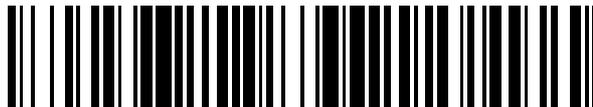


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 369 946**

51 Int. Cl.:  
**C12N 15/10** (2006.01)  
**C07K 16/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **08706942 .3**  
96 Fecha de presentación: **27.02.2008**  
97 Número de publicación de la solicitud: **2121920**  
97 Fecha de publicación de la solicitud: **25.11.2009**

54 Título: **PROCEDIMIENTO PARA CLONACIÓN DE ANTICUERPOS ANÁLOGOS.**

30 Prioridad:  
**01.03.2007 DK 200700316**  
**05.03.2007 US 904772 P**

45 Fecha de publicación de la mención BOPI:  
**09.12.2011**

45 Fecha de la publicación del folleto de la patente:  
**09.12.2011**

73 Titular/es:  
**SYMPHOGEN A/S**  
**ELEKTROVEJ BUILDING 375**  
**2800 KGS. LYNGBY, DK**

72 Inventor/es:  
**MEIJER, Per-Johan;**  
**NIELSEN, Lars, Soegaard y**  
**KASTRUP, Jesper**

74 Agente: **Carpintero López, Mario**

**ES 2 369 946 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

Procedimiento para clonación de anticuerpos análogos

**Campo técnico de la invención**

5 La presente invención se refiere a un procedimiento para enlazar pares análogos de secuencias que codifican para  $V_H$  y  $V_L$  a partir de una población de células enriquecidas en marcadores de antígenos de superficie particulares. El procedimiento de enlace implica un procedimiento de amplificación molecular múltiplex capaz de enlazar secuencias nucleótidas de interés en relación con la amplificación, en particular reacción en cadena de polimerasa (PCR múltiplex). El procedimiento es particularmente ventajoso para la generación de bibliotecas de pares análogos así como bibliotecas combinatorias de secuencias codificantes de región variable para inmunoglobulina. La invención  
10 también se refiere a procedimientos para generación de anticuerpos quiméricos humanos/no humanos y bibliotecas de expresión generadas por tales procedimientos.

**Antecedentes de la invención**

15 El documento WO 2005/042774 (Symphogen) desvela un procedimiento para enlazar secuencias nucleótidas de interés, en particular pares análogos de secuencias que codifican para  $V_H$  y  $H_L$  utilizando un procedimiento molecular múltiplex. Las secuencias de interés preferiblemente se amplifican y enlazan a partir de células solas aisladas después de dilución limitada u otras técnicas de separación de células. La referencia desvela varias maneras de enriquecer una población de células que contienen linfocitos para obtener una población de células plasmáticas que son particularmente adecuadas para el procedimiento de amplificación molecular múltiplex.

20 El documento WO 2007/003041 desvela un procedimiento para aislar una célula de secreción de anticuerpos que tiene una alta probabilidad de secretar un anticuerpo específico para un antígeno deseado, que comprende las etapas: a) proporcionar una muestra de un animal que se ha inmunizado recientemente con el antígeno deseado, en el que la muestra comprende una población de células que secretan anticuerpos que secretan anticuerpos específicos para el antígeno deseado; b) aumentar la concentración de células que secretan anticuerpos en la muestra; y c) aislar una célula que secreta anticuerpos. La concentración de células que secretan anticuerpos en la  
25 muestra se puede aumentar por medio del enriquecimiento de las células o por medio de la depleción de las células en la muestra con marcadores de superficie.

**Sumario de la invención**

30 La presente invención está enfocada en procedimientos para generar bibliotecas de secuencias que codifican para inmunoglobulina a partir de animales no humanos y procedimientos para generar, en algunas etapas adaptadas para clonación y cribado de alto rendimiento, bibliotecas de vectores que codifican para anticuerpos quiméricos que comprenden regiones constantes humanas y regiones variables no humanas.

En un primer aspecto, la invención se refiere a un procedimiento para producir una biblioteca de pares análogos que comprenden secuencias que codifican para la región variable enlazada, comprendiendo dicho procedimiento:

- a) proporcionar una fracción de células que comprenden linfocitos de un donante roedor;
- 35 b) obtener una población de células solas aisladas, distribuyendo células de dicha fracción celular individualmente en una pluralidad de recipientes, en los que por lo menos una subpoblación de las células expresadas están enriquecidas y expresan antígenos CD43 y CD138 o antígenos MHCII y B220; y
- 40 c) amplificar y llevar a cabo el enlace de las secuencias codificantes de la región variable contenidas en dicha población de células solas aisladas por amplificación, en un procedimiento de amplificación molecular múltiplex, de secuencias nucleótidas de interés utilizando una plantilla derivada de una sola célula aislada o una población de células isogénicas y llevar a cabo el enlace de las secuencias nucleótidas de interés amplificadas.

Este procedimiento proporciona una biblioteca de anticuerpos de pares análogos o fragmentos de anticuerpos.

45 En otro aspecto, la invención se refiere a un procedimiento para enlazar de forma aleatoria una pluralidad de secuencias nucleótidas no contiguas de interés, en el que dichas secuencias nucleótidas de interés comprenden secuencias codificantes para la región variable y el enlace genera una biblioteca combinatoria de pares de secuencias codificantes para la región variable, comprendiendo dicho procedimiento:

- a) Amplificar, en un procedimiento de amplificación molecular múltiplex, secuencias nucleótidas de interés usando una plantilla derivada de una población de células genéticamente diversas, en el que las células genéticamente diversas se derivan de una fracción celular que comprende linfocitos de un donante roedor, y en el que al menos una subpoblación de las células está enriquecida y expresa los antígenos CD43 y CD138 o los antígenos MHCII y B220; y
- 50 b) Llevar a cabo el enlace de las secuencias nucleótidas de interés amplificadas en la etapa a).

Este procedimiento proporciona preferiblemente una biblioteca combinatoria de dominios variables de inmunoglobulina de cadena pesada y ligera combinados aleatoriamente.

55 Las pruebas experimentales que se proporcionan en la presente solicitud establecen que las poblaciones de células aisladas de esplenocitos de ratón y que son positivas para los antígenos de superficie enumerados proporcionan buen material de partida para clonación de secuencias que codifican anticuerpos utilizando un procedimiento de amplificación molecular múltiplex. Los procedimientos de la invención se pueden aplicar fácilmente a otras especies

que expresan los ortólogos de CD43 y CD138 o MHCII y B220, en particular, se pueden aplicar los procedimientos a otros roedores, tal como ratas.

El procedimiento proporciona varias ventajas sobre la generación alternativa, por ejemplo de hibridoma. Cuando se prepara un hibridoma de un ratón inmunizado, las líneas de células establecidas codifican un repertorio de isotipos de anticuerpos diferentes. Las líneas de células de hibridoma posteriormente necesitan ser cribadas para función (por ejemplo, unión a un antígeno específico, eficacia en neutralización del patógeno) y para el isotipo de anticuerpo. Por lo tanto se requiere un procedimiento de cribado de dos etapas para seleccionar un hibridoma con un isotipo de anticuerpo particular y con un efecto particular en un análisis funcional. Finalmente, con el fin de generar una línea de células productoras, el anticuerpo secretado por el hibridoma seleccionado necesita ser clonado y secuenciado antes de que se pueda transferir a una línea de células productoras.

Mediante el uso del procedimiento de la presente invención, se determinan isotipos de anticuerpo por los cebadores usados para la amplificación molecular múltiple y por lo tanto no hay necesidad de que se determinen. El isotipo de anticuerpo también se puede determinar (y se puede cambiar) por el enlace o empalmes subsiguientes de dominios constantes a secuencias variables clonadas. Además, el procedimiento de la invención proporciona una biblioteca de polinucleótidos que se pueden secuenciar y/o insertar fácilmente en vectores, tales como los vectores de expresión, de transferencia, de presentación o lanzadera, de manera que una vez que se ha seleccionado un anticuerpo particular, se clone de antemano, y su secuencia se conozca de antemano y pueda ser transferido con facilidad a un vector de expresión apropiado para producción de anticuerpo.

Se espera que las células clasificadas de acuerdo con el protocolo proporcionado suministren una fuente de anticuerpos de alta afinidad, potencialmente con afinidades en el intervalo picomolar. Los anticuerpos monoclonales de hibridoma no poseen afinidades en el intervalo picomolar y necesitarán ser madurados por afinidad sintéticamente para alcanzar estas afinidades.

De acuerdo con una realización, estos procedimientos comprenden además determinar antes de la amplificación molecular múltiple que la población de células que comprenden linfocitos comprende células que expresen los antígenos CD43 y CD138 o los antígenos MHCII y B220, preferiblemente CD43 y CD138. También los procedimientos pueden incluir enriquecer la fracción de células que comprenden linfocitos para una población de linfocitos que expresan los antígenos CD43 y CD138 o MHCII y B220 antes de la amplificación molecular múltiple.

Preferiblemente, los procedimientos comprenden además aislar de la población que comprende linfocitos células que expresan los antígenos CD43 y CD138 o los antígenos MHCII y B220 antes de la amplificación molecular múltiple. En una realización preferida, las células aisladas o las subpoblación de células son altas en CD138/altas en CD43 o intermedias en CD138/altas en CD43 en relación a la fracción de células que comprenden linfocitos. De manera más preferible, las células aisladas o la subpoblación de células son altas en CD138/altas en CD43 en relación a la fracción de células que comprenden linfocitos.

El enriquecimiento o aislamiento preferiblemente comprende un procedimiento de clasificación automatizado tal como MACS o FACS.

También se desvela en la presente memoria descriptiva un procedimiento para generar un vector que codifica para un anticuerpo quimérico con regiones constantes humanas y regiones variables no humanas, comprendiendo dicho procedimiento:

a) proporcionar una fracción de células que comprenden linfocitos de un animal no humano;

b) obtener una población de células solas aisladas, que comprende distribuir células de dicha fracción celular individualmente en una pluralidad de recipientes;

c) amplificar y llevar a cabo el enlace de los ácidos nucleicos codificantes para la región variable contenidos en dicha población de células solas aisladas por amplificación, en un procedimiento de amplificación molecular múltiple, usando dichos ácidos nucleicos una plantilla derivada de una sola célula aislada o una población de células isogénicas; y llevando a cabo el enlace de los ácidos nucleicos amplificados que codifican para las regiones variables de las cadenas pesada y ligera;

d) llevar a cabo el enlace de las regiones variables amplificadas a las regiones constantes humanas; e

e) insertar el ácido nucleico obtenido en un vector.

Preferiblemente, el animal no humano es un ratón. En la medida en que los procedimientos de la invención se aplican a células de ratón, los procedimientos se denominan: Mouse-Symplex<sup>MR</sup> o mSymplex<sup>MR</sup>.

Este procedimiento proporciona un procedimiento novedoso para generación de bibliotecas de anticuerpos quiméricos humanos/no humanos. Esto se hace posible al combinar la amplificación molecular múltiple y la clonación subsiguiente en una estructura principal del vector con ligación y/o empalme de dominios constantes de cadena pesada y ligera humanos. Tradicionalmente, en un procedimiento para generar anticuerpos quiméricos humanos/no humanos, la quimerización ha sido la última etapa después de que se ha establecido y cribado el hibridoma y el anticuerpo codificado se ha clonado. La quimerización puede alterar la especificidad y/o afinidad de unión de un anticuerpo y por lo tanto existe el riesgo de que un buen anticuerpo monoclonal de ratón pierda su eficacia cuando se quimeriza en un anticuerpo humano/de ratón.

Mediante el suministro de un procedimiento que genera directamente un anticuerpo de anticuerpos quiméricos, se puede llevar a cabo el cribado sobre productos que no necesitan ser modificados adicionalmente antes del desarrollo preclínico y clínico.

Se pueden proporcionar regiones humanas constantes en una etapa de amplificación molecular o se pueden proporcionar como parte de un vector-estructural principal en la cual las regiones variables se clonan después de amplificación molecular. En una realización preferida, el procedimiento comprende una etapa de amplificación adicional en el que un polinucleótido que codifica para una cadena ligera constante humana o un fragmento de la misma con una superposición capaz de proporcionar enlace a la cadena ligera variable se agrega a la mezcla de PCR junto con un conjunto cebador capaz de amplificación de un constructo que comprende en orden: una cadena VH murina, un engarce, una cadena VL murina y una cadena ligera constante humana.

En otra realización, el procedimiento comprende una etapa de amplificación adicional, en el que el polinucleótido que codifica para la cadena pesada constante humana o un fragmento del mismo con una superposición capaz de proporcionar enlace a la cadena pesada variable se agrega a la mezcla de PCR junto con un conjunto de cebadores capaces de amplificación de un constructo que comprende en orden: una cadena pesada constante humana, una cadena VH murina, un engarce y una cadena VL murina.

En consecuencia, también se proporciona una biblioteca de vectores que codifican para anticuerpos quiméricos, cada miembro de anticuerpo está constituido por secuencias que codifican para la región variable de inmunoglobulina no humana y regiones constantes de cadena pesada y ligera de inmunoglobulina humana.

Preferiblemente, los vectores son vectores de expresión que habilitan la expresión de los miembros de anticuerpos de la biblioteca para cribado subsiguiente. De manera más preferible, el vector de expresión es para expresión de mamífero.

Los vectores de la biblioteca se pueden obtener por un procedimiento de la invención.

En una realización, la región constante de cadena ligera es una región constante kappa.

Las secuencias no humanas pueden ser de rata, oveja, cabra, conejo, cobaya u otro animal adecuado para el cual se han descrito protocolos de inmunización, para los cuales está disponible la información de secuencia para permitir el diseño de cebadores adecuados y para los cuales las técnicas de clasificación de células adecuadas permiten la clasificación de células plasmáticas para amplificación molecular múltiplex de una sola célula para enlazar pares análogos de secuencias de región variable. En una realización, las secuencias no humanas son de origen de primate no humano, tal como mono cynomolgous, mono Rhesus, chimpancé o macaco. Preferiblemente, las secuencias no humanas son de roedor, tal como de ratón o rata. En otra realización, las secuencias no humanas son secuencias de conejo.

Preferiblemente, las regiones variables de los anticuerpos son pares análogos.

También se desvela en la presente memoria descriptiva una biblioteca secundaria que codifica para anticuerpos que presentan las especificidades de unión deseadas dirigidas contra una diana particular, seleccionada de una biblioteca preparada de acuerdo con la invención.

### **Descripción de las figuras**

Figura 1. PCR murino – mSymplex<sup>MR</sup>: RT-PCR de extensión de superposición múltiplex para la amplificación y enlace análogo de los genes de anticuerpo de cadena pesada y ligera a partir de una sola célula. Las mezclas de cebador ilustrativas utilizadas para RT-PCR y las PCR incrustadas se desvelan con detalle en la tabla 2 y tabla 3 (o tabla 5).

Figura 2. Clonación de repertorio murino: un grupo de productos PCR mSymplex<sup>MR</sup> que codifican pares de genes VHNL a partir de células plasmáticas solas se empalman a la cadena ligera constante kappa humana que codifica para el gen mediante empalme por extensión de superposición. El grupo de genes, que codifican para anticuerpos quiméricos humanos/de ratón completos, se insertan en un vector de expresión seguido por una inserción de una casete promotora bi-direccional (2xCMV). Las mezclas de cebador utilizadas para el empalme kappa humana se desvelan con detalle en la tabla 6.

Figuras 3A. Clasificación de esplenocitos de ratón. (A) Para aislar células plasmáticas definidas como altas en CD43, altas en CD138, positivas a PI (yoduro de propidio) o células muertas se excluyen en el panel izquierdo inferior (Not P1). Después las células plasmáticas se clasifican como altas en CD43, altas en CD138 en el panel derecho inferior (P2). Finalmente, los dobletes se excluyen en la gráfica SSC-H, SSC-W en el panel derecho superior (P3). Las células positivas para las tres clasificaciones se clasifican en placas ELISPOT. (B) Para aislamiento de blastocitos plasmáticos definidos como intermedios en MHCII, intermedios en B220, positivos para PI (yoduro de propidio) o células muertas se excluyen en el panel izquierdo inferior (Not P1). Después blastocitos plasmáticos se clasifican como intermedios en MHCII, intermedios en B220, en el panel derecho inferior (P2). Se excluyen los dobletes en SSC-H, SSC-W en el panel derecho superior de la gráfica (P3) y finalmente las células se clasifican por tamaño en el panel izquierdo superior (P4). Las células positivas para los cuatro criterios se clasifican en placas ELISPOT.

Figura 4. Clasificación de esplenocitos de ratón. En primer lugar se excluyen las células PI positivas o muertas en el panel izquierdo inferior (P1). En la gráfica por puntos derecha superior, se muestran CD138 PE y CD43 FITC. Se establecen cuatro criterios en las poblaciones de células fenotípicas diferentes: P2 es intermedio a CD138, alto en CD43, P3 es alto en CD138, alto en CD43. P4 es alto en CD138, negativo en CD43. P5 es intermedio en CD138, bajo en CD43. Se clasificaron 10.000 células positivas para P1 y cada uno de los cuatro criterios en tubos de ensayo y se congelaron para evaluación por Symplex.

Figura 5. Electroforesis en gel de productos de PCR para titulación PCR Symplex de lisado celular de las cuatro fracciones clasificadas. P2, P3, P4 y P5 son los criterios clasificados de acuerdo con la figura 4. M son marcadores de peso molecular.

Figura 4. Una representación esquemática del vector de expresión de anticuerpo de longitud completa de mamífero 00-VP-002. Amp y Amp pro, gen de resistencia a ampicilina y su promotor; origen pUC, origen pUC de replicación; CMV, promotor de mamífero que activa la expresión de la cadena ligera y la cadena pesada; líder IGHV, líder de cadena pesada humana genómica; mezclado H, pieza intermedia que se intercambia para la secuencia que codifica para la región variable de cadena pesada; IGHG1, secuencia que codifica para la región constante de cadena pesada de inmunoglobulina genómica isotipo G1 (la secuencia se muestra en el apéndice 1); B-globina A de conejo, secuencia polyA de globina beta de conejo; líder IGKV, líder kappa murino; mezclado L, pieza intermedia que se intercambia por la secuencia que codifica para la cadena ligera; SV40 term, secuencia terminadora de virus simio 40; FRT, sitio objetivo de reconocimiento de Flp; Neo, gen de resistencia a neomicina; SV40 polyA, secuencia de señal polyA de virus simio 40.

Figura 5. Análisis de un repertorio de anticuerpos quiméricos anti-hEGFR. Análisis de grupo de la diferencia de absorbancia a 450-620 nm. Los sobrenadantes se agrupan por reactividad, se indica por el número (1 a 4) seguido por el número de clon. El color gris oscuro indica una disminución en el número de células metabólicamente activas mientras que el gris claro indica un incremento en el número de células metabólicamente activas. El color negro indica sobrenadantes sin efectos sobre el número de células metabólicamente activas.

### **Descripción de la invención**

La presente invención establece proporcionar posibilidades adicionales para uso del procedimiento de amplificación y enlace descrito en el documento WO 2005/042774 para proporcionar colecciones de vectores de anticuerpos a partir de animales no humanos. Estas mejoras permiten la clonación de las secuencias codificantes para el anticuerpo quimérico humano/no humano con pares análogos de regiones variables que se van a acoplar a un formato de alto rendimiento. Esto se obtiene básicamente al proporcionar un nuevo material de partida para los procedimientos de amplificación y enlace y en proporcionar procedimientos para generación de bibliotecas de anticuerpos quiméricos humanos/no humanos con pares análogos de regiones variables.

Un aspecto de la invención es un procedimiento para enlazar secuencias variables de cadena pesada y ligera que comprende amplificar, en un procedimiento de amplificación molecular múltiple, las secuencias nucleótidas relevantes utilizando una plantilla derivada de una sola célula aislada, una población de células isogénicas o una población de células genéticamente diversas y llevar a cabo un enlace subsiguiente de las secuencias amplificadas.

### **Definiciones**

La expresión “par análogo” desvela un par original de ácidos nucleicos no contiguos de interés que están contenidos dentro o derivados de una sola célula. En realizaciones preferidas, un par análogo comprende dos secuencias que codifican para la región variable las cuales, juntas, codifican para un dominio variable de proteína de unión y secuencias de gen las cuales se derivan de la misma célula. Por lo tanto, cuando se expresan ya sea como una proteína de unión completa o como un fragmento estable de las mismas, conserva la afinidad de unión y especificidad de la proteína de unión expresada originalmente de esta célula. Un par análogo puede estar comprendido, por ejemplo, de una secuencia que codifica para la cadena pesada variable de anticuerpo asociada con una secuencia que codifica para la cadena ligera variable de la misma célula o una cadena  $\alpha$  de receptor de linfocito T asociada con una cadena  $\beta$  que codifica para una secuencia de la misma célula. Una biblioteca de pares análogos es una colección de los pares análogos.

La expresión “polimerasa de inicio caliente” desvela polimerasas que son inactivas o que tienen muy baja actividad a temperaturas utilizadas para transcripción inversa. Dichas polimerasas necesitan ser activadas por altas temperaturas (90 a 95°C) para que se vuelvan funcionales. Esto es una ventaja, por ejemplo, en procedimientos RT-PCR de una sola etapa dado que evita la interferencia de la polimerasa con la reacción de transcriptasa inversa.

El término “población isogénica de células” desvela una población de células genéticamente idénticas. En particular, una población isogénica de células derivadas por expansión clonal de una sola célula aislada es de interés de la presente invención.

La expresión “sola célula aislada” desvela una célula que ha sido separada físicamente de una población de células que corresponden a “una sola célula en un solo recipiente”. Cuando se distribuye una población de células individualmente entre una pluralidad de recipientes, se obtiene una población de células solas aisladas. Como se especifica en la sección titulada “fuentes de plantillas”, la proporción de recipientes con una sola célula no es necesariamente el 100% con el fin de denominarla una población de células solas.

Los términos derivados de “enlace” o “unión” con relación a una amplificación desvelan la asociación de las secuencias amplificadas de ácido nucleico que codifican para las secuencias de ácido nucleico de interés en un solo segmento. En relación con pares análogos, un segmento comprende secuencias de ácido nucleico que codifican para un dominio variable, por ejemplo, una región variable de cadena pesada de anticuerpo asociada con una secuencia que codifica para una región variable de cadena ligera de anticuerpo, derivada de la misma célula. El enlace puede obtenerse simultáneamente con la amplificación o como una etapa inmediata después de la amplificación. No hay requerimientos acerca de la forma o funcionalidad del segmento, puede ser lineal, circular, de cadena sencilla o de cadena doble. Tampoco el enlace es necesariamente permanente, una de las secuencias de ácido nucleico de interés se puede aislar del segmento si se desea, una de la secuencia que codifica para la región variable se puede aislar, por ejemplo, de un segmento de par análogo. No obstante, en la medida en que las regiones variables originales constituyen el par análogo no estén barajadas con otras regiones variables, aún se les considera un par análogo, aunque no unidas juntas en un solo segmento. El enlace es preferiblemente un enlace fosfodiéster nucleotídico. No obstante, el enlace también se puede obtener por diferentes procedimientos de reticulación química.

