar un análisis rápido del producto de ligación mediante secuenciación de nanoporos en tiempo real o espectroscopía de masas, por ejemplo.

Ejemplo 5: Ensayos LISH para diagnóstico en inmunooncología

La inmunooncología es un campo que evoluciona rápidamente y ensayos clínicos recientes demuestran espectaculares beneficios para los pacientes (Couzin-Frankel, J. *Science* (2013) 342; 1432-1433). La inmunooncología, por ejemplo, puede utilizar fármacos conocidos como inmunoterapias que se dirigen al sistema inmunológico del organismo para ayudar a combatir el cáncer. Las respuestas pueden variar ampliamente entre individuos con tumores aparentemente similares, lo que sugiere que las características del microambiente tumoral subyacen o socavan la eficacia de un régimen inmunoterapéutico. La interfase tumor-sistema inmunitario es a menudo una arquitectura tisular muy compleja (p. ej., estructura), donde numerosos tipos de células, receptores de superficie y productos secretados se unen para gobernar la respuesta inmunitaria antitumoral de un paciente. Por lo tanto, las mediciones altamente multiplexadas de la expresión génica en especímenes microdisecados se podrían utilizar para estratificar poblaciones de pacientes y guiar la intervención terapéutica. Es importante destacar que la metodología LISH simple descrita en este ejemplo no requiere instrumental ni experiencia especializados y podría integrarse inmediatamente en los flujos de trabajo de patología clínica actuales. A medida que la secuenciación de ADN de alto rendimiento se vuelve cada vez más ubicua y económica, la plataforma LISH y las técnicas derivadas se adoptarán ampliamente para aplicaciones clínicas y de investigación de rutina.

La ligación basada en Rnl2 de sondas químicas que contienen dirribonucleótidos en ARN fijado tendrá aplicaciones más allá de las presentadas aquí. Por ejemplo, los linfocitos frecuentemente se fijan y permeabilizan antes de la inmunotinción. En una realización, LISH se realiza antes de la tinción y clasificación para permitir la expresión génica en subpoblaciones de células, o potencialmente incluso en células individuales de interés. Además, la realización de LISH con alto rendimiento en cultivos celulares fijados podría utilizarse para la detección basada en la expresión de genes multiplexados. La capacidad de ligar eficazmente ADN amplificable in situ sobre secuencias diana de ARN fijado también permitirá o facilitará diversas tecnologías emergentes. Por ejemplo, los métodos *in situ* para secuenciar transcritos diana (Lee JH y otros, *Science* (2014); 343(6177):1360-3), la realización de ligación de proximidad (Frei AP et al., *Nature Methods* (2016); 13(3): 269-75) o sondas de inversión molecular circularizantes en moldes de ARN fijado (O'Roak BJ et al., *Science* (2012); 338(6114): 1619-22), pueden beneficiarse todos de los enfoques descritos en el presente documento.

Otras realizaciones

Si bien la invención se ha descrito junto con la descripción detallada de la misma, se pretende que la descripción anterior ilustre y no limite el alcance de la invención, que está definido por el alcance de las reivindicaciones adjuntas.

La bibliografía de patentes y científica a la que se hace referencia en el presente documento establece el conocimiento que está disponible para los expertos en la técnica.

REIVINDICACIONES

1. Un método que comprende:

5 aplicar una o más sondas multipartitas a una muestra fijada seleccionada del grupo que consiste en una célula, un órgano, un tejido y cualquier combinación de los mismos, en donde cada una de las una o más sondas multipartitas incluye al menos dos subsondas en donde las al menos dos subsondas se reasocian con una secuencia diana contigua en un ácido nucleico diana dentro de la muestra fijada;

reasociar al menos una de las una o más sondas multipartitas aplicadas con al menos un ARN diana dentro de la muestra fijada sin extracción de ARN;

10 ligar enzimáticamente al menos un extremo 3' a al menos un extremo 5-fosforilado de las al menos dos subsondas asociadas con la al menos una sonda multipartita reasociada dentro de la muestra fijada para crear un proxy de ácido nucleico diana que puede ser detectado;

liberar el proxy de ácido nucleico diana del ácido nucleico diana; y

amplificar el proxy de ácido nucleico diana.

15 2. El método de la reivindicación 1, en donde las subsondas comprenden sitios de unión de cebadores adjuntos para facilitar la amplificación posterior del proxy de ácido nucleico diana.

3. El método de una cualquiera de las reivindicaciones 1 a 2, en donde al menos una de las subsondas es un oligonucleótido de ADN que tiene: (i) una terminación 3' de al menos dos bases de ARN.

4. El método de una cualquiera de las reivindicaciones 1 a 3, en donde la muestra es una muestra de tejido fijada con formalina e incluida en parafina (FFPE).

20 5. El método de la reivindicación 4, en donde el espécimen de tejido fijado se procesa en secciones que tienen un espesor de 1 a 100 micrómetros.

6. El método de una cualquiera de las reivindicaciones 1 a 3, en donde la muestra:

(i) es fijada utilizando reactivos sin formalina;

(ii) está compuesta por células fijadas en suspensión;

25 (iii) es un cultivo de tejido fijado; o

(iv) es una muestra en una forma altamente degradada antes de la fijación.

7. El método de una cualquiera de las reivindicaciones 1 a 6, en donde el proxy de ácido nucleico diana se libera del ácido nucleico diana mediante una endonucleasa o se recupera desnaturalizando el proxy de ácido nucleico diana que contiene las al menos dos subsondas ligadas.

30 8. El método de la reivindicación 1, en donde el proxy de ácido nucleico diana amplificado se analiza utilizando una técnica seleccionada del grupo que consiste en Secuenciación de Próxima Generación (NGS), Secuenciación Profunda, secuenciación basada en espectrometría de masas o análisis de longitud, análisis de secuencia o longitud de fragmentos de ADN mediante electroforesis en gel o electroforesis capilar, reacción en cadena de la polimerasa (PCR) cuantitativa e hibridación sobre sondas de detección inmovilizadas.

35 9. El método de una cualquiera de las reivindicaciones 1 a 8, en donde cada una de las al menos dos subsondas de la pluralidad de sondas multipartitas tiene:

(i) 10-200 nucleótidos de longitud; o

(ii) 15-40 nucleótidos de longitud.

40 10. El método de una cualquiera de las reivindicaciones 1 a 9, en donde cada una de la pluralidad de sondas multipartitas incluye:

(i) dos subsondas; o

(ii) tres subsondas.

45 11. El método de una cualquiera de las reivindicaciones 1 a 10, en donde la muestra proviene de un sujeto que se sospecha que padece o corre riesgo de desarrollar una infección viral, una infección bacteriana o una infección fúngica; y en donde el ácido nucleico diana se selecciona del grupo que consiste en un ácido nucleico viral, un ácido nucleico bacteriano y un ácido nucleico fúngico.

