

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 928 255**

51 Int. Cl.:

**C12Q 1/682** (2008.01)

**C12Q 1/6841** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.03.2018 PCT/EP2018/055422**

87 Fecha y número de publicación internacional: **18.10.2018 WO18188856**

96 Fecha de presentación y número de la solicitud europea: **06.03.2018 E 18707928 (0)**

97 Fecha y número de publicación de la concesión europea: **13.07.2022 EP 3610039**

54 Título: **Etiquetado de sondas oligonucleotídicas mediante ligazón de vías múltiples**

30 Prioridad:

**13.04.2017 EP 17166584**

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

**16.11.2022**

73 Titular/es:

**DEUTSCHES KREBSFORSCHUNGSZENTRUM  
STIFTUNG DES ÖFFENTLICHEN RECHTS  
(100.0%)  
Im Neuenheimer Feld 280  
69120 Heidelberg, DE**

72 Inventor/es:

**LIU, HAI-KUN y  
CHENG, YONG-SHENG**

74 Agente/Representante:

**SÁNCHEZ SILVA, Jesús Eladio**

ES 2 928 255 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

Etiquetado de sondas oligonucleotídicas mediante ligazón de vías múltiples

## 5 Campo de la invención

La presente invención proporciona un método novedoso para el etiquetado múltiple de sondas de ácido nucleico mediante ligazón de acuerdo con el conjunto de reivindicaciones adjuntas. El método usa una reacción catalizada por ligasa para conectar una sonda de ácido nucleico con moléculas portadoras de la etiqueta de ácido nucleico múltiple preparada previamente, en presencia de un oligonucleótido adaptador complementario estabilizador. El método permite un etiquetado fácil, económico y rápido de múltiples sondas con múltiples etiquetas diferentes. De esta manera, los costos y el esfuerzo para la generación de ensayos de hibridación fluorescente in situ (smFISH) de molécula única se redujeron significativamente, lo que permitió un código de barras de sondas de ácido nucleico multicolor combinatorio. La invención proporciona además métodos para la generación de bibliotecas FISH y kits de etiquetado que comprenden las nuevas herramientas de la invención.

## Descripción

La invención se define por las reivindicaciones. Cualquier tema que quede fuera del alcance de las reivindicaciones se proporciona solo con fines informativos.

Desde la invención de la hibridación in situ (ISH) por Joseph G. Gall y Mary-Lou Pardue hace más de 40 años, esta poderosa técnica ha transformado enormemente la investigación básica (es decir, la expresión génica, etc.) y el diagnóstico para la detección de biomarcadores. Solo a través de la ISH, los investigadores pudieron estudiar la expresión génica y diagnosticar el nivel de biomarcadores con información espacial. Hoy en día, la ISH y sus derivados se han convertido en el caballo de batalla en una variedad de campos, que incluyen histología, biología de célula única, etc. Incluso se realizaron esfuerzos en la escala ómica para tratar de tener un atlas completo de expresión génica, tal como el Atlas Allen Brain, basado en ISH. Con los avances en la química de fluoróforos y los hardware de microscopía, la hibridación de fluorescencia in situ (FISH) ha ganado más atención debido a una mejora en la señal de fondo baja superior y una señal espacial localizada en escala nm (frente a la señal difusiva generada en la reacción enzimática de la ISH) y podría servir como una buena alternativa para la ISH, especialmente en futuras aplicaciones en la medicina de precisión donde se requiera la cuantificación digital y espacio-temporal de biomarcadores.

Desde los años 90, Robert Singer y sus colegas han usado de manera innovadora la FISH para obtener imágenes in situ de transcritos de ARNm individuales a nivel de molécula única (smFISH), desde entonces la expresión génica o la detección de biomarcadores podrían realizarse de forma digital y de molécula única (figura 1). Sin embargo, la primera biblioteca de sondas se realizó a partir de cinco oligos específicos para genes diferentes, etiquetados con un pentafluoróforo, sintetizados de acuerdo a un diseño, lo cual todavía es muy costoso hoy en día y hace que esta poderosa invención no sea aplicable para todos los genes. A partir de 2008, Arjun y sus colegas inventaron un método de etiquetado combinado de fluoróforo único, que se basa en la química de conjugación clásica de NHS (N-hidroxisuccinimida), el cual reduce considerablemente el costo de la biblioteca de sondas. Y este nuevo etiquetado ahora tiene licencia de Biosearch Technologies Inc., por lo que los usuarios clínicos y de investigación podrían usar esta técnica a un precio de ~760 euros por 100-400 reacciones. Este costo excepcionalmente alto para el uso de la biblioteca comercial de sondas de FISH es una limitación fundamental de la smFISH, así como también su bajo rendimiento, típicamente el enfoque smFISH permite sondear solo unos pocos genes a la vez. Este bajo rendimiento se debe a la falta de sondas distinguibles con las que etiquetar las células y al costo de producir grandes cantidades de sonda etiquetada necesarias para una tinción de alta eficiencia. Por lo tanto, se necesitan mejoras en la generación de sondas smFISH y mejorar la eficiencia de detección.

La solicitud EP núm. 16190862.9 describe un método basado en la ligazón para generar oligonucleótidos etiquetados. Sin embargo, el método liga una secuencia de sonda con un segundo oligo etiquetado múltiple. El etiquetado múltiple de un oligo puede, sin embargo, dar lugar a problemas de espaciado debido a interferencia, tal como la extinción entre los restos del etiquetador.

El documento US2006/003333 A1 describe un método basado en ligazón para producir una sonda de ácido nucleico marcada que implica un oligonucleótido de unión que comprende un extremo 3' complementario a un oligonucleótido de sonda y un extremo 5' complementario a un ácido nucleico etiquetado.

El problema anterior se resuelve mediante un método para producir una sonda de oligonucleótido etiquetada o modificada de otra manera, el método comprende las etapas de:

(a) Proporcionar un oligonucleótido de secuencia de sonda que comprende (i) una secuencia de sonda que comprende una secuencia de nucleótidos complementaria a un ácido nucleico objetivo y (ii) una secuencia de nucleótidos etiqueta predeterminada complementaria a una primera secuencia de nucleótidos adaptadora,

- (b) Proporcionar un primer y un segundo, opcionalmente un tercer o más oligonucleótidos portadores de la etiqueta que comprenden al menos una porción de etiqueta u otra porción funcional, en donde el primer oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una segunda secuencia de nucleótidos adaptadora, el segundo oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una tercera secuencia de nucleótidos adaptadora, opcionalmente, el tercer o más oligonucleótidos portadores de la etiqueta consisten en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una cuarta o más secuencias de nucleótidos adaptadoras,
- (c) Proporcionar un oligonucleótido adaptador, que comprende en secuencia directa la primera, la segunda, la tercera y, opcionalmente, la cuarta o más secuencias de nucleótidos adaptadoras;
- (d) Poner en contacto en condiciones de hibridación el oligonucleótido de secuencia de sonda, el primero y el segundo, opcionalmente el tercero o más oligonucleótidos portadores de la etiqueta y el oligonucleótido adaptador, para formar un complejo, en donde, en el complejo, un grupo 3' OH libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta o un grupo fosfato libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente de un grupo 3' OH libre del primer oligonucleótido portador de la etiqueta, un grupo 3' OH libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta o un grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo 3' OH libre del primer oligonucleótido portador de la etiqueta y, opcionalmente, un grupo 3' OH libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del tercer o más oligonucleótidos portadores de la etiqueta o un grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo 3' OH libre del tercer o más oligonucleótidos portadores de la etiqueta,
- (e) Hacer reaccionar el complejo para formar un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del oligonucleótido de secuencia de sonda y el primer oligonucleótido portador de la etiqueta, un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del primer oligonucleótido portador de la etiqueta y el segundo oligonucleótido portador de la etiqueta y opcionalmente un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del segundo oligonucleótido portador de la etiqueta y el tercer o más oligonucleótidos portadores de la etiqueta, mediante el uso de una ligasa en condiciones de ligazón, para formar la sonda oligonucleotídica etiquetada o modificada de otra manera,
- (f) Opcionalmente, retirar el oligonucleótido adaptador.

El método inventivo puede usarse tanto para la unión de oligonucleótidos portadores de la etiqueta al extremo 3' como al extremo 5' del oligonucleótido de secuencia de sonda. Por lo tanto, es preferible un método de la invención en una modalidad en donde en el oligonucleótido de secuencia de sonda la secuencia de sonda está en 5' de la secuencia de nucleótidos etiqueta predeterminada y en el oligonucleótido adaptador, la primera y la segunda secuencia de nucleótidos adaptadora están en dirección 3' a 5' en secuencia directa y en el complejo en la etapa (d) el grupo 3' OH libre del oligonucleótido de secuencia de sonda está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta, el grupo 3' OH libre del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta y, opcionalmente, el grupo 3' OH libre del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del tercer o más oligonucleótidos portadores de la etiqueta. En una modalidad alternativa de la invención, se proporciona un método en donde en el oligonucleótido de secuencia de sonda la secuencia de sonda está en 3' de la secuencia de nucleótidos etiqueta predeterminada y en el oligonucleótido adaptador, la primera y la segunda secuencia de nucleótidos adaptadora están en dirección 5' a 3' en secuencia directa y en el complejo en la etapa (d) el grupo fosfato libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente del grupo 3' OH libre del primer oligonucleótido portador de la etiqueta, el grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo 3' OH libre del segundo oligonucleótido portador de la etiqueta y opcionalmente, el grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo 3' OH libre del tercer o más oligonucleótidos portadores de la etiqueta.