- La expresión "amplificación molecular múltiplex" desvela la amplificación simultánea de dos o más secuencias diana en la misma reacción. Los procedimientos de amplificación adecuados incluyen reacción en cadena de polimerasa (PCR) (documento U.S. 4.683.202), reacción en cadena de ligasa (LCR), (Wu y Wallace, 1989, Genomics 4, 560-9), técnica de amplificación por desplazamiento de cadena (SDA) (Walker et al., 1992, Nucl. Acids Res. 20, 1691-6), replicación de secuencia autosostenida (Guatelli et al., 1990, Proc. Nat. Acad. Sci. U.S.A., 87, 1874-8) y amplificación de secuencia basada en ácido nucleico (NASBA) (Compton J., 1991, Nature 350, 91-2). Los dos últimos procedimientos de amplificación involucran reacciones isotérmicas basadas en transcripción isotérmica, las cuales producen tanto ARN de cadena sencilla (ARNss) así como ADN de cadena doble (ADNs).
- El término "PCR múltiplex" desvela una variante de PCR en la cual se amplifican simultáneamente dos o más secuencias diana, al incluir más de un conjunto de cebadores en la misma reacción, por ejemplo un conjunto de cebadores adaptados para amplificación de la región variable de cadena pesada y un conjunto de cebadores adaptados para amplificación de la región variable de cadena kappa en la misma reacción de PCR. De manera adicional o alternativa se puede combinar con estos conjuntos de cebadores un conjunto de cebador adaptado para amplificación de la región variable de cadena lambda.
- El término "RT-PCR multiplex" desvela una reacción de PCR múltiplex la cual está precedida por una etapa de transcripción inversa (RT). La RT-PCR múltiplex se puede realizar como un procedimiento de dos etapas con una etapa RT separada antes de PCR múltiplex, o como un procedimiento de una sola etapa en donde todos los componentes tanto de RT como de PCR múltiplex se combinan en un solo tubo.
- La expresión "PCR de superposición-extensión múltiplex" y "RT-PCR de superposición-extensión múltiplex" implica que la PCR múltiplex o la RT-PCR múltiplex se realizan utilizando una mezcla de cebador de superposición de una extensión múltiplex para amplificar las secuencias diana, por lo que se habilita la amplificación simultánea y el enlace de las secuencias diana.
- La expresión "una pluralidad de recipientes" desvela cualquier objeto (o colección de objetos) los cuales permiten la separación física de una sola célula de una población de células. Estos pueden ser tubos de ensayo, placas de multipocillos (por ejemplo placas de 96 pocillos, de 384 pocillos, placas de microtitulación o placas de multipocillos diferentes), disposiciones, microdisposiciones, microchips, geles o una matriz de gel. Preferiblemente, el objeto se aplica para amplificación por PCR.
- La expresión "proteína policlonal" o "policlonalidad", como se utiliza en la presente, se refiere a una composición de proteína que comprende moléculas de proteína diferentes pero homólogas, preferiblemente seleccionadas de la superfamilia de inmunoglobulinas. De esta manera, cada molécula de proteína es homóloga a otras moléculas de la composición pero también contiene una o más cadenas de secuencia polipeptídica variable, las cuales están caracterizadas por diferencias en la secuencia de aminoácidos entre los elementos individuales de la proteína policlonal. Los ejemplos conocidos de las proteínas policlonales incluyen moléculas de anticuerpo o inmunoglobulina, receptores de células T y receptores de células B. Una proteína policlonal puede estar constituida por un subconjunto definido de moléculas de proteínas las cuales se han definido por una característica común tal como la actividad de unión compartida hacia el objetivo deseado, por ejemplo un anticuerpo policlonal que muestra especificidad de unión hacia el antígeno diana deseado.
- La expresión "una población de células genéticamente diversas", como se utiliza en la presente, se refiere a una población de células en donde las células individuales en la población difieren entre sí a nivel genómico. La población de células genéticamente diversas es, por ejemplo, una población de células derivadas de un donante o una fracción de tales células, por ejemplo una fracción de células que contienen linfocitos B o linfocitos T.
- La expresión "conjunto cebador" se utiliza de manera intercambiable con el término "par de cebador" y desvela dos o más cebadores los cuales juntos son capaces de cebar la amplificación de una secuencia nucleotídica de interés (por ejemplo, un miembro de un par análogo). Un conjunto de cebador de la presente invención se puede diseñar para cebar una familia de secuencias nucleótidas que contienen secuencias que codifican para la región variable. Los ejemplos de diferentes familias son las cadenas ligeras kappa de anticuerpos, las cadenas ligeras lambda, las regiones variables de cadena pesada y las regiones variables del receptor de células T  $\alpha$ ,  $\beta$ ,  $\gamma$ , o  $\delta$ . Un conjunto de cebador para la amplificación de una familia de secuencias nucleótidas que contienen secuencias que codifican para la región variable con frecuencia constituye una pluralidad de cebadores en donde varios cebadores pueden ser cebadores degenerados.
- El término "identidad de secuencia" se expresa como un porcentaje el cual indica el grado de identidad entre las secuencias de ácido nucleico sobre la longitud de la más corta de las dos secuencias. Se puede calcular como  $(N_{ref} - N_{dif}) \times 100\% / N_{ref}$ , en donde  $N_{ref}$  es el número de residuos en la más corta de las secuencias y en donde  $N_{dif}$  es el número total de residuos no idénticos en una  $N_{ref}$  de longitud alineada para coincidencia óptima entre las dos secuencias. Por lo tanto, la secuencia de ADN AGTCAGTC (SEC ID N° 32) tendrá una identidad de secuencia del 75% con la secuencia **TAATCAATCGG** (SEC ID N°: 33) ( $N_{dif}=2$  y  $N_{ref}=8$ ) (el subrayado muestra la alineación óptima y lo que está en negrita indica los dos residuos no idénticos de los 8).
- El término "aleatoriamente" o "aleatorio" con respecto a un enlace se refiere a un enlace de secuencias nucleótidas las cuales no se derivan de la misma célula sino que se enlazan transversalmente entre una población de células genéticamente diversas. Si las secuencias nucleótidas de interés son secuencias que codifican para la región variable, esto resultará en una biblioteca combinatoria de secuencias enlazadas. Por otra parte, si las secuencias nucleótidas de interés codifican para una proteína heteromérica no diversa, las secuencias enlazadas aleatoriamente aparecerán similares a las secuencias enlazadas de una sola célula.
- La expresión "plantilla derivada de una sola célula aislada", con respecto a la transcripción inversa, se refiere a los ácidos nucleicos dentro de una célula aislada como tal. Los ácidos nucleicos pueden estar, por ejemplo, en forma de

ARN, ARNm, ADN o ADN genómico. Los ácidos nucleicos se pueden aislar de la célula o aún estar con los contenidos remanentes de la célula, en donde la célula está en una forma intacta o en una forma lisada.

5 El término "CD43" se refiere a un antígeno de superficie de ratón que se conoce bajo numerosos sinónimos incluyendo el antígeno 3E8, A630014B01Rik, el antígeno de diferenciación de células B LP-3, Cd43, el antígeno CD43, Galgp, sialoglucoproteína de leucocitos, leucosialina, precursor de leucosialina, Ly48, Ly-48, sialoforina así como los marcadores de superficie ortólogos de otros animales.

El término "CD138" se refiere a un antígeno de superficie de ratón conocido bajo numerosos sinónimos incluyendo Syndecan-1, AA408134, AA409076, CD138, syn-1, Synd, Synd1, SYND1, Synd-1, el precursor Syndecan-1 así como marcadores de superficie ortólogos de otros animales.

10 El término "MHCII" se refiere a un antígeno de superficie de ratón conocido bajo numerosos sinónimos incluyendo antígenos CD74, CLIP, DHLAG, cadena  $\gamma$  de antígeno de histocompatibilidad clase II H-2, HLADG, HLA-DR-GAMMA, cadena invariable asociada al antígeno Ia, Ia-GAMMA, li, cadena invariable asociada a MHC clase II así como otros marcadores de superficie ortólogos de otros animales.

15 El término "B220" se refiere a un antígeno de superficie de ratón conocido bajo numerosos sinónimos incluyendo B220, Cd45, CD45, antígeno CD45, CD45R, L-CA, precursor de antígeno común de leucocitos, loc, Ly-5, antígeno común linfocítico Ly-5, Lyt-4, T200 así como marcadores de superficie ortólogos de otros animales.

### **Descripción detallada de la invención**

#### **Procedimiento de Amplificación y Enlace**

20 Una característica de la presente invención reduce el número de tubos necesarios para amplificar las secuencias nucleótidas de interés, utilizando una variante de PCR en la cual se amplifican simultáneamente en el mismo tubo dos o más secuencias diana, al incluir más de un conjunto de cebadores, por ejemplo, todos los cebadores necesarios para amplificar secuencias que codifican para la región variable, en la misma reacción. Generalmente, este enfoque se conoce como reacción en cadena de polimerasa múltiplex (PCR múltiplex).

25 Una característica adicional de la presente invención es que dos o más secuencias diana amplificadas por PCR múltiplex están enlazadas en proximidad cercana con el procedimiento de amplificación. En particular, los pares análogos de las secuencias que codifican para la región variable están enlazadas por este procedimiento.

30 Una realización de la presente invención aprovecha que la mezcla de cebador múltiplex se puede diseñar para trabajar en un procedimiento de PCR de superposición/extensión, lo que resulta en una amplificación y enlace simultáneos de las secuencias nucleótidas de interés. Esta técnica de PCR de superposición-extensión múltiplex sirve para reducir el número de reacciones necesarias para aislar y enlazar secuencias nucleótidas de interés, en particular pares análogos de regiones variables enlazadas.

35 Otras realizaciones de la presente invención aplican enlace por ligación o por recombinación como una alternativa al enlace por PCR por superposición-extensión múltiplex. En estos procedimientos, el enlace no se realiza simultáneamente con la amplificación PCR múltiplex, sino como una etapa inmediata después de la amplificación. No obstante, el enlace aún se puede realizar en el mismo tubo en donde se realizó la PCR múltiplex.

40 Una PCR de superposición-extensión múltiplex requiere la presencia de dos o más conjuntos de cebadores (una mezcla de cebador múltiplex) en donde por lo menos un cebador de cada conjunto está equipado con una cola de superposición-extensión. Las colas de superposición-extensión permiten el enlace de los productos generados por cada uno de los conjuntos de cebador durante la amplificación. Dicha mezcla de cebador se denomina una mezcla de cebador por superposición-extensión múltiplex. La PCR de superposición-extensión múltiplex difiere de la PCR de superposición-extensión convencional en que las secuencias que se van a enlazar se generan simultáneamente en el mismo tubo por lo que se proporciona enlace inmediato de las secuencias diana durante la amplificación, sin ninguna purificación intermedia.

45 Una característica adicional de la presente invención es la etapa de transcripción inversa (RT) que precede a la PCR múltiplex o a la amplificación por PCR de superposición-extensión múltiplex, utilizando una plantilla derivada de una sola célula aislada o de una población de células isogénicas.

50 Una característica adicional de la presente invención es el uso de secuencias nucleótidas derivadas de una sola célula aislada o de una población de células isogénicas como plantilla para la amplificación por PCR múltiplex. Preferiblemente, el ARN de una sola célula es sometido a transcripción inversa en ADNc antes de la PCR múltiplex. Para la amplificación de algunas secuencias de ácido nucleico de interés se puede utilizar ADN genómico como una alternativa al ARNm. Mediante la utilización de células solas aisladas o una población de células isogénicas derivadas por expansión clonal de una sola célula aislada como fuente de plantilla, es posible evitar el mezclado de secuencias nucleótidas que codifican para una proteína heteromérica de interés, con secuencias nucleótidas derivadas de células diferentes dentro de una población de células. Esto es de importancia si uno desea obtener la composición original de las secuencias de interés. Especialmente para la generación de un par análogo de secuencias que codifican para la región variable, el uso de una sola célula aislada o una población de células isogénicas como fuente de plantilla es una característica importante.

60 De manera adicional, la presente invención facilita la generación de bibliotecas de secuencias de interés de ácido nucleico enlazadas, en particular bibliotecas combinatorias y bibliotecas de pares análogos de regiones variables. Además, la presente invención utiliza ácidos nucleicos derivados de células solas, preferiblemente en forma de ARN que no necesita ser aislado de los contenidos de células remanentes antes de que pueda ser utilizado como plantilla.

Una realización de la presente invención abarca el enlace de una pluralidad de secuencias nucleótidas no contiguas de interés. El procedimiento comprende amplificar, en un procedimiento de amplificación PCR múltiplex o RT-PCR múltiplex, secuencias nucleótidas de interés utilizando una plantilla derivada de una sola célula aislada o una población de células isogénicas y llevar a cabo el enlace de las secuencias nucleótidas amplificadas de interés. Además, el procedimiento comprende una etapa opcional de realizar una amplificación adicional de los productos enlazados. Una realización adicional de la presente invención abarca un procedimiento para producir una biblioteca de pares análogos que comprende secuencias que codifican para la región variable enlazada. El procedimiento comprende proporcionar una fracción de células que contiene linfocitos de un donante, la cual opcionalmente está enriquecida para una población linfocítica particular de dicha fracción celular o en donde una población linfocítica particular se ha aislado de dicha fracción celular. Además, una población de las células solas aisladas se obtiene al distribuir células de la fracción celular que contiene linfocitos, o la fracción de células enriquecidas, individualmente de entre una pluralidad de recipientes. Se realiza la amplificación molecular múltiplex (amplificación RT-PCR múltiplex) de las secuencias que codifican para la región variable contenida en la población de células solas aisladas y el enlace de pares de secuencias que codifican para la región variable, en donde un par individual de secuencias de la región variable se deriva de una sola célula, dentro de la población de células solas aisladas. Además, la técnica comprende dos etapas opcionales: en la primera etapa, la sola célula aislada individual en la población de células solas se expande a una población de células isogénicas antes de realizar la amplificación RT-PCR múltiplex, obteniendo así una pluralidad de recipientes con una población diversa de células isogénicas (una población de células isogénicas en un recipiente). La segunda etapa opcional abarca realizar una amplificación adicional de las secuencias que codifican para la región variable enlazada.

En realizaciones preferidas de la presente invención, un elemento individual de dicha biblioteca de pares análogos comprendida de una secuencia que codifica para la región variable de la cadena ligera de inmunoglobulina se asocia con una secuencia que codifica para la región variable de la cadena pesada de inmunoglobulina, que se origina de la misma célula o de secuencias que codifican para un dominio que une receptor de células T, constituido en la región variable de cadena  $\alpha$  asociada con la región variable de cadena  $\beta$  o una región variable de cadena  $\gamma$  asociada con la región variable de cadena  $\delta$ , en donde las regiones variables asociadas se originan de la misma célula.

La amplificación RT-PCR múltiplex de la presente invención se puede realizar ya sea como un procedimiento de dos etapas, en donde se realiza transcripción inversa (RT) separada de la amplificación por PCR múltiplex (o amplificación molecular múltiplex alternativa) o como un procedimiento de una sola etapa, en donde las etapas de amplificación RT y PCR múltiplex se realizan con los mismos cebadores en un tubo.

La transcripción inversa (RT) se realiza con una enzima que contiene actividad de transcriptasa inversa lo que resulta en la generación de ADNc a partir de ARN total, ARNm o ARN específico diana a partir de una sola célula aislada. Los cebadores los cuales se pueden utilizar para la transcripción inversa son, por ejemplo, cebadores oligo-dT, hexámeros aleatorios, decámeros aleatorios, otros cebadores aleatorios o cebadores que son específicos para las secuencias nucleótidas de interés.

El procedimiento de amplificación RT-PCR múltiplex de dos etapas permite que el ADNc generado en la etapa RT se distribuya a más de un recipiente lo que permite el almacenamiento de una fracción de plantilla antes de seguir con la amplificación. De manera adicional, la distribución de ADNc a más de un tubo permite el desempeño de más de una amplificación PCR múltiplex de ácido nucleico derivado de la misma plantilla. Aunque esto resulta en un número aumentado de reacciones separadas, abre la posibilidad de disminuir la complejidad de la mezcla de cebador múltiplex si esto es lo que se desea. Este enfoque de dos etapas se puede aplicar, por ejemplo, para amplificar y enlazar una región variable de cadena pesada y las secuencias que codifican para la región variable de cadena ligera  $\kappa$  en un tubo, y la región variable de cadena pesada y las secuencias que codifican para la región variable de cadena ligera  $\lambda$  en un tubo diferente utilizando la misma plantilla. Una sola célula habitualmente expresa sólo una de las cadenas ligeras. No obstante, con frecuencia será más fácil realizar las reacciones simultáneamente en vez de esperar el resultado de una de las reacciones antes de realizar la otra. Además, la amplificación tanto de  $\kappa$  como de  $\lambda$  sirve como un control negativo interno, dado que se esperará que únicamente  $\kappa$  o  $\lambda$  se amplifiquen de una sola célula.

En el procedimiento RT-PCR múltiplex de una sola etapa, la transcripción inversa y la amplificación por PCR múltiplex se lleva a cabo en el mismo recipiente. Todos los componentes necesarios para realizar tanto la transcripción inversa como la PCR múltiplex en una sola etapa inicialmente se agregan en los recipientes y se lleva a cabo la reacción. Generalmente, no hay necesidad de agregar componentes adicionales una vez que se ha iniciado la reacción. La ventaja de amplificación RT-PCR múltiplex de una sola etapa es que reduce el número de etapas necesarias para generar las secuencias nucleótidas enlazadas de la presente invención aún más. Esto es particularmente útil cuando se realiza RT-PCR múltiplex en una disposición de células solas, en donde la misma reacción necesita llevarse a cabo en una pluralidad de recipientes. La RT-PCR múltiplex de una sola etapa se realiza utilizando cebadores inversos presentes en la mezcla de cebador múltiplex necesaria para la amplificación por PCR múltiplex como cebadores para la transcripción inversa también. Generalmente, la composición necesaria para RT-PCR múltiplex de una sola etapa comprende una plantilla de ácido nucleico, una enzima con actividad de transcriptasa inversa, una enzima con actividad de ADN polimerasa, una mezcla de desoxinucleósido trifosfato (mezcla dNTP que comprende dATP, dCTP, dGTP y dTTP) y una mezcla de cebador múltiplex. La plantilla de ácido nucleico preferiblemente es ARN total o ARNm derivado de una sola célula aislada ya sea en una forma purificada, en forma de un lisado de la célula o aún dentro de la célula intacta. Generalmente, la composición exacta de la mezcla de reacción requiere cierta optimización para cada mezcla de cebador múltiplex que se va a utilizar con la presente invención. Esto se aplica a los procedimientos RT-PCR múltiplex de dos etapas como de una sola etapa.

Para algunas reacciones de RT-PCR múltiplex de una sola etapa, puede ser una ventaja agregar componentes adicionales durante la reacción, por ejemplo, la adición de polimerasa después de la etapa RT. Otros componentes pueden ser, por ejemplo, una mezcla de dNTP o una mezcla de cebador múltiplex posiblemente con una composición de cebador diferente. Esto se puede considerar como RT-PCR múltiplex de un tubo, la cual

generalmente tiene las mismas ventajas que la RT-PCR de una sola etapa, dado que también limita el número de tubos necesarios para obtener los productos enlazados deseados.

Las secuencias nucleótidas de interés, amplificadas por RT-PCR múltiplex, se pueden enlazar entre sí por varios procedimientos, tal como RT-PCR de superposición-extensión múltiplex, ligación o recombinación utilizando mezclas de cebador múltiplex diferentes. Preferiblemente, la amplificación RT-PCR múltiplex y los procedimientos de enlace son procedimientos de una sola etapa o de dos etapas. No obstante, el procedimiento de enlace también se puede realizar como un procedimiento de etapas múltiples utilizando, por ejemplo, un fragmento de relleno para enlazar las secuencias de ácido nucleico de interés, ya sea con PCR, ligación o recombinación. El fragmento de relleno puede contener elementos cis, elementos promotores o una secuencia codificante relevante o secuencia de reconocimiento. En una realización preferida, el procedimiento de enlace se realiza en el mismo recipiente como amplificación RT-PCR múltiplex.

En una realización de la presente invención, el enlace de una pluralidad de secuencias nucleótidas no contiguas de interés se realiza en asociación con la amplificación PCR múltiplex, utilizando una mezcla de cebador de superposición de una extensión múltiplex. Esto resulta en la amplificación combinada y el enlace de las secuencias diana. Generalmente, la composición necesaria para PCR de superposición-extensión múltiplex comprende una plantilla de ácido nucleico, una enzima con actividad de ADN polimerasa, una mezcla de trifosfato de desoxinucleósido (mezcla dNTP que comprende dATP, dCTP, dGTP y dTTP) y una mezcla de cebador de superposición-extensión múltiplex.

En una realización particular de la presente invención, el enlace de una pluralidad de secuencias nucleótidas no contiguas de interés se realiza por RT-PCR de superposición-extensión múltiplex utilizando una plantilla derivada de una sola célula aislada o una población de células isogénicas. Además, el procedimiento comprende una etapa opcional de realizar una amplificación molecular adicional de productos enlazados. Preferiblemente, la RT-PCR de superposición-extensión múltiplex se realiza como una reacción de una sola etapa/un tubo.

Una mezcla de cebador de superposición-extensión múltiplex de la presente invención comprende por lo menos dos conjuntos de cebador capaces de cebar la amplificación y el enlace de por lo menos dos secuencias que codifican para la región variable, por ejemplo la amplificación y el enlace de secuencias a partir de las familias de la región variable de cadena pesada de inmunoglobulina con las familias de la cadena ligera  $\kappa$  o  $\lambda$  o amplificación y enlace de secuencias a partir de las familias de receptor de células T  $\alpha$ ,  $\beta$ ,  $\gamma$  o  $\delta$ .

En otra realización de la presente invención, la pluralidad de secuencias nucleótidas de interés, amplificadas por RT-PCR múltiplex, se enlazan por ligación. Para obtener esto, la mezcla de cebador múltiplex utilizada para RT-PCR múltiplex, se diseña de manera que las secuencias diana amplificadas se pueden escindir con enzimas de restricción apropiadas y se puede realizar enlace covalente por ligación de ADN (el diseño de cebador se desvela en la sección "*mezclas y diseños de cebadores*"). Después de la amplificación RT-PCR múltiplex con la mezcla de cebador múltiplex, las enzimas de restricción necesarias para formar extremos compatibles de las secuencias diana se agregan a la mezcla junto con la ligasa. Antes de esta etapa no se necesita purificación de los productos de PCR, aunque se puede realizar purificación. La temperatura de reacción de la escisión de restricción y ligación combinadas está aproximadamente entre 0 y 40°C. No obstante, si la polimerasa de la reacción PCR múltiplex aún está presente en la mezcla, se prefiere una temperatura de incubación inferior a la temperatura ambiente, de manera más preferible temperaturas entre 4 y 16°C.

En otra realización adicional de la presente invención, la pluralidad de secuencias nucleótidas de interés, amplificadas por RT-PCR múltiplex, se enlazan por recombinación. En este enfoque, las secuencias diana amplificadas se pueden unir utilizando sitios de recombinación idénticos. El enlace después se realiza al agregar las recombinasas que facilitan la recombinación. Algunos sistemas de recombinasa adecuados son recombinasa F1p con una diversidad de sitios FRT, recombinasa Cre con una variedad de sitios lox, integrasa  $\Phi$ C31, la cual lleva a cabo la recombinación entre el sitio attP y el sitio attB, el sistema de  $\beta$ -recombinasa-seis así como el sistema Gin-gix. El enlace por recombinación se ha ilustrado para dos secuencias nucleótidas ( $V_H$  unida con  $V_L$ ) (Chapal, N. et al. 1997 Bio Techniques 23, 518-524).

En una realización preferida de la presente invención, las secuencias nucleótidas de interés comprenden secuencias que codifican para la región variable y el enlace genera un par análogo de secuencias que codifican para la región variable. Dicho par análogo puede comprender una o más secuencias que codifican para la región constante además de las regiones variables. Preferiblemente, las regiones constantes son de origen humano y el par análogo de región variable es de origen diferente, tal como ratón, rata o conejo.

En una realización incluso más preferida de la presente invención, las secuencias nucleótidas de interés comprenden secuencias que codifican para la región variable de inmunoglobulina y el enlace genera un par análogo de secuencias que codifican para la región variable de cadena ligera y región variable de cadena pesada. Dicho par análogo puede comprender una o más secuencias que codifican para la región constante además de las regiones variables. Además, dicho par análogo se puede aislar de la plantilla derivada de células de la línea de linfocitos B enriquecida de una fracción celular que contiene linfocitos, tal como sangre entera, células mononucleares o leucocitos.

En otra realización de la presente invención, las secuencias nucleótidas de interés comprenden secuencias que codifican para la región variable TcR y el enlace genera un par análogo de secuencias que codifican para la región variable de la cadena  $\alpha$  y la región variable de la cadena  $\beta$  o secuencias que codifican para la región variable de la cadena  $\gamma$  y región variable de la cadena  $\delta$ . Dicho par análogo puede comprender una o más secuencias que codifican para la región constante además de las regiones variables. Además, dicho par análogo se puede aislar de una plantilla derivada de células de la línea de linfocitos T enriquecida de una fracción de células que contienen linfocitos, tal como sangre entera, células mononucleares o leucocitos.

Otro aspecto de la presente invención es utilizar RT-PCR múltiplex con una población de células genéticamente diversas como fuente de plantilla. La mayor parte de las secuencias que codifican para proteína heteromérica no varía de una célula a otra como en el caso con las secuencias que codifican para la región variable de proteínas de unión. De esta manera, cuando se utiliza la presente invención para la clonación de las secuencias que codifican para proteínas heteroméricas no variables no hay necesidad de realizar un aislamiento inicial de células solas.

En esta realización de la presente invención, una pluralidad de secuencias nucleótidas no contiguas de interés están unidas aleatoriamente por un procedimiento que comprende, realizar amplificación por RT-PCR múltiplex de secuencias nucleótidas de interés utilizando una plantilla derivada de una población de células genéticamente diversas y llevar a cabo el enlace de las secuencias nucleótidas amplificadas de interés. Además, el procedimiento comprende una etapa opcional de realizar una amplificación adicional de los productos enlazados. Al igual que con el enfoque de células solas, el enlace se puede realizar utilizando una mezcla de cebador de superposición-extensión múltiplex para la amplificación, o de manera alternativa por ligación o recombinación. Preferiblemente, la plantilla derivada de la población de células no está contenida estrictamente dentro de las células. La población de células puede, por ejemplo, lisar.

La aplicación del procedimiento de enlace aleatorio en una población de células que expresan proteínas de unión variantes permite la generación simplificada de bibliotecas combinatorias de secuencias que codifican para la región variable. Preferiblemente, la población de células constituye células que expresan proteínas que se unen a la región variable, tales como linfocitos B, linfocitos T, células de hibridoma, células plasmáticas, plasmablastos o una mezcla de estas células.

La población de células en la realización mencionada anteriormente se puede permeabilizar o lisar, por ejemplo, sin purificación adicional o los ácidos nucleicos de plantilla se pueden aislar de las células por procedimientos convencionales. Se prefiere el procedimiento RT-PCR múltiplex de una sola etapa. No obstante, también se puede utilizar el procedimiento de dos etapas en la realización.

Una manera eficaz de incrementar la especificidad, sensibilidad y el rendimiento de los procedimientos de enlace RT-PCR múltiplex es mediante la realización de una amplificación molecular adicional de las secuencias nucleótidas enlazadas obtenidas de RT-PCR múltiplex seguido por enlace mediante ligación o recombinación o enlace utilizando RT-PCR de superposición-extensión múltiplex. Esta amplificación adicional preferiblemente se realiza con amplificación PCR utilizando una mezcla de cebador adaptada para amplificar las secuencias de ácido nucleico enlazadas de interés. La mezcla de cebador utilizada puede ser los cebadores externos de la mezcla de cebador múltiplex o una mezcla de cebador de superposición-extensión múltiplex, lo que significa que los cebadores que se reasocian con el extremo 5' y el extremo 3' más externos de la cadena con sentido de las secuencias que codifican para la región variable enlazada, con lo que se permite la amplificación de la totalidad del producto enlazado. Los cebadores externos también se pueden describir como los cebadores de la mezcla de cebador de superposición-extensión múltiplex que no contienen colas de extensión de superposición. De manera alternativa, un conjunto de cebador incrustado o semi-incrustado se puede utilizar para la amplificación adicional de las secuencias nucleótidas enlazadas. La PCR incrustada sirve especialmente para incrementar la especificidad del procedimiento así como para incrementar la cantidad de producto enlazado. Para la presente invención, se considera que una PCR semi-incrustada (como se desvela en la sección intitulada *mezclas y diseños de cebadores*) funciona también como una PCR incrustada. Por lo tanto, se desea, aunque no es necesario para la presente invención, realizar una amplificación PCR adicional de los productos enlazados a partir de RT-PCR de superposición-extensión múltiplex o de los productos enlazados por ligación o recombinación, preferiblemente utilizando PCR incrustada o PCR semi-incrustada.