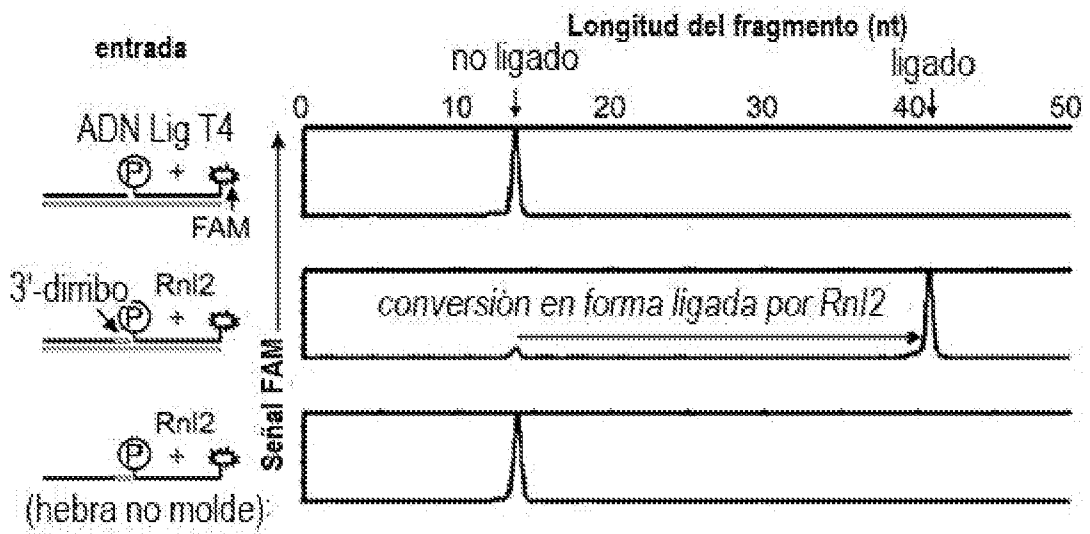


FIG. 1A

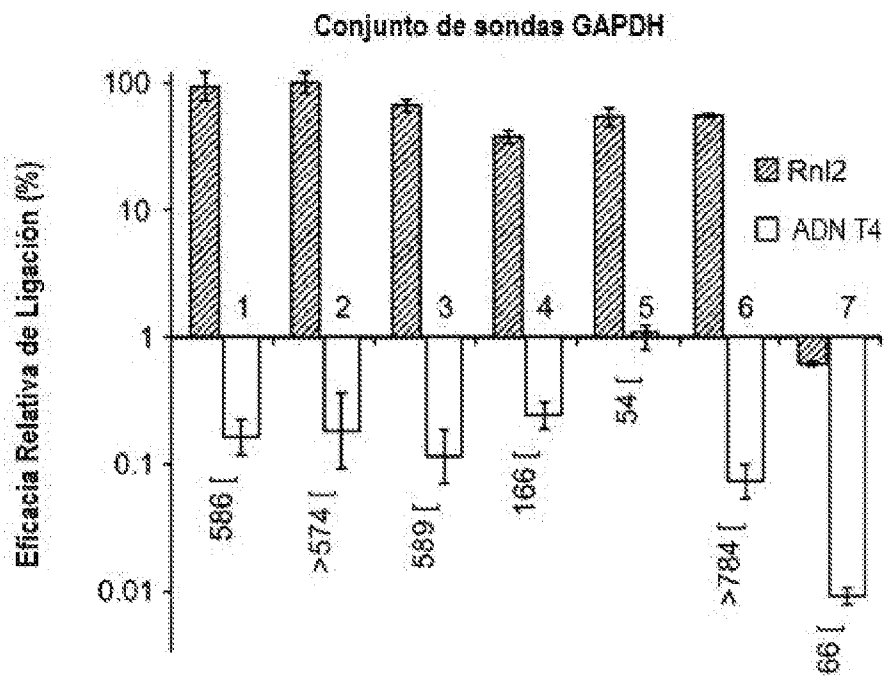


FIG. 1B

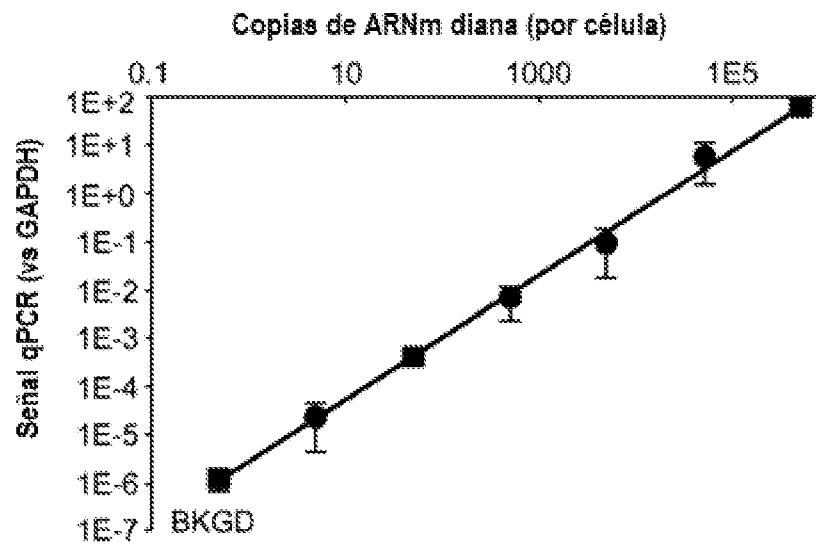


FIG. 1C

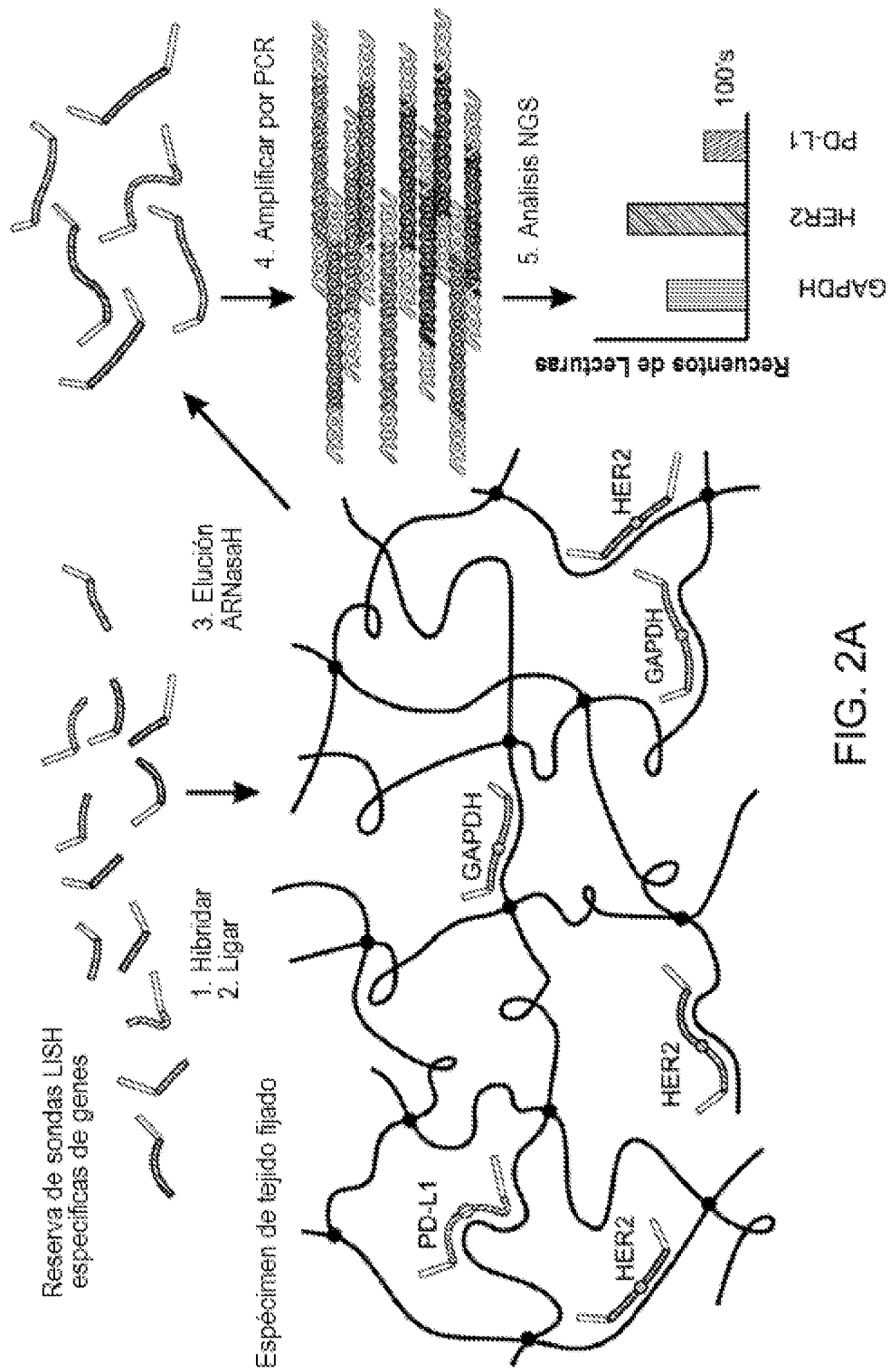