Los métodos y compuestos de la invención permiten un etiquetado rápido, económico y sencillo de las sondas de ácido nucleico. La idea de la invención es proporcionar un sistema que permita al médico etiquetar cualquier sonda de oligonucleótido dada (oligonucleótido de secuencia de sonda) en el estante sin la necesidad de unir la porción de la etiqueta a la secuencia de sonda individual, lo cual es costoso y por lo tanto perjudica la generación de bibliotecas de muchas sondas individuales etiquetadas. Con este fin, la invención proporciona ahora un sistema que comprende al menos un primer y un segundo, opcionalmente más preparado previamente, oligonucleótido portador de la etiqueta, en donde cada oligonucleótido portador de la etiqueta comprende una etiqueta diferente u otra modificación. A continuación, la invención proporciona un oligonucleótido adaptador con una secuencia complementaria a la(s) secuencia(s) de nucleótidos etiqueta(s) predeterminada(s) en el oligonucleótido de secuencia de sonda y en los oligonucleótidos portadores de la etiqueta para ensamblar estas moléculas en un complejo donde pueden fusionarse entre sí mediante ligazón simple. De esta forma, la invención proporciona un enfoque sencillo

para unir múltiples etiquetas diferentes a sondas de secuencia de nucleótidos mediante ligazón. Mediante el uso de solo unas pocas, preferentemente solo una etiqueta u otra modificación para cada uno de los oligonucleótidos portadores de la etiqueta ligados, cada una de estas etiquetas también está suficientemente separada entre sí para evitar interferencias.

5 El método de la invención en la etapa (b) comprende proporcionar un primero y un segundo, opcionalmente un tercer o más (un cuarto, quinto, sexto, 50<sup>mo</sup>... 100<sup>mo</sup>, etc.), oligonucleótidos portadores de la etiqueta, en donde el primer oligonucleótido portador de la etiqueta tiene una secuencia de nucleótidos etiqueta predeterminada complementaria a una segunda secuencia de nucleótidos adaptadora, el segundo oligonucleótido portador de la  
10 etiqueta tiene una secuencia de nucleótidos etiqueta predeterminada complementaria a una tercera secuencia de nucleótidos adaptadora, opcionalmente, el tercer o más oligonucleótido portador de la etiqueta tiene una secuencia de nucleótidos etiqueta predeterminada complementaria a una cuarta o más secuencias de nucleótidos adaptadoras y en donde en la etapa (c) el oligonucleótido adaptador comprende en secuencia directa la primera, la segunda, la  
15 tercera, opcionalmente la cuarta o más secuencias de nucleótidos adaptadoras. De este modo, el oligonucleótido de secuencia de sonda de la invención puede etiquetarse fácilmente con múltiples etiquetas u otras porciones.

En el contexto de la invención, cuando se hace referencia a un oligonucleótido o sonda específica, o a otras expresiones que describen ácidos nucleicos, el experto reconocerá que la expresión no se refiere a moléculas únicas, sino a una única especie de moléculas. En la aplicación práctica en el laboratorio, el experto usa  
20 preparaciones de una multitud de moléculas de una especie de, por ejemplo, una secuencia de oligonucleótidos. En tal población de moléculas, normalmente todos los nucleótidos son idénticos, excepto cuando los oligonucleótidos se produjeron para contener una región de secuencia aleatoria, por ejemplo, durante la síntesis de oligonucleótidos. Esto puede lograrse al polimerizar el oligonucleótido con una mezcla igual de más de un tipo de nucleótido, que luego se acoplan al azar. En este caso, la especie de oligonucleótidos en la mezcla contiene una secuencia  
25 "redundante".

El término "nucleósido" y "nucleótido" pretenden incluir aquellas porciones que contienen no solo las bases de purina y pirimidina conocidas, sino también otras bases heterocíclicas que se modificaron. Tales modificaciones incluyen purinas o pirimidinas metiladas, purinas o pirimidinas aciladas, ribosas alquiladas u otros heterociclos. Además, el  
30 término "nucleótido" incluye las porciones que contienen hapteno o etiquetas fluorescentes y pueden contener no solo azúcares ribosa y desoxirribosa convencionales, sino también otros azúcares. Los nucleósidos o nucleótidos modificados también incluyen modificaciones en la porción azúcar, por ejemplo, en donde uno o más de los grupos hidroxilo se reemplazan con átomos de halógeno o grupos alifáticos, se funcionalizan como éteres, aminas, ácido nucleico peptídico (PNA), ácido nucleico bloqueado o similares.

El término "ácido nucleico" se refiere a un polímero de cualquier longitud, por ejemplo, mayor de aproximadamente 2 bases, mayor de aproximadamente 10 bases, mayor de aproximadamente 100 bases, mayor de aproximadamente 500 bases, mayor de 1000 bases, hasta aproximadamente 10 000 o más (por ejemplo, 100 000 000 o más) bases  
40 compuestas de nucleótidos, por ejemplo, desoxirribonucleótidos o ribonucleótidos y pueden producirse enzimática o sintéticamente, que pueden hibridar con ácidos nucleicos naturales de una manera específica de secuencia análoga a la de dos ácidos nucleicos naturales, por ejemplo, pueden participar en interacciones de apareamiento de bases de Watson-Crick. Los nucleótidos de origen natural incluyen guanina, citosina, adenina y timina/uracilo (G, C, A y T/U, respectivamente).

El término "oligonucleótido", como se usa en la presente descripción, denota un multímero de nucleótido monocatenario de aproximadamente 2 a aproximadamente 500 nucleótidos. Los oligonucleótidos pueden ser sintéticos o pueden prepararse enzimáticamente. Los oligonucleótidos pueden contener monómeros de ribonucleótidos (es decir, pueden ser oligorribonucleótidos), monómeros de desoxirribonucleótidos o una  
50 combinación de los dos. Los oligonucleótidos pueden ser de 10 a 20, 11 a 30, 31 a 40, 41 a 50, 51 a 60, 61 a 70, 71 a 80, 80 a 100, 100 a 150, 150 a 200 o 200 a 250 o hasta 500 nucleótidos de longitud.

El término "hibridación" se refiere a la unión específica de un ácido nucleico a un ácido nucleico complementario a través del apareamiento de bases de Watson-Crick. En consecuencia, el término "hibridación in situ" se refiere a la unión específica de un ácido nucleico a un ácido nucleico complementario dentro de una célula o en un cromosoma  
55 intacto. Los términos "hibridación" y "unión", con respecto a los ácidos nucleicos, se usan indistintamente.

El término "condiciones de hibridación" pretende significar aquellas condiciones de tiempo, temperatura y pH y las cantidades y concentraciones necesarias de reactantes y reactivos, suficientes para permitir que al menos una parte de las secuencias complementarias se hibriden entre sí. Como se conoce bien en la técnica, las condiciones de  
60 tiempo, temperatura y pH requeridas para lograr la hibridación dependen del tamaño de las moléculas de oligonucleótidos a hibridar, el grado de complementariedad entre los oligonucleótidos a hibridar y la presencia de otros materiales en la mezcla de reacción de hibridación, sales, etc. Las condiciones reales necesarias para cada etapa de la hibridación se conocen bien en la técnica o pueden determinarse sin experimentación muy específica.

El término "condiciones de hibridación in situ" como se usa en la presente descripción, se refiere a condiciones que permiten la hibridación de un ácido nucleico con un ácido nucleico complementario, por ejemplo, una secuencia de

nucleótidos en una molécula de ARN o ADN y un oligonucleótido complementario, en una célula. Las condiciones de hibridación in situ adecuadas pueden incluir tanto condiciones de hibridación como condiciones de lavado opcionales, tales condiciones incluyen temperatura, concentración de reactivos desnaturalizantes, sales, tiempo de incubación, etc. Tales condiciones se conocen en la técnica.

En algunas modalidades de la invención, el oligonucleótido de secuencia de sonda, el oligonucleótido portador de la etiqueta y el oligonucleótido adaptador son moléculas de ácido nucleico monocatenario, preferentemente ARN y/o ADN. Además, todas las moléculas de ácido nucleico pueden contener residuos de ácido nucleico modificados o no naturales, tales como residuos con modificaciones O-metilo. Sin embargo, se prefieren específicamente los oligonucleótidos de ADN que son más fáciles y económicos de sintetizar.

