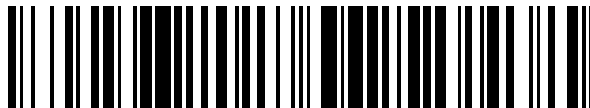


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 541 934**

51 Int. Cl.:

**G01N 33/53** (2006.01)

**G01N 33/68** (2006.01)

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.02.2009 E 09710996 (1)**

97 Fecha y número de publicación de la concesión europea: **15.04.2015 EP 2243030**

54 Título: **Métodos para la identificación de un anticuerpo o una diana**

30 Prioridad:

**11.02.2008 EP 08151276**

**11.02.2008 US 27507**

**15.04.2008 US 45039**

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

**28.07.2015**

73 Titular/es:

**MORPHOSYS AG (100.0%)**  
**Lena-Christ-Strasse 48**  
**82152 Planegg-Martinsried, DE**

72 Inventor/es:

**ENZELBERGER, MARKUS**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 541 934 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**

Métodos para la identificación de un anticuerpo o una diana

**CAMPO**

5 Esta descripción se refiere a métodos para la identificación de un anticuerpo, una molécula diana o un agente mediante el análisis de los datos de la secuencia de inmunoglobulinas en una muestra biológica y mediante la determinación de las cadenas VH y VL más abundantes presentes en dicha muestra biológica. También se proporcionan los materiales utilizados para ello.

**ANTECEDENTES**

10 Los métodos actuales para la identificación de un anticuerpo o una molécula diana implican procesos laboriosos de aislamiento de anticuerpos a partir de linfocitos B humanos activados. Por ejemplo, el aislamiento de anticuerpos completamente humanos a partir de linfocitos B de pacientes inmunizados o con cáncer se considera una ruta ventajosa para los anticuerpos completamente humanos. Diversas compañías ofrecen servicios comerciales para el aislamiento de linfocitos B individuales procedentes de seres humanos. Estos linfocitos B se immortalizan o se recupera la información genética de las inmunoglobulinas de las células individuales. Tales métodos pueden implicar técnicas de alto rendimiento que son laboriosas y costosas, incluyendo técnicas para el aislamiento de información genética célula a célula, immortalizando miles de células y escrutando su rendimiento respectivo sobre el tejido diana.

15 Desde hace poco tiempo, diversas compañías ofrecen servicios de secuenciación de todo el genoma, o aparatos que se pueden utilizar para realizar las tareas respectivas. Esto incluye el sistema 454 de Roche, el sistema Solexa de Illumina y el sistema Heliscope de Helicos Biosciences. Helicos, por ejemplo, puede secuenciar 2x10<sup>9</sup> bases en 20 24 horas con un solo aparato, manteniendo también una distribución cuantitativa de las secuencias diana.

25 El documento de Patente de Estados Unidos nº 7.288.249 describe un método para la identificación de un antígeno que se expresa diferencialmente en la superficie de dos o más poblaciones celulares distintas. La inmunización activa los linfocitos B para producir una combinación VH-VL que se une al inmunógeno para proliferar (expansión clonal) y para secretar el anticuerpo correspondiente. Sin embargo, el procedimiento de acuerdo con el documento US 7.288.249 implica la clonación de los genes VH y VL (los genes VH y VL se clonan separadamente mediante reacción en cadena de la polimerasa (PCR)), y los genes VH y VL se recombinan aleatoriamente en genotecas de fagos (es decir, no hay una selección de los genes VH y VL más abundantes), entre las cuales se busca a continuación los clones que se unen a un antígeno, tal y como se describe en Winter et al., Ann. Rev. Immunol., 12: 433-455 (1994). El ácido nucleico que codifica segmentos génicos variables de anticuerpo (incluyendo los segmentos VH y VL) se recupera a partir de las células de interés y se amplifica. En el caso de genotecas de VH y VL reordenadas, el ADN deseado se obtiene mediante el aislamiento de ADN genómico o ARNm a partir de linfocitos, seguido por una reacción en cadena de la polimerasa (PCR) con cebadores que se emparejan con los extremos 5' y 3' de genes de VH y VL reordenados. Para encontrar anticuerpos útiles, la genoteca de fagos con anticuerpos no activados se escruta en busca de células cancerosas vivas.