FIG. 2A

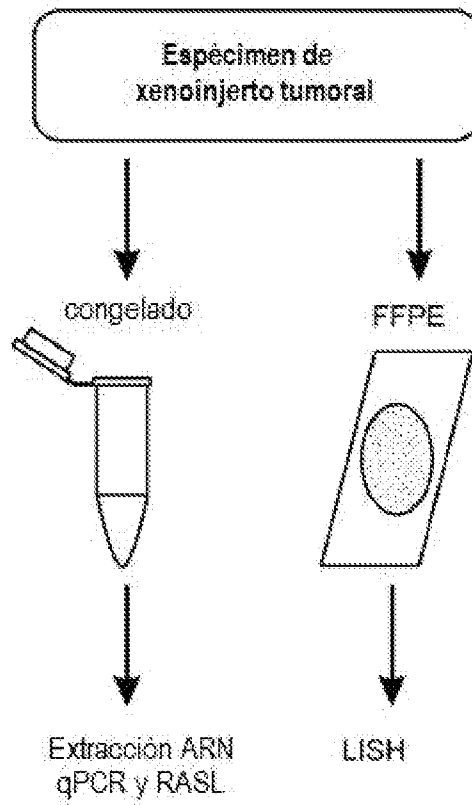


FIG. 2B

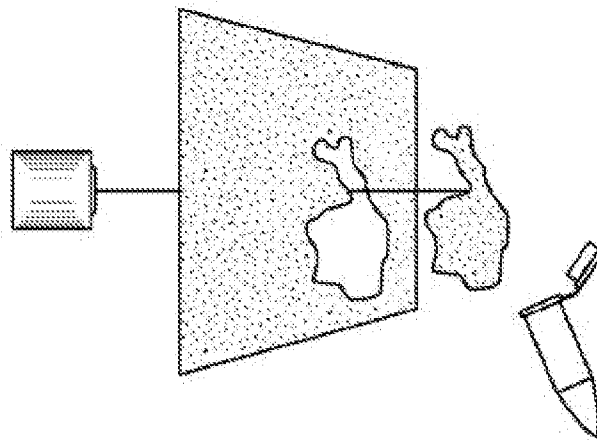
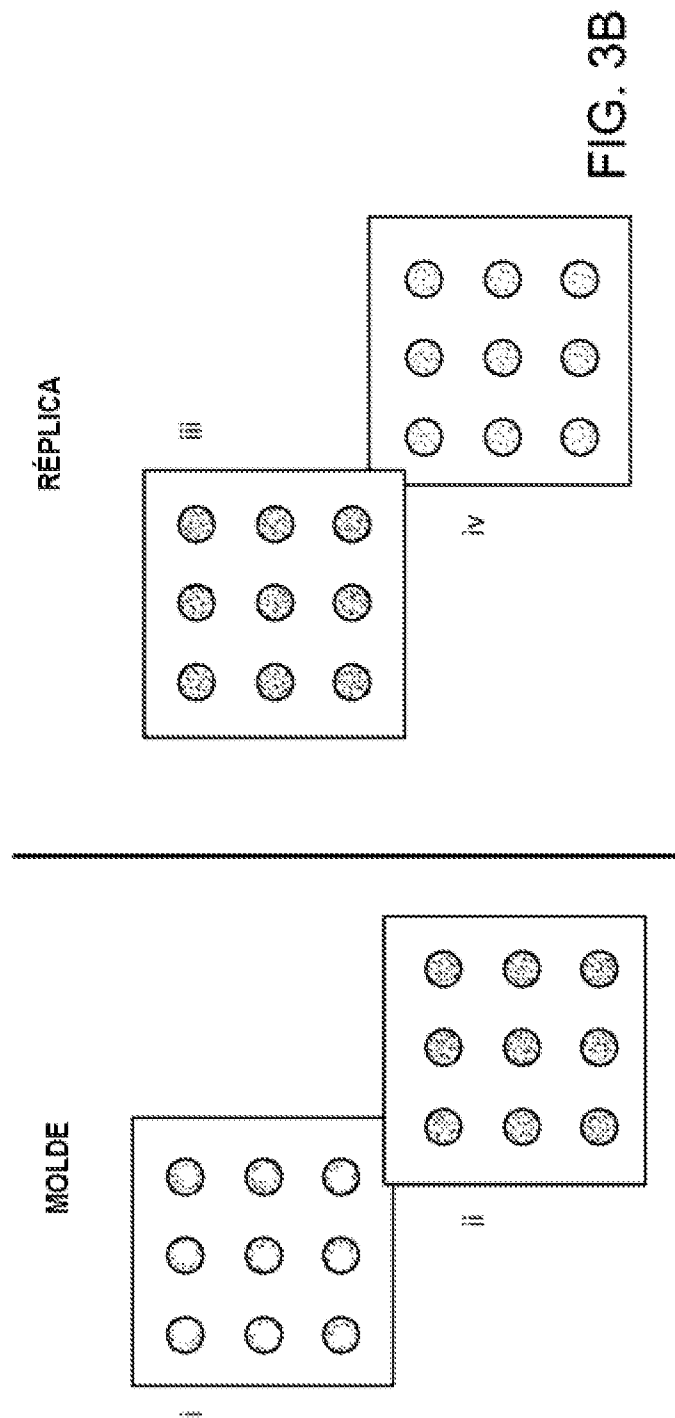
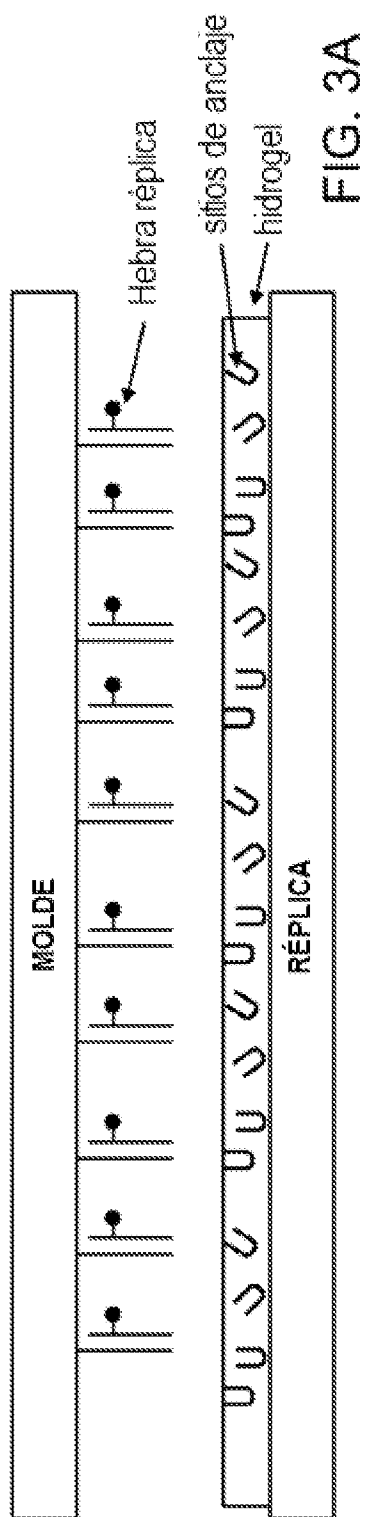