En algunas modalidades, se prefiere que el oligonucleótido de secuencia de sonda no comprenda una modificación del grupo OH del extremo 3' y/o del grupo fosfato del extremo 5' del polinucleótido, en particular es preferible que el oligonucleótido de secuencia de sonda no comprenda un grupo amino del extremo. En el contexto de la invención, es probable que el oligonucleótido de secuencia de sonda se sintetice químicamente de acuerdo con los procedimientos de síntesis de oligonucleótidos del estado de la técnica. Dichos procedimientos incluyen a menudo el uso de grupos de enmascaramiento o bloqueo, que preferentemente se eliminan todos antes de aplicar el oligonucleótido de secuencia de sonda al método de la invención. Además, la síntesis de oligonucleótidos a menudo da como resultado un grupo OH en el extremo 5' prima. En algunas modalidades en las que el oligonucleótido portador de la etiqueta está ligado al extremo 5' prima del oligonucleótido de secuencia de sonda, es necesario unir un grupo fosfato al extremo 5' prima para permitir la reacción de ligazón.

El oligonucleótido portador de la etiqueta comprende al menos, en algunas modalidades, dos o más, preferentemente tres, cuatro o cinco, porciones de etiquetado u otras porciones funcionales.

Sin embargo, en modalidades más preferidas, el oligonucleótido portador de la etiqueta comprende solo uno, o no más de dos o tres, de dichas porciones etiquetadas u otras porciones funcionales. De acuerdo con la invención, cada oligonucleótido de secuencia de sonda puede etiquetarse con múltiples etiquetas mediante el uso de una pluralidad de oligonucleótidos portadores de la etiqueta que están todos ligados al oligonucleótido de secuencia de sonda en tándem mediante el método de la invención. A continuación, cada uno de los oligonucleótidos portadores de la etiqueta de la pluralidad de oligonucleótidos portadores de la etiqueta se marca con una etiqueta o modificación diferente. Esta modalidad permite un sondeo simultáneo de múltiples secuencias objetivo tales como especies de ARNm en una célula.

El oligonucleótido portador de la etiqueta puede ser de cualquier longitud adecuada para la ligazón con el oligonucleótido de secuencia de sonda. Algunos oligonucleótidos portadores de la etiqueta preferidos de la invención tienen una longitud de 5 a 200 nucleótidos, preferentemente de 5 a 100 nucleótidos, con mayor preferencia de 5 a 30 u 8 a 20, con mayor preferencia aproximadamente 15. La secuencia del oligonucleótido portador de la etiqueta en modalidades preferidas debería permitir una unión específica de la secuencia de nucleótidos etiqueta predeterminada a la secuencia adaptadora respectiva.

El oligonucleótido portador de la etiqueta en una modalidad preferida comprende un grupo fosfato 5' prima (libre) para permitir una ligazón catalizada por enzima con el oligonucleótido de secuencia de sonda. En otra modalidad, el oligonucleótido portador de la etiqueta comprende un grupo OH terminal de extremo 3' prima (en caso de que el etiquetado del oligonucleótido de secuencia de sonda se realice en el extremo 5').

El término "oligonucleótido adaptador" generalmente se refiere a un oligonucleótido que comprende en una disposición en tándem al menos dos, preferentemente más, secuencias adaptadoras que son complementarias a las secuencias de nucleótidos etiqueta predeterminadas en el oligonucleótido de secuencia de sonda y el(los) oligonucleótido(s) portador(es) de etiqueta. Por tanto, las secuencias adaptadoras también están predeterminadas. Es en algunas modalidades que las secuencias de las secuencias de nucleótidos etiqueta predeterminadas y, por complementariedad, también las secuencias adaptadoras correspondientes, en el oligonucleótido de secuencia de sonda y cada uno de los oligonucleótidos portadores de la etiqueta son diferentes entre sí. En una modalidad donde el etiquetado del oligonucleótido de secuencia de sonda se realiza en su extremo 3', el oligonucleótido adaptador comprende en orden consecutivo desde su extremo 3' a 5' la primera, luego la segunda secuencia adaptadora. Si se usan oligonucleótidos portadores de la etiqueta adicionales, las secuencias adaptadoras se añaden consecutivamente al extremo 3' del oligonucleótido adaptador (tercero, cuarto, quinto, etc.). El orden será de 5' a 3' si el etiquetado se realiza en el extremo 5' del oligonucleótido de secuencia de sonda. De esta forma, cuando se ponen en contacto en condiciones de hibridación, el oligonucleótido de secuencia de sonda, todos los oligonucleótidos portadores de la etiqueta y el oligonucleótido adaptador forman un complejo bicatenario en el que la cadena compuesta por el oligonucleótido de secuencia de sonda y todos los oligonucleótidos portadores de la etiqueta comprenden "muescas" ligables entre cada oligonucleótido individual. Una muesca ligable es un espacio entre dos oligonucleótidos lo suficientemente estrecho como para permitir que un grupo 3'OH y 5'fosfato se involucre en una reacción de ligasa catalizada por enzimas para formar un enlace covalente y, por lo tanto, una cadena principal de fosfodiéster. En algunas modalidades, los extremos 3' y 5' del oligonucleótido adaptador pueden

modificarse para que no puedan participar en una reacción de ligazón para evitar subproductos no deseados durante la ligazón.

El oligonucleótido adaptador comprende tres o más secuencias adaptadoras, cada una de las cuales tiene al menos 3, preferentemente 4, 5, 6, 7, 8 o más posiciones de ácido nucleico, preferentemente de 8 a 20, de 10 a 20, de 12 a 18, con mayor preferencia aproximadamente 15. En algunas modalidades, cada secuencia adaptadora consiste en una secuencia complementaria inversa (cuando se ve de 5' a 3') a la respectiva secuencia de nucleótidos etiqueta predeterminada. El grado de complementariedad entre la secuencia adaptadora del oligonucleótido adaptador y la secuencia de nucleótidos etiqueta predeterminada correspondiente del oligonucleótido portador de la etiqueta o del oligonucleótido de secuencia de sonda se selecciona para permitir la hibridación estable de las dos secuencias en condiciones adecuadas para realizar una ligazón. En este contexto, el grado de complementariedad inversa es preferentemente al menos 90 %, con la máxima preferencia 95 % o 100 %.

De acuerdo con la presente invención, el oligonucleótido adaptador no comprende un saliente degenerado en el extremo 3' y/o 5' y tampoco en el extremo que cuando se forma el complejo es capaz de hibridarse con la secuencia de sonda en el oligonucleótido de secuencia de sonda.

En algunas modalidades, el método de la invención puede comprender además una etapa de purificación del producto ligado del oligonucleótido de secuencia de sonda y del(de los) oligonucleótido(s) portador(es) de la etiqueta. El producto de ligazón forma la sonda de oligonucleótido etiquetada o modificada de otra manera y, por tanto, el producto del proceso inventivo. La purificación puede realizarse por cualquier medio conocido por el experto en la técnica para purificar sondas de ácido nucleico, e incluye, pero no se limita a, purificación mediante electroforesis en gel, tal como PAGE. Con la máxima preferencia, el producto ligado monocatenario primero se purifica y, posteriormente, en una segunda etapa, se hibrida de nuevo con el oligonucleótido adaptador para estabilizar la sonda antes de su uso.

El término "ligasa", como se usa en la presente descripción, se refiere a una enzima que se usa comúnmente para unir polinucleótidos entre sí o para unir los extremos de un único polinucleótido. Una ligasa de la invención se selecciona preferentemente entre ligasas de polinucleótidos bicatenarios dependientes de ATP, ligasas de ARN o de ADN bicatenario dependientes de NAD<sup>+</sup> y ligasas de polinucleótidos monocatenarios y preferentemente se seleccionan entre ligasas bacterianas, tales como ADN ligasa de E. coli, ADN ligasa Taq, ADN ligasa termoestable Ampligase®, ligasas de fagos, tal como ADN T3 ligasa, ADN T4 ligasa, ADN T7 ligasa y mutantes de las mismas, incluidas ligasas de fusión que contienen un dominio de unión al ADN y una ligasa, tal como ADN ligasa Sso7-T3, ADN ligasa Sso7-T4, ADN ligasa Sso7-T7, ADN ligasa Sso7-Taq, ADN ligasa Sso7-E. coli, ADN ligasa Sso7-Ampligasa, ADN ligasa Sso7, ARN T4 ligasa 1, ARN T4 ligasa 2 y ARN T4 ligasa truncada y mutada (K227Q). Con la máxima preferencia se usa ADN T4 ligasa, porque esta enzima es robusta, barata y eficiente.

El término "objetivo" se refiere a una entidad biológica que puede separarse espacialmente, hibridarse con una sonda y visualizarse. Las células, los cromosomas individuales y el material depositado en una matriz son ejemplos de objetivos. En el contexto de la invención, el ácido nucleico objetivo es un ácido nucleico monocatenario objetivo, que comprende una secuencia que tiene al menos una, preferentemente múltiples, regiones de unión para sondas de ácido nucleico como se produce mediante el método de la presente invención. En modalidades preferidas, el ácido nucleico objetivo es un ARN mensajero (ARNm).