La amplificación adicional se puede realizar directamente utilizando una fracción o la totalidad del producto de reacción de RT-PCR de superposición-extensión múltiplex o el producto de ligación o el producto de recombinación, o una fracción de cualquiera de estos productos o utilizando productos enlazados, purificados parcialmente, de cualquiera de estas reacciones, por ejemplo al realizar electroforesis en gel de agarosa de los productos enlazados, y al cortar el fragmento correspondiente del tamaño esperado de las secuencias que codifican para la región variable enlazada. Para productos enlazados por RT-PCR de superposición-extensión múltiplex, la amplificación adicional preferiblemente se realiza directamente sobre una fracción de la reacción de RT-PCR de superposición-extensión múltiplex, puesto que de esta manera se puede ayudar en el enlace de las secuencias diana individuales que no están enlazadas en la primera reacción.

### **Secuencias de interés**

Las secuencias nucleótidas de interés de la presente invención se pueden seleccionar a partir de secuencias que codifican para diferentes subunidades o dominios, los cuales, cuando se expresan, forman una proteína o parte de una proteína. Dichas proteínas que están constituidas por al menos dos subunidades no idénticas se conocen como proteínas heteroméricas. Las proteínas heteroméricas son comunes en todas las clases de especies. Algunas de las clases a las cuales pertenecen las proteínas son, por ejemplo, enzimas, inhibidores, proteínas estructurales, toxinas, proteínas de canal, proteínas G, proteínas receptoras, proteínas de la superfamilia de inmunoglobulinas, proteínas de transporte, etc. Las secuencias nucleótidas que codifican para las proteínas heteroméricas son no contiguas, lo que significa, por ejemplo, que se originan de genes diferentes, o moléculas de ARNm diferentes. No obstante, el término no contiguo, como se utiliza en la presente invención, también significa secuencias nucleótidas que codifican para dominios de la misma proteína, en donde los dominios están separados por secuencias nucleótidas los cuales no son de interés.

En una realización de la presente invención, las secuencias nucleótidas de interés contienen secuencias que codifican para la región variable de la superfamilia de inmunoglobulina, tales como inmunoglobulinas (anticuerpos), receptores de células B y receptores de células T (TcR). Especialmente, las secuencias que codifican para la región variable de inmunoglobulinas son de interés. Las secuencias que codifican para la región variable comprenden anticuerpos de longitud completa así como Fab, Fv, scFv y combinaciones o fragmentos de las secuencias que

codifican para la región variable, por ejemplo regiones determinadoras de complementariedad (CDR), genes de unión o genes V o combinaciones de estos. Generalmente, la presente invención se puede aplicar a cualquier combinación de secuencias que codifican para la región variable y fragmentos de las mismas. La presente solicitud permite el enlace de únicamente los dominios variables de las secuencias que codifican para Fv o scFv que generan cadenas pesada y ligera. O bien, el enlace de la totalidad de la cadena ligera con la región variable de cadena pesada + dominio de la región constante C<sub>H1</sub> + partes de la región de bisagra, lo que genera Fab, Fab' o F(ab)<sub>2</sub>. Además, es posible agregar cualquier región de los dominios de la región constante de cadena pesada a la cadena pesada variable, con lo que se generan secuencias que codifican para anticuerpo de longitud completa o secuencias que codifican para anticuerpos truncados. En un aspecto de la invención, las secuencias variables no humanas se enlazan a regiones constantes humanas para generar anticuerpos quiméricos completos humanos/no humanos, preferiblemente anticuerpos quiméricos con regiones constantes humanas.

En una realización adicional de la presente invención las secuencias que codifican para la región variable comprenden un tipo de secuencias que codifica para la cadena ligera de inmunoglobulina ( $\kappa$  o  $\lambda$ ) y una secuencia que codifica para la región variable de cadena pesada de inmunoglobulina. Esto se obtiene al seleccionar cebadores que amplifican únicamente un isotipo de cadena ligera y pesada. El isotipo también se puede determinar por enlace o empalme de regiones constantes humanas de uno o más isotipos particulares de cadena pesada y ligera.

Las secuencias que codifican para la región variable derivada de los receptores de células T (TcR) también son de interés. Dichas secuencias que codifican para TcR comprenden secuencias que codifican para las cadenas  $\alpha$  y  $\beta$  de longitud completa o las cadenas  $\gamma$  y  $\delta$  así como TcR solubles o únicamente los dominios variables de estas cadenas o las proteínas de fusión de cadena sencilla de los mismos (por ejemplo la cadena sencilla  $\alpha\beta$  o la cadena sencilla  $\gamma\delta$ ).

### **Fuentes de plantillas**

Una característica de la presente invención es la capacidad de enlazar secuencias nucleótidas derivadas de una sola célula aislada, una población de células isogénicas o una población genéticamente diversa de células que no han sido separadas en recipientes solos.

Una característica preferida de la presente invención es el uso de células solas aisladas o una población de células isogénicas como fuente de plantilla, dado que se evita el mezclado de las secuencias de ácido nucleico de interés, en particular las secuencias que codifican para la región variable. Esto es de importancia si no desea obtener un par original, por ejemplo, de secuencias que codifican para la región variable.

Otra característica preferida de la presente invención es la obtención de una sola célula o una población de células solas a partir de una fracción celular que comprende linfocitos tales como linfocitos B, linfocitos T, células plasmáticas y/u otras diversas etapas de desarrollos de estas líneas celulares. Otras poblaciones de células que expresan proteínas de unión para la superfamilia de inmunoglobulina también se pueden utilizar para obtener células solas. Las líneas de células tales como células de hibridoma, líneas de células de la línea de linfocitos B o de linfocitos T o líneas de células inmortalizadas por virus o células derivadas de donante que participan en la respuesta inmunitaria también son aplicables en la presente invención. Las fracciones de células que contienen linfocitos derivados de donantes se pueden obtener de tejido natural o fluido el cual tiene alta concentración de las células, por ejemplo sangre, médula ósea, nódulos linfáticos, tejido del bazo, tejido de las amígdalas o de infiltraciones en y alrededor de los tumores o infiltraciones de tejido inflamatorio. Preferiblemente, en el caso de animales no humanos, se utiliza el tejido del bazo o de médula ósea. Los donantes pueden ser aquellos que previamente no han sido expuestos o pueden ser hiperinmunes con respecto al objetivo deseado. Para el aislamiento de proteínas que unen antígeno con especificidades de unión hacia un objetivo deseado, se prefieren los donantes hiperinmunes. Los donantes hiperinmunes pueden ser donantes inmunizados con el objetivo o fragmentos del objetivo, o pueden ser pacientes en convalecencia o individuos quienes no están sanos los cuales están desarrollando una respuesta inmunitaria natural hacia el objetivo, por ejemplo pacientes autoinmunes, pacientes con cáncer, pacientes con enfermedades infecciosas, por ejemplo en pacientes con VIH, pacientes con hepatitis A, B o C, pacientes con SARS, etc., o pacientes con enfermedades crónicas. No obstante, en una realización particularmente preferida, el donante es un animal no humano que ha sido inmunizado con un autoantígeno humano tal como una proteína humana implicada en cáncer tal como EGFR.

Para uso en la presente invención, las células donantes pueden ser de la misma especie que las especies que van a ser tratadas con los productos que se pueden obtener de las secuencias nucleótidas enlazadas de la presente invención. Preferiblemente, una célula donantea es un animal doméstico, una mascota o un humano. Una característica especial de la presente invención es que permite la generación de bibliotecas de anticuerpos quiméricos humanos/no humanos para generación de anticuerpos quiméricos para uso en tratamiento para humanos. Se prefiere un enfoque como tal cuando los anticuerpos se dirigen contra lo que se denominan antígenos propios, es decir, antígenos humanos.

El donante también puede ser un animal transgénico, en particular un ratón transgénico. Los animales transgénicos que portan los loci de inmunoglobulina humana se desvelan en la patente de EE.UU. número 6.111.166 y Kuroiwa, Y. et al, Nature Biotechnology; 2002; 20: 889-893. Dichos animales transgénicos son capaces de producir inmunoglobulinas humanas. Por lo tanto, los anticuerpos completamente humanos contra una diana específica se pueden generar por técnicas habituales de inmunización de dichos animales transgénicos. Esto permite la generación de bibliotecas que codifican para proteínas de unión con especificidades hacia dianas más difíciles tales como antígenos humanos para los que existe una respuesta de anticuerpo humano natural nula o limitada. Dichos animales transgénicos se pueden desarrollar de igual manera para producir receptores de células T humanas.

En una realización adicional de la presente invención, la fracción de células que contienen linfocitos está constituida por sangre entera, médula ósea, células mononucleares o leucocitos que se obtienen de un donante. Las células mononucleares se pueden aislar de sangre, médula ósea, nódulos linfáticos, bazo, infiltraciones alrededor de células

cancerosas e infiltraciones inflamatorias. Las células mononucleares se pueden aislar por técnicas de centrifugación de densidad, por ejemplo gradientes Ficoll. Si las células mononucleares se aíslan de muestras constituidas por tejido, el tejido se desintegra antes de que se realice la centrifugación en gradiente. La desintegración se puede llevar a cabo, por ejemplo, por procedimientos mecánicos tales como triturado, electroporación y/o por procedimientos químicos tales como tratamientos enzimáticos. El aislamiento de los leucocitos se puede llevar a cabo directamente de los donantes utilizando leucoferesis. Las preparaciones crudas, por ejemplo de médula ósea o de tejido, las cuales contienen linfocitos, también se pueden utilizar en la presente invención. Tales preparaciones necesitarán ser desintegradas, por ejemplo como se desvela en lo anterior con el fin de facilitar distribución de células solas.

Una característica adicional de la presente invención es el enriquecimiento de una fracción de células que contienen linfocitos, por ejemplo sangre entera, células mononucleares, leucocitos o médula ósea, con respecto a una población linfocítica particular, tales como las células de la línea de linfocitos B o de linfocitos T. El enriquecimiento de linfocitos B se puede realizar, por ejemplo, utilizando clasificación de células por esferas magnéticas (MACS) o clasificación de células activadas por fluorescencia (FACS) aprovechando las proteínas marcadoras de superficie celular específicas de línea tales como CD19 u otros marcadores específicos de líneas de células B tales como B220. El enriquecimiento de linfocitos T se puede realizar, por ejemplo, utilizando un marcador de superficie celular tal como CD3 u otros marcadores específicos de líneas de células T.

Una característica preferida de la presente invención es clasificar los linfocitos B enriquecidos adicionalmente con el fin de adquirir células plasmáticas, antes de distribuir las células individualmente entre una pluralidad de recipientes. El aislamiento de células plasmáticas generalmente se realiza por clasificación MACS o clasificación FACS utilizando marcadores de superficie tales como CD19. Se pueden utilizar también otros marcadores de superficie específicos de células plasmáticas o combinaciones de los mismos, por ejemplo CD138, CD43, CD19, MHC-II, dependiendo la selección exacta del marcador de la fuente de células plasmáticas, por ejemplo bazo, amígdalas, sangre o médula ósea. Por supuesto, la selección exacta de los marcadores de superficie también depende de las especies de las cuales se aíslan las células.

En un aspecto de la invención, los marcadores utilizados para clasificación y/o selección de células son CD43 y CD138 o MHCII y B220 u ortólogos. Preferiblemente, la combinación de marcadores es CD43 y CD138 y preferiblemente las células seleccionadas tienen una expresión intermedia o elevada de estos marcadores en relación a la población de células que comprenden linfocitos a partir de la cual se seleccionan o aíslan. De manera más preferible, el nivel de expresión de CD43 y CD138 es relativamente alto respecto a la población de células que comprenden linfocitos a partir de la cual se seleccionan o aíslan.

Las células plasmáticas también se pueden obtener de una población de células no enriquecida que contienen linfocitos que se obtienen de cualquiera de estas fuentes. Las células plasmáticas aisladas de sangre algunas veces denominadas células plasmáticas tempranas o plasmablastos. En la presente invención, estas células también se denominan células plasmáticas. Se desean células plasmáticas para el aislamiento de pares análogos de secuencias que codifican para inmunoglobulina debido a su mayor frecuencia de estas células lo que produce anticuerpos específicos de antígeno que reflejan la inmunidad adquirida hacia el antígeno deseado y la mayor parte de las células han experimentado hipermutación somática y por lo tanto codifican para anticuerpos de alta afinidad. Además, las concentraciones de ARNm en células plasmáticas son elevadas en comparación con la población de linfocitos B remanentes, y por lo tanto el procedimiento de transcripción inversa es más eficaz cuando se utilizan células plasmáticas solas. Como una alternativa al aislamiento de células plasmáticas, los linfocitos B de memoria se pueden aislar de una fracción de células que contengan linfocitos utilizando un marcador de superficie celular tal como CD27 e IgG.

Una característica alternativa de la presente invención es la selección de linfocitos B enriquecidos para especificidad de antígeno antes de distribuir las células entre una pluralidad de recipientes. El aislamiento de linfocitos B específicos de antígeno se realiza al poner en contacto linfocitos B enriquecidos con el antígeno o los antígenos deseados permitiendo la unión del antígeno a inmunoglobulina expuesta en la superficie, seguido por aislamiento de aglutinantes. Esto se puede realizar, por ejemplo, por biotinación del antígeno o los antígenos deseados seguido por técnicas adecuadas de clasificación de células. Las células plasmáticas así como los linfocitos B, células mononucleares no enriquecidas, leucocitos, sangre entera, médula ósea o preparaciones de tejidos se pueden someter a aislamiento con respecto a la especificidad de antígeno si esto es lo que se desea.

Otra característica de la presente invención es clasificación de linfocitos T enriquecidos (por ejemplo células CD3 positivas) utilizando marcadores de superficie tales como, por ejemplo, CD27 para obtener una fracción de linfocitos T de memoria. Los linfocitos T también se pueden seleccionar por especificidad a antígeno MHC utilizando complejos de MHC-péptido (por ejemplo Callan, M.F. *et al.* 1998. *J. Exp. Med.* 187, 1395-1402; Novak, E. J. *et al.* 1999. *J. Clin. Invest* 104, R63-R67).

Como una alternativa a clasificar células que expresan ciertos marcadores de superficie, es decir, una selección positiva, es concebible que las células que NO expresen los marcadores se supriman de la composición de células, dejando células que en realidad expresan los marcadores.

Una característica adicional de la presente invención es la inmortalización de cualquiera de las fracciones de células aisladas descritas en lo anterior (por ejemplo linfocitos B, células plasmáticas, células de memoria o linfocitos T). La inmortalización se puede llevar a cabo, por ejemplo, con virus de Epstein-Barr (Traggiai, E., *et al.*, 2004. *Nat Med* 10, 871-875) antes de la distribución celular. De manera alternativa, las células solas aisladas se pueden inmortalizar y expandir antes de transcripción inversa. Traggiai *et al.*, *Nat Med.* Agosto 2004; 10(8):871-5.

Una característica adicional de la presente invención es la distribución de una población de células deseadas (por ejemplo células de hibridoma, líneas de células de las líneas de linfocitos B o de linfocitos T, células de sangre entera, células de médula ósea, células mononucleares, leucocitos, linfocitos B, células plasmáticas, linfocitos B

específicos de antígeno, linfocitos B de memoria, linfocitos T, linfocitos T específicos para péptido/MHC o linfocitos T de memoria), individualmente, en una pluralidad de recipientes con el fin de obtener una población de células solas aisladas. Este aislamiento de células solas se refiere a la separación física de las células de una población de células de manera tal que un solo recipiente contiene una sola célula, o una microdisposición, chip o matriz de gel se carga de manera que produce células solas. Las células se pueden distribuir directamente en multitud de recipientes tales como disposiciones de recipientes solos, por dilución limitante. Los recipientes solos utilizados en la presente invención preferiblemente son aquellos aplicables en PCR (por ejemplo tubos de PCR y placas de PCR de 96 pocillos o 384 pocillos o disposiciones más grandes de recipientes). No obstante se pueden utilizar también otros recipientes. Cuando se distribuyen células solas en una gran cantidad de recipientes solos (por ejemplo placas de 384 pocillos), se obtiene una población de células solas. Dicha distribución se puede realizar, por ejemplo, al suministrar un volumen en un recipiente solo que en promedio abarca una concentración de células de uno, 0,5 o 0,3 células, obteniendo así recipientes que en promedio contienen una sola célula o menos. Dado que la distribución de células por dilución limitante es un evento estadístico, una fracción de los recipientes estará vacía, una fracción mayor contendrá una sola célula y una fracción menor contendrá dos o más células. Cuando están presentes dos o más células en un recipiente puede producirse cierto mezclado de las secuencias que codifican para la región variable entre las células presentes en el recipiente. No obstante, dado que es un acontecimiento menor, no afectará la utilidad general de la presente invención. De manera adicional, las combinaciones de secuencias que codifican para la región variable las cuales no poseen la afinidad de unión y especificidad deseada muy probablemente no se seleccionarán y por lo tanto se eliminan durante el procedimiento de cribado. Por lo tanto acontecimientos menores de mezclado no afectarán de manera significativa la biblioteca final de la presente invención.

Existen alternativas a la distribución de células por dilución limitante utilizando, por ejemplo, clasificadores de células tales como máquinas FACS o robots que se pueden programar para suministrar con precisión células solas en recipientes solos. Estas alternativas son preferibles, dado que son menos laboriosas y más eficientes para obtener de manera uniforme una distribución de células solas en recipientes solos.

Los procedimientos de enriquecimiento, clasificación y aislamiento descritos anteriormente se realizan de manera que la mayor parte de las células se mantienen intactas. La ruptura de las células durante el enriquecimiento y clasificación puede resultar en el mezclado de las secuencias que codifican para la región variable. No obstante, no se espera que esto sea un problema dado que se espera que la frecuencia de ruptura sea baja. El lavado y posible tratamiento con ribonucleasa de las células antes de la distribución en recipientes solos eliminará cualquier ARN que se pudiera haber fagado durante el procedimiento.

Además, cuando se consideran las descripciones anteriores de cómo distribuir células con el fin de obtener una población de células solas en una población de recipientes solos, no debe interpretarse como una característica que se requiere de manera absoluta en donde cada recipiente debe contener una sola célula. En vez de esto, indica que la mayor parte de los recipientes contienen células solas, por ejemplo, un número de recipientes con dos o más células es inferior al 25% de la cantidad total de células distribuidas, o incluso mejor, si es inferior al 10%.

Una característica adicional de la presente invención es el desempeño de una transcripción inversa utilizando una plantilla derivada de células distribuidas individualmente entre una pluralidad de recipientes.

Para el propósito de transcripción inversa (RT), de acuerdo con la presente invención, los ácidos nucleicos dentro de una sola célula que sirven como fuente de plantilla para la RT se considera que se derivan de una sola célula aunque no necesariamente han estado separados de los contenidos remanentes de la sola célula.

Cuando la distribución final de las células solas a sus recipientes solos se ha realizado, las células solas se pueden expandir con el fin de obtener una población de células isogénicas antes de transcripción inversa. Este procedimiento proporciona más ARNm para ser utilizado como plantilla lo cual puede ser importante si se va a amplificar y enlazar una diana rara. No obstante, las células deben permanecer genéticamente idénticas con respecto al gen diana durante la expansión. Las células aisladas o la población de células isogénicas se puede mantener intacta o lisada en la medida en que la plantilla para la transcripción inversa no se degrade. Preferiblemente, las células se lisan con el fin de facilitar la posterior transcripción inversa y amplificación por PCR.

En una realización diferente de la presente invención, el procedimiento RT-PCR de superposición-expresión múltiplex descrito o RT-PCR múltiplex seguido por enlace por ligación o recombinación también se puede utilizar sobre una plantilla derivada de una población genéticamente diversa de células las cuales no han sido separadas en recipientes solos, sino que permanecen juntas como un grupo de células. Este procedimiento se puede utilizar para la generación de bibliotecas combinatorias. Dicho enfoque no requerirá la distribución de células solas. No obstante, las células las cuales se pueden utilizar en este enfoque son las mismas que las descritas para la solución de sola célula, por ejemplo la población (grupo) de linfocitos B o linfocitos T clasificados. Cuando se realizan RT-PCR de superposición-extensión múltiplex de una sola etapa o RT-PCR múltiplex de una sola etapa seguido por enlace por ligación o recombinación de células, es preferible lisan las células antes de la reacción y, si se desea se puede aislar ARN total o ARNm del lisado. La sensibilidad de la RT-PCR de superposición de una expresión múltiplex de una sola etapa de la presente invención permite el uso de una cantidad muy baja de plantilla, por ejemplo una cantidad de plantillas que corresponden al lisado de una sola célula.

#### **Mezclas y diseño de cebadores**

Las mezclas de cebador de la presente invención comprenden por lo menos cuatro cebadores que forman conjuntos de cebador dos por dos, los cuales son capaces de amplificar por lo menos dos secuencias diana diferentes de interés. Las mezclas de dos o más de tales conjuntos de cebador constituyen una mezcla de cebador múltiplex. Preferiblemente, una mezcla múltiplex comprende por lo menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 ó 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 ó 150 conjuntos de cebador (pares de cebador). En particular, para la amplificación de las secuencias que codifican para la región variable, puede un conjunto de cebador individual dentro de la mezcla de cebador múltiplex constituir varios más de dos cebadores.

Preferiblemente, un conjunto de cebador individual comprende por lo menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280 o 300 cebadores. Preferiblemente, el número total de cebadores en una mezcla de cebador múltiplex es por lo menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 45, 50, 60, 70, 80, 90, 100, 125, 150 ó 200 y como máximo desde 225, 250, 275, 300, 325, 350, 375 ó 400 cebadores.

Todos los cebadores de la presente invención comprenden una región específica de genes, y preferiblemente todos los cebadores están equipados adicionalmente con una cola de cebador en el extremo 5' del cebador, es decir, secuencias no codificantes del extremo 5' las cuales se fusionan al extremo 3' de la parte de cebador específica del gen. Dicha cola de cebador tiene aproximadamente de 6 a 50 nucleótidos de longitud, pero también puede ser más grande, si se desea. Después de la amplificación se agregan las colas de cebador a las secuencias diana.

Las colas de cebador de la presente invención son, por ejemplo, colas de clonación y colas de enlace tales como colas adaptadas para enlace por ligación, colas adaptadas para enlace por recombinación o colas de superposición-extensión.

Las colas de clonación pueden ser de 6 a 20 nucleótidos de largo o más grandes y comprenden sitios de restricción y/o sitios de recombinación los cuales son útiles para la introducción del producto enlazado en un vector apropiado.

Para permitir el enlace por ligación, los conjuntos de cebador de la mezcla de cebador múltiplex se diseñan de tal forma que una parte (uno o varios cebadores directos o inversos) del primer conjunto de cebadores está equipado con una cola de enlace que contiene un sitio de restricción que, tras la ruptura, será compatible con el sitio de restricción localizado en la cola de enlace de una parte del segundo conjunto de cebador. Para el enlace de dos o más secuencias diana, la segunda parte del segundo conjunto de cebador está equipada con un sitio de restricción que, tras la ruptura, será compatible con un sitio de restricción que se localiza en una parte del tercer conjunto de cebador. Este segundo sitio de restricción localizado en el segundo conjunto de cebador debe ser no compatible con el del primer conjunto de cebador. Se pueden enlazar un número considerable de secuencias diana al diseñar conjuntos de cebadores de esta manera. Los sitios de restricción con una frecuencia baja o sin presentación en las secuencias diana son los que se deben seleccionar. Además, es preferible que los sitios de restricción compatibles no sean idénticos, de manera que el sitio de ligación se vuelva resistente a la escisión para las enzimas de restricción particulares utilizadas. Esto activará la reacción hacia el enlace de una secuencia diana una con la secuencia diana dos, puesto que el enlace entre secuencias diana idénticas se podrán escindir por medio de las enzimas de restricción. Los pares adecuados de sitios de restricción son, por ejemplo, SpeI con XbaI (de manera alternativa, NheI o AvrII pueden sustituir uno o ambos de estos), NcoI con BspHI, EcoRI con MfeI o PstI con NsiI. Para enlace, se puede localizar por ejemplo SpeI en la secuencia diana uno, se puede localizar XbaI en la secuencia diana dos, se puede localizar NcoI en el otro extremo de la secuencia diana dos y BspHI en la secuencia diana tres, y así sucesivamente. Para simplificar el procedimiento más, es una ventaja si las enzimas de restricción funcionan en el mismo tampón.

Para permitir el enlace por recombinación de los conjuntos de cebador de una mezcla de cebador múltiplex puede diseñarse, por ejemplo, como se ilustra en el artículo de Chapal (1997 BioTechniques 23, 518-524).

Para permitir que el enlace de las secuencias nucleótidas de interés en la misma etapa que la amplificación por PCR múltiplex, las colas adaptadas para PCR por superposición-expresión se agregan a por lo menos un cebador de cada conjunto de cebador de la mezcla de cebador múltiplex, generando así una mezcla de cebador de superposición de una extensión múltiplex.