FIG. 2C



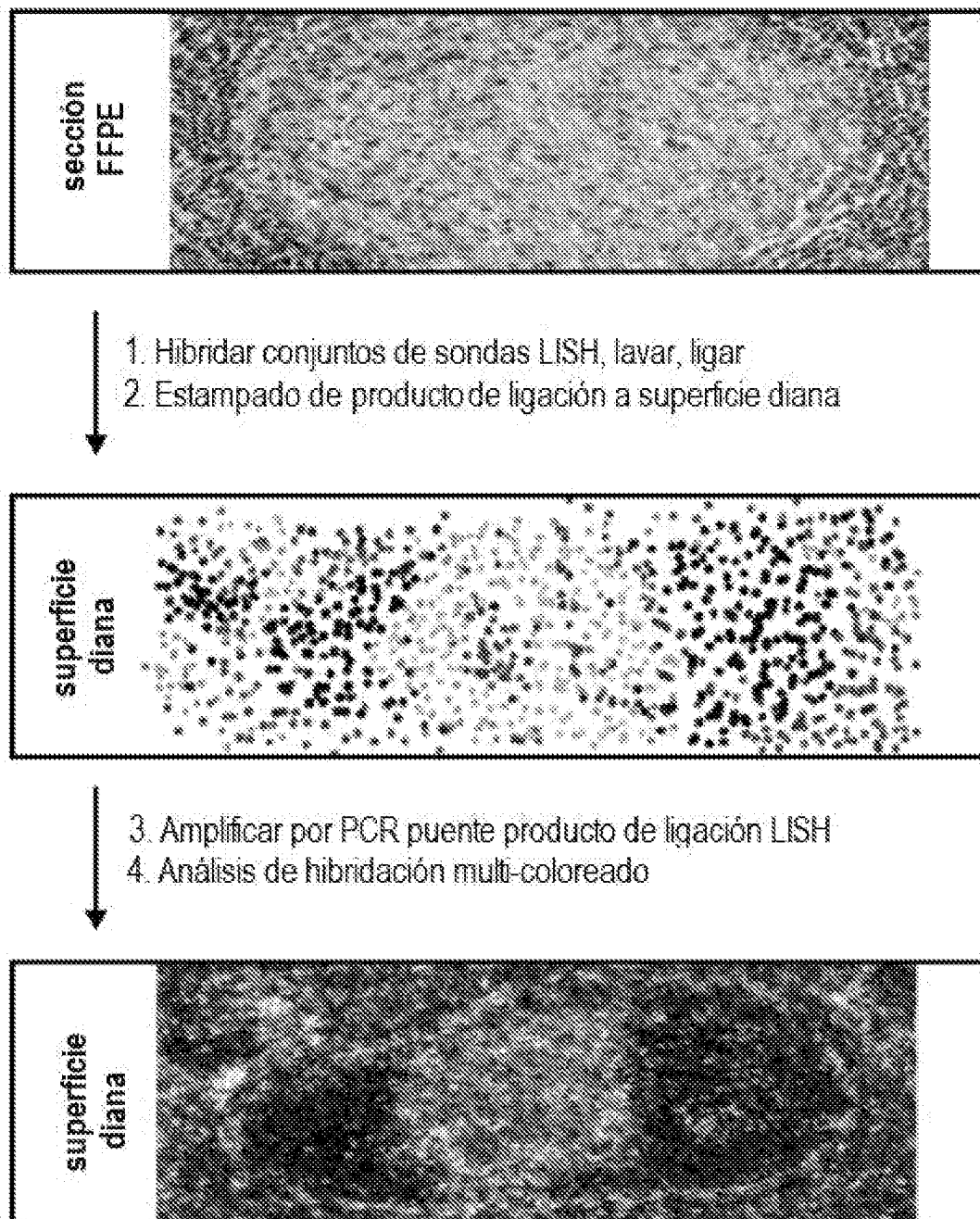


FIG. 4

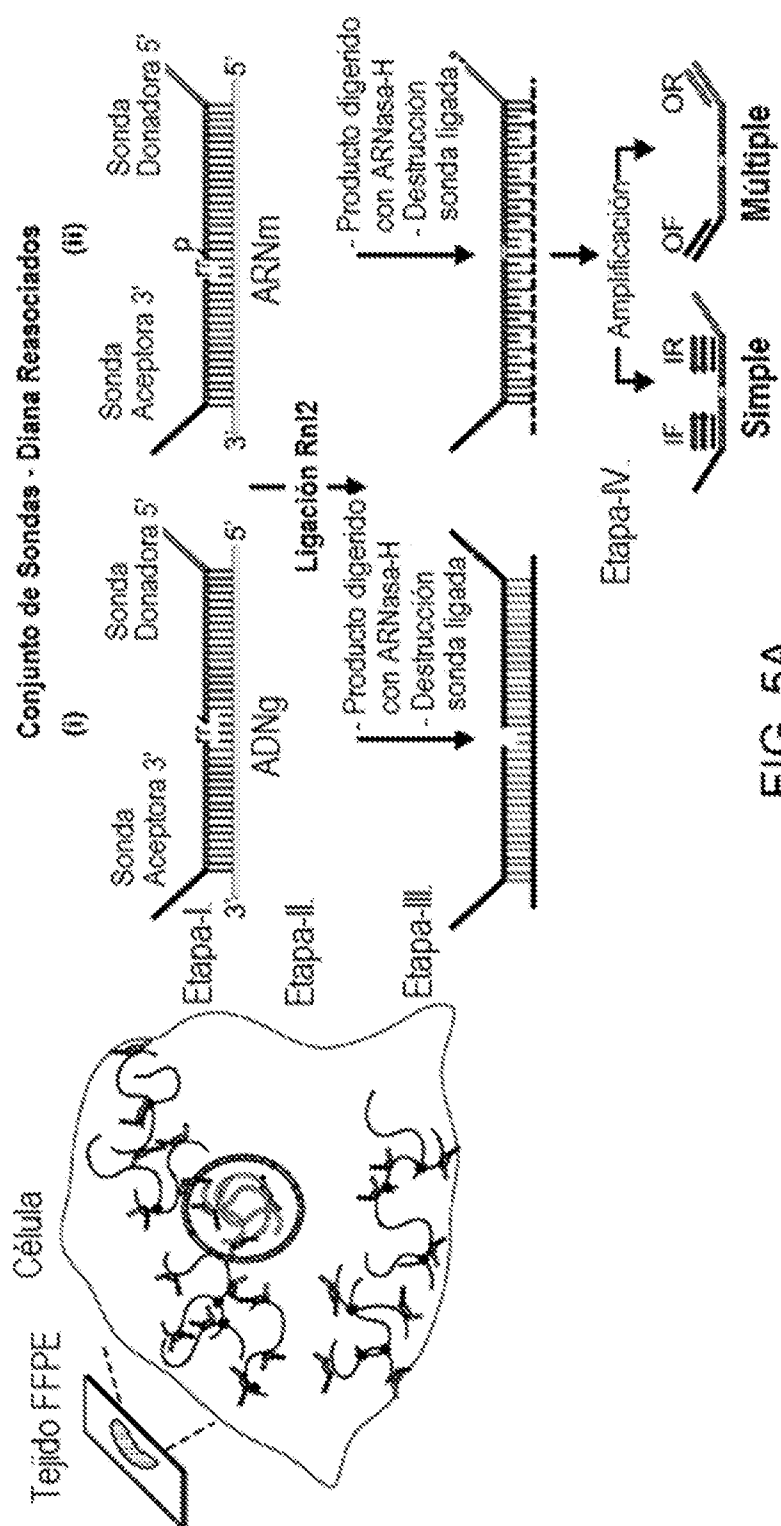


FIG. 5A