35 El documento de Patente de Estados Unidos nº 6.897.028 describe un método para la identificación de dianas moleculares en el que una proteína se une a un ligando, el ligando se escruta frente a una genoteca de péptidos o proteínas en donde los miembros peptídicos o proteínicos de la genoteca se seleccionan entre los productos de expresión de una genoteca de ADNc obtenida a partir de una célula y de fragmentos de esos productos de expresión. El procedimiento también implica la determinación de la secuencia de ácido nucleico que codifica los miembros que se han separado de la genoteca y la traducción de estas secuencias de ácidos nucleicos en secuencias peptídicas y la 40 identificación de la proteína.

45 El documento de Publicación nº 20060141532 (Solicitud con nº de Serie 11/286917) describe métodos para la identificación y el diseño de péptidos inmunogénicos mediante el uso de un protocolo para determinar la secuencia de aminoácidos de ciertas regiones VH o VL de un anticuerpo anti-idiotípico, tal y como se describe en Iwasaki, et al. Eur. J. Immunol., 24: 2874-2881, 1994. La secuencia de aminoácidos del antígeno se determina por técnicas convencionales de análisis de aminoácidos o mediante secuenciación química; la secuencia de aminoácidos de las regiones VH y/o VL del anticuerpo anti-idiotípico se determina mediante secuenciación del ADN genómico o el ADNc que codifica la región respectiva, de acuerdo con métodos conocidos en la técnica.

50 El documento WO2005/094159 describe el aislamiento de péptidos de unión a partir de linfocitos immortalizados y el análisis de estos péptidos de unión, por lo general inmunoglobulinas del tipo IgM, para estudiar la unión selectiva a tejido tumoral, pero no a tejido sano. El péptido de unión puede inhibir la proliferación de tumores.

55 El documento WO03/052416 describe una metodología para el aislamiento y la secuenciación de secuencias candidatas de VH y VL procedentes de linfocitos B. Las inmunoglobulinas codificadas por dichas secuencias candidatas de VH y VL son potencialmente útiles en el tratamiento de infecciones. La metodología descrita en el documento WO03/052416, al igual que todos los demás métodos descritos en la técnica anterior, requiere la manipulación de los ácidos nucleicos antes de la secuenciación. En particular, son necesarias etapas de clonación más largas lo que hace imposible poner en práctica el método a una escala tal y como se describe en la presente invención.

5 Como en el caso del documento WO03/052416, todos los métodos mencionados anteriormente implican un procedimiento de aislamiento laborioso de los ácidos nucleicos que codifican una diana y/o una secuenciación genómica mediante técnicas de rendimiento muy elevado. Sin embargo, sería ventajoso un método menos laborioso, menos costoso y/o más factible para la identificación de una molécula diana, analizando datos de la secuencia de las cadenas VH y VL y determinando la cadena VH y VL más dominante.

### **COMPENDIO**

10 Las realizaciones de esta descripción se refieren en general a métodos para la identificación de un anticuerpo, una molécula diana o un agente mediante el análisis de los datos de la secuencia de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas presentes en una muestra biológica. Algunas de las realizaciones en el presente documento se refieren a la determinación de las cadenas VH y VL más abundantes, presentes en una muestra, utilizando un algoritmo predeterminado implementado por ordenador, sintetizando polinucleótidos de las cadenas VH y VL más abundantes para la expresión en un vector, y sometiendo a ensayo los anticuerpos expresados para identificar la molécula diana. Otras realizaciones se refieren a la determinación de las cadenas VH y VL más abundantes presentes en una muestra, utilizando un algoritmo predeterminado implementado por ordenador, sintetizando polinucleótidos de las cadenas VH y VL más abundantes para la expresión en un vector, y sometiendo a ensayo cuál de los anticuerpos expresados se une a una cierta molécula diana o tejido diana. En ciertas realizaciones, la muestra es una muestra biológica que no está seleccionada o enriquecida previamente.

20 Una realización proporciona métodos de identificación de un anticuerpo, una diana o un agente en una muestra, que comprenden: a) la obtención de los ADNc de los ARNm que codifican inmunoglobulinas en una muestra, obteniendo de este modo una mezcla de ADNc; b) la secuenciación de cada una de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas, obteniendo de este modo datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas presentes en la muestra; c) la determinación de las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo predeterminado implementado por ordenador; d) la síntesis de polinucleótidos de las cadenas VH y VL más abundantes