Las colas de superposición de una extensión típicamente son más grandes, que varían de 8 a 75 nucleótidos de longitud y pueden contener sitios de restricción o sitios de recombinación los cuales permiten la inserción subsiguiente de elementos reguladores tales como promotores, sitios de unión ribosómica, secuencias de terminación o secuencias enlazantes tales como en un scFv. La cola de superposición-extensión también puede contener un codón de detención si esto se desea así. Generalmente existen tres tipos de colas de superposición-extensión, como se ilustran en la figura 1 del documento WO 2005/042774. En el tipo I, las colas de superposición-extensión de los conjuntos de cebadores se superponen únicamente entre sí. No necesariamente la totalidad de los nucleótidos de dos colas de superposición-extensión son complementarias entre sí. En un aspecto de la presente invención, los nucleótidos complementarios representan entre el 60% y 85% de la cola de superposición-extensión. En las colas de superposición-extensión del tipo II, 4 a 6 de los nucleótidos 5' son complementarios para la región específica de gen de la secuencia diana adyacente. En las colas de superposición-extensión tipo III, la totalidad de la superposición es complementaria a la secuencia diana adyacente. Las colas de superposición-extensión del tipo I y II se prefieren cuando los elementos reguladores y similares posteriormente se van a insertar entre las secuencias diana enlazadas. Las colas de superposición-extensión del tipo II se prefieren si las secuencias diana se van a enlazar por un enlazante definido como se observa con scFv. Las colas de superposición-extensión tipo III se prefieren si las secuencias diana que se van a enlazar están en marco entre sí. El diseño de colas de superposición-extensión depende de las características de secuencia tales como la longitud, contenido relativo de GC (% de GC), presencia de los sitios de restricción, palíndromos, temperatura de fusión, la parte específica de genes en la cual están acoplados, etc. Las colas de superposición-extensión deben ser de una longitud de entre 8 y 75 nucleótidos, preferiblemente tienen una longitud de 15 a 40 nucleótidos. Incluso de manera más preferible tienen una longitud de 22 a 28 nucleótidos. El uso de colas de superposición-extensión muy grandes (50 a 75 nucleótidos) puede favorecer el enlace de los productos elaborados por cada conjunto de cebador. No obstante, la proporción entre la longitud de la cola de superposición-extensión y la región específica de gen probablemente necesitará ajustarse cuando se utilicen colas de superposición-extensión muy grandes. La preferencia de % de GC depende de la longitud de la cola de superposición-extensión. Puesto que las colas más cortas tienen un área más corta cuando son complementarias, necesitan un % GC superior para soportar la interacción en comparación con colas más grandes. Otros principios de diseño de cebador deben observarse de igual manera, por ejemplo, la dimerización de cebador y la formación de horquillas debe minimizarse. Tampoco debe permitirse el acoplamiento por cebado en falso. Además, se sabe que la ADN polimerasa Taq con frecuencia agrega una adenosina (A) en el extremo 3' de una

cadena de ADN recién sintetizada y esto se puede albergar en un diseño de cola de superposición-extensión al permitir que las colas de superposición-extensión acomoden una adición A que no está en la plantilla 3'.

La selección de cebadores que tengan la cola de enlace, por ejemplo, la cola de superposición-extensión, cola adaptada para enlace por ligación o cola adaptada para enlace por recombinación, define el orden y la dirección del enlace de las secuencias diana. No es esencial para la invención si es el cebador directo o el cebador inverso de un conjunto de cebador o posiblemente ambos cebadores, directo o inverso, que estén equipados con la cola de enlace. No obstante, debe otorgarse alguna consideración a esto de cualquier manera puesto que el orden y dirección de las secuencias diana en el producto final puede ser de relevancia, por ejemplo, para la inserción de elementos reguladores tales como promotores y secuencias de terminación o para el enlace en marco de las secuencias diana individuales.

Para el enlace de dos secuencias nucleótidas de interés, la cola de enlace se puede agregar ya sea al cebador inverso o al cebador directo de cada conjunto de cebador utilizado para la amplificación por PCR de cada secuencia diana.

La presente invención ilustra la adición de colas de superposición-extensión y colas adaptadas para enlace por ligación a los cebadores directos mVH y mVK de cada conjunto. Esto resulta en una dirección de enlace de los productos que es 5' a 5' (cabeza a cabeza y bidireccional). No obstante, las colas de enlace también se pueden agregar al cebador inverso de cada conjunto. Esto resulta en una dirección de enlace del producto que es 3' a 3' (cola a cola y bidireccional). Una tercera opción es agregar las colas de enlace al cebador inverso del primer conjunto de cebadores y al cebador directo del segundo conjunto de cebadores, o viceversa. Esto resulta en una orientación 3' a 5' (cabeza a cola y unidireccional).

Cuando se enlazan más de dos secuencias nucleótidas de interés parte de los conjuntos de cebadores necesitan tener colas de enlace en los cebadores tanto directo como inverso, de manera que una cola sea complementaria a una cola del conjunto de cebador precedente y que la otra cola sea complementaria a uno de los cebadores del conjunto de cebadores subsiguiente. Este principio es válido para la totalidad de los conjuntos de cebadores que amplifican secuencias diana que van a estar enlazadas entre dos secuencias diana adicionales.

El diseño de la parte de cebador específica de gen generalmente debe cumplir las reglas de diseño de cebador conocidas tales como minimización de dimerización de cebador, formación de horquilla y reasociación no específica. Además, deben evitarse cuando sea posible nucleótidos G o C múltiples así como las bases 3'. La temperatura de fusión ( $T_m$ ) de las regiones específicas de gen en un conjunto de cebador preferiblemente deben ser iguales entre sí más/menos 5°C. En la presente invención, son deseables valores  $T_m$  entre 45°C y 75°C y valores  $T_m$  de aproximadamente 60°C son óptimos para la mayor parte de las aplicaciones. Ventajosamente, el diseño de cebador inicial se puede ayudar por medio de programas de ordenador desarrollados para esta tarea. No obstante, los diseños de cebador generalmente necesitan pruebas de laboratorio y optimización sistemática. Esto se puede llevar a cabo, por ejemplo, al analizar el tamaño, el polimorfismo de longitud de fragmentos de restricción (RFLP) y el secuenciado de los productos de amplificación obtenidos utilizando los conjuntos de cebador. El uso de posiciones degeneradas dentro de cebadores es una solución útil cuando se amplifican secuencias con regiones variables o cuando se busca por miembros de familia nuevos que pertenezcan a una clase de proteínas especificada. Los números de posiciones degeneradas también pueden requerir optimización.

Una característica de la presente invención, son las mezclas de cebadores constituidas por al menos dos conjuntos de cebadores que son capaces de cebar la amplificación y promover el enlace de por lo menos dos secuencias nucleótidas de interés. Las mezclas de cebadores de la presente invención son capaces de cebar la amplificación de por lo menos dos subunidades o dominios a partir de proteínas heteroméricas, por ejemplo que pertenecen a la clase de enzimas, inhibidores, proteínas estructurales, toxinas, proteínas de canal, proteínas G, proteínas receptoras, proteínas de superfamilia de inmunoglobulina, proteínas de transporte, etc., preferiblemente inmunoglobulinas.

Una característica adicional de la presente invención es una mezcla de cebador de superposición-extensión múltiples que comprende conjuntos de cebador en donde por lo menos un miembro del conjunto de cebador de cada conjunto de cebadores comprende una cola de superposición-extensión capaz de hibridizar con la cola de superposición-extensión de un miembro del conjunto de cebador de un segundo conjunto de cebadores.

Las colas de superposición-extensión permiten el enlace inmediato de los nucleótidos de interés durante la amplificación por PCR mediante cola de superposición-extensión múltiple al equipar cada producto individual que surge de los conjuntos de cebador con una cola que es complementaria a un producto adyacente. No obstante, esto no significa que el enlace necesariamente se produzca durante esta primera amplificación por PCR. En base en los ajustes de la reacción, la mayor parte del enlace real se puede realizar durante una amplificación adicional con los cebadores externos de la primera amplificación PCR (amplificación PCR múltiple).

Una característica adicional de la presente invención es un conjunto de cebador diseñado para amplificar una familia de secuencias nucleótidas que contienen secuencias que codifican para la región variable. Los ejemplos de las familias son las cadenas ligeras  $\kappa$  (por ejemplo VK1-19 en ratones), las cadenas ligeras  $\lambda$  (por ejemplo VL1-8 en ratones) y las cadenas pesadas variables (por ejemplo VH1-15 en ratones) a partir de inmunoglobulinas y las regiones variables  $\alpha$ ,  $\beta$ ,  $\gamma$  o  $\delta$  de TcR. Un conjunto de cebador para la amplificación de una familia de secuencias nucleótidas que contienen secuencias que codifican para la región variable con frecuencia comprenden una pluralidad de cebadores en donde varios cebadores pueden ser cebadores degenerados. La amplificación de familias de secuencias que codifican para la región variable de la cadena ligera de inmunoglobulina se realiza, por ejemplo, utilizando un conjunto de cebador constituido por una pluralidad de cebadores complementarios al extremo 5' de la región variable de la cadena  $\kappa$  (uno o varios cebadores mVK) o la secuencia líder  $\kappa$  y/o la cadena  $\lambda$  o la secuencia líder  $\lambda$  (cebadores directos) junto con la región constante  $\kappa$  (cebadores mKappara1) y/o cebadores  $\lambda$  (cebadores inversos) o una pluralidad de tales cebadores. De manera alternativa, los cebadores de la región de

unión de cadena ligera se pueden utilizar como cebadores inversos en vez de los cebadores de la región constante. De manera alternativa se pueden utilizar cebadores directos que se reasocian en la región UTR que precede a la secuencia líder de la cadena ligera variable. Igualmente, las familias de las secuencias que codifican para la región variable de cadena pesada de inmunoglobulina se pueden amplificar con un conjunto de cebador que utiliza diversas combinaciones de cebador. Por ejemplo, una pluralidad de cebadores complementarios al extremo 5' de la región variable de cadena pesada (uno o varios cebadores mVH) o la secuencia líder de esta región (cebadores directos) junto con una pluralidad de cebadores de la región de unión de la cadena pesada o uno o varios cebadores de la región constante de la cadena pesada (cebadores inversos). El cebador mCH puede ser específico para el isotipo y en principio se puede utilizar cualquier cebador mCH, también uno que pueda resultar en una cadena pesada de longitud completa. Preferiblemente, se utiliza un cebador mCH que no amplifica la cadena pesada de longitud completa para permitir la adición de una región constante de cadena pesada humana. De manera alternativa, se pueden utilizar cebadores directos que se reasocian en la región UTR que precede a la secuencia líder de la cadena pesada variable.

El uso de cebadores directos que se reasocian en la secuencia líder en vez del extremo 5' de la región variable es particularmente útil si la hibridación cruzada debida a un alto grado de similitud de secuencia se observa para los cebadores de la región variable. Las mutaciones debidas a hibridación cruzada utilizando cebadores líderes se eliminarán de la proteína final debido a que las secuencias líder se escinden durante el procesamiento de proteínas dentro de las células.

Una característica de la presente invención son los cebadores los cuales se reasocian en el extremo 3' de la secuencia codificante líder que precede a una secuencia que codifica para la región variable y su uso para la amplificación de secuencias que codifican para la región variable.

En una realización de la presente invención, la mezcla de cebador por superposición-extensión múltiplex utilizada para PCR por superposición-extensión múltiplex y posiblemente para la etapa de transcripción inversa también, comprende:

a) por lo menos un cebador mKappar1 o hmJK complementario a la cadena directa de una secuencia que codifica para la región de cadena ligera de inmunoglobulina;

b) por lo menos un cebador mVK complementario a la cadena antisentido de una secuencia que codifica para la región variable de la cadena ligera de inmunoglobulina o la secuencia líder de la región variable de la cadena ligera y capaz de formar un conjunto de cebador con uno o varios de los cebadores en la etapa a);

c) por lo menos un cebador mCHrev1, mHCrev1-ext o mJH complementario a la cadena directa de una secuencia que codifica para el dominio de cadena pesada de inmunoglobulina; y

d) por lo menos un cebador mVH complementario a la cadena antisentido de una secuencia que codifica para la región variable de la cadena pesada de inmunoglobulina o la secuencia líder de la región variable de cadena pesada y capaz de formar un conjunto de cebador con uno o varios de los cebadores en la etapa c).

En una realización de la presente invención, los cebadores de la cadena ligera están adaptados para amplificar las secuencias que codifican para la región variable de la cadena ligera tanto  $\kappa$  como  $\lambda$ .

En una realización adicional de la presente invención, los cebadores mVK de inmunoglobulina tienen cola de enlace, preferiblemente en forma de colas de superposición-extensión complementarias. Esto genera secuencias que codifican para la región variable que están enlazadas de manera cabeza a cabeza. Para el enlace de secuencias que codifican para la región variable de una manera cabeza a cola, ya sea el cebador mKappar1 y mVH que contienen las colas de enlace o los cebadores mVK y mCHrev1 que contienen las colas de enlace, preferiblemente en forma de colas de superposición-extensión complementarias. Para el enlace de las secuencias que codifican para la región variable de manera cola a cola, los cebadores mCH y mKappar1 contienen colas de enlace, preferiblemente en forma de colas de superposición-extensión complementarias.

De manera preferencial, las mezclas de cebador múltiplex, que incluyen mezclas de cebador por superposición-extensión múltiplex, de la presente invención comprenden dos conjuntos de cebadores. De esta manera, una mezcla de cebador múltiplex comprende por lo menos cuatro cebadores diferentes. En un aspecto adicional de la presente invención una mezcla de cebador múltiplex comprende más de cuatro cebadores diferentes. Una mezcla de cebador múltiplex de la presente invención se utiliza para la amplificación de secuencias diana en un recipiente único. Por ejemplo, las regiones variables de la cadena  $\kappa$ ,  $\lambda$  y pesada se pueden amplificar en el mismo recipiente.

La presente invención también abarca cebadores para una amplificación por PCR adicional de los productos enlazados obtenidos por RT-PCR múltiplex seguido por enlace por ligación o recombinación o por RT-PCR de superposición-extensión múltiplex. Esta amplificación por PCR adicional se puede realizar utilizando una mezcla de cebador adaptada para amplificar las secuencias diana enlazadas. Dicha mezcla de cebador puede comprender los cebadores externos o la mezcla de cebadores múltiplex o una mezcla de cebador de superposición-extensión múltiplex, lo que significa que los cebadores que se reasocian al extremo 5' más externo y al extremo 3' de la cadena con sentido de las secuencias nucleótidas enlazadas, por lo que se habilita selectivamente la amplificación de la totalidad del producto enlazado. Este procedimiento generalmente sirve para incrementar la cantidad de producto enlazado que se obtiene de RT-PCR múltiplex seguido por enlace por ligación o recombinación o a partir de RT-PCR de superposición-extensión múltiplex.

De manera alternativa, un conjunto de cebador el cual se incrusta en comparación con los cebadores externos utilizados en RT-PCR múltiplex primaria o en una reacción RT-PCR de superposición-extensión múltiplex se pueden utilizar para amplificación adicional de las secuencias nucleótidas enlazadas. En la presente invención, dicho conjunto de cebador se denomina un conjunto de cebador incrustado. El diseño de los cebadores incrustados

5 generalmente observa las mismas reglas de diseño que para los cebadores específicos de gen descritos previamente, excepto que ceban parcial o totalmente 3' respecto a la posición de reasociación de los cebadores externos utilizados en RT-PCR múltiplex o RT-PCR de superposición-extensión múltiplex. El producto que resulta de una PCR incrustada por lo tanto puede ser más corto que el producto enlazado que se obtiene por RT-PCR múltiplex seguido por enlace por ligación o recombinación o por RT-PCR de superposición-extensión múltiplex. Además de incrementar la cantidad de producto enlazado, la PCR incrustada sirve además para incrementar la especificidad general, especialmente de la tecnología RT-PCR de superposición-extensión múltiplex. No obstante, debe hacerse notar que no todas las mezclas de cebador múltiplex/mezclas de cebador de superposición-extensión múltiplex que se han descrito previamente son adecuadas para combinación con un conjunto de cebador incrustado cuando se realiza la amplificación adicional. En tales casos, los cebadores externos de la mezcla de cebador múltiplex/mezcla de cebador por superposición-extensión múltiplex se puede utilizar para la amplificación adicional o una PCR semi-incrustada se puede aplicar como se desvela posteriormente.

En una realización de la presente invención se utiliza una mezcla de cebadores  $J_L$  y  $J_H$  como cebadores incrustados para la amplificación adicional de las secuencias que codifican para la región variable de inmunoglobulina enlazada.

15 Los conjuntos de cebador incrustado de la presente invención también pueden estar constituidos por uno o varios cebadores inversos (o directos) externos de la primera mezcla de cebador múltiplex/mezcla de cebador por superposición-extensión múltiplex y un segundo cebador incrustado que ceba 3' respecto a la posición de reasociación de uno o varios de los cebadores externos directos (o inversos) de la primera mezcla de cebador múltiplex/mezcla de cebador de superposición-extensión múltiplex. El uso de los conjuntos de cebadores para una amplificación por PCR adicional generalmente se conoce como una PCR semi-incrustada. La PCR semi-incrustada se puede aplicar, por ejemplo, si es difícil diseñar un cebador incrustado en una región específica, por ejemplo, para las secuencias de región variable, debido a que el cebador puede necesitar reasociarse en regiones determinadoras de complementariedad (las CDR, por sus siglas en inglés). Además, se puede utilizar la PCR semi-incrustada cuando es deseable mantener un extremo de las secuencias enlazadas intacto, por ejemplo con propósitos de clonación.

#### **Optimización de PCR por superposición-extensión múltiplex**

30 Los parámetros de la etapa de PCR de superposición-extensión múltiplex tanto del procedimiento de dos etapas como el de una sola etapa se pueden optimizar en varios parámetros (véase, por ejemplo, Henegariu, O. et al. 1997. BioTechniques 23, 504-511; Markoulatos, P. et al. 2002. J. Clin. Lab. Anal. 16, 47-51). Generalmente se aplican los mismos parámetros de optimización para RT-PCR múltiplex, aunque la relación entre los cebadores externo e interno es menos importante para una reacción como tal.

##### **a. Concentración de Cebador**

35 La concentración de los cebadores que transportan la cola de superposición-extensión (por ejemplo los cebadores  $V_H$  y  $V_L$ ) preferiblemente es menor que la concentración de los cebadores externos sin la cola de superposición-extensión (por ejemplo los cebadores  $J_H$  y  $\kappa$ ).

40 Si una de las secuencias diana se amplifica con una eficiencia menor que las otras, por ejemplo como resultado de un % mayor de GC, es posible igualar la eficacia de amplificación. Esto se puede llevar a cabo utilizando una concentración mayor del conjunto de cebador la cual media la amplificación con baja eficiencia o disminuyendo la concentración de los otros cebadores. Por ejemplo, las secuencias que codifican para las regiones variables de cadena pesada tienden a tener un % de GC mayor y por lo tanto una eficiencia menor de amplificación en comparación con las regiones variables de cadena ligera. Esto indica hacia el uso de cebadores  $V_L$  en una concentración menor en comparación con los cebadores  $V_H$ .

45 Además, cuando se utiliza una gran cantidad de cebadores, la concentración de cebador total puede ser un problema. El límite superior se determina experimentalmente por experimentos de titulación. Para el sistema de PCR "AmpliTaQ Gold" de Applied Biosystems, se ha encontrado que el límite superior es de 1,1  $\mu\text{M}$  de la concentración total de oligonucleótido, para otros sistemas, no obstante puede ser tan elevada como 2,4  $\mu\text{M}$ . Dicho límite superior de la concentración oligonucleotídica total influye en la concentración máxima de cebadores individuales. Si la concentración de cebador individual es demasiado baja, es probable que provoque una sensibilidad pobre de PCR.

50 También se ha encontrado que la calidad de los cebadores oligonucleotídicos es importante para la PCR de superposición-extensión múltiplex. Los oligonucleótidos purificados por HPLC han producido los mejores resultados.

##### **b. Condiciones de ciclado por PCR:**

De manera preferencial, las condiciones de ciclado son las siguientes:

Desnaturalización:	10-30 s	94°C	
Reasociación:	30-60 s	50-70°C	Aproximadamente 5°C por debajo de los cebadores
$T_m$			
Extensión:	1 min x EPL	65-72°C	EPL es el producto esperado, longitud en kb.
Número de ciclo:	30-80		

Extensión final: 10 min 65-72°C

Para la RT-PCR de superposición-extensión múltiplex de una sola etapa se construyen las siguientes etapas en el programa de ciclado antes del ciclado por amplificación indicado antes:

Transcripción inversa: 30 min 42-60°C Estas condiciones también se utilizan cuando se realiza transcripción inversa separada

Activación de polimerasa: 10-15 min 95°C Las polimerasas de inicio en caliente son favorables en RT-PCR de una sola etapa. Activación de acuerdo con el fabricante.

5

Es posible optimizar todos estos parámetros. Especialmente es importante la temperatura de reasociación. Por lo tanto, inicialmente todos los conjuntos de cebador individuales que van a constituir la mezcla de cebador final se deben probar por separado con el fin de identificar la temperatura y tiempo de reasociación óptimos así como los tiempos de elongación y desnaturalización. Esto proporcionará una buena idea acerca del intervalo dentro del cual se pueden optimizar estos parámetros para la mezcla de cebador por superposición-extensión múltiplex.

10

Los problemas con una sensibilidad pobre de PCR, por ejemplo debido a una concentración de cebador baja o una concentración de plantilla se pueden superar utilizando una gran cantidad de ciclos térmicos. Un número alto de ciclos térmicos estará constituido por 35 y 80 ciclos, de manera preferible aproximadamente 40 ciclos.

15

Además, tiempos de extensión más grandes pueden mejorar los procedimientos de PCR de superposición-extensión múltiplex. Los tiempos de extensión grandes están constituidos por 1,5-5 min x EPL, en comparación con una extensión normal de 1 min.

#### c. Uso de Adyuvantes

20

Las reacciones de PCR múltiplex se pueden mejorar de manera significativa mediante el uso de un aditivo de PCR tal como DMSO, glicerol, formamida o betaína lo cual relaja el ADN y por lo tanto facilita la desnaturalización de la plantilla.

#### d. dNTP y MgCl<sub>2</sub>

25

La calidad y concentración del trifosfato de desoxinucleósido (dNTP) es importante para la PCR por superposición-extensión múltiplex. La mejor concentración de dNTP está entre 200 y 400 μM de cada dNTP (dATP, dCTP, dGTP y dTTP), por encima de lo cual la amplificación se inhibe rápidamente. El hecho de disminuir las concentraciones de dNTP (100 μM de cada dNTP) es suficiente para obtener amplificación por PCR. Los concentrados de dNTP son sensibles a los ciclos de recalentamiento/congelamiento. Después de tres a cinco de dichos ciclos, la PCR múltiplex con frecuencia no funciona bien. Para evitar tales problemas se pueden elaborar alícuotas pequeñas de dNTP y se mantienen congeladas a -20°C.

30

La optimización de la concentración de Mg<sup>2+</sup> es crítica dado que la mayor parte de las ADN polimerasas son enzimas que dependen de magnesio. Además de la ADN polimerasa, los cebadores de plantilla de ADN y los dNTP que unen Mg<sup>2+</sup>. Por lo tanto, la concentración óptima de Mg<sup>2+</sup> dependerá de la concentración de dNTP, el ADN plantilla y la composición del tampón de muestra. Si los cebadores y/o los tampones de ADN plantilla contienen quelantes tales como EDTA o EGTA, el óptimo aparente de Mg<sup>2+</sup> se puede alterar. Una concentración excesiva de Mg<sup>2+</sup> estabiliza ADN de cadena doble y evita la desnaturalización completa de ADN, lo que reduce el rendimiento. Mg<sup>2+</sup> en exceso también puede estabilizar reasociación espuria de cebador a sitios de plantilla incorrectos, por lo que disminuye la especificidad. Por otra parte, una concentración inadecuada de Mg<sup>2+</sup> reduce la cantidad de producto. Un buen equilibrio entre dNTP y MgCl<sub>2</sub> es aproximadamente 200 a 400 μM de dNTP (de cada uno) a 1,5 a 3 mM de MgCl<sub>2</sub>.

35

#### e. Concentración de Tampón de PCR

40

Generalmente los tampones basados en KCl son suficientes para PCR de superposición-extensión múltiplex; no obstante, los tampones basados en otros componentes tales como (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, MgSO<sub>4</sub>, Tris-HCl, o combinaciones de los mismos también se pueden utilizar para que funcionen con PCR de superposición-extensión múltiplex. Los pares de cebador involucrados en la amplificación de productos más grandes funcionan mejor a menores concentraciones de sal (por ejemplo 20 a 50 mM de KCl), mientras que los pares de cebadores involucrados en la amplificación de productos cortos funcionan mejor a concentraciones de sal más elevadas (por ejemplo KCl 80 a 100 mM). El hecho de incrementar la concentración de tampón a 2X en vez de 1X puede mejorar la eficiencia de la reacción múltiplex.

45

#### f. ADN polimerasa

50

La presente invención se ilustra con polimerasa Taq. De manera alternativa otros tipos de ADN polimerasa resistentes al calor incluyen, por ejemplo, Pfu, Phusion, Pwo, Tgo, Tth, Vent, Deep-vent, las cuales pueden ser utilizadas. Las polimerasas sin o con actividad de exonucleasa 3' a 5' también se pueden utilizar solas o combinadas entre sí.

#### Vectores y Bibliotecas

El enlace de las secuencias nucleótidas de interés de acuerdo con la presente invención produce un segmento nucleótido que comprende las secuencias nucleótidas enlazadas que codifican para las regiones variables de inmunoglobulinas. Además, las bibliotecas de dichas secuencias de ácido nucleico enlazadas se producen por los procedimientos de la presente invención, en particular bibliotecas de secuencias que codifican para la región variable no humana unidas o empalmadas a las secuencias de la región constante humana (cadena pesada y ligera).

Una característica de la presente invención es la inserción de un segmento que contiene secuencias nucleótidas enlazadas de interés o una biblioteca de secuencias nucleótidas enlazadas de interés, generadas por un procedimiento de la presente invención, en vectores adecuados. Las bibliotecas pueden ser bibliotecas combinatorias o de manera más preferible, bibliotecas de pares análogos de secuencias que codifican para la región variable. Los sitios de restricción generados por los cebadores externos, cebadores incrustados o cebadores semi-incrustados se diseñan preferiblemente para que coincidan con los sitios de restricción apropiados del vector de elección. Las secuencias de ácido nucleico enlazadas de interés también se pueden insertar en vectores por recombinación, si uno de los cebadores semi-incrustados, incrustados o cebadores externos está equipado con un sitio de recombinación adecuado y el vector de elección contiene uno también.

Básicamente, no hay limitaciones respecto a los vectores que se puedan utilizar como vehículos de los productos generados por uno de los procedimientos de enlace RT-PCR múltiplex de la presente invención. Los vectores de elección pueden ser aquellos adecuados para amplificación y expresión en células que incluyen, por ejemplo, células de bacterias, levaduras, otros hongos, células de insectos, células vegetales o células de mamífero. Los vectores se pueden utilizar para facilitar etapas de clonación adicionales, el desplazamiento entre sistemas de vector, la presentación del producto insertado en el vector, la expresión del producto insertado y/o la integración en el genoma de una célula hospedadora.