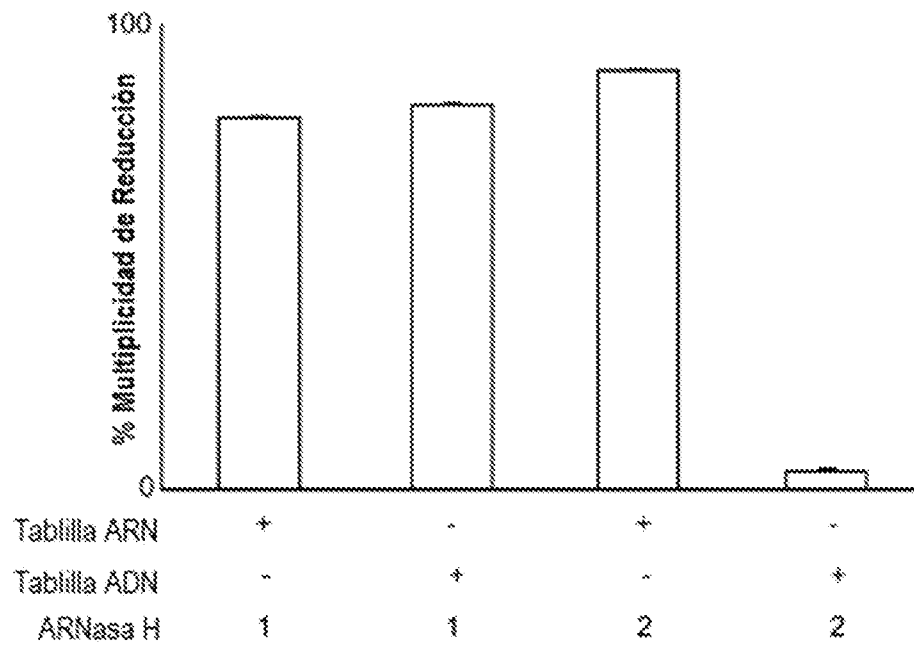


FIG. 5B

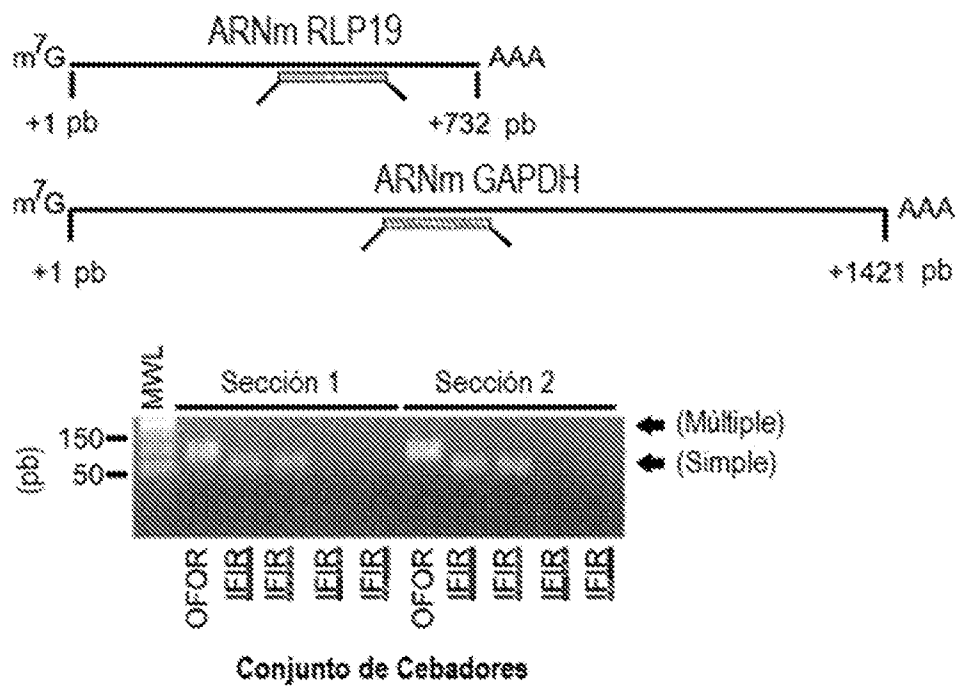


FIG. 5C

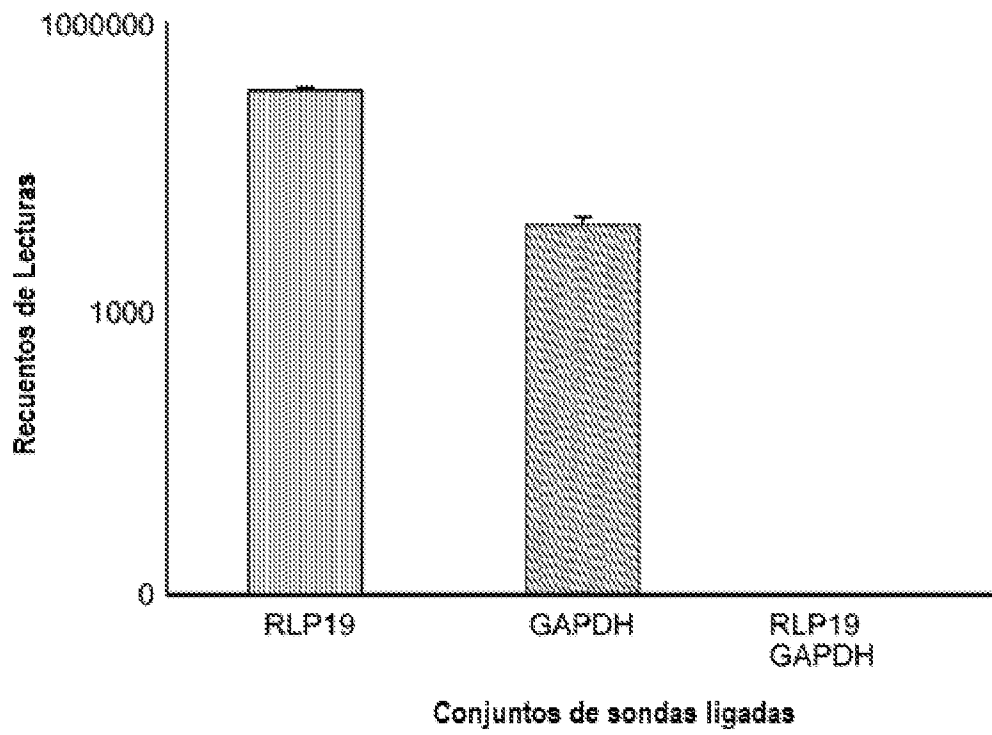