Una etiqueta en el contexto de la invención es una porción detectable que puede producir una señal directa o indirectamente. Un ejemplo de una porción detectable que produce una señal directamente es una molécula fluorescente. Las porciones detectables que producen una señal indirectamente incluyen porciones que producen una señal después de la exposición a reactivos de detección, tales como sustratos o anticuerpos, etc. Una porción detectable que produce una señal directamente puede detectarse opcionalmente por medios indirectos, tales como el uso de un anticuerpo etiquetado que se une a la porción. En determinados casos, una señal puede tener una longitud de onda particular que es detectable por un fotodetector, por ejemplo, un microscopio óptico, un espectrofotómetro, un microscopio fluorescente, un lector de muestras fluorescentes o un clasificador de células activado por fluorescencia, etc. En el contexto de la invención la porción de etiqueta puede ser cualquier porción que permita la detección de la presencia o ausencia de la porción. Las etiquetas adecuadas incluyen colorantes fluorescentes que incluyen colorantes de xanteno, por ejemplo, colorantes de fluoresceína y rodamina, tales como isotiocianato de fluoresceína (FITC), 6-carboxifluoresceína (comúnmente conocida por las abreviaturas FAM y F), 6-carboxi-2',4',7',4,7-hexaclorofluoresceína (HEX), 6-carboxi-4',5'-dicloro-2',7'-dimetoxifluoresceína (JOE o J), N,N,N',N'-tetrametil-6-carboxirrodamina (TAMRA o T), 6-carboxi-X-rodamina (ROX o R), 5-carboxirrodamina-6G (R6G<sup>5</sup> o G<sup>5</sup>), 6-carboxirrodamina-6G (R6G<sup>6</sup> o G<sup>6</sup>) y rodamina 110; colorantes de cianina, por ejemplo colorantes Cy3, Cy5 y Cy7; cumarinas, por ejemplo, umbeliferona; colorantes de bencimida, por ejemplo, Hoechst 33258; colorantes de fenantridina, por ejemplo, Rojo Texas; colorantes de etidio; colorantes de acridina; colorantes de carbazol; colorantes de fenoxazina; colorantes de porfirina; colorantes de polimetina, por ejemplo colorantes de cianina tales como Cy3, Cy5, etc.; colorantes BODIPY y colorantes de quinolina. Los fluoróforos específicos de interés que se usan comúnmente en algunas aplicaciones incluyen: pireno, cumarina, dietilaminocumarina, FAM, fluoresceína clorotriazinilo, R110, eosina, JOE, R6G, tetrametilrodamina, TAMRA, lisamina, ROX, naftofluoresceína, rojo Texas, naftofluoresceína, Cy3 y Cy5, etc. Los pares de etiquetas fluorescentes distinguibles adecuados útiles en

los métodos objeto incluyen Cy-3 y Cy-5 (Amersham Inc., Piscataway, N.J.), Quasar 570 y Quasar 670 (Biosearch Technology, Novato Calif.), preferentemente los siguientes cuatro colorantes de Alexa: Alexa fluor 488, Alexa fluor 555, Alexa fluor 594 y Alexa fluor 647 (Molecular Probes, Eugene, Oreg.); colorantes Atto: Atto488, Atto565, Atto594, Atto647N, Atto750 (Atto-Tec GmbH); BODIPY V-1002 y BODIPY V1005 (Molecular Probes, Eugene, Oreg.), POPO-3 y TOTO-3 (Molecular Probes, Eugene, Oreg.) y POPRO3 TOPRO3 (Molecular Probes, Eugene, Oreg.). Pueden encontrarse más etiquetas detectables distinguibles adecuadas en Kricka y otros (Ann Clin Biochem. 39:114-29, 2002).

La frase "etiquetas distinguibles" o "etiquetas de diferentes colores" o cualquier equivalente gramatical de las mismas se refiere a etiquetas que pueden detectarse y medirse de forma independiente, incluso cuando las etiquetas se mezclan. En otras palabras, las cantidades de etiqueta presente (por ejemplo, la cantidad de fluorescencia) para cada una de las etiquetas pueden determinarse por separado, incluso cuando las etiquetas se colocan (por ejemplo, en el mismo tubo o en la misma molécula dúplex o en la misma célula). Las etiquetas anteriores pueden usarse como etiquetas distinguibles. Algunos pares de etiquetas fluorescentes distinguibles preferidas incluyen Cy-3 y Cy-5 (Amersham Inc., Piscataway, NJ), Quasar 570 y Quasar 670 (Biosearch Technology, Novato Calif.), Alexafluor555 y Alexafluor647 (Molecular Probes, Eugene, Oreg.), BODIPY V-1002 y BODIPY V1005 (Molecular Probes, Eugene, Oreg.), POPO-3 y TOTO-3 (Molecular Probes, Eugene, Oreg.) y POPRO3 y TOPRO3 (Molecular Probes, Eugene, Oreg.) y preferentemente FAM y ATTO550. Pueden encontrarse más etiquetas detectables distinguibles adecuadas en Kricka y otros (Ann Clin Biochem. 39:114-29, 2002).

El término "predeterminado" se refiere a algo que se conoce antes de su uso.

Preferentemente, el oligonucleótido de secuencia de sonda que va a usarse en los métodos de la invención tiene una longitud de entre 20 y 300 nucleótidos. En dependencia de qué tipo de ácido nucleico objetivo va a detectarse por la sonda producida de acuerdo con la invención, se selecciona la longitud de la sonda. Aunque para la aplicación de sondas smFISH la longitud de las sondas se selecciona por debajo de 500 bases, otras aplicaciones pueden requerir el uso de sondas de ácido nucleico más largas. Dado que la longitud de secuencia de sonda no es importante para el proceso de etiquetado, la presente invención no se limitará a ninguna longitud específica. Sin embargo, se prefiere una secuencia de sonda para su uso en smFISH. En el oligonucleótido de secuencia de sonda, la secuencia de nucleótidos etiqueta predeterminada tiene una longitud de al menos 5 ácidos nucleicos, preferentemente 6, 7, 8, 9, 10 o 15 o más, con mayor preferencia de 5 a 50, de 5 a 20, con mayor preferencia de 10 a 20, de 12 a 18, con la máxima preferencia aproximadamente 15 ácidos nucleicos.

El problema se resuelve además mediante un método para generar una biblioteca de sondas de hibridación fluorescente in situ (smFISH) de molécula única, que comprende producir al menos dos sondas de oligonucleótido etiquetadas con fluorescencia de acuerdo con un método de etiquetado de la invención descrito anteriormente, en donde las al menos dos sondas de oligonucleótido con fluorescencia son capaces de unirse a un ácido nucleico objetivo.

Algunas modalidades de la invención pertenecen al método anterior de generación de bibliotecas en donde las al menos dos sondas de oligonucleótido etiquetadas con fluorescencia son capaces de unirse al ácido nucleico objetivo en ubicaciones (secuencias) diferentes, preferentemente no superpuestas. Las al menos dos sondas de oligonucleótido etiquetadas con fluorescencia en una biblioteca de la invención son al menos 10, preferentemente al menos 30, con mayor preferencia de 30 a 150 y con la máxima preferencia aproximadamente 100 sondas de oligonucleótido etiquetadas con fluorescencia y en donde cada una de dichas sondas de oligonucleótido etiquetadas con fluorescencia es capaz de unirse al ácido nucleico objetivo, preferentemente en posiciones no superpuestas.

En algunas modalidades preferidas de este aspecto, al menos dos de las al menos dos sondas de oligonucleótido etiquetadas con fluorescencia en la biblioteca están etiquetadas con múltiples porciones de etiquetado, sin embargo, en donde preferentemente cada sonda de oligonucleótido etiquetada en la biblioteca comprende una etiqueta idéntica o una combinación de etiquetas idénticas. Una biblioteca de sondas de ácido nucleico en el contexto de la invención indicará un conjunto de sondas para sondear y detectar una molécula de ácido nucleico, por ejemplo, un ARNm en un enfoque por smFISH.

En otro aspecto de la invención, se proporciona un método para sondear una secuencia objetivo de moléculas de ácido ribonucleico mensajero (ARNm) en la célula, dicha secuencia objetivo que incluye múltiples regiones de unión a la sonda no superpuesta, que comprende sumergir dicha célula en un exceso de al menos dos sondas oligonucleotídicas, en donde cada sonda oligonucleotídica está marcada de forma múltiple con la misma combinación de al menos dos etiquetas fluorescentes de diferentes colores y cada uno contiene una secuencia de ácido nucleico que es complementaria a una región de unión de sonda diferente de dicha secuencia objetivo; lavar dicha célula fijada para eliminar las sondas no unidas y detectar la fluorescencia de dichas sondas.

Las etiquetas fluorescentes de diferentes colores en el contexto de la invención son porciones fluorescentes que tienen diferentes longitudes de onda de excitación y/o señal y, por lo tanto, permiten una detección individual.