Los vectores de clonación y lanzadera son preferiblemente vectores bacterianos. No obstante también se pueden aplicar otros tipos de vectores en los procedimientos de clonación y desplazamiento.

Los vectores de presentación pueden ser, por ejemplo, vectores de fago o vectores fagémidos que se originan a partir de la clase de los bacteriófagos filamentosos fd, M13 o f1. Tales vectores son capaces de facilitar la presentación de una proteína que incluye, por ejemplo una proteína de unión o un fragmento de la misma sobre la superficie de un bacteriófago filamentosos. Los vectores de presentación adecuados para presentación en ribosomas, ADN, células de levadura o células de mamífero también se conocen en la técnica. Estos comprenden, por ejemplo, vectores virales o vectores que codifican para proteínas quiméricas.

Los vectores de expresión existen para todas las especies mencionadas y se puede seleccionar uno dependiendo completamente de la proteína que se va a expresar. Algunos vectores de expresión adicionalmente son capaces de integrarse en el genoma de una célula hospedadora ya sea por integración aleatoria o por integración específica de sitio utilizando sitios de recombinación apropiados. Los vectores de expresión se pueden diseñar para proporcionar secuencias codificantes adicionales que, cuando el producto enlazado se injerta en marco con estas secuencias, permite la expresión de una proteína más grande, por ejemplo un anticuerpo monoclonal de longitud completa, cuando se introduce en una célula hospedadora apropiada. Esta inserción en marco también puede facilitar la expresión de proteínas quiméricas que facilitan la presentación en la superficie de un bacteriófago filamentosos o célula. En un sistema de presentación por bacteriófago, las secuencias nucleótidas enlazadas de interés se pueden insertar en marco a una secuencia que codifica para una proteína de recubrimiento tal como pIII o pVIII (Barbas, C.F. et al. 1991. Proc. Natl. Acad. Sci. USA 88, 7978-7982; Kang, A.S. et al. 1991. Proc. Natl. Acad. Sci. USA 88, 4363-4366).

En una realización de la presente invención, los segmentos individuales de las secuencias nucleótidas enlazadas de interés están constituidos por una secuencia que codifica para la región variable de cadena pesada de inmunoglobulina asociada con una secuencia que codifica para la región variable de cadena ligera de una especie, insertado en un vector que contiene una o varias secuencias que codifican para uno o más dominios constantes de inmunoglobulina humana, preferiblemente ambas regiones constantes de la cadena ligera y pesada humana. La inserción se realiza de manera que la región variable de la cadena pesada enlazada y/o las secuencias que codifican para la región variable de cadena ligera se insertan en marco en las secuencias que codifican para la región constante. La inserción puede generar, por ejemplo, un vector de expresión Fab o F(ab')<sub>2</sub>, un vector de expresión de anticuerpo de longitud completa o un vector de expresión que codifica para un fragmento de un anticuerpo de longitud completa. De manera preferible un vector como tal es un vector de expresión adecuado para expresión (por ejemplo vectores de *E. coli*, fagémido o de mamífero) y las secuencias que codifican para la cadena pesada de la región constante se seleccionan de las clases de inmunoglobulinas humanas IgG1, IgG2, IgG3, IgG4, IgM, IgA1, IgA2, IgD o IgE, permitiendo así la expresión de un Fab o un anticuerpo recombinante de longitud completa. Además de las secuencias que codifican para la cadena pesada constante, el vector también puede contener una secuencia que codifica para la cadena ligera constante que se selecciona de las cadenas humanas  $\lambda$  o  $\kappa$ . Esto se prefiere en la generación de anticuerpos quiméricos como en las secuencias nucleótidas enlazadas en estos casos únicamente codifican para las secuencias que codifican para la región variable de inmunoglobulina (las Fv) a partir de especies no humanas.

En una realización alternativa, una o varias de las regiones constantes humanas se dividen o enlazan a las regiones variables no humanas en una etapa del procedimiento de amplificación molecular por adición a los recipientes de una región constante humana que proporciona una superposición con la secuencia no humana y cebadores apropiados que aseguran la amplificación de una o varias de las regiones tanto variable como constante en marco. De esta manera, la cadena  $\kappa$  o  $\lambda$  constante humana se puede agregar y/o se puede agregar la cadena pesada constante humana. Mediante el uso de este procedimiento no hay necesidad de proporcionar un sitio de restricción dentro de la secuencia codificante, lo cual es una ventaja. En otra realización de la presente invención, los segmentos individuales de las secuencias nucleótidas enlazadas están constituidos de una secuencia que codifica para la región variable de la cadena  $\alpha$  de TcR asociada con una secuencia que codifica para la región variable de la

- cadena  $\beta$  o una secuencia que codifica para la región variable de la cadena  $\gamma$  asociada con una secuencia que codifica para la región variable de la cadena  $\delta$ . Preferiblemente, estas secuencias enlazadas se insertan en un vector que contiene secuencias que codifican para uno o más dominios constantes TcR. La inserción se lleva a cabo de manera que las secuencias que codifican para la región variable enlazada insertada estén en marco con las secuencias que codifican para la región constante TcR correspondiente. En una realización adicional, el vector es un vector de expresión quimérico que comprende secuencias que codifican para una cremallera de leucina en marco con las regiones constantes TcR. Se ha demostrado que las construcciones incrementan la estabilidad de las TcR solubles (Willcox, B.E. et al. 1999, Protein Sci 8, 2418-2423).
- Las bibliotecas de pares análogos de la presente invención se pueden introducir en vectores por dos enfoques diferentes. En el primer enfoque, los pares análogos solos se insertan individualmente en un vector adecuado. Esta biblioteca de vectores se puede mantener separada o se puede acumular. En el segundo enfoque, todos los pares análogos se acumulan antes de la inserción del vector, seguido por la inserción en masa en vectores adecuados que generan una biblioteca acumulada de vectores. Dicha biblioteca de vectores comprende una gran diversidad de pares de secuencias que codifican para la región variable.
- Un aspecto de la presente invención es una biblioteca de anticuerpos con pares análogos de secuencias que codifican para la región variable enlazada. Preferiblemente, los anticuerpos individuales de la biblioteca comprenden una secuencia que codifica para la región variable de la cadena ligera de inmunoglobulina asociada con una secuencia que codifica para la región variable de la cadena pesada de una especie y regiones constantes humanas.
- Otra biblioteca preferida de pares análogos comprende secuencias que codifican para la región TcR enlazada en donde cada secuencia que codifica para la región TcR individual comprende una secuencia que codifica para la región variable de cadena  $\alpha$  asociada con una secuencia que codifica para la región variable de cadena  $\beta$  y/o una secuencia que codifica para la región variable de cadena  $\gamma$  TcR asociada con una secuencia que codifica para la región variable de la cadena  $\delta$ .
- Una realización de la presente invención es una biblioteca secundaria de pares análogos de secuencias que codifican para la región variable enlazada los cuales codifican para las especificidades de unión deseadas dirigidas contra una diana particular. Preferiblemente, estos pares análogos comprenden una región variable de cadena ligera de inmunoglobulina unida y secuencias que codifican para la región variable de cadena pesada, secuencias que codifican para la región variable de cadena  $\alpha$  TcR y la región variable de cadena  $\beta$  y/o una región variable de cadena  $\gamma$  TcR y secuencias que codifican para la región variable de cadena  $\delta$ .
- Una realización adicional es una biblioteca secundaria que se selecciona de una biblioteca original de pares análogos de secuencias que codifican para la región variable, como se desvela a través de la invención.
- Una realización preferida de la presente invención es una biblioteca o una biblioteca secundaria que codifica para pares análogos de inmunoglobulinas quiméricas de longitud completa que se seleccionan de las clases de inmunoglobulina humana IgA1, IgA2, IgD, IgE, IgG1, IgG2, IgG3, IgG4 o IgM.
- Otra característica preferida de la presente invención es una biblioteca o biblioteca secundaria que codifica para pares análogos solubles y estables de TcR.
- Una característica de la presente invención es la diversidad de las bibliotecas, las cuales están constituidas por 5, 10, 20, 50, 100, 1000,  $10^4$ ,  $10^5$  o  $10^6$  anticuerpos de pares análogos diferentes.
- En una realización adicional de la presente invención, dichas bibliotecas de pares análogos de secuencias que codifican para la región variable enlazada se pueden obtener por un procedimiento que comprende las etapas que se desvelan en la presente. Esta biblioteca también se puede denominar como biblioteca original.

### **Cribado y Selección**

- La biblioteca original de pares de secuencias que codifican para la región variable unida aislada de un donante, utilizando uno de los procedimientos de la presente invención se espera que representa una diversidad de proteínas de unión en las cuales algunas serán irrelevantes, es decir, no se unen a un objetivo deseado, en particular para bibliotecas combinatorias. Por lo tanto, la presente invención abarca el enriquecimiento y cribado, para una biblioteca secundaria que codifica para un subconjunto de diversidades de especificidades de unión dirigidas contra una diana particular.
- Para bibliotecas de pares análogos, la diversidad de la biblioteca se espera que represente la diversidad que se encuentra en el material donante, con únicamente un número menor de regiones variables unidas aleatoriamente. Así, puede no ser necesaria una etapa de enriquecimiento antes del cribado en búsqueda de afinidades de unión específicas de diana en una biblioteca constituida por pares análogos.
- En una realización adicional de la presente invención, el procedimiento de generación de una biblioteca de pares de secuencias que codifican para la región variable enlazada, comprende además generar una biblioteca secundaria al seleccionar un subconjunto de pares de secuencias de región variables enlazadas que codifican para proteínas de unión con una especificidad deseada hacia la diana. La selección de secuencias que codifican para la región variable enlazada también se denomina una biblioteca de pares análogos específicos para la diana.
- En una realización preferida de la presente invención, la biblioteca de pares análogos específicos para la diana de secuencias que codifican para una región variable se transfiere a un vector de expresión. El vector de expresión puede ser un vector de expresión mamífero, un vector de expresión de levadura, un vector de expresión de hongo, un vector de expresión vegetal, un vector de expresión bacteriano, dependiendo del tipo de célula utilizada para el cribado. Preferiblemente, el vector de expresión es mamífero.

Los análisis inmunológicos son generalmente adecuados para la selección de secuencias que codifican para la región variable de inmunoglobulina, específicas para diana. Los análisis son bien conocidos en la técnica y constituyen, por ejemplo, los análisis FMAT, FLISA, ELISPOT, ELISA, análisis de membrana (por ejemplo las transferencias Western), disposiciones sobre filtros o FACS. Los análisis se pueden realizar de manera directa, utilizando los polipéptidos producidos a partir de las secuencias que codifican para la región variable de inmunoglobulina. De manera alternativa, los inmunoanálisis se pueden realizar en combinación con, o después de procedimientos de enriquecimiento tales como presentación de fagos, presentación de ribosomas, presentación de superficie bacteriana, presentación de levadura, presentación de virus eucarióticos, presentación de ARN o presentación covalente (revisado en FitzGerald, K., 2000. *Drug Discov. Today* 5, 253-258). Tanto las bibliotecas de expresión Fab análogas como las bibliotecas de expresión de anticuerpos de longitud completa análogos se pueden someter a cribado, por lo que se genera una biblioteca secundaria de clones positivos. Los análisis de cribado y procedimientos de enriquecimiento también son adecuados para fragmentos Fv o scFv o bibliotecas combinatorias de regiones variables enlazadas.

Además del cribado inmunológico, una característica especial de la invención es que permite el uso de diversos tipos de cribado funcional para seleccionar clones secretantes de anticuerpo con las propiedades que se desean. Los análisis de cribado incluyen pero no se limitan a análisis de proliferación, análisis de inactivación de virus, análisis de destrucción de células, etc. Preferiblemente, los análisis funcionales se pueden llevar a cabo en un formato de alto rendimiento utilizando sobrenadantes de células transfectadas con vectores de expresión de la invención.

En una realización preferida de la presente invención, la selección de una biblioteca secundaria de pares análogos específicos para diana o pares combinatorios de secuencias que codifican para la región variable se realiza mediante el uso de un análisis de cribado de alto rendimiento. El análisis de cribado de alto rendimiento puede ser, pero no se limita a análisis de ELISA realizado con equipo semiautomatizado o completamente automatizado. También puede ser un análisis de membrana en el cual las bacterias se colocan robóticamente y se distribuyen en grupos sobre una membrana apropiada en la parte superior de placas de agar generando disposiciones de colonias que expresan moléculas que unen antígeno. Las moléculas se secretan a través de la membrana sobre una segunda membrana subyacente, recubierta con antígeno, la cual se puede desarrollar por separado y se puede utilizar para identificar clones que secreten moléculas que unen antígeno, hacia la diana que se desea (de Wildt, R.M.) et al. 2000. *Nat. Biotechnol.* 18, 898-994).

Cuando una biblioteca secundaria de pares análogos o pares combinatorios de clones que unen antígeno se ha seleccionado por una tecnología apropiada, es posible realizar un análisis adicional por secuenciado de ADN de la región variable de cadena ligera de inmunoglobulina enlazada y las secuencias que codifican para la región variable de cadena pesada. En primer lugar, el secuenciado de ADN proporcionará información acerca de la diversidad de biblioteca tal como el origen de la línea germinal, la distribución de la familia y la maduración dentro de las regiones de CDR. El análisis permitirá la selección de clones los cuales representan una diversidad amplia descartando clones repetidos. En segundo lugar, el secuenciado de ADN mostrará mutaciones introducidas durante el procedimiento de aislamiento.

Cuando se analizan secuencias que codifican para la región variable existen tres tipos de mutaciones a considerar cuando se determina si una mutación es aceptable: i) el tipo más frecuente de mutaciones que resultan de cebado cruzado, en donde un cebador de gen V, debido a similitudes de secuencia, ceba una secuencia de línea germinal a la cual no es totalmente homóloga. Los cambios introducidos son principalmente sustituciones de codones que se presentan de manera natural en una posición particular. Debido al alto grado de homología de secuencia entre las secuencias de gen V. Algunos de estos cambios pueden ser significativos, algunas veces sin contraparte natural. Tales cambios pueden afectar potencialmente la inmunogenicidad de la región variable al crear epitopes nuevos. Tales cambios se pueden identificar fácilmente y posteriormente se pueden reparar utilizando técnicas convencionales de biología molecular o los clones se pueden excluir de la biblioteca; ii) errores creados por la ADN polimerasa Taq los cuales son identificados más fácilmente en las secuencias que codifican para la región constante y pueden ser eliminados fácilmente. No obstante, las mutaciones inducidas por Taq por supuesto también estarán presentes en las secuencias que codifican para la región variable en donde son indistinguibles de las mutaciones somáticas que se producen de manera natural, las cuales también son el resultado de mutaciones aleatorias en las secuencias codificantes de la región variable. Considerando que las mutaciones son no sistemáticas y únicamente afectan pares particulares de maneras distintas, parece razonable descartar los cambios.

En una realización adicional de la presente invención, la biblioteca secundaria de los pares analizados de secuencia específicos de diana y posiblemente de la región variable de cadena ligera de inmunoglobulina enlazada y las secuencias que codifican para la región variable de cadena pesada se transfieren a un vector de expresión de mamífero. La transferencia se puede llevar a cabo en cualquiera de los vectores descritos en la sección previa habilitando la expresión de un anticuerpo recombinante de longitud completa. Si el cribado se realiza con una biblioteca de expresión de anticuerpo de longitud completa análoga de mamífero dicha transferencia puede no ser necesaria.

En otra realización de la presente invención, la biblioteca original se genera a partir de una fracción de células que contiene linfocitos la cual está enriquecida con linfocitos T. Los pares de secuencias que codifican para la región variable enlazada que constituyen la biblioteca original se pueden seleccionar para codificar un subconjunto de pares de secuencias de región variable enlazadas, constituidos de las cadenas  $\alpha$  y  $\beta$  y/o  $\gamma$  y  $\delta$  que codifican para proteínas de unión con especificidad deseada a la diana, lo que genera una biblioteca secundaria de pares análogos o pares combinatorios. Los receptores de linfocitos T específicos para antígeno posteriormente se pueden identificar de un acumulado de células transfectadas utilizando metodología convencional tal como tinción con complejos tetraméricos de MHC-péptido (por ejemplo Callan, M.F. et al. 1998. *J. Exp. Med.* 187, 1395-1402; Novak, E.J. et al. 1999. *J. Clin. Invest* 104, R63-R67), al medir las respuestas celulares en forma de liberación de IL-2 o por medios más sofisticados tales como técnicas de presentación en levadura o presentación retroviral.

**Células hospedadoras y expresión**

Las bibliotecas de la presente invención se pueden transferir a vectores adecuados para expresión y producción de proteínas codificadas a partir de las secuencias de ácido nucleico enlazadas de interés, en particular proteínas de unión que contienen región variable o fragmentos de las mismas. Tales vectores se desvelan en la sección de vectores y bibliotecas, y proporcionan la expresión, por ejemplo, de anticuerpos de longitud completa, fragmentos Fab, fragmentos Fv, scFv, TcR unidos a membrana o solubles o fragmentos TcR de una especie de elección.

Una característica de la presente invención es la introducción en una célula hospedadora de una biblioteca o una biblioteca secundaria de vectores de pares análogos de secuencias que codifican para la región variable enlazada o un clon único que codifica para un par análogo de secuencias que codifican para la región variable enlazada, para amplificación y/o expresión. Las células hospedadoras se pueden seleccionar de bacterias, levaduras, otros hongos, células de insecto, células vegetales o células de mamífero. Para propósitos de expresión se prefieren células de mamífero, tales como células de ovario de hámster chino (CHO), células COS, células BHK, células de mieloma (por ejemplo células Sp2/0, NS0), NIH 3T3, células de fibroblastos o humanas inmortalizadas tales como células HeLa, células HEK 293 o PER.C6.

La introducción de vectores en células hospedadoras se puede llevar a cabo por numerosos procedimientos de transformación o transfección conocidos por los expertos en la técnica, que incluyen precipitación con fosfato de calcio, electroporación, microinyección, fusión de liposomas, fusión de eritrocitos fantasma, fusión de protoplastos, infección viral y similares. Es bien conocido en la producción de anticuerpos monoclonales de longitud completa, fragmentos Fab, fragmentos Fv y fragmentos scFv.

La producción de anticuerpos policlonales recombinantes para ser utilizados para tratamiento es un área muy nueva. Se ha descrito una tecnología de fabricación policlonal recombinante en la solicitud PCT WO 2004/061104. En resumen, esta tecnología implica la generación de una colección de células, adecuadas como una línea de fabricación de células. La siguiente descripción de la técnica se realiza para una biblioteca de pares análogos, no obstante, es también aplicable para una biblioteca combinatoria. Las células individuales en la colección de células son capaces de expresar un elemento distinto de la proteína de unión policlonal recombinante, por ejemplo, de una biblioteca de pares análogos. Con el fin de asegurar que las células individuales expresen un par análogo único y no varios pares análogos de la proteína de unión policlonal, las secuencias de ácido nucleico que codifican para los pares análogos se introducen en un sitio único, específico de sitio en el genoma de cada célula individual. Esta es una característica importante de la colección de células dado que evita el mezclado de las cadenas pesada y ligera expresado de cada célula sino también debido a que genera células que son virtualmente idénticas entre sí, excepto por pequeñas diferencias en las regiones variables de los pares análogos individuales. Este rasgo permitirá un crecimiento sin desviaciones de la colección de células sobre el período de tiempo necesario para la producción. Para asegurar integración específica de sitio sola, se debe utilizar una línea de células hospedadoras con únicamente un sitio de integración, estas están disponibles comercialmente, por ejemplo, las células CHO Flp-In de Invitrogen que contienen un sitio FRT único. Los vectores apropiados para esta línea de células que contienen un sitio FRT correspondiente se introducen en el genoma utilizando la Flp recombinasa. Existen otras recombinasas varias conocidas, por ejemplo, Cre,  $\beta$ -recombinasa, Gin, Pin, PinB, PinD, R/RS, integrasa  $\lambda$  o fago  $\Phi$ C31 integrasa que se pueden utilizar en combinación con sus sitios de recombinación correspondientes. Además, los vectores apropiados contienen un marcador de selección que permite la selección de los integrantes específicos de sitio.

La generación de una línea de células de fabricación policlonal en la producción de una proteína policlonal recombinante a partir de la línea celular se puede obtener por varias estrategias diferentes de transfección y fabricación.

Una manera es utilizar una biblioteca de vectores que se unen mezclándolos en una sola composición, para la transfección de una línea de células hospedadoras con un solo sitio de integración por célula. Este procedimiento se denomina transfección a granel o transfección en granel. Generalmente, el diseño del vector y la célula hospedadora descrito previamente asegurarán que una línea de células policlonales capaz de crecimiento sin desviaciones se obtenga mediante la selección apropiada. Un congelado de la línea de células policlonales se generará antes del inicio de la elaboración de la proteína policlonal recombinante.

Otra manera es utilizar una biblioteca de vectores divididos en fracciones que contienen aproximadamente 5 a 50 vectores individuales en la biblioteca en una composición, para transfección. Preferiblemente, una fracción de la biblioteca está constituida por 10 a 20 vectores individuales. Cada composición después se transfecta en una alícuota de células hospedadoras. Este procedimiento se denomina transfección semi a granel. El número de alícuotas transfectadas dependerá del tamaño de la biblioteca y el número de vectores individuales en cada fracción. Si la biblioteca está constituida por, por ejemplo, 100 pares análogos distintos los cuales están divididos en fracciones que contienen 20 elementos distintos en una composición, 5 alícuotas de células hospedadoras necesitan ser transfectadas con una composición de biblioteca que constituya una fracción distinta de la biblioteca original. Las alícuotas de las células hospedadoras se seleccionan para integración específica de sitio. Preferiblemente, las alícuotas distintas se seleccionan por separado. No obstante, también se pueden acumular antes de la selección. Las alícuotas se pueden analizar por su diversidad clonal y únicamente aquellas con diversidad suficiente se utilizarán para generar un concentrado de biblioteca de par análogo policlonal. Para obtener la línea de células policlonales deseadas para fabricación, las alícuotas se pueden mezclar antes de generar el concentrado por congelamiento, inmediatamente después de que han sido recuperadas del concentrado o después de un tiempo breve de proliferación y adaptación. Opcionalmente, las alícuotas de células se mantienen separadas durante la producción y la composición de proteína policlonal se ensambla al combinar los productos de cada alícuota en vez de las alícuotas de células antes de la producción.

Una tercera manera es un procedimiento de alto rendimiento en el cual las células hospedadoras se transfectan por separado utilizando los vectores individuales que constituyen la biblioteca de los pares análogos. Este procedimiento se denomina transfección individual. Las células hospedadoras sometidas a transfección individual preferiblemente

- 5 se seleccionan para integración específica de sitio, por separado. Los clones de células individuales generados por selección se pueden analizar con respecto al tiempo de proliferación y preferiblemente, aquellos con velocidades de crecimiento similares se utilizan para generar un concentrado de biblioteca de par análogo policlonal. Los clones de células individuales se pueden mezclar para obtener la línea de células policlonales deseadas antes de generar el concentrado, inmediatamente después de que han sido recuperados del concentrado o después de un tiempo breve de proliferación y adaptación. Este enfoque puede eliminar cualquier posible desviación de secuencia residual durante la transfección, integración y selección. De manera alternativa, las células hospedadoras transfectadas individualmente se mezclan antes de que se realice la selección, y esto permitirá el control de la desviación de la secuencia debido a transfección.
- 10 Una característica compartida en las estrategias de fabricación indicadas en lo anterior es que todos los pares análogos individuales que constituyen la proteína policlonal recombinante se pueden producir en uno o en un número limitado de biorreactores. La única diferencia es la etapa en la cual uno selecciona generar la colección de células que constituyen la línea de células de fabricación policlonal.
- 15 Una realización de la presente invención es una población de células hospedadoras que comprende una biblioteca análoga o una biblioteca secundaria de pares enlazados de secuencias que codifican para la región variable.
- En una realización adicional, una población de células hospedadoras comprende una biblioteca que se obtiene de una población de células solas aisladas que constituyen linfocitos, utilizando la amplificación RT-PCR múltiplex seguido por enlace por ligación o recombinación o la tecnología RT-PCR de superposición-extensión múltiplex de la presente invención, para enlazar pares análogos.
- 20 Otra realización de la presente invención es una población de células hospedadoras que comprende una biblioteca combinatoria o una biblioteca secundaria de pares enlazados de secuencias que codifican para la región variable.
- Una población de células hospedadoras de acuerdo con la presente invención abarcará una población diversa de células que corresponden a la diversidad de la biblioteca, con las cuales las células se han transformado/transfectado. Preferiblemente, cada célula de la población de células únicamente constituye un par análogo de la biblioteca completa de pares análogos, y ningún miembro individual de la biblioteca de pares análogos excede más del 50%, de manera más preferida el 25% o de manera mucho más preferida el 10% del número total de elementos individuales expresados a partir de la población de células hospedadoras.
- 25 En una realización preferida de la presente invención, la población de células hospedadoras son células de mamífero.
- 30 Una población de células hospedadoras como se desvela en lo anterior se puede utilizar para la expresión de una proteína de unión policlonal recombinante dado que las células individuales de la población constituyen secuencias que codifican para la región variable de diversidad diferente.
- Una realización de la presente invención es una proteína policlonal recombinante expresada de una población de células hospedadoras que comprende una biblioteca de vectores que codifican para pares análogos diversos de secuencias que codifican para la región variable enlazada, en donde la biblioteca se puede obtener por el procedimiento de la presente invención. Típicamente, una proteína policlonal recombinante de la presente invención está constituida por al lo menos 2, 5, 10, 20, 50, 100, 1000,  $10^4$ ,  $10^5$  ó  $10^6$  proteínas comprendidas de pares análogos diferentes.
- 35 Una realización preferida de la presente invención es una inmunoglobulina policlonal recombinante que se expresa de una población de células hospedadoras que comprende una biblioteca de vectores que codifica para los pares análogos diversos de las secuencias que codifican para la región variable de cadena pesada y la región variable de cadena ligera.
- 40 Otra realización preferida de la presente invención es TcR policlonal recombinante que se expresa de una población de células hospedadoras que comprenden una biblioteca de vectores que codifican para pares análogos diversos de la región variable de cadena  $\alpha$  TcR unida con secuencias que codifican para la región variable de cadena  $\beta$  y/o la región variable de cadena  $\gamma$  TcR unida con secuencias que codifican para la región variable de la cadena  $\delta$ .
- 45 Otra realización de la presente invención es una célula hospedadora adecuada para producción de una proteína monoclonal. En particular, un anticuerpo monoclonal constituido por un par análogo de una región variable de cadena ligera con una región variable de cadena pesada o un TcR monoclonal constituido por un par análogo de una región variable  $\alpha$  con una región variable  $\beta$  o una región variable  $\delta$  con una región variable  $\gamma$ . Preferiblemente, dicha línea de células de producción monoclonal no es una línea de células de hibridoma.
- 50 Dicho anticuerpo monoclonal o TcR se puede generar al agregar las siguientes etapas al procedimiento de enlace de una pluralidad de secuencias nucleótidas no continuas de interés a) al insertar las secuencias de ácido nucleico enlazadas en un vector; b) al introducir el vector en una célula hospedadora; c) al cultivar células hospedadoras bajo condiciones adecuadas para expresión; y d) al obtener el producto proteínico expresado del vector insertado en la célula hospedadora. Preferiblemente, el vector introducido en la célula hospedadora codifica para un par análogo individual de secuencias que codifican para la región variable.
- 55

### **Aplicaciones de la invención**

- 60 Una de las aplicaciones principales de la presente invención es el enlace de pares análogos de secuencias que codifican para la región variable, especialmente secuencias que codifican para la región variable de las cadenas pesada y ligera de inmunoglobulina o secuencias que codifican para la región variable de la cadena  $\alpha$  y  $\beta$  de TCR o  $\gamma$  y  $\delta$ , mediante un procedimiento de alto rendimiento para la generación de bibliotecas de pares análogos. Además de la generación de bibliotecas de pares análogos, la RT-PCR múltiplex seguida de enlace por ligación o

recombinación o la técnica de RT-PCR de superposición-extensión múltiple de la presente invención se pueden utilizar en la generación de bibliotecas combinatorias de anticuerpos quiméricos humanos/no humanos al realizar la técnica en una población de células genéticamente diversas, lisados celulares a partir de los cuales una población de células o sobre ARN purificado de la población de células. Las bibliotecas, bibliotecas secundarias o clones solos de una de estas bibliotecas facilitan la expresión de proteínas policlonales o monoclonales. Especialmente se pueden obtener anticuerpos monoclonales o policlonales de las bibliotecas de la presente invención.