FIG. 5D

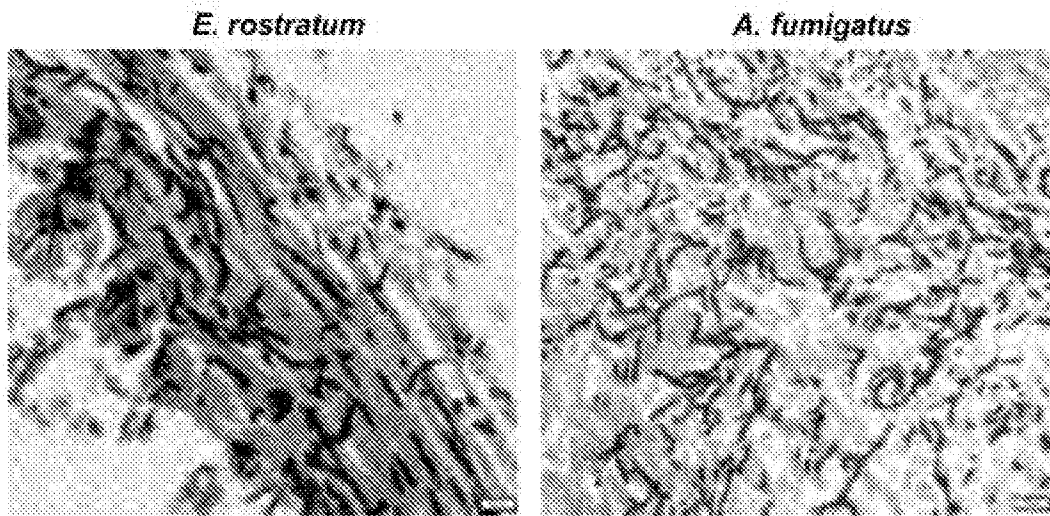


FIG. 6A

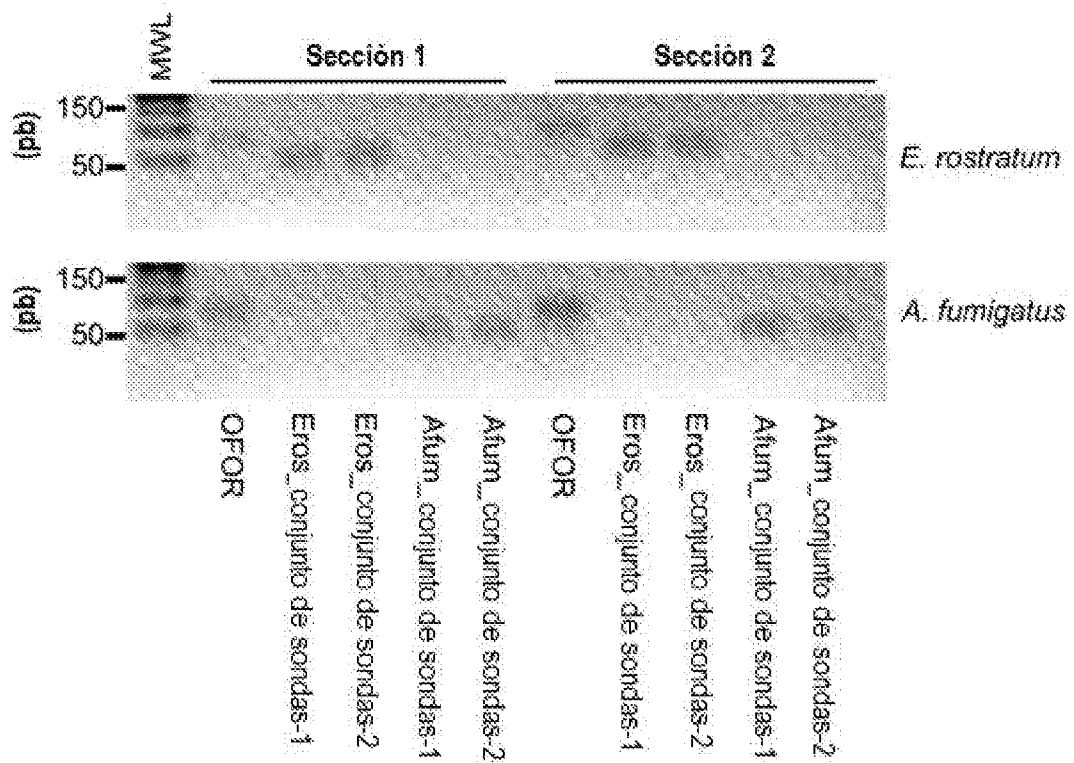


FIG. 6B