Otro aspecto de la invención se refiere además a un método para sondear al menos dos secuencias objetivo de moléculas de ácido ribonucleico mensajero (ARNm) simultáneamente en una célula, que incluyen cada una de dichas secuencias objetivo múltiples regiones de unión de sonda no superpuesta, que comprende sumergir dicha célula en un exceso de conjuntos de sondas, un conjunto de sondas para cada secuencia objetivo, en donde cada conjunto de sondas comprende al menos dos sondas oligonucleotídicas y cada sonda oligonucleotídica de un conjunto de sondas está etiquetada de forma múltiple con una combinación idéntica de al menos dos etiquetas fluorescentes de diferentes colores, para proporcionar un código de barras de color para cada secuencia objetivo y en donde la combinación de etiquetas fluorescentes de diferentes colores es diferente entre cada conjunto de sondas y en donde cada sonda oligonucleotídica en un conjunto de sondas contiene una secuencia de ácido nucleico que es complementaria a una región de unión de sonda diferente de dicha secuencia objetivo; lavar dicha célula fijada para eliminar las sondas no unidas y detectar la fluorescencia de dichas sondas.

En el contexto de la invención, la célula es preferentemente una célula fijada y permeabilizada.

En este aspecto, la(s) "sonda(s) de oligonucleótido" se preparan preferentemente de acuerdo con un método de etiquetado de la invención como se describió anteriormente en la presente descripción.

El número de regiones de unión de la sonda en los ARNm objetivo y las sondas puede, en algunas modalidades, ser de 40 a 60. En otras modalidades, las al menos 30 sondas tienen secuencias complementarias al objetivo que tienen una longitud de 7-40 nucleótidos, con la máxima preferencia tienen una longitud de 15-30 nucleótidos.

Cada sonda puede añadirse a la célula en una concentración de 0,2-10 nanogramos por microlitro.

Las células fijadas se preparan preferentemente mediante fijación con formaldehído.

La detección incluye preferentemente la imagenología con un microscopio de fluorescencia de campo amplio.

Además, se proporciona un kit de etiquetado para etiquetar una sonda oligonucleotídica, el kit que comprende un primer y un segundo, opcionalmente un tercer o más, oligonucleótidos portadores de la etiqueta, en donde el primer oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una segunda secuencia de nucleótidos adaptadora, el segundo oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una tercera secuencia de nucleótidos adaptadora, opcionalmente, el tercer o más oligonucleótidos portadores de la etiqueta consisten en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una cuarta o más secuencias de nucleótidos adaptadoras y un oligonucleótido adaptador que comprende en secuencia directa una primera, la segunda, la tercera, opcionalmente la cuarta o más secuencias de nucleótidos adaptadoras, en donde la primera secuencia de nucleótidos adaptadora es complementaria a una secuencia de nucleótidos etiqueta predeterminada en un oligonucleótido de secuencia de sonda (o sonda oligonucleotídica).

El kit de etiquetado de la invención puede comprender además una ligasa, opcionalmente junto con tampones o reactivos para su uso. Las ligasas preferidas se describieron anteriormente en la presente descripción.

El kit de etiquetado de la invención puede comprender además instrucciones para su uso.

La presente invención se describirá ahora aún más en los siguientes ejemplos con referencia a las figuras y secuencias adjuntas, sin embargo, sin limitarse a las mismas. En las figuras:

Figura 1: Etiquetado con múltiples fluoróforos basado en la invención. (a) Etiquetado con fluoróforo basado en ligazón de múltiples vías. (b) tinción in situ con sondas preparadas con la invención y prehibridadas con el oligonucleótido adaptador para evitar la extinción de la fluorescencia de múltiples colores. (c) Expresión de Gapdh en Hepa 1-6 con la sonda FISH de Gapdh conjugada con Atto488, Atto565 y Atto647N, sin estabilización de oligonucleótido adaptador. (d) ARNm de Gapdh visualizado como puntos individuales mediante una sonda prehibridada con oligonucleótido adaptador en (c).

Figura 2: Código de barras combinatorio de 3 colores de FISH para la detección in vivo de genes de desarrollo en cerebro embrionario de ratón. (a) Esquema de codificación de colores para FISH a partir de 3 colores base. (b) Detección de 7 genes en la zona ventricular del cerebro embrionario de ratón E12.5 - Barra de escala de 5 µm. (c) Cuantificación de transcritos de ARNm para estos 7 genes de desarrollo.

## Ejemplos

En la solicitud de patente anterior EP 16190862.9, se había usado una ADN T4 ligasa (T4DL) para producir de manera eficiente y rentable sondas etiquetadas fluorescentes para smFISH. Sin embargo, debido a la química de los oligos preetiquetados fluorescentes, existen crecientes dificultades y costos asociados para producir múltiples oligos etiquetados con fluoróforos mediante técnicas de etiquetado y síntesis de oligos de última generación. Con el

objetivo de lograr la rentabilidad y superar el desafío técnico, se ha desarrollado una versión alternativa del método anterior. El fundamento detrás del método de la presente invención es aprovechar la capacidad de ligazón de múltiples vías de T4DL (ADN T4 ligasa) u otras ADNdc ligasas tales como ADN T7 ligasa. El conjunto de oligos de smFISH sin etiquetar se sintetiza con una secuencia de nucleótidos etiqueta predeterminada común y oligos etiquetados fluorescentes simples/múltiples (oligonucleótidos portadores de la etiqueta) y un oligonucleótido adaptador. Después de una reacción de ligazón en un solo recipiente, la sonda smFISH se conjugará a 3 oligos fluorescentes desde el extremo 5 al 3, como se muestra en la figura 1A. El espacio entre las etiquetas individuales es de aproximadamente 15 nucleótidos.

Con la estabilización por el oligo adaptador, pueden obtenerse puntos claros individuales en los 3 canales para las sondas etiquetadas de Gapdh con una copia de Atto488, Atto565 y Atto647N (figura 1c y 1d). Con la capacidad de etiquetado múltiple del método, los inventores podrían asignar diversas combinaciones de colores a un panel de genes y decodificar los puntos al contar su aparición en los canales. La capacidad evolucionada de etiquetado de múltiples fluoróforos con smFISH amplía la smFISH convencional con un mecanismo de código de barras de color combinatorio autónomo. Cada combinación de colores está unida covalentemente con la sonda individual, por lo que la estequiometría del fluoróforo es invariable entre sondas. Durante la adquisición de imágenes, la relación de intensidad entre los fluoróforos será independiente del brillo de los puntos de FISH. La capacidad del código de barras simplemente aumenta con los exponenciales del canal (elección del fluoróforo) número  $n$  (el número teórico de combinaciones es la suma de todas las combinaciones de colores es  $2^n - 1$ ). Si la imagen de puntos de FISH es lo suficientemente precisa y puede determinarse la relación relativa entre la intensidad máxima en cada canal, las combinaciones pueden ser mayores.

Una de las aplicaciones más interesantes para smFISH es la exploración del patrón de expresión génica múltiple en muestras de tejido. Solo con 3 colores base, las combinaciones de colores pueden usarse para detectar 7 genes en una ronda de hibridación. Aquí, los inventores usaron muestras de crio-secciones de embrión de ratón de día embrionario 12.5 (E12.5) para visualizar la heterogeneidad del tejido de 7 genes relacionados con el desarrollo (figura 2a). La detección simultánea de 7 genes muestra el patrón de organización celular de la expresión génica del desarrollo (figura 2b). El agrupamiento jerárquico muestra 4 grupos con varios grados de expresión de marcadores de células madre (figura 2c).

## Métodos

### Preparación de cultivos celulares y cortes de tejidos

Se aislaron células madre neurales primarias (NSC) de la zona subventricular de cerebro de ratón C57BL/6J macho adulto de 8 semanas. Las piezas de tejido aisladas se cortaron aún más en un tamaño de 1 mm<sup>3</sup> y se digirieron con solución de Accutase (Sigma, A6964) a 37 °C durante 20 minutos. Las células se centrifugaron durante 5 minutos a 250 g, luego se resuspendieron y se cultivaron en medio sin suero (DMEM/F-12 HEPES (ThermoFisher, 31330095) suplementado con glucosa al 0,6 %, HEPES 20 mM (reserva 1 M, Sigma, H0887), progesterona 0,2 mM, putrescina 0,06 mM, B27 al 2 %, 20 ng/ml de EGF, 10 ng/ml de FGF, ITSS al 0,1 %, penicilina/estreptomina 1×, heparina 0,36 U/ml (Sigma, H3149), NaHCO<sub>3</sub> al 1,2 % (p/v)). Para el cultivo de NSC en placa recubierta o cubreobjetos, la placa o el cubreobjetos se recubrieron con solución salina tamponada con fosfato (PBS) suplementada con 1 mg/ml de poli-1-lisina (Sigma, P7280), 0,0025 mg/ml de laminina (Roche, 11243217001) a 37 °C durante la noche y luego se secó bajo la campana de cultivo celular con radiación UV. Se cultivaron células Hepa 1-6 de ratón en medio DMEM con suero bovino fetal al 10% y penicilina/estreptomina 1 ×. Las células Hepa 1-6 se cultivaron directamente en cubreobjetos sin recubrimiento. Se cortaron crio-secciones de tejido embrionario de cerebro de ratón de 6 a 10 μm de embrión de ratón C57BL/6J de día embrionario 12.5 incrustado en Tissue-Tek O.C.T. (Sakura, 4583).