Es bien conocido el uso de anticuerpos monoclonales recombinantes en el diagnóstico, tratamiento y profilaxis. Los anticuerpos monoclonales y policlonales recombinantes generados por la presente invención tendrán las mismas aplicaciones que los productos de anticuerpos generados por tecnologías existentes. En particular, una composición farmacéutica que comprende una inmunoglobulina recombinante policlonal como ingrediente activo, combinada con por lo menos un excipiente farmacéuticamente aceptable se puede elaborar por medio de la presente invención. De manera más preferible son las composiciones farmacéuticas en donde la inmunoglobulina recombinante policlonal está constituida por pares análogos de secuencias que codifican para la región variable. Dichas composiciones farmacéuticas de inmunoglobulinas recombinantes policlonales se pueden utilizar como medicamentos. La inmunoglobulina recombinante policlonal de la composición puede ser específica o reactiva contra una diana de enfermedad predeterminado y la composición de esta manera se puede utilizar para el tratamiento, disminución o prevención de enfermedades tales como cáncer, infecciones, enfermedades inflamatorias, alergia, asma y otras enfermedades respiratorias, enfermedades autoinmunes, malos funcionamientos inmunológicos, enfermedades cardiovasculares, enfermedades en el sistema nervioso central, enfermedades metabólicas y endocrinas, rechazo de trasplantes o embarazo no deseado, en mamíferos tales como un humano, un animal doméstico o una mascota.

## Ejemplos

### Ejemplo 1:

Se inmunizan subcutáneamente a ratones Balb/c con 50 µg de toxoide tetánico (TT) en adyuvante completo de Freund. Se refuerza a los ratones en el día 14 con 50 µg de TT en adyuvante incompleto de Freund. Después de 30 días adicionales se refuerza a los ratones con 50 µg de TT en adyuvante incompleto de Freund. Tres días después del último refuerzo se sacrifican los ratones y se extirpan los bazo y se transfieren a un tubo que contiene 30 ml de RPMI 1640 p/FCS al 10% a 4°C. El tejido se transfiere a un separador de células de 74 µm (Corning, 136350-3479) en un recipiente de 10 cm. Con la parte trasera del émbolo de una jeringa se maceran los bazo a través del filtro. El filtro se enjuaga con 10 ml de RPMI 1640, solución FCS al 10%. Se separa el filtro y el recipiente se llena con 20 ml de RPMI 1640 frío, FCS al 10%. Las células se transfieren a un tubo de 50 ml y se centrifugan a 300 x g a 2-8°C durante 5 minutos. Las células se resuspenden en 5-10 ml de RPMI 1640 4°C con p/FCS 1% y se filtran a través de un filtro de jeringa de 50 µm (Becton Dickinson, 340603). Las células se sedimentan y se resuspenden en FCS; DMSO al 10% para obtener una densidad celular de  $2 \times 10^7$  células/ampolla y se congelan.

Se recalientan viales congelados con esplenocitos en una suspensión de células solas de ratones Balb/c inmunizados con toxoide tetánico a 37°C y se transfiere a un tubo de 15 ml aún con hielo presente. Se agrega a gotas 10 ml de RPMI enfriado con hielo, FCS al 10% al tubo, mientras se forma un remolino. Después de un lavado en 10 ml de FACS PBS, se agregan 5 ml de PBS, FCS al 2% antes de filtrar las células a través de un equipo Filcon 50 µm (Becton Dickinson catálogo N° 340603). Se sedimentan las células y se resuspenden en 1 ml de PBS, FCS al 2% (volumen final) y posteriormente se tiñen ya sea anti-CD43 FITC diluido 1:100 (BD catálogo N° 553270) y anti-CD138 PE diluido 1:40 (BD catálogo N° 553714) en 1 ml de PBS, FCS al 2% o con anti-B220 APC (BD catálogo N° 553092) diluido 1:40 y anti-MHC II FITC (BD catálogo N° 5535547) diluido 1:200. Las células se incuban a 4°C durante 20 min en la oscuridad. Finalmente, las células se lavan 2 veces con 2 ml de PBS, FCS al 2% y se agregan a 15 ml de PBS, FCS al 2% (suero bovino fetal). Justo antes de la clasificación se agrega PI a 1:100 y las células se clasifican en un recuento de acontecimientos de aproximadamente 1000 - 2000 células/segundo. Los criterios para ambas tinciones se desvelan en la figura 3.

Figura 3A: Se excluyen células PI positivas (muertas) en el panel izquierdo inferior (P1). Después las células plasmáticas se clasifican como altas en CD43, altas en CD138 en el panel derecho inferior (P2). Finalmente, se excluyen los dobletes en la gráfica SSC-H, SSC-W en el panel derecho superior (P3). Las células positivas para los tres criterios se clasifican en placas ELISPOT, de acuerdo con la Tabla 1.

### ELISPOT:

Las placas de 96 pocillos de fondo de nitrocelulosa (placas HA, Millipore, Bedford, MA) se prehumedecen con PBS (también los pocillos de bloqueo) antes de recubrirse con 100 µl de toxoide tetánico (TT) 25 µg/ml o un anticuerpo de oveja anti-IgG murino (Jackson Immuno Research, catálogo N° 515-005-062) en PBS. El mismo volumen de PBS está presente en los pocillos control. La placa se deja reposar a 4°C. Al día siguiente, los pocillos se lavan 3 veces en PBS antes de ser bloqueados con 200 µl de RPMI + polvo de leche desnatada al 2% y se dejan reposar a 4°C. La placa se mueve al incubador (37°C, CO<sub>2</sub> al 5%, humedad al 100%) 1h antes de la adición de las células. Se agregan 100 µl de células en RPMI completo a pocillos de TT-, anti-IgG recubiertos y pocillos solo bloqueados. El medio sin células se incuban como control. La placa se mueve a un incubador (37°C, CO<sub>2</sub> al 5%, 100% de humedad). Al día siguiente la placa se lava seis veces para eliminar las células, 3 veces en tampón: PBS + Tween 20 al 0,01% y 3 veces en PBS. Posteriormente, a los pocillos se les agrega anticuerpo de cabra anti-IgG de ratón conjugado con HRP (Caltag M30007) diluido 1:3000 en RPMI + polvo de leche desnatada al 2% (100 µl/pocillo). Después de 2 h de incubación a 37°C se lavan los pocillos tres veces en PBS + Tween 20 al 0,01% seguido por tres veces solamente con PBS. Los puntos después se revelan con 100 µl de sustrato cromogénico que consiste de H<sub>2</sub>O<sub>2</sub> al 0,015% Y 0,3 mg/ml de 3-amino-9-etilcarbazol en acetato de sodio 0,1M, pH 5,1. Se detiene el desarrollo de color al lavar con agua corriente después de 5 min.

Al considerar los resultados en la Tabla 1, las células plasmáticas (PC; altas en CD43, altas en CD138, criterio P3 en la figura 3A) específico para TT están presentes en aproximadamente el 2% mientras que los blastos plasmáticos

específicos para TT (PB; positivos para B220 y MHCII, criterio P4 en la figura 3B) están presentes en aproximadamente el 4%. Esto ilustra la superioridad de PC a PB en la producción de anticuerpos específicos después de inmunización de TT.

**Tabla 1**

		Número de puntos (número de células clasificadas)		
Recubrimiento: PBS	PC	0 (1,001)	ND	ND
	PB	0 (13,976)	0 (1500)	0 (150)
Recubrimiento: TT	PC	22 (1,002)	2 (250)	0 (50)
	PB	54 (13,257)	0 (1500)	0 (150)
Recubrimiento: α-IgG	PC	55 (1,000)	6 (250)	1 (50)
	PB	48 (5000)	8 (1500)	0 (150)

5

**Ejemplo 2**

Un vial congelado de esplenocitos se tiñe como se desvela en el Ejemplo 1.

Cuatro fenotipos diferentes se clasifican de cuatro maneras. Los criterios de clasificación se muestran en la figura 4. En primer lugar, se excluyen células PI positivas o células muertas en el panel izquierdo inferior (PI). P2 es intermedio a CD138, alto en CD43. P3 es alto en CD138, alto en CD43. P4 es alto en CD138, negativo CD43. P5 es intermedio en CD138 y bajo en CD43. Se clasifican 10,000 células positivas para P1 y cada uno de los cuatro criterios en tubos de ensayo y se congelan para evaluación por Symplex de ratón.

10

Las fracciones P2, P3, P4 y P5 se clasifican a granel en tubos, cada uno con 10,000 células. Los tubos se centrifugan y resuspenden en medio de Eagle modificado por Dulbecco que contiene 2 U/μl de inhibidor de ribonucleasa (RNasin, Promoga cat. N° N2511) a una concentración de 250 células/μl, 10 μl por tubo y se congela a -80°C.

15

Se realiza amplificación Symplex de ratón en una serie de dilución de cada uno de los 4 conjuntos de linfocitos clasificados para comparar el contenido de ARNm para anticuerpo IgG-k. Los principios básicos de las reacciones son:

20

- Primero se realiza una reacción RT en donde la cadena pesada y ligera se ceban por cebadores de región constantes específicos.

- En segundo lugar se realiza una reacción múltiple utilizando cebadores 5' VH y VK que cubren todas las regiones variables y se equipan con partes sobresalientes complementarias que facilitan la formación de bandas de superposición. Los cebadores 3' se localizan en la región constante de la cadena pesada y ligera.

25

- Finalmente se realiza una reacción incrustada amplificando únicamente VH y VK utilizando cebadores JH y JK.

- El producto de reacción final está constituido por VH y VK conjugado extremo 5' a extremo 5' y conectado con un enlazante. El tamaño debe ser de aproximadamente 700 pares de bases.

30

Para la reacción RT-PCR múltiple combinada se utiliza el conjunto de cebadores que se muestran en la Tabla 2, empleándose el equipo RT-PCR Qiagen OneStep esencialmente de acuerdo con las instrucciones del fabricante. Las células congeladas se recalientan en hielo, se resuspenden y centrifugan. En cada serie de diluciones se utiliza lisado celular que corresponde a 100, 32, 10, 3,2, 1, 0,32, 0,1 y 0 células. Los volúmenes de reacción totales son de 20 μl. Las condiciones de los ciclos son:

35

- 55°C, 30 min.
- 95°C, 15 min.
- 94°C, 30 seg. } 35 ciclos
- 60°C, 30 seg. }
- 72°C, 5 min
- 72°C, 10 min.

40

La reacción incrustada se realiza con el conjunto de cebadores que se muestran en la Tabla 3 utilizando polimerasa FastStart (Roche) y los activos suministrados, esencialmente de acuerdo con las instrucciones del fabricante. Se utiliza 1 μl de producto de reacción de RT-PCR para la reacción incrustada en un volumen total de 20 μl. Las condiciones de reacción son:

- 95°C, 30 seg. } 35 ciclos
- 60°C, 30 seg. }
- 72°C, 90 seg.
- 72°C, 10 min.

5

Finalmente se analizan 10 µl de cada producto de reacción final en 1 gel de *imples* al 1%.

10 A partir de los resultados de *imples* sobre lisados de células titulados (figura 5) es evidente que se pueden enlazar regiones variables de cadena pesada y ligera con células a partir de P3 hasta aproximadamente 0,1 célula y a partir de P2 desde aproximadamente 3,2 células. Las otras clasificaciones fueron menos eficaces con el enlace el cual únicamente era posible desde aproximadamente 32 células y más. En conclusión, alto en CD43 y alto en CD138 (P3) es más útil para *Symplex<sup>MR</sup>* a nivel de sola célula mientras que alto en CD43 e intermedio en CD138 se puede utilizar, pero es menos eficaz.

Tabla 2: Conjunto de cebadores *imples* de ratón utilizados para RT combinada y reacción múltiplex

Cebador	Concentración (Nm)	Secuencia	SEC ID N°:
mHCrev1	0,2	GACSGATGGGCCCTTGTTGG	1
mKappar	0,2	GCTGTAGGTGCTGTCTTTGC	2
<b>conjunto Mvh</b>			
Mvh A	0,04	TATTCCCATGGCGCGCCSAGGTCCARCTGCARCAGY	3
Mvh B	0,04	TATTCCCATGGCGCGCCGARGTGMAGCTKGTGKAGTC	4
Mvh C	0,04	TATTCCCATGGCGCGCCSAGGTGCAGCTKMAGGAGTC	5
Mvh 8	0,04	TATTCCCATGGCGCGCCCAGGTTACTCTGAAAGAGTC	6
Mvh 9	0,04	TATTCCCATGGCGCGCCCAGATCCAGTTGGTGCAGTC	7
<b>conjunto Mvk</b>			
Mvk D	0,04	GGCGCGCCATGGGAATAGCTAGCCGAYATCCAGATGA	8
Mvk E	0,04	GGCGCGCCATGGGAATAGCTAGCCRACATTGTGMTGA	9
Mvk F	0,04	GGCGCGCCATGGGAATAGCTAGCCSAMATTGKCTSA	10
Mvk 1-2	0,04	GGCGCGCCATGGGAATAGCTAGCCGATRRTTGTGATGA	11

*W=A/T, R=A/G, S=G/C, Y=C/T, K=G/T, M=A/C, H=ACT, B=GCT; Conc. – Concentración final.*

15 Tabla 3: Cebadores para PCR incrustada *imples* de ratón

Cebador	Concentración Nm	Secuencia	SEC ID N°:
<b>conjunto Mjh</b>			
Mjh1	200	GGAGGCGCTCGAGACGGTGACCGTGGTCCC	12
Mjh2	200	GGAGGCGCTCGAGACTGTGAGAGTGGTGCC	13
Mjh3	200	GGAGGCGCTCGAGACAGTGACCAGAGTCCC	14
Mjh4	200	GGAGGCGCTCGAGACGGTGACTGAGGTTCC	15
<b>conjunto Mjk</b>			
Mjk1	200	GATGGTGCAGCCACAGTTTCGTTTGATTCCAGCTTGGTG	16

(cont.)			
mJK2	200	GATGGTGCAGCCACAGTTCGTTTTATTTCCAGCTTGGTC	30
mJK4	200	GATGGTGCAGCCACAGTTCGTTTTATTTCCAACCTTTGTC	31
mJK5	200	GATGGTGCAGCCACAGTTCGTTTCAGCTCCAGCTTGGTC	17

### Ejemplo 3. Clonación de anticuerpos anti-EGFR

#### Inmunizaciones

5 Se utilizaron ratones hembra BALB/c, cepa A o C57B16 (8-10 semanas de edad) para inmunizaciones por inyecciones con diferentes proteínas purificadas además de células que sobreexpresan EGFR.

10 Se utilizaron para algunas de las inmunizaciones proteínas EGFR disponibles comercialmente (R&D systems, catálogos #1095-ER o Sigma#E3641). Para otras de las inmunizaciones se utilizaron EGFR humano recombinante y EGFRvIII producido como proteínas de fusión que consiste del ECD de EGFR o EGFRvIII y hormona del crecimiento humana (hGH) que también incluye el sitio de separación del virus de la roya de tabaco (TEV) además de una etiqueta His. En algunos casos, el ECD de EGFR se aisló por separación con TEV-proteasa y purificación subsiguiente en una columna de níquel.

15 La línea de células de cáncer de cabeza y cuello humanas, HN5 (Easty DM, Easty GC, Carter RL, Monaghan P, Butler LJ. Br J Cancer. 1981 Jun;43(6):772-85. Ten human carcinoma cell lines derived from squamous carcinomas of the head and neck.) que expresa aproximadamente  $10^7$  receptores/célula se utilizó para inmunizaciones basadas en células. Las células se cultivan en medio DMEM suplementado con FBS (suero bovino fetal) al 10%, glicerol 3 mM, piruvato de sodio 5 mM y penicilina/estreptomicina al 1%. Antes de cada inmunización las células se lavaron en PBS, se tripsinizaron con TrypLE y se resuspendieron en medio de cultivo. Subsiguientemente, las suspensiones celulares se lavaron dos veces en PBS por centrifugación a 250 x g durante 5 min, disociándose y por resuspensión en 15 ml de PBS estéril.

20 Las células o el antígeno se diluyeron en PBS y después se mezclaron 1:1 con adyuvante de Freund. Se utiliza el adyuvante para potenciar y modular la respuesta inmunitaria. Para las primeras inmunizaciones se utilizó adyuvante completo de Freund (CFA, por sus siglas en inglés) mientras que para las inmunizaciones subsiguientes se utilizó adyuvante incompleto de Freund (IFA, por sus siglas en inglés). IFA es una emulsión aceite en agua constituida por aceites minerales y CFA es IFA al cual se le ha agregado la especie *Mycobacterium* seca e inactivada por calor. 25 Ambos adyuvantes tienen un efecto de depósito. CFA da lugar a una persistencia a largo plazo de la respuesta inmunitaria y se utiliza para inmunizaciones por primera vez para reforzar la respuesta inmunitaria, e IFA se utiliza para inmunizaciones subsiguientes. Las emulsiones se prueban al agregar una gota de la superficie de un vidrio con agua. Si la gota permanece como una gota, la emulsión es estable y las inyecciones se pueden llevar a cabo. Únicamente se administraron a los ratones emulsiones estables.

30 Dependiendo del programa (véase Tabla 4), se utilizaron para cada inyección 25-100  $\mu$ g de antígeno o  $10^7$  células. En total, los ratones recibieron 4 inyecciones. A todos los ratones se les inyectó ya sea con 300  $\mu$ l o 200  $\mu$ l de emulsión. Dependiendo del programa, las inyecciones se realizaron subcutáneamente (s.c.), por vía intraperitoneal (i.p.) o intravenosa (i.v.).

35 Al finalizar, se sacrificó a los ratones por dislocación cervical y se extirparon los bazo y se transfirieron a un aparato para destrucción celular de 74  $\mu$ m (Corning#136350-3479). Se maceraron las células a través del filtro, se resuspendieron en RPMI 1640 frío con FBS al 10% y se centrifugaron a 300 x g durante 5 minutos. El sedimento celular se resuspende en RPMI 1640 con FBS al 1%, se filtra a través de un filtro de jeringa de 50  $\mu$ m (BD# 340603) y se recolecta por centrifugación. El sedimento celular se criopreserva después de la resuspensión en FCS con DMSO al 10% y las células congeladas se almacenan a -80°C hasta clasificación por FACS.

#### 40 Clasificación FACS de células plasmáticas murinas

Los viales con los esplenocitos congelados se recalientan a 37°C y se transfieren a un tubo de ensayo de 15 ml aún con hielo presente. Se agrega a gotas al tubo mientras se forma un remolino 10 ml de RPMI enfriado con hielo, FBS al 10% (suero bovino fetal). Después de un lavado en 10 ml de PBS de FACS, se agregan 5 ml de PBS de FCS antes de filtrar las células a través de Filcon 50  $\mu$ m. Las células después se sedimentan y resuspenden en 1 ml de PBS con FBS al 2% (volumen final) y se tiñen con anti-CD43-FITC y anti-CD138-PE, de acuerdo con la dilución específica hasta una concentración final de aproximadamente 5  $\mu$ g/ml. Se incuban las células a 4°C durante 20 min en la oscuridad. Posteriormente se lavan las células dos veces con 2 ml de tampón FACS. Se agregan hasta 15 ml de PBS de FACS. Se agrega yoduro de propidio (PI) 1:100 y las células posteriormente se clasifican en placas de PCR de 96 pocillos que contienen tampón de reacción de PCR (véase más adelante) y se sedimentan durante 2 min a 400 x g antes de que las placas se congelen a -80°C. Las células plasmáticas se clasifican como positivas para CD43/positivas para CD-138.

#### Enlace de pares análogos $V_H$ y $V_L$ .

55 El enlace de las secuencias codificantes para  $V_H$  y  $V_L$  se realiza sobre células solas clasificadas como células plasmáticas, facilitando el pareado análogo de las secuencias codificantes  $V_H$  y  $V_L$ . El procedimiento utiliza un procedimiento de PCR de dos etapas en base a RT-PCR de superposición-extensión múltiplex de una etapa seguido por PCR incrustada. Las mezclas de cebador utilizadas en el presente ejemplo solo amplifican cadenas ligeras  $\kappa$ . No

obstante, se pueden agregar a la mezcla de cebador múltiplex, cebadores capaces de amplificar las cadenas ligeras  $\lambda$  y una mezcla de cebador para PCR incrustado, se desea. Si se agregan cebadores  $\lambda$ , el procedimiento de clasificación se debe adaptar de manera que no se excluyan células positivas  $\lambda$ . El principio para el enlace de las secuencias análogas  $V_H$  y  $V_L$  se ilustra en la figura 1.

- 5 Las placas de 96 pocillos producidas se recalientan y las células clasificadas sirven como plantilla para la RT-PCR de superposición-extensión múltiplex. El tampón de clasificación agregado a cada pocillo antes de la clasificación de células solas contiene tampón de reacción (OneStep RT-PCR Buffer; Qiagen), los cebadores para RT-PCR (véase Tabla 2 anterior) e inhibidor de ribonucleasa (RNasin, Promega). Esto se suplementa con la mezcla de enzima RT-PCR OneStep (dilución 25 x; Qiagen), y la mezcla de dNTP (200  $\mu$ M de cada uno) para obtener la concentración final dada en un volumen de reacción de 20  $\mu$ l. Las placas se incuban durante 30 min a 55°C para permitir la transcripción inversa del ARN de cada célula. Después de RT, las placas se someten al siguiente ciclo de PCR: 10 min a 94°C, 35 x (40 seg a 94°C, 40 seg a 60°C, 5 min a 72°C), 10 min a 72°C.

15 Las reacciones de PCR se realizaron en un equipo H20BIT Thermal cycler con un equipo Peel Seal Basket para 24 placas de 96 pocillos (ABgene) para facilitar un alto rendimiento. Las placas de PCR se clasifican a -20°C después de los ciclos.

- 20 Para la etapa de PCR incrustada, se preparan placas de PCR de 96 pocillos con la siguiente mezcla en cada pocillo (reacciones de 20  $\mu$ l) para obtener la concentración final dada: tampón FastStart 1x (Roche), mezcla dNTP (200  $\mu$ M de cada uno), mezcla de cebador incrustado (véase la Tabla 5), ADN polimerasa Phusion (0,08 U; Finnzymes) y combinación de enzimas FastStart High Fidelity (0,8 U; Roche). Como plantilla para la PCR incrustada se transfiere 1  $\mu$ l de las reacciones de PCR de superposición-extensión múltiplex. Las placas PCR incrustadas se someten al siguiente procedimiento de termociclado: 35 x (30 seg a 95°C, 30 seg a 60°C, 90 seg a 72°C), 10 min a 72°C.

Las reacciones seleccionadas aleatoriamente se analizan en un gel de agarosa al 1% para verificar la presencia de un fragmento de superposición-extensión de aproximadamente 890 pares de bases (pb).

Las placas se almacenan a -20°C hasta el procesamiento adicional de los fragmentos de PCR.