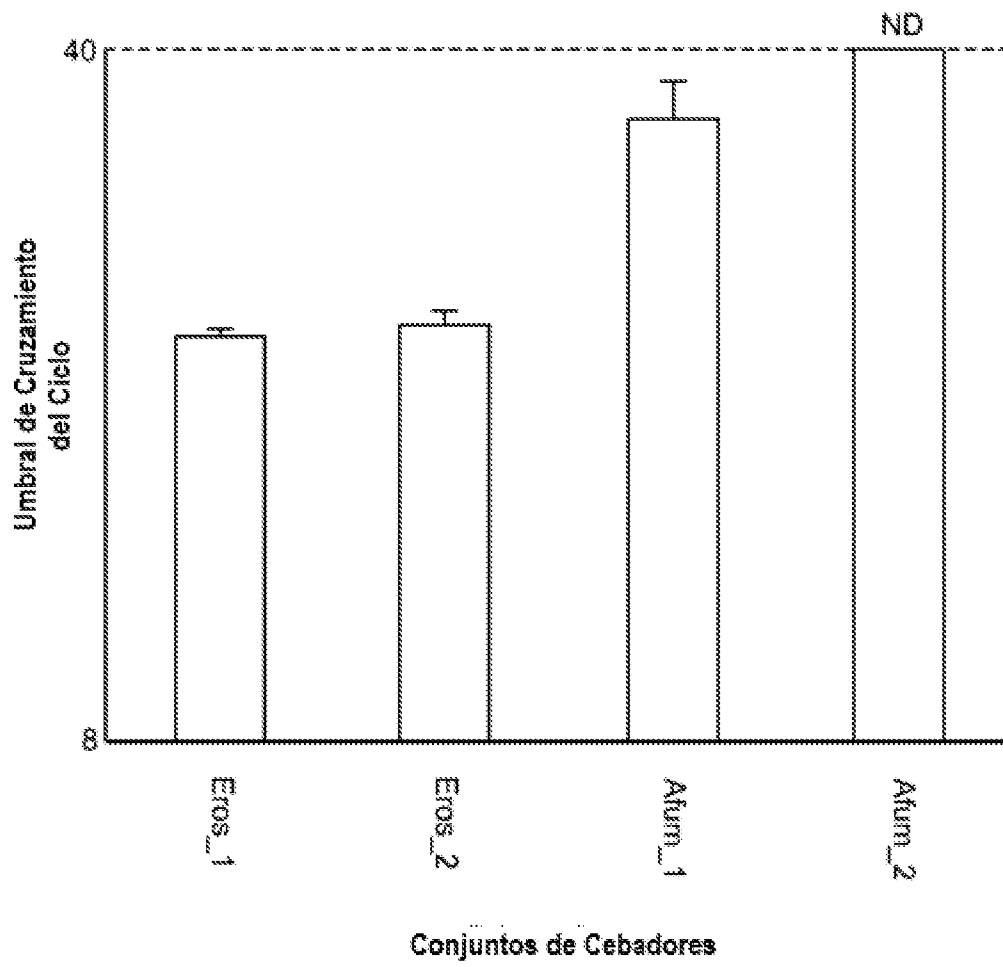


FIG. 6C

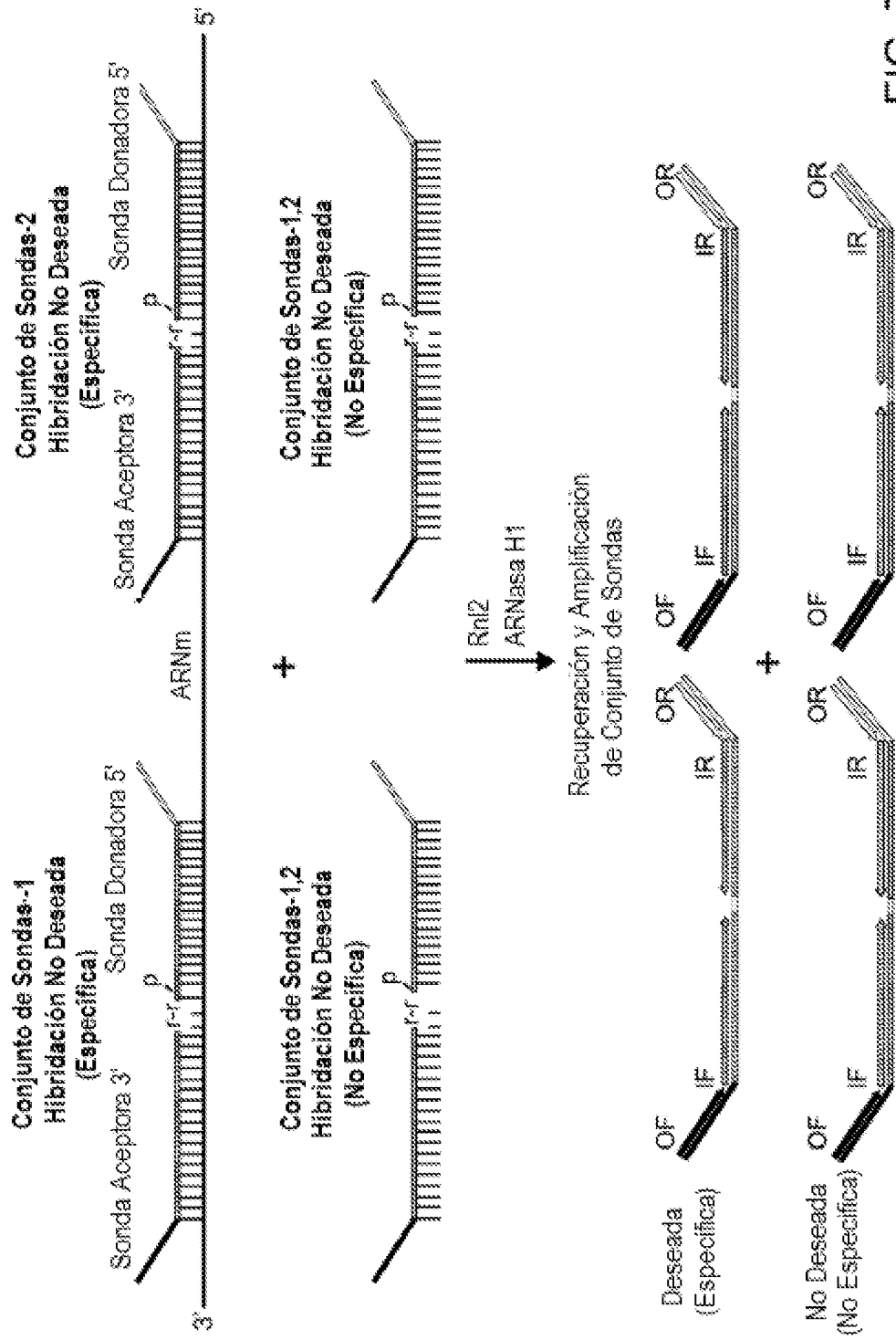


FIG. 7A

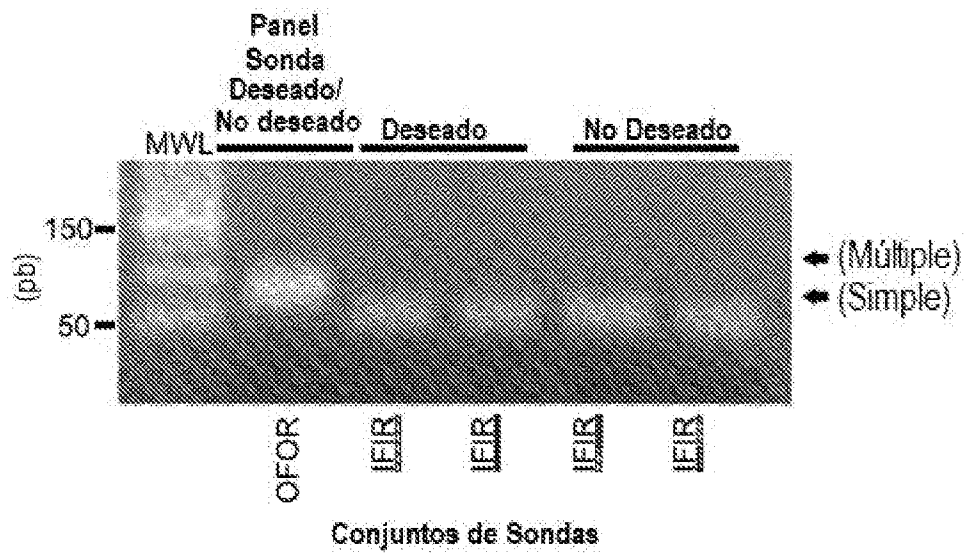