Las células NSC recubiertas, las células Hepa 1-6 adherentes o las crio-secciones se fijaron con formaldehído al 4 % en PBS durante 10 minutos y luego se extinguieron con glicina 135 mM en PBS durante 10 minutos a temperatura ambiente. A continuación, las células fijadas se lavaron una vez con PBS y se permeabilizaron en etanol al 70 % durante la noche a 4 °C. Toda el agua usada para los tampones relacionados con FISH se trató con pirocarbonato de dietilo (DEPC). Después de la permeabilización, las células se almacenaron en crioprotector (glicerol al 25 %, etilenglicol al 25 %, tampón de fosfato 0,1 M, pH 7,4) a -20 °C hasta la tinción de FISH.

### Diseño de la sonda

Las sondas de smFISH basadas en el diseño convencional<sup>3</sup> se implementaron en un script R para seleccionar las sondas primero con cebador<sup>3</sup> para obtener todas las sondas posibles sin una estructura secundaria demasiado fuerte de la secuencia de ADNc de entrada mediante el uso de la condición estándar para seleccionar el cebador correcto en cebador<sup>3</sup>. Luego se seleccionaron sondas de smFISH no superpuestas con un espacio de 2 pb. Para las sondas de HuluFISH, se calcularon adicionalmente todas las sondas del cebador<sup>3</sup> para la eficiencia de hibridación con el paquete DECIPHER en R en las condiciones usadas para la tinción. Y las sondas se filtraron para tener una eficiencia de hibridación superior a 0,9 (máximo 1) y luego se seleccionaron sondas de HuluFISH no superpuestas como antes. Las secuencias adaptadoras se generaron aleatoriamente y se controlaron para una estructura

secundaria fuerte por UNAFold. Las secuencias aprobadas se contrastaron contra una base de datos local de transcritos humanos y de ratón (versión del conjunto 87) para una coincidencia exacta menor o igual a 15 pb.

Etiquetado y purificación de la sonda de HuluFISH

5

El método de etiquetado de la invención se denomina "HuluFISH".

Las agrupaciones de oligos de sonda de FISH y los oligos adaptadores se sintetizaron a partir de Sigma con la calidad más baja para la purificación (desalinización). Para el gen individual, los oligos se agruparon para tener una concentración de oligos total de 100  $\mu$ M. Los oligos fluorescentes se compraron de Eurofins genomics con varios colorantes, incluidos los colorantes Atto, colorantes Alexa o colorantes Cy. Para el etiquetado de la sonda sin secuencia etiqueta adicional, la ligazón se realizó en tampón de ADN T4 ligasa (NEB, B0202S), con oligo adaptador 30  $\mu$ M con secuencia degenerativa, grupo de oligo de sonda de FISH sin etiquetar 3  $\mu$ M y oligo etiquetador fluorescente, PEG8000 al 25 %, 30 U/ $\mu$ l de ADN T4 ligasa (NEB, M0202M). A continuación, la reacción de ligazón se incubó en un termociclador, con 12 ciclos 10 segundos a 37 °C/5 minutos a 16 °C. Para el etiquetado de la sonda con una secuencia etiqueta, la ligazón se realizó como antes con algunas modificaciones, tales como 16,7  $\mu$ M de todos los componentes de oligos y el conjunto de oligos, 50 U/ $\mu$ l de ADN T4 ligasa. Luego, la mezcla de ligazón se dejó en la oscuridad a temperatura ambiente durante 2 horas. El producto de ligazón se concentró con 9 volúmenes de butanol y se centrifugó como sedimento a 20 000 g, 15 minutos a 4 °C. El sedimento de oligo etiquetado con colores se lavó una vez con etanol al 100 % y se centrifugó para eliminar el etanol, luego se volvió a solubilizar en tampón de carga (urea 8M, TBE 1  $\times$  (Carl Roth, A118.1), azul de bromofenol al 0,01 % y cianol de xileno). Con 5 minutos de desnaturalización a 90 °C, los oligos se cargaron en gel de urea-PAGE al 15 % (urea 8M, TBE 1  $\times$ , Rotiphorese Gel 30 al 15 % (Carl Roth, 3029.2), persulfato de amonio al 0,05 %, tetrametiletilendiamina al 0,05 %) recorrida a 300 V durante 30 minutos. La condición de corrida era generalmente 300 V, 30 minutos o hasta que el azul de bromofenol llegara al final. Las bandas de gel con conjugados de colorante fluorescente-oligo se escindieron bajo la luz ambiental. Las piezas de gel se homogeneizaron manualmente con un mortero de microtubos (Sigma, Z359947-100EA) y luego se extrajeron con 500  $\mu$ l de tampón TE 10 mM (pH 8,5, tris(hidroximetil)aminometano (Tris) 10 mM, ácido etilendiaminotetraacético (EDTA) 1 mM) a temperatura ambiente durante la noche, protegido de la luz mediante envoltura en papel de aluminio. Los oligos extraídos en TE se concentraron de nuevo con butanol y se lavaron una vez con etanol como antes. Los sedimentos finales se secaron en la oscuridad a temperatura ambiente durante 5-10 minutos y luego se volvieron a solubilizar en H<sub>2</sub>O. La concentración se determinó mediante nanodrop one (ThermoFisher) como ADNsc.

35

Imágenes y tinción de la sonda de FISH

La mezcla de sondas de HuluFISH se ajustó a 10 nM para cada oligo único en tampón de hibridación (SSC (citrato de sodio y solución salina) 2  $\times$ , sulfato de dextrano al 10 %, formamida al 10 %, 1 mg/ml de ARNt (Roche, 10109541001), complejo de ribonucleósido vanadilo 2 mM (NEB, S1402S), 0,2 mg/ml de BSA). La sonda Gapdh-Quasar570 se adquirió de Biosearch Technology, se resuspendió y se realizó la tinción según las instrucciones del fabricante. La hibridación se realizó en un baño de agua a 30 °C durante la noche, con la muestra boca abajo sobre el parafilm. Las células en cubreobjetos o secciones de tejido en el portaobjetos de vidrio se lavaron con tampón de lavado (SSC 2  $\times$ , formamida al 10 %, Tween-20 al 0,1 %) a 37 °C durante 6  $\times$  10 minutos. La última etapa de lavado incluyó 0,5  $\mu$ g/ml de DAPI (4',6-diamidino-2-fenilindol) para la tinción de núcleos. La muestra se montó en ProLong Gold Antifade (ThermoFisher, P10144) y se curó durante la noche. A continuación, se tomaron imágenes de la muestra en un microscopio de campo amplio (Zeiss Cell Observer) con 200 ms, 950 ms y 5000 ms para el canal de 405 nm, 488 nm y 561 nm, o en un microscopio confocal con Airyscan (Zeiss LSM800, equipado con láser de 405, 488, 561 y 647 nm) con la potencia máxima del láser en cada canal. La muestra se escaneó con tecnología Airyscan con la configuración óptima proporcionada por el programa informático Zeiss.

50

Análisis de imágenes

Excepto por el contorno nuclear definido manualmente en ImageJ, todo el análisis de imágenes se realizó en R y se basó principalmente en el paquete EImage. Todos los valores de umbral de intensidad se basaron en las unidades arbitrarias generadas a partir de Zeiss Airyscan y, por lo tanto, no se especifican en la siguiente descripción. La identificación de puntos de FISH se basó en la identificación y alineación de máximos locales 2D. Inicialmente, para cada cuadro, se identificaron máximos 2D por encima de un valor de umbral bajo. Cada máximo local 2D consideró su proyección en los cortes z cercanos para la alineación: aquellos que caen dentro de 0,08  $\mu$ m se asignaron al mismo punto de FISH. Los píxeles con intensidades máximas (pseudo máximos 3D) para los puntos de FISH identificados se extrajeron para su posterior análisis.

60

La relación señal-ruido (SNR) y el contraste se generaron de forma adaptativa para cada punto de FISH individual. Con este fin, los valores de píxeles (fondo local) se tomaron de un cuadrado centrado alrededor de los pseudo máximos 3D, que excluyen todas las regiones circulares que cubren la PSF (función de dispersión de puntos) para máximos 2D en el mismo plano. El contraste se define como la relación entre la intensidad máxima y la media de sus valores de fondo locales; SNR, como se define tradicionalmente, es igual a la intensidad máxima dividida por la desviación estándar de los valores de fondo locales.

65

Para la decodificación de colores en muestras con sonda Hulu para múltiples genes, la presencia del fluoróforo en cada canal se determinó inicialmente por separado. Se asignó una codificación de color doble o triple cuando los puntos de FISH de diferentes canales se colocalizaban dentro de 0,08  $\mu\text{m}$ . La asignación de un solo color requería un umbral con una intensidad más alta, dado que había tres copias de fluoróforos en la sonda Hulu de un color único. Los núcleos se segmentaron manualmente en la imagen proyectada de máxima intensidad en ImageJ. Sin la ayuda de la inmunotinción de membrana, cada punto de FISH identificado se asignó a sus núcleos más cercanos.