- 25 Los repertorios de los pares codificantes  $V_H$  y  $V_L$  enlazados de PCR incrustados se acumulan, sin la mezcla de pares de donantes diferentes y se purifican por electroforesis en gel de agarosa al 1% preparativa. La secuencia que codifica para la cadena ligera constante  $\kappa$  humana se divide por extensión de superposición a la región codificante  $V_L$  de los productos de PCR acumulados de los pares codificantes  $V_H$  y  $V_L$  enlazados (figura 2). La secuencia codificante de cadena ligera constante  $\kappa$  humana se amplifica de un plásmido que contiene la secuencia codificante de un anticuerpo humano con la cadena ligera  $\kappa$  en una reacción que contiene: enzima Phusion (2 U; Finnzymes), tampón Phusion 1x, mezcla de dNTP (200  $\mu$ M de cada uno), cebador hKCforw-v2 y cebador Kappa3' (tabla 6) y la plantilla de plásmido pLL138 (10 ng/ $\mu$ l) en un volumen total de 50  $\mu$ l. La reacción se somete al siguiente tratamiento termocíclico: 25 x (30 seg a 95°C, 30 seg a 55°C, 45 seg a 72°C), 10 min a 72°C. El fragmento de PCR resultante se purifica por electroforesis preparativa en gel de agarosa al 1%.
- 30
- 35 Cada uno de los fragmentos de PCR acumulados y purificados de cada repertorio se divide al fragmento de PCR amplificado y purificado de la región que codifica para la parte constante  $\kappa$  humana (apéndice 1) por el siguiente empalme por PCR de superposición y extensión (volumen total de 50  $\mu$ l) que contiene: un fragmento de región que codifica para la parte constante  $\kappa$  humana (1,4 ng/ $\mu$ l), un fragmento de PCR acumulado y purificado (1,4 ng/ $\mu$ l), ADN polimerasa Phusion (0,5 U; Finnzymes) y una combinación de enzimas FasStart High Fidelity (0,2 U; Roche);
- 40 tampón FastStart 1x (Roche), mezcla dNTP (200  $\mu$ M de cada uno), cebador mhKCreV y conjuntos de cebadores mJH (véase la Tabla 6). La reacción se somete al siguiente procedimiento de termociclado: 2 min a 95°C, 25 x (30 seg a 95°C, 30 seg a 55°C, 1 min a 72°C), 10 min a 72°C. El fragmento de PCR resultante (aproximadamente 1070 pares de bases) se purifica por electroforesis preparativa en gel de agarosa al 1%.

#### **Inserción de pares codificantes $V_H$ y $V_L$ análogos en un vector de cribado**

- 45 Con el fin de identificar anticuerpos con especificidad de unión por EGFR, las secuencias codificantes  $V_H$  y  $V_L$  obtenidas se expresan como anticuerpos de longitud completa. Esto implica la inserción del repertorio de pares codificantes  $V_H$  y  $V_L$  en un vector de expresión y transfección en una célula hospedadora.

50 Se utiliza un procedimiento de clonación de dos etapas para generación de un repertorio de vectores de expresión que contienen los pares codificantes  $V_H$  y  $V_L$  enlazados. Estadísticamente, si el repertorio de los vectores de expresión contiene diez veces la cantidad de plásmidos recombinantes que el número de productos de PCR para  $V_H$  y  $V_L$  pareados utilizados para generación del repertorio de cribado, existe el 99% de probabilidad de que estén representados todos los pares de genes únicos. Por lo tanto, si se obtienen 400 fragmentos del gen V de superposición-extensión, se genera para cribado un repertorio de por lo menos 4000 clones.

- 55 En resumen, el producto de PCR purificado de los repertorios de los pares codificantes  $V_H$  y  $V_L$  enlazados, divididos en la región codificante constante  $\kappa$  humana se escinden con ADN endonucleasas *XhoI* y *NotI* en los sitios de reconocimiento introducidos en las partes terminales de los productos de PCR. Los fragmentos escindidos y purificados se ligan en un vector de expresión IgG de mamífero con *XhoI/NotI*, OO-VP-002 (figura 6) por procedimientos de ligación estándar. La mezcla de ligación se somete a electroforación en *E. coli* y se agrega a placas 2xYT que contienen el antibiótico apropiado y se incuban a 37°C durante la noche. El repertorio amplificado de vectores se purifica de células recuperadas de las placas utilizando procedimientos de purificación de ADN convencional o estándar (Qiagen). Los plásmidos se preparan para inserción de fragmento promotor-líder por escisión utilizando endonucleasas *AscI* y *NheI*.
- 60

Los sitios de restricción para estas enzimas se localizan entre los pares de genes que codifican para  $V_H$  y  $V_L$ . Después de la purificación del vector se inserta un fragmento promotor-líder de mamífero bidireccional digerido con

Ascl-NheI dentro de los sitios de restricción Ascl y NheI por procedimientos de ligación estándar. El vector ligado se amplifica en *E. coli* y el plásmido se purifica utilizando procedimientos estándar o convencionales. El repertorio generado de vectores de cribado se transforma en *E. coli* por procedimientos convencionales. Las colonias que se obtienen se consolidan en placas maestras de 384 pocillos y se almacenan. El número de colonias distribuidas excede al número de productos de PCR introducidos en por lo menos tres veces, proporcionando así el 95% de probabilidad para la presencia de todos los pares de genes V únicos obtenidos.

Se prepara ADN plasmídico de los clones seleccionados en lo anterior y células CHO-S FreeStyle (Invitrogen) se transfectan a escalas de 2 ml para expresión de anticuerpos (de acuerdo con las instrucciones del fabricante). Los sobrenadantes se cosechan 96 horas después de la transfección.

#### 10 **Cribado para unión al dominio extracelular de EGFR**

En general, el cribado se realiza como un procedimiento de dos etapas. Las bibliotecas de anticuerpos se criban para reactividad para proteína EGFR recombinante en ELISA después de lo cual se utiliza FMAT (FLISA) como un enfoque basado en células, con la línea de células NR6wtEGFR (Batra *et al*, 1995, Cell Growth Differ, 6(10):1251-9), para detección de anticuerpos para EGFR que se unen a la superficie celular que expresa EGFR. Para las bibliotecas 101 y 108/109 (Tabla 4) se realiza la prueba de ELISA con EGFR recombinante que representa el dominio extracelular de EGFR.

En resumen, las placas de ELISA Nunc maxisorb (No. de catálogo 464718) se recubren con 1 µg/ml de proteína (producida en el laboratorio) diluida en PBS a 4°C durante la noche. Antes del bloqueo en 50 µl de leche-PBS-T al 2% se lavan las placas una vez con PBS + Tween 20 al 0,05% (PBS-T). Las placas se lavan una vez con PBS-T, 20 µl de leche-PBS-T al 2% y se agregan sobrenadantes de 5 µl de transfectantes CHO-S FreeStyle (véase antes) y se incuban durante 1,5 horas a temperatura ambiente, después de lo cual las placas se lavan una vez con PBS-T 20 µl por pocillo. El anticuerpo secundario (de cabra anti-IgG humano conjugado a HRP, Jackson N° de catálogo 109-035-097) se diluye 1:10000 en leche-PBS-T al 2% que se agrega para detectar el anticuerpo unido a los pocillos y se incuban durante 1 hora a temperatura ambiente. Las placas se lavan una vez en PBS-T antes de la adición de 25 µl de sustrato (Kem-en-tec Diagnostics, Catálogo N° 4390) que se incuban durante 5 minutos. Se agregan 25 µl de ácido sulfúrico 1M y después de la incubación se detiene la reacción. La señal específica se detecta en un lector ELISA a 450 nm.

Para la detección por FMAT basado en células de anticuerpos anti-EGFR, se mantienen células SKBR-3 (ATCC #HTB-30) o NR6wtEGFR en medio de cultivo. Las células se cuentan y diluyen a 125,000 células/ml con el anticuerpo de cabra anti IgG humano (H-L) conjugado a Alexa-647 (Molecular probes N° A21445, lote N° 34686A) diluido 1:40,000. Un total de 20 µl de esta suspensión se transfiere a placas Nunc de fondo transparente de 384 pocillos. Posteriormente se agregan a las células 10 µl de sobrenadante de transfección. La señal FMAT de la reacción se miden después de 6-10 horas de incubación.

Los datos del cribado indican que 221 (4,8%) de los clones totales son positivos en la forma de ELISA. Una cantidad de 93 (2,0%) de estos clones también son positivos en FMAT. En total 220 (4,8%) de los clones son positivos en la prueba FMAT y entre estos 127 (220-93) positivos de manera única para el antígeno de superficie celular. Se criba la biblioteca 111 de una manera similar, pero dado que el procedimiento de inmunización se realiza para generar anticuerpos específicos para el receptor EGFR mutante por supresión EGFRvIII, los cribados por ELISA incluyen análisis para detectar tanto EGFR de tipo silvestre como EGFRvIII. Se identificaron siete clones como específicos para EGFRvIII en la prueba de ELISA y, de manera interesante, estos clones fueron negativos para tinción de células que expresan wtEGFR en FMAT. Se identificaron 13 clones como positivos para wtEGFR en FMAT y ELISA pero no para EGFRvIII, los cuales fueron únicos para esta biblioteca en comparación con las bibliotecas 101 y 108/109. Todos los clones positivos para ELISA se seleccionaron para análisis adicional.

#### 10 **Análisis de secuencia y selección de clon**

Los clones identificados como específicos para EGFR en ELISA se recuperaron de las placas maestras originales (formato de 384 pocillos) y se consolidaron en placas nuevas. El ADN se aisló de los clones y se envió para secuenciado de ADN de los genes V. Las secuencias se alinearon y todos los clones únicos se seleccionaron. Las alineaciones múltiples de las secuencias obtenidas mostraron la condición única de cada clon particular y se permitió para identificación de anticuerpos únicos. Después del análisis de secuencia de 220 clones, se identificaron 70 grupos de secuencias de anticuerpos genéticamente distintos. Cada grupo de secuencias relacionadas probablemente se han derivado a través de hipermutaciones somáticas de un clon precursor común. En general, uno o dos clones de cada grupo se seleccionan para validación de secuencia y especificidad.

#### 10 **Validación de secuencia y especificidad**

Con el fin de validar los clones que codifican para anticuerpo, se prepara ADN plasmídico y se realiza para expresión transfección de células CHO-S FreeStyle (Invitrogen) a escala de 2 ml. El sobrenadante se cosecha 96 horas después de la transfección. Se calcularon los niveles de expresión con la prueba de ELISA anti-IgG estándar y se determinó la especificidad por ELISA específico para EGFR o para EGFRvIII. Se demostró que el 85% de los clones tienen la especificidad y secuencia correctas.

#### 10 **Cribado para efectos antiproliferativos**

El daño celular inevitablemente resultará en pérdida de la capacidad de las células para mantener y proporcionar energía para las funciones celulares metabólicas y el crecimiento. Los análisis de actividad metabólica se basan en esta premisa. Habitualmente se mide la actividad mitocondrial. El reactivo de proliferación celular WST-1 (Roche Cat. N° 11 644 807 001) es un sustrato listo para usarse el cual mide la actividad metabólica de células viables. Después se supone que la actividad metabólica se correlaciona con el número de células viables. En este ejemplo

se utiliza el análisis WST-1 para medir el número de células metabólicamente activas después del tratamiento con sobrenadantes de cultivo celular que contienen diferentes anticuerpos anti-EGFR.

Antes de realizar el análisis WST-1 se transfieren diferentes volúmenes de sobrenadantes de 2 ml (0, 10, 25, 50 y 150 µl) a pocillos apropiados en una placa de 96 pocillos.

5 Después se lavan células HN5 con PBS1x y se separan por tripsinización con 3 ml de solución de tripsina. Después se agregan a las células 17 ml de medio completo y se sedimentan por centrifugación a 300 x g (1200 rfc) durante 5 min. Se separa el sobrenadante y las células se resuspenden en DMEM + FBS 0.5%. Se cuentan las células y se ajusta su concentración y se agregan 1500 células de los pocillos con sobrenadantes de manera que cada pocillo contiene 200 µl de medio en total. Las placas se incuban durante 4 días en un incubador humidificado a 37°C.  
 10 Después se agregan 20 µl de reactivo WST-1 por pocillo y las placas se incuban durante una hora a 37°C. Las placas después se transfieren a un agitador de placa orbital y se dejan otra hora. Se mide la absorbancia a 450 y 620 nm (longitud de onda de referencia) en un lector ELISA. La diferencia en los niveles de células metabólicamente activas (MAC) se calcula como el porcentaje de los sobrenadantes control, como sigue:

$$15 \quad \% \text{ de MAC} = \left( 1 - \frac{(DO_{exp.} - DO_{media})}{DO_{no \text{ tratado}} - DO_{media}} \right) \times 100$$

Estos valores después se utilizan como la base del análisis de grupos jerárquicos supervisados (agrupados en base a la reactividad en ELISA) realizadp utilizando el programa libre Cluster y TreeView.

20 Es preferible ser capaces de cribar para la búsqueda de anticuerpos funcionales en la etapa temprana en el procedimiento de selección de anticuerpos. Los sobrenadantes de cultivo de 83 transfecciones de 2 ml se utilizan para el cribado para funciones inhibitoras de crecimiento en un análisis de proliferación realizado utilizando células HN5 en FBS al 0,5%. Los resultados se visualizan por análisis de grupo jerárquico simple. Como se puede ver en el análisis de grupo (figura 7) se encontró que varios sobrenadantes disminuyen el número de células HN5 metabólicamente activas (gris oscuro) de una manera que depende de la concentración (grupo 2). De manera similar, algunos sobrenadantes incrementaron el número de células HN5 metabólicamente activas (gris claro) de una manera que depende de la concentración (grupos 1, 3 y 4). Una observación interesante es que los sobrenadantes, los cuales disminuyeron el número de células HN5 metabólicamente activas, presentaron una reactividad 2 (flechas negras) mientras que los sobrenadantes los cuales incrementaron el número de células HN5 metabólicamente activas presentaron una reactividad 1 (flechas grises). Los sobrenadantes con reactividad 2 son positivos para las pruebas de ELISA para wtEGFR como para EGFRvIII, mientras los sobrenadantes con reactividad 1 únicamente tienen reactividad hacia wtEGFR. De esta manera, el análisis puede proporcionar relaciones entre la reactividad de anticuerpo en ELISA y la funcionalidad en análisis celulares.

### Reparación de clon

35 Cuando se utiliza una solución de PCR múltiplex, se espera cierto grado de cebado cruzado dentro de la familia o entre la familia de genes V debido a degeneración del cebador y al alto grado de homología. El cebado cruzado introduce aminoácidos que no se encuentran de manera natural en la infraestructura de inmunoglobulina con varias consecuencias potenciales, por ejemplo, cambios estructurales e inmunogenicidad aumentada, todo lo que resulta en una actividad terapéutica disminuida.

40 Con el fin de eliminar estos inconvenientes y asegurar que los clones seleccionados reflejen la respuesta inmunitaria humoral natural, se corrigen las mutaciones de cebado cruzado en un procedimiento que se denomina reparación de clon.

45 En la primera etapa del procedimiento de reparación de clon, la secuencia V<sub>H</sub> se amplifica por PCR con un conjunto de cebador que contiene la secuencia que corresponde al gen V<sub>H</sub> del clon de interés del cual se origina, por lo que se corrige cualquier mutación introducida por cebado cruzado. El fragmento de PCR se digiere con XhoI y Ascl y se liga de regreso en un vector de expresión de mamífero digerido con XhoI/Ascl (figura 6) utilizando procedimientos convencionales de ligación. El vector ligado se amplifica en *E. coli* y el plásmido se purifica por procedimientos convencionales. La secuencia V<sub>H</sub> se secuencia para verificar la corrección y el vector se digiere con NheI/NotI para prepararlo para inserción de la cadena ligera.

50 En la segunda etapa, la cadena ligera completa se amplifica por PCR con un conjunto de cebador que contiene la secuencia que corresponde al gen V<sub>L</sub> del clon de interés del cual se origina, por lo que se corrige cualquier mutación introducida por cebado cruzado. El fragmento de PCR se digiere con NheI/NotI y se liga en el vector que contiene V<sub>H</sub> preparado en lo anterior. El producto de ligación se amplifica en *E. coli* y el plásmido se purifica por procedimientos estándar o convencionales. Posteriormente se secuencia la cadena ligera para verificar la corrección.

55 En el caso en donde la región constante κ de un clon seleccionado contiene mutaciones, introducidas durante la amplificación de los genes, se sustituye por una región constante no mutada. Esto se realiza en una PCR de superposición en donde el gen V<sub>L</sub> reparado (amplificado sin la región constante) se fusiona a una región constante con la secuencia correcta (obtenida en una PCR separada). La secuencia completa se amplifica y clona en el vector que contiene V<sub>H</sub> como se desvela en lo anterior y la cadena ligera reparada se secuencia para verificar la corrección.

Tabla 4 Procedimientos de inmunización utilizados para generar material de partida para clonación anti-EGFR

Procedimiento, grupo de ratones	Inicio	Inyección 1	Inyección 2	Inyección 3	Inyección 4	Finalización
101	Balb/c	Día 1, 25 µg de rhEGFR (R&D systems 1095-ER) CFA s.c.	Día 35, 25 µg de rhGH-EGFR (Symphogen) IFA s.c.	Día 56, 25 µg de rhEGFR (Symphogen) IFA s.c.	Día 70, 25 µg de rhEGFR* (Symphogen) IFA s.c.	Día 73
108	Balb/c	Día 1, 1 x 10 <sup>7</sup> células HN5 CFA i.p.	Día 28, 25 µg de rhEGFR* (Symphogen) IFA s.c.	Día 42, 1 x 10 <sup>7</sup> células HN5 IFA i.p.	Día 56, 25 µg de rhEGFR* (Symphogen) IFA s.c.	Día 59
109	Balb/c	Día 1, 1 x 10 <sup>7</sup> células HN5 CFA i.p.	Día 28, 25 µg de rhEGFR* (Symphogen) IFA s.c.	Día 42, 1 x 10 <sup>7</sup> células HN5 IFA i.p.	Día 56, 25 µg de rhEGFR* (Symphogen) PBS i.v.	Día 59
111	Balb/c	Día 1, 25 µg de rhEGFR* (Symphogen) CFA s.c.	Día 28, 25 µg de rhEGFR+ rhEGFRvIII* * (Symphogen) IFA s.c.	Día 42, 25 µg de rhEGFR+ rhEGFRvIII* * (Symphogen) IFA s.c.	Día 56, 25 µg de rhEGFR+ rhEGFRvIII* * (Symphogen) IFA s.c.	Día 59
118	Balb/c	Día 1, 1 x 10 <sup>7</sup> células HN5 CFA i.p.	Día 29, 100 µg de rhGH-EGFR (Symphogen) IFA s.c.	Día 44, 1 x 10 <sup>7</sup> células HN5 IFA i.p.	Día 58, 25 µg de rhEGFR (Sigma) E3641 IFA s.c.	Día 61
119	Balb/c	Día 1, 1 x 10 <sup>7</sup> células HN5 CFA i.p.	Día 29, 100 µg de rhGH-EGFR (Symphogen) IFA s.c.	Día 44, 1 x 10 <sup>7</sup> células HN5 IFA i.p.	Día 58, 25 µg de rhEGFR (Sigma) E3641 IFA s.c.	Día 61

Tabla 5 Conjunto de cebador incrustado

Nombre del cebador	Concentración (nM)	Secuencia	Sec. No
mHCrev1-	0,2	GGACAGGGMTCCAKAGTTCCADKT	18
<b>conjunto hmJK</b>			
hmJK1-v2	0,2	GACAGATGGTGCAGCCACAGTTCGTTTGATTTCCAG	19
hmJK2-v2	0,2	GACAGATGGTGCAGCCACAGTTCGTTTTATTTCCAGC	20
hmJK4-v2	0,2	GACAGATGGTGCAGCCACAGTTCGTTTTATTTCCAAC	21
hmJK5-v2	0,2	GACAGATGGTGCAGCCACAGTTCGTTTCAGCTCCAG	22

K = G/T, M=A/C, D=AGT; Conc. - Concentración final.

5 Tabla 6 Conjunto de cebador de empalme constante Kappa

Cebador	Concentración	Secuencia	Sec. No
<b>Amplificación constante kappa humana</b>			
hKCforw-	0,2	GAAGTGTGGCTGCACCATCTGTG	23
Kappa3'	0,2	ACCGCCTCCACCGCGCCGCTTATTAACACTCTCCCCTGTT G	24

(cont.)

**Empalme por extensión de superposición**

mhKCreV	0,2	ACCGCCTCCACCGGCGGCCGCTTATTAACACTCT CCCCTGTTGAAGCTCTT	25
<b>conjunto mJH</b>			
mJH1	0,2	GGAGGCGCTCGAGACGGTGACCGTGGTCCC	12
mJH2	0,2	GGAGGCGCTCGAGACTGTGAGAGTGGTGCC	13
mJH3	0,2	GGAGGCGCTCGAGACAGTGACCAGAGTCCC	14
mJH4	0,2	GGAGGCGCTCGAGACGGTGACTGAGGTTCC	15

**Apéndice 1, Secuencia de la región constante de anticuerpo**

>Región IGKC humana (Sec. Nº 26)

t t c a t c t t c c c g c c a t c t g a t g a g c a g t t g a a a t c t g g a a c t g c c t c t g t t g t g t g c c t g c t g a a t a a c t  
t c t a t c c c a g a g a g g c c a a a g t a c a g t g g a a g t g g a t a a c g c c c t c c a a t c g g g t a a c t c c c a g g a g a g  
t g t c a c a g a g c a g g a c a g c a a g g a c a g c a c c t a c a g c c t c a g c a g c a c c c t g a c g c t g a g c a a a g c a g a c  
t a c g a g a a a c a c a a a g t c t a c g c c t g c g a a g t c a c c c a t c a g g g c c t g a g c t c g c c c g t c a c a a a g a g c t  
t c a a c a g g g g a g a g t g t t a a t a a g c g g c c g c c g g t g g a g g c g g t

5 >Región IGKC humana (Sec. Nº 27)

TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDS  
KDSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

Exón1 1..298

Intrón 299..689

10 Exón 2 690..734

Intrón 735..852

Exón 3 853..1182

Intrón 1183..1279

Exon 4 1280..1602

15 >Secuencia genómica de dominio constante IGHG1 humana (Sec. Nº 28)

ES 2 369 946 T3

agtgcctccaccaagggcccacgtcttccccctggcaccctcctccaagagcacctctgggggacag  
cggccctgggctgctggtcaaggactacttcccgaaccgggtgacgggtcgtggaactcaggcgcct  
gaccagcggcgtgcacaccttcccggctgtcctacagtcctcaggactctactccctcagcagcgtggtg  
accgtgccctccagcagcttgggacaccagacctacatctgcaacgtgaatcacaagcccagcaacca  
aggtggacaagagagtgggtgagagggccagcagggaggagggtgtctgctggaagccaggctcagcg  
ctcctgctggacgcatcccggctatgcagtcctcagtcagggcagcaaggcaggccccgtctgctctt  
caccggaggcctctgcccgcccaactcatgctcagggagagggtcttctggtttttccccaggctctg  
ggcaggcacaggctaggtgcccctaaccaggccctgcacacaaagggcagggtgctgggctcagacctg  
ccaagagccatataccgggaggacctgcccctgacctaaagcccaagccaaactctcactccc  
tcagctcggacaccttctctcctccagattccagtaactcccaatcttctctctgagagcccaatct  
tgtgacaaaactcacacatgcccacgtgcccaggtaagccagccaggcctcgccctccagctcaaggc  
gggacaggtgccctagagttagctgcatccagggacaggccccagccgggtgctgacacgtccacctcca  
tctcttctcagcacctgaactcctgggggaccgtcagctcttctcttcccccaaaacccaaggcac  
cctcatgatctcccggacccctgaggtcacatgctggtggtggacgtgagccacgaagacctgaggtc  
aagttcaactggtacgtggacggcgtggaggtgcataatgccaagacaaagccggggaggagcagta  
acagcagctaccgtgtggtcagcgtcctcaccgtcctgcaccaggactggctgaatggcaaggagta  
gtgcaaggtctccaacaaagccctcccagccccatcgagaaaaccatctccaaagccaaaggtgggacc  
cgtggggtgcgagggccacatggacagagggcggctcggcccaccctctgacctgagagtgacctgta  
ccaacctctgtccctacagggcagccccgagaaccacaggtgtacacctgccccatcccgggaggaga  
tgaccaagaaccaggtcagcctgacctgctggtcaaaggcttctatcccagcagatcgccgtggagtg  
ggagagcaatgggagccgggagaactacaagaccacgctcccgtgctggactccgacggctccttc  
ttcctctatagcaagctcaccgtggacaagagcaggtggcagcaggggaacgtcttctcatgctccgta  
tgcagaggtctctgcacaaccactacacgcagaagagcctctcctgtccccgggtaaatga

>IGHG1 (Sec. N° 29)

SASTKGPSVFPLAPSSKSTSGGTAALGLVKDYFPEPVTVSWNSGALTSVHTFPAVLQSSGLYSLSSVVT  
VPSSSLGTQTYICNVNHKPSNTKVDKRVKPKSCDKHTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRT  
EVTCTVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA  
LPAPIEKTIISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPV  
LDSGDGFFLYSKLTVDKSRWQQGNVFSQSVMEALHNHYTQKSLSLSPGK

**REIVINDICACIONES**

1. Un procedimiento para producir una biblioteca de pares análogos que comprenden secuencias que codifican para la región variable enlazada, comprendiendo dicho procedimiento:
  - a) proporcionar una fracción de células que comprende linfocitos de un donante;
  - 5 b) obtener una población de células solas aisladas distribuyendo células de la fracción celular individualmente en una pluralidad de recipientes, en donde por lo menos una subpoblación de células están enriquecidas y expresan antígenos CD43 y CD138 o antígenos MHCII y B220; y
  - 10 c) amplificar y llevar a cabo el enlace de las secuencias codificantes de región variable contenidas en dicha población de células solas aisladas por amplificación, en un procedimiento de amplificación molecular múltiplex, secuencias nucleótidas de interés utilizando una plantilla derivada de una sola célula aislada o una población de células isogénicas; y llevar a cabo el enlace de las secuencias nucleótidas de interés amplificadas.
2. El procedimiento de conformidad con la reivindicación 1, en el que la fracción de células que contienen linfocitos comprende esplenocitos, sangre entera, médula ósea, células mononucleares o leucocitos, preferiblemente esplenocitos o médula ósea, más preferiblemente esplenocitos; y opcionalmente en el que la fracción de células que contienen linfocitos o línea de linfocitos B se enriquece en plasmocitos o plasmablastos.
3. El procedimiento de conformidad con las reivindicaciones 1 ó 2, en el que las secuencias nucleótidas de interés comprenden secuencias que codifican para una región variable de inmunoglobulina y el enlace genera un par análogo de una secuencia que codifica para una región variable de cadena ligera asociada con una secuencia que codifica para una región variable de cadena pesada.
- 20 4. Un procedimiento de enlace aleatoriamente de una pluralidad de secuencias nucleótidas no contiguas de interés, en el que dichas secuencias nucleótidas de interés comprenden secuencias que codifican para una región variable y el enlace genera una biblioteca combinatoria de pares de secuencias que codifican para una región variable, preferiblemente en el que dichas secuencias nucleótidas de interés comprenden secuencias que codifican para la región variable de inmunoglobulina y el enlace genera una biblioteca combinatoria de pares de secuencias que codifican para una región variable de cadena ligera y la región variable de cadena pesada, comprendiendo dicho procedimiento:
  - a) amplificar, en un procedimiento de amplificación molecular múltiplex, secuencias nucleótidas de interés utilizando una plantilla derivada de una población de células genéticamente diversas;
  - 30 en el que las células genéticamente diversas se derivan de una fracción de células que comprenden linfocitos de un donante roedor;
  - y en el que al menos una subpoblación de las células están enriquecidas en y expresan los antígenos CD43 y CD138 o los antígenos MHCII y B220; y
  - b) llevar a cabo el enlace de las secuencias nucleótidas de interés amplificadas en la etapa a).
- 35 5. El procedimiento de cualquiera de las reivindicaciones precedentes, que comprende además determinar, antes de la amplificación molecular múltiplex que la población de células que comprenden linfocitos comprende células que expresan los antígenos CD43 y CD138 o los antígenos MHCII y B220, preferiblemente CD43 y CD138.
6. El procedimiento de cualquiera de las reivindicaciones precedentes, que comprende además enriquecer dicha fracción de células que comprenden linfocitos para una población de linfocitos que expresan los antígenos CD43 y CD138 o MHCII y B220 antes de amplificación molecular múltiplex.
- 40 7. El procedimiento de cualquiera de las reivindicaciones precedentes, que comprende además aislar de dicha población que comprende linfocitos células que expresan los antígenos CD43 y CD138 o los antígenos MHCII y B220 antes de amplificación molecular múltiplex.
8. El procedimiento de cualquiera de las reivindicaciones precedentes, en el que el enriquecimiento o aislamiento comprende un procedimiento de clasificación automatizado MACS, FACS u otro.
- 45 9. El procedimiento de cualquiera de las reivindicaciones precedentes, en el que el donante es un ratón.
10. El procedimiento de la reivindicación 9, en el que el ratón es un ratón transgénico.
11. El procedimiento de conformidad con cualquiera de las reivindicaciones precedentes, en el que dicho procedimiento de amplificación molecular múltiplex es una amplificación RT-PCR múltiplex, y en el que:
  - 50 dicha amplificación RT-PCR múltiplex es un procedimiento de dos etapas que comprende una etapa de transcripción inversa separada antes de la amplificación PCR múltiplex, o
  - en el que dicha amplificación RT-PCR múltiplex se realiza en una etapa única que comprende inicialmente agregar la totalidad de los componentes necesarios para realizar tanto transcripción inversa como amplificación por PCR múltiplex en un solo recipiente.
- 55 12. El procedimiento de conformidad con cualquiera de las reivindicaciones precedentes, en el que dicho enlace de las secuencias nucleótidas de interés se lleva a cabo en asociación con la amplificación PCR múltiplex, utilizando una mezcla de cebador de superposición-extensión múltiplex.