FIG. 7B

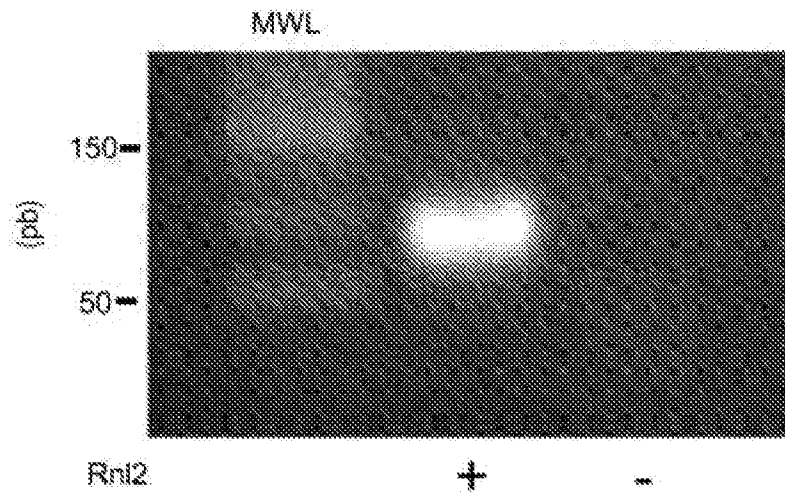


FIG. 7C

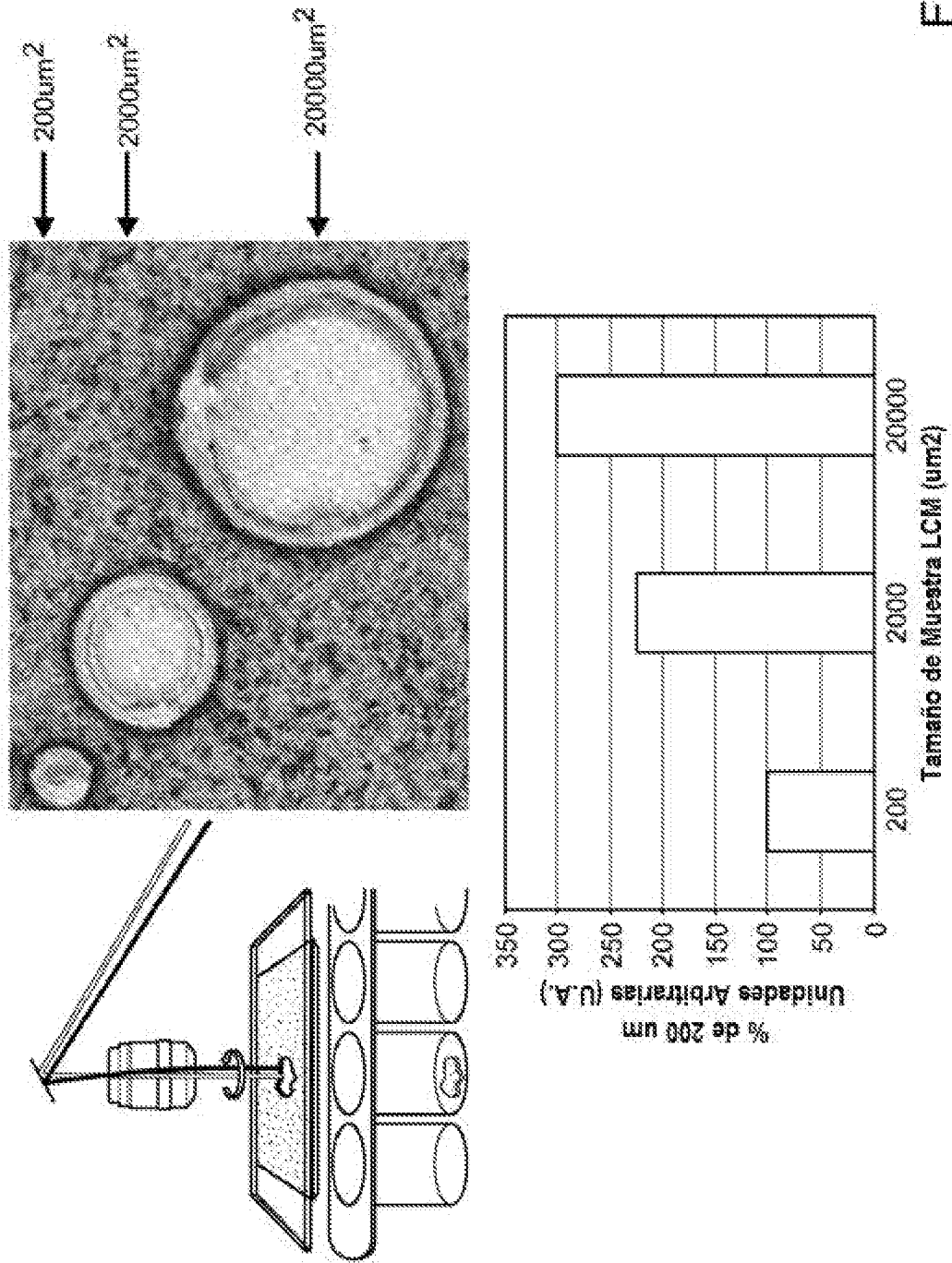


FIG. 8