#### Referencias

1. Gall, J. G. y Pardue, M. L. Formation and Detection of Rna-Dna Hybrid Molecules in Cytological Preparations. *Proc. Natl. Acad. Sci.* 63, 378-383 (1969).
2. Femino, A. M., Fay, F. S., Fogarty, K. y Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* 280, 585-590 (1998).
3. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. y Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Meth* 5, 877-879 (2008).
4. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. y Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11, 360-361 (2014).
5. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. y Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* aaa6090 (2015). doi:10.1126/science.aaa6090
6. Lubeck, E. y Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* 9, 743-748 (2012).
7. Levesque, M. J. y Raj, A. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat. Methods* 10, 246-248 (2013).
8. Raj, A. y Tyagi, S. in *Methods in Enzymology* 472, 365-386 (Elsevier, 2010).
9. Lyubimova, A. y otros Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* 8, 1743-1758 (2013).
10. Tsanov, N. y otros smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* gkw784 (2016). doi:10.1093/narjgkw784
11. Untergasser, A. y otros Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40, e115-e115 (2012).
12. Wright, E. S., Yilmaz, L. S., Corcoran, A. M., Okten, H. E. y Noguera, D. R. Automated Design of Probes for rRNA-Targeted Fluorescence In Situ Hybridization Reveals the Advantages of Using Dual Probes for Accurate Identification. *Appl. Environ. Microbiol.* 80, 5124-5133 (2014).
13. Roy, R., Hohng, S. y Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* 5, 507-516 (2008).
14. Zhuang, X. y otros Fluorescence quenching: A tool for single-molecule protein-folding study. *Proc. Natl. Acad. Sci.* 97, 14241-14244 (2000).
15. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31,3406-3415 (2003).
16. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. y Huber, W. EBImage-an-R package for image processing with applications to cellular phenotypes. *Bioinformatics* 26, 979-981 (2010).

## REIVINDICACIONES

1. Un método para producir una sonda oligonucleotídica etiquetada o modificada de otra manera, el método comprende las etapas de:

5 a. Proporcionar un oligonucleótido de secuencia de sonda que comprende (i) una secuencia de sonda que comprende una secuencia de nucleótidos complementaria a un ácido nucleico objetivo y (ii) una secuencia de nucleótidos etiqueta predeterminada complementaria a una primera secuencia de nucleótidos adaptadora,

10 b. Proporcionar un primer y un segundo, opcionalmente un tercer o más oligonucleótidos portadores de la etiqueta que comprenden al menos una porción de etiqueta u otra porción funcional, en donde el primer oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una segunda secuencia de nucleótidos adaptadora, el segundo oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una tercera secuencia de nucleótidos adaptadora, opcionalmente, el tercer o más oligonucleótidos portadores de la etiqueta consisten en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una cuarta o más secuencias de nucleótidos adaptadoras,

15 c. Proporcionar un oligonucleótido adaptador, que comprende en secuencia directa la primera, la segunda, la tercera y, opcionalmente, la cuarta o más secuencias de nucleótidos adaptadoras;

20 d. Poner en contacto en condiciones de hibridación el oligonucleótido de secuencia de sonda, el primero y el segundo, opcionalmente el tercero o más oligonucleótidos portadores de la etiqueta y el oligonucleótido adaptador, para formar un complejo, en donde, en el complejo, un grupo 3' OH libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta o un grupo fosfato libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente de un grupo 3' OH libre del primer oligonucleótido portador de la etiqueta, un grupo 3' OH libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta o un grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo 3' OH libre del segundo oligonucleótido portador de la etiqueta y, opcionalmente, un grupo 3' OH libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo fosfato libre (desbloqueado) del tercer o más oligonucleótidos portadores de la etiqueta o un grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente de un grupo 3' OH libre del tercer o más oligonucleótidos portadores de la etiqueta,

25 e. Hacer reaccionar el complejo para formar un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del oligonucleótido de secuencia de sonda y el primer oligonucleótido portador de la etiqueta, un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del primer oligonucleótido portador de la etiqueta y el segundo oligonucleótido portador de la etiqueta y opcionalmente un enlace covalente entre el grupo 3'-OH y el grupo 5'-fosfato del segundo oligonucleótido portador de la etiqueta y el tercer o más oligonucleótidos portadores de la etiqueta, mediante el uso de una ligasa en condiciones de ligazón, para formar la sonda oligonucleotídica etiquetada o modificada de otra manera,

30 f. Opcionalmente, retirar el oligonucleótido adaptador.

- 45 2. El método para producir una sonda oligonucleotídica etiquetada o modificada de otra manera de acuerdo con la reivindicación 1, en donde

50 en el oligonucleótido de secuencia de sonda, la secuencia de sonda está en 5' de la secuencia de nucleótidos etiqueta predeterminada, y en el oligonucleótido adaptador, la primera y la segunda secuencia de nucleótidos adaptadora están en dirección 3' a 5' en secuencia directa, y en el complejo en la etapa (d) el grupo 3' OH libre del oligonucleótido de secuencia de sonda está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta, el grupo 3' OH libre del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta y, opcionalmente, el grupo 3' OH libre del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo fosfato libre (desbloqueado) del tercer o más oligonucleótidos portadores de la etiqueta.

- 60 3. El método para producir una sonda oligonucleotídica etiquetada o modificada de otra manera de acuerdo con la reivindicación 1, en donde

65 en el oligonucleótido de secuencia de sonda la secuencia de sonda está en 3' de la secuencia de nucleótidos etiqueta predeterminada y en el oligonucleótido adaptador, la primera y la segunda secuencia de nucleótidos adaptadora están en dirección 5' a 3' en secuencia directa y

- 5 en el complejo en la etapa (d) el grupo fosfato libre (desbloqueado) del oligonucleótido de secuencia de sonda está muy cerca espacialmente del grupo 3' OH libre del primer oligonucleótido de secuencia de sonda, el grupo fosfato libre (desbloqueado) del primer oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo 3' OH libre del segundo oligonucleótido portador de la etiqueta y opcionalmente, el grupo fosfato libre (desbloqueado) del segundo oligonucleótido portador de la etiqueta está muy cerca espacialmente del grupo 3' OH libre del tercer o más oligonucleótidos portadores de la etiqueta.
- 10 4. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde cada una de las secuencias de nucleótidos adaptadoras comprende secuencias de nucleótidos distintas.
- 15 5. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde cada uno de los oligonucleótidos portadores de la etiqueta se etiqueta de manera diferente o se modifica de otra manera.
- 20 6. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde el oligonucleótido portador de la etiqueta comprende dos o más, preferentemente tres, cuatro o cinco, porciones de etiquetado.
7. El método de acuerdo con la reivindicación 6, en donde las dos o más, preferentemente tres, cuatro o cinco, porciones de etiquetado, comprenden al menos dos o más porciones de etiquetado diferentes u otras porciones funcionales.
- 25 8. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, que comprende purificar el producto ligado del oligonucleótido de secuencia de sonda y el(los) oligonucleótido(s) portador(es) de la etiqueta.
- 30 9. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde el oligonucleótido de secuencia de sonda tiene una longitud de entre 20 a 300 nucleótidos.
- 35 10. Un método para generar una biblioteca de sondas de hibridación fluorescente in situ (smFISH) de molécula única, que comprende producir al menos dos sondas de oligonucleótido etiquetadas con fluorescencia de acuerdo con un método de una cualquiera de las reivindicaciones 1 a 9, en donde las al menos dos sondas de oligonucleótido con fluorescencia son capaces de unirse a un ácido nucleico objetivo.
- 40 11. El método de acuerdo con la reivindicación 10, en donde las al menos dos sondas de oligonucleótido etiquetadas con fluorescencia son al menos 10, preferentemente al menos 30, con mayor preferencia 30 a 150 y con la máxima preferencia aproximadamente 100 sondas oligonucleotídicas etiquetadas con fluorescencia y en donde cada una de dichas sondas de oligonucleótido etiquetadas con fluorescencia es capaz de unirse al ácido nucleico objetivo.
- 45 12. Un kit que comprende un primer y un segundo, opcionalmente un tercer o más, oligonucleótidos portadores de la etiqueta, en donde el primer oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una segunda secuencia de nucleótidos adaptadora, el segundo oligonucleótido portador de la etiqueta consiste en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una tercera secuencia de nucleótidos adaptadora, opcionalmente, el tercer o más oligonucleótidos portadores de la etiqueta consisten en una secuencia de nucleótidos etiqueta predeterminada, en donde la secuencia de nucleótidos etiqueta predeterminada es al menos 90 % complementaria a una cuarta o más secuencias de nucleótidos adaptadoras y un oligonucleótido adaptador que comprende en secuencia directa una primera, la segunda, la tercera, opcionalmente la cuarta o más secuencias de nucleótidos adaptadoras, en donde la primera secuencia de nucleótidos adaptadora es complementaria a una secuencia de nucleótidos etiqueta predeterminada en un oligonucleótido de secuencia de sonda (o sonda oligonucleotídica).
- 50