13. El procedimiento de conformidad con la reivindicación 12, en el que la mezcla de cebador de superposición-extensión múltiplex comprende:
- a) por lo menos un cebador mKappa1 o hmJK complementario a la cadena directa de una secuencia que codifica para la región de cadena ligera de inmunoglobulina;
  - 5 b) por lo menos un cebador mVK complementario a la cadena antisentido de una secuencia que codifica para la región variable de cadena ligera de inmunoglobulina o la secuencia líder de la región variable de cadena ligera y capaz de formar un conjunto de cebador con uno o varios de los cebadores en la etapa a);
  - c) por lo menos un cebador mCHrev1, mCHrev1-ext o mJH complementario a la cadena directa de una secuencia que codifica para el dominio de cadena pesada de inmunoglobulina; y
  - 10 d) por lo menos un cebador mVH complementario a la cadena antisentido de una secuencia que codifica para la región variable de cadena pesada de inmunoglobulina o la secuencia líder de región variable de cadena pesada, y capaz de formar un conjunto de cebador con uno o varios de los cebadores en la etapa c).
14. El procedimiento de conformidad con cualquiera de las reivindicaciones precedentes, que comprende además insertar las secuencias nucleótidas enlazadas o una biblioteca de pares análogos en un vector; y que además comprende opcionalmente las etapas de:
- a) introducir un vector que codifica para un segmento de secuencias nucleótidas enlazadas en una célula hospedadora;
  - b) cultivar dichas células hospedadoras bajo condiciones adaptadas por expresión; y
  - c) obtener el producto de proteína expresado del vector insertado en dicha célula hospedadora.
15. El procedimiento de conformidad con la reivindicación 14, en el que las secuencias nucleótidas enlazadas o los elementos individuales de la biblioteca de pares análogos comprenden una secuencia que codifica para una región variable de la cadena pesada de inmunoglobulina asociada con una secuencia que codifica para una región variable de la cadena ligera y dichas secuencias se insertan en marco en un vector que contiene las secuencias que codifican uno o más dominios constantes de inmunoglobulina o fragmentos de los mismos.
16. El procedimiento de conformidad con la reivindicación 15, que además comprende transferir a dicho par análogo o biblioteca de pares análogos específicos de diana de secuencias que codifican para una región variable a un vector de expresión de mamífero, en el que el vector de expresión de mamífero opcionalmente codifica para uno o más dominios de región constante seleccionados de las clases de inmunoglobulina humana IgA1, IgA2, IgD, IgE, IgG1, IgG2, IgG3, IgG4, IgM, cadena ligera kappa o cadena pesada lambda.

30

1/7

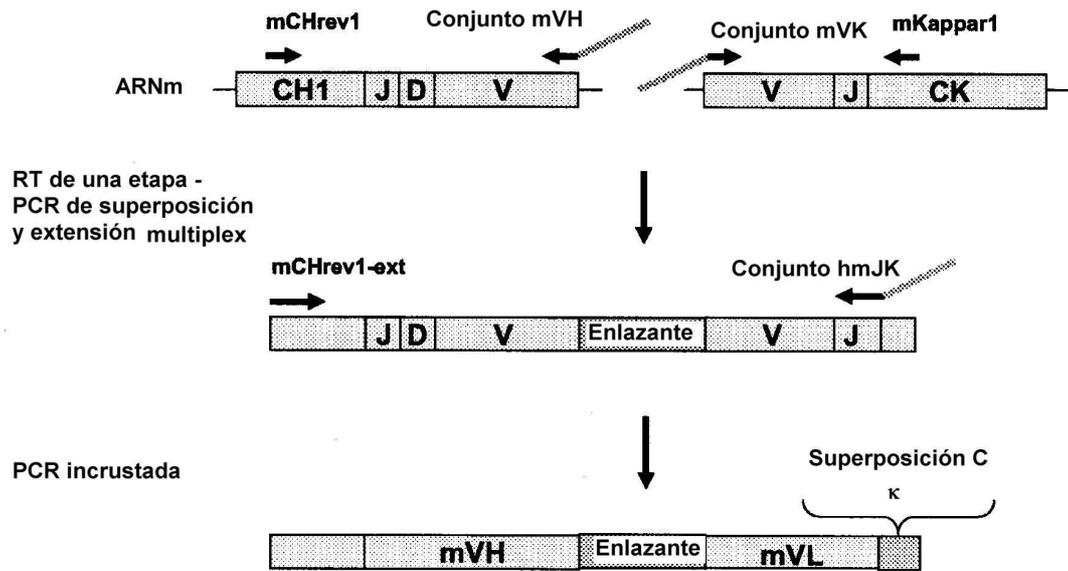


Fig. 1

2/7

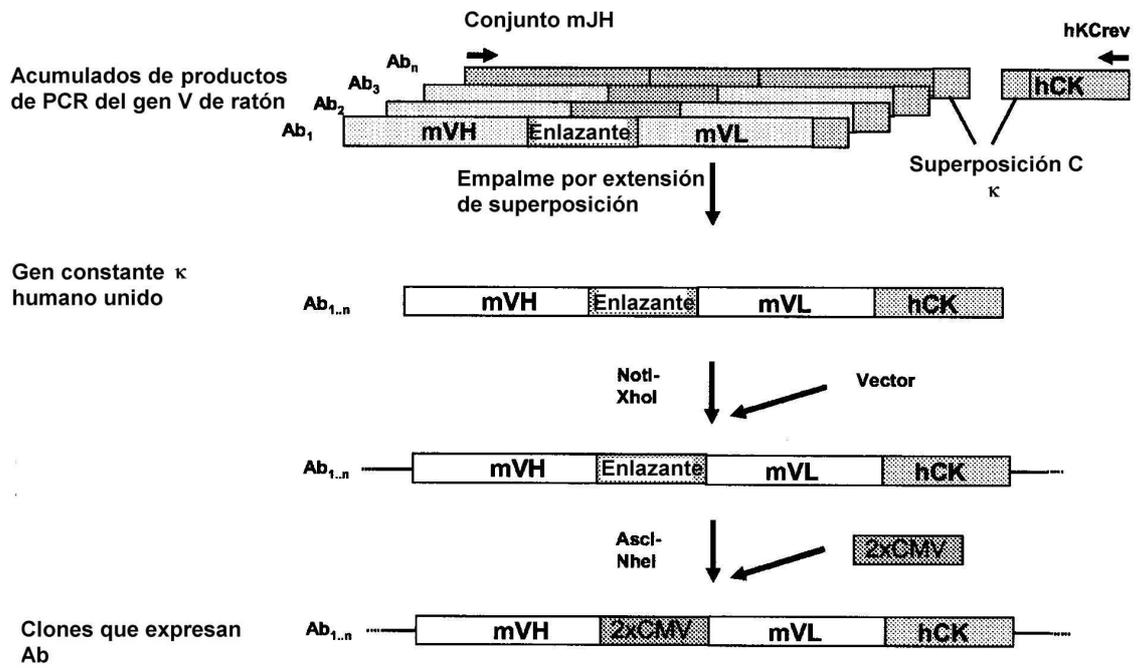


Fig. 2

3/7

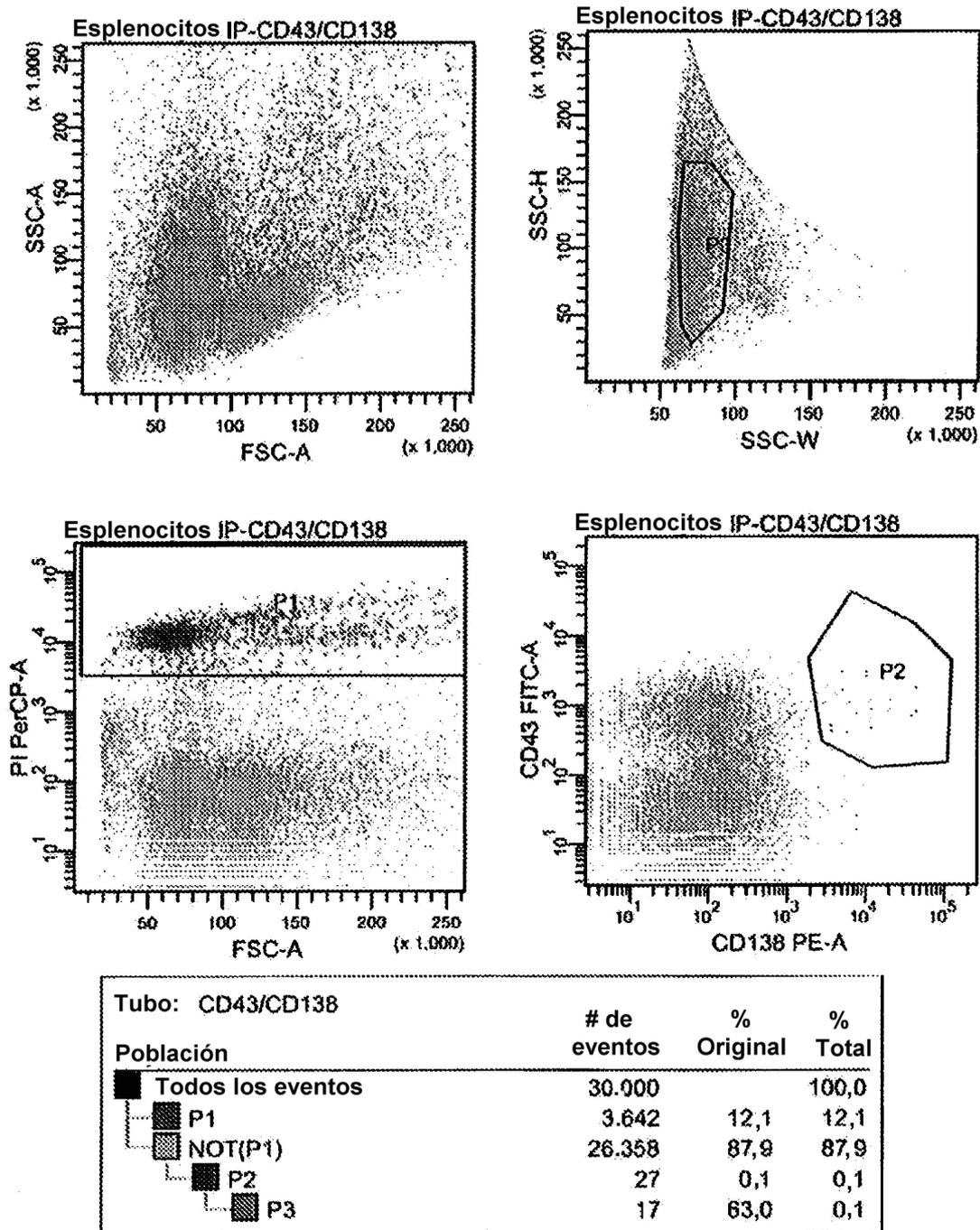


Fig. 3A

4/7

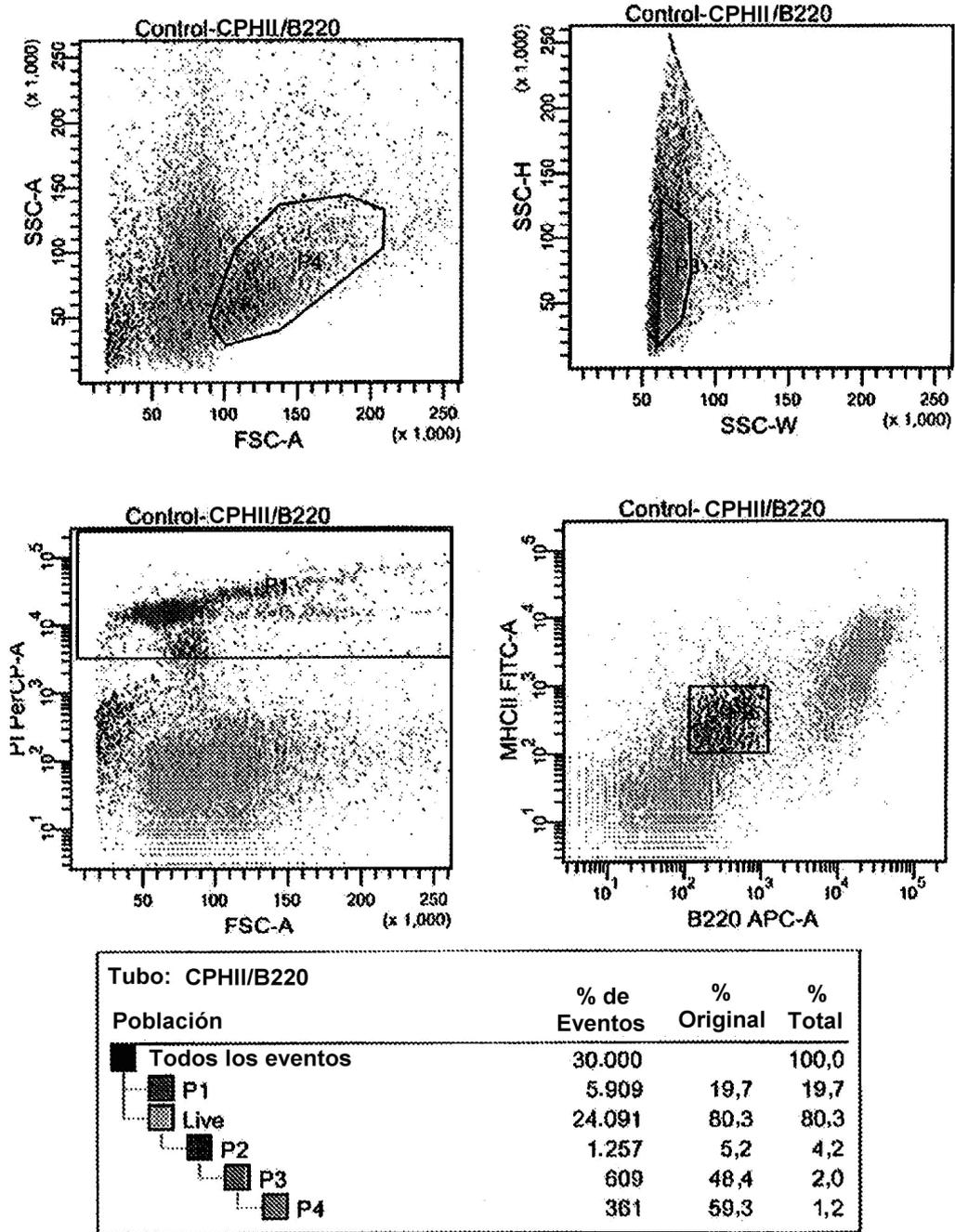


Fig. 3B

5/7

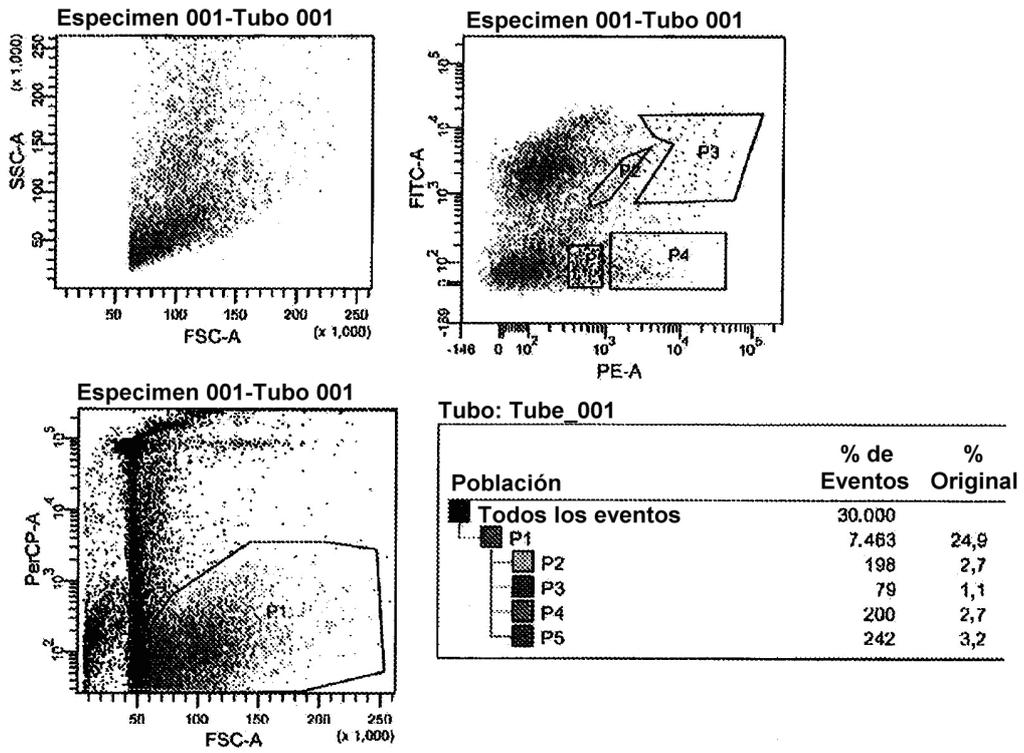


Fig. 4

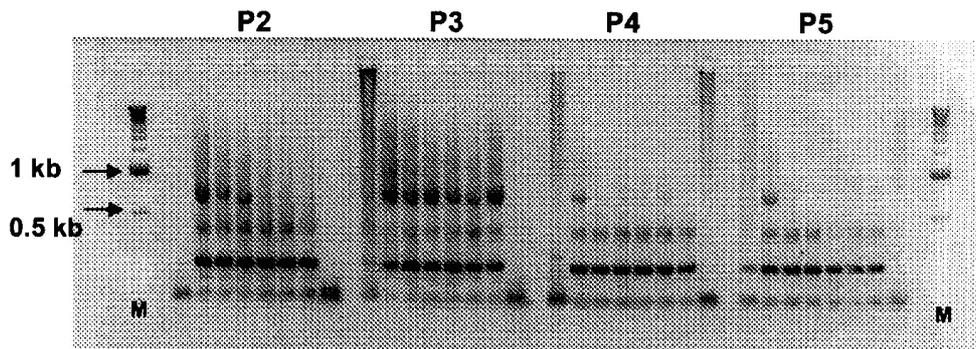


Fig. 5

6/7

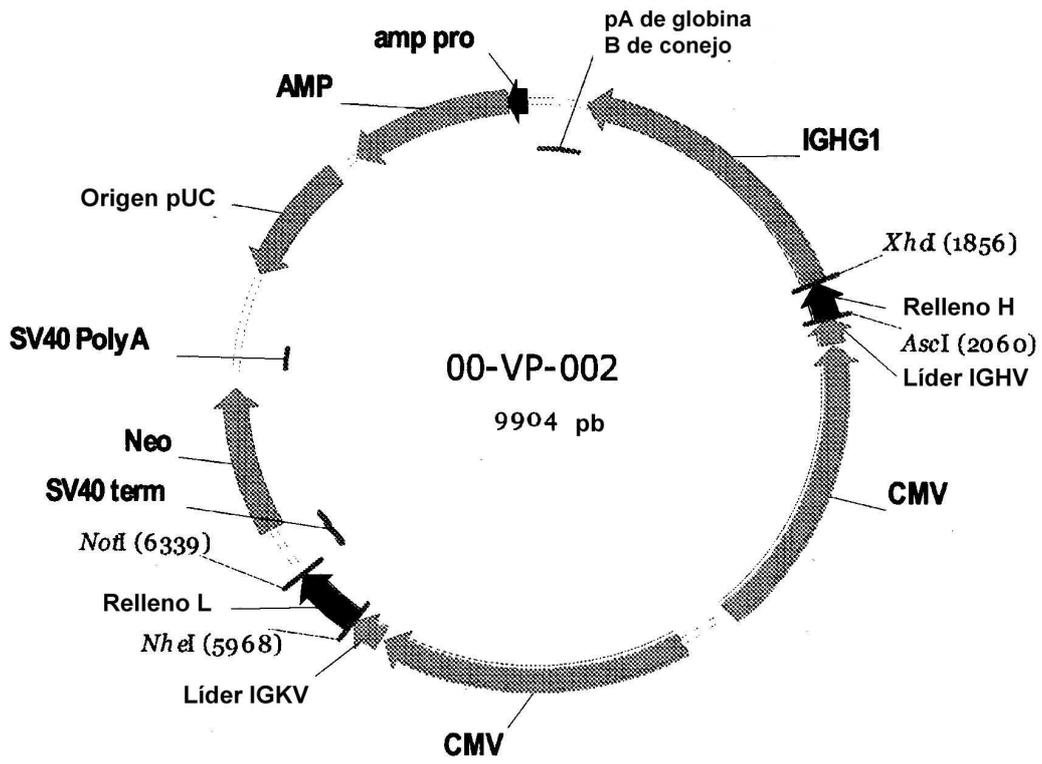


Fig. 6

7/7

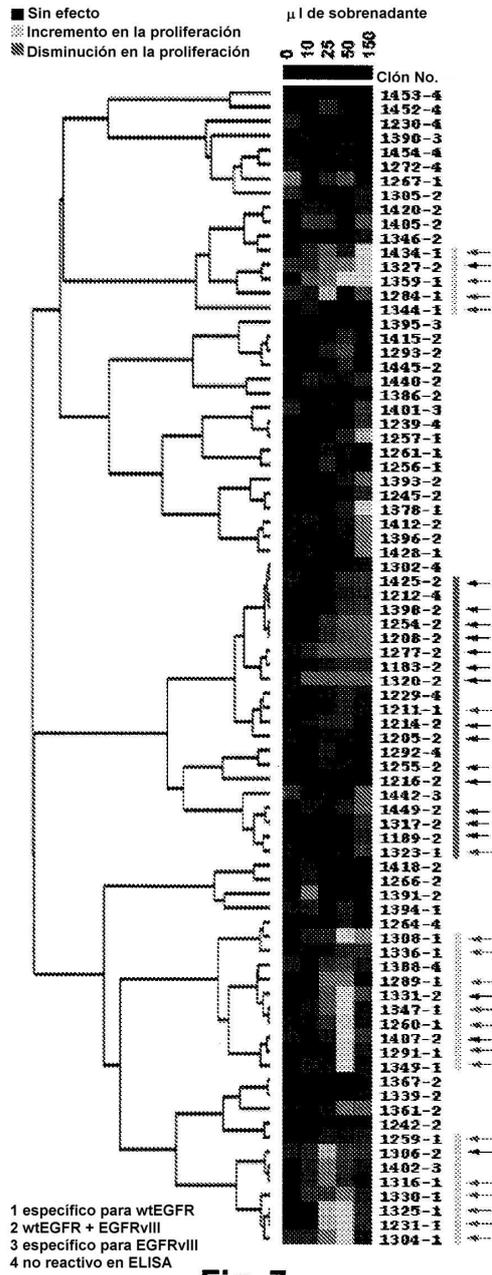


Fig. 7

Fig. 7