Figura 1:

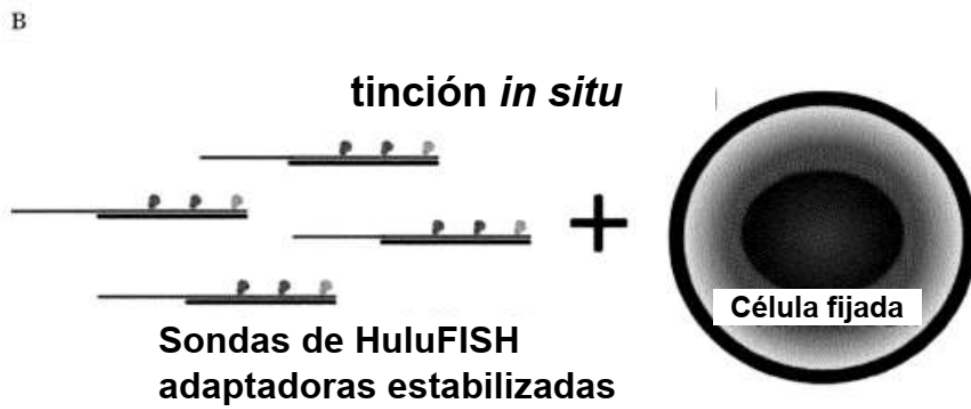
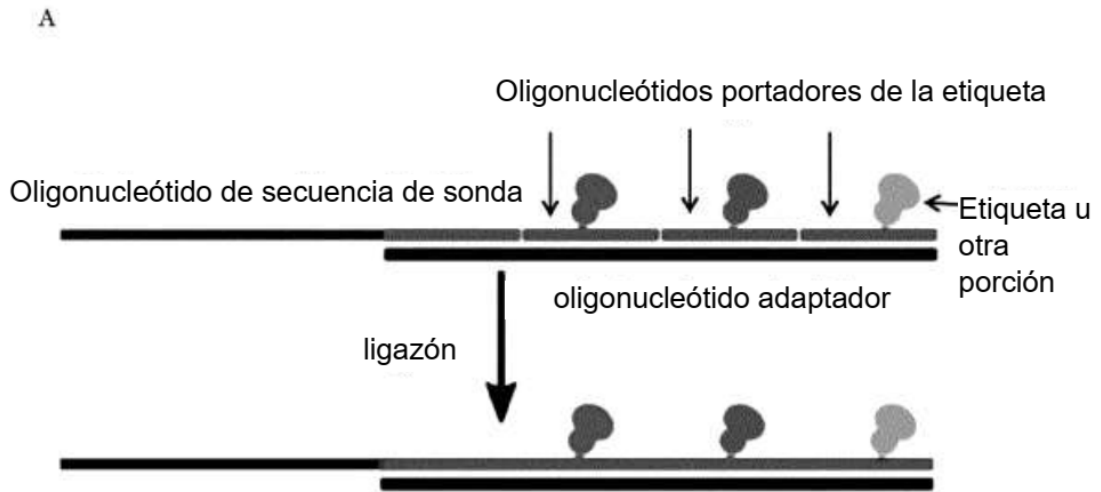


Figura 1 cont.:

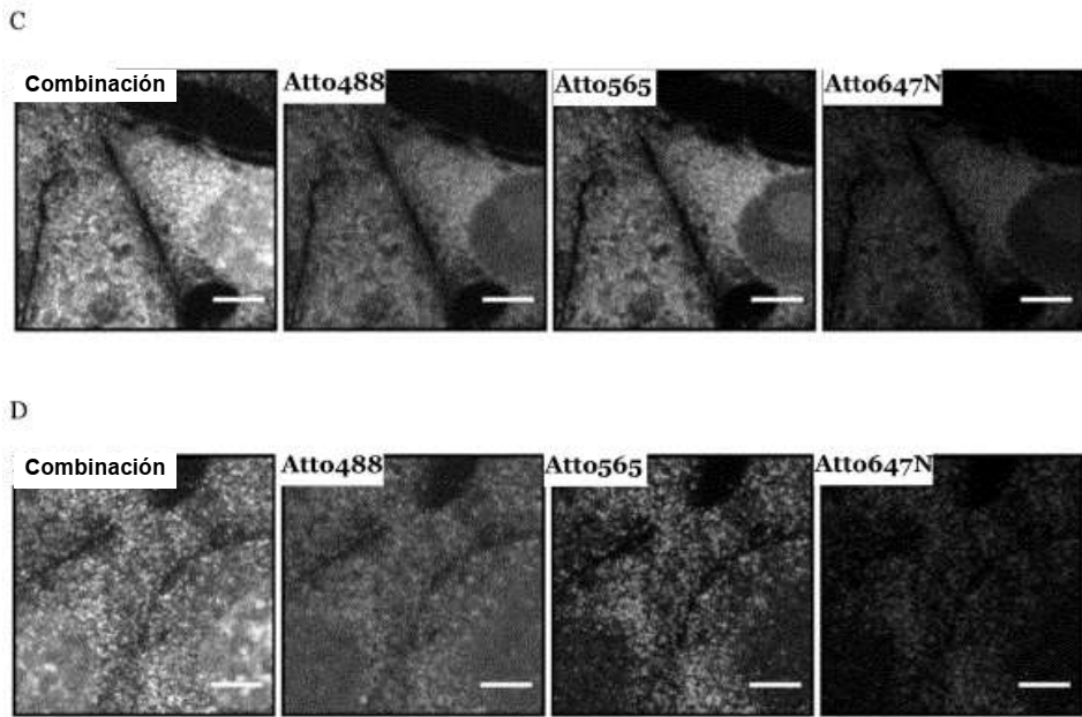


Figura 2:

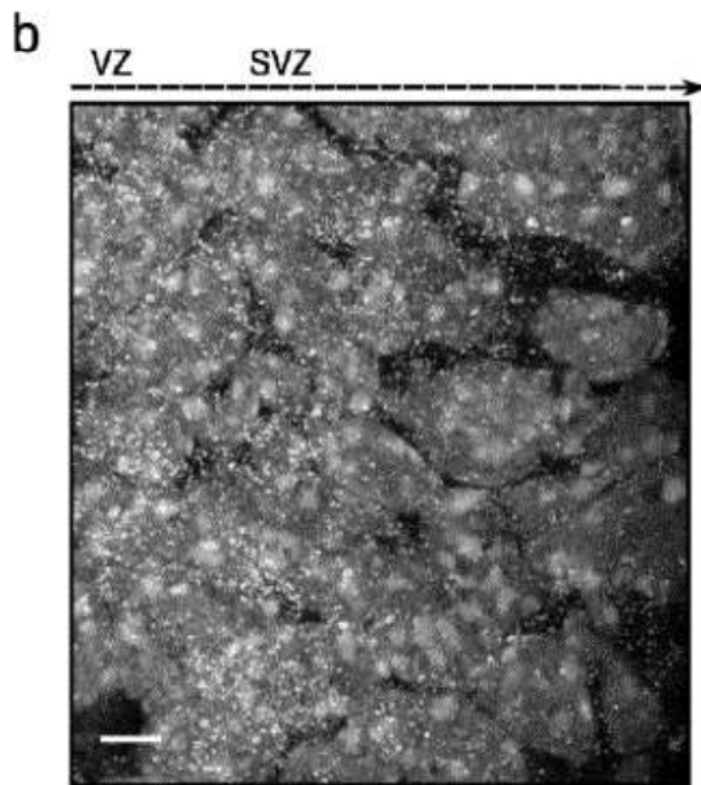
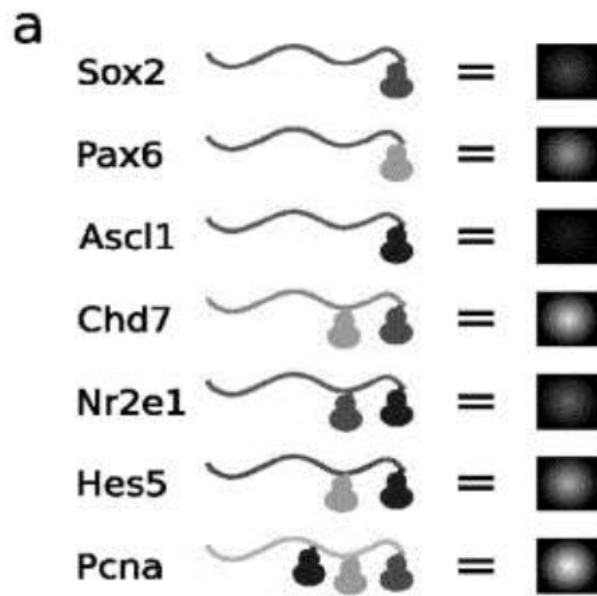


Figura 2 cont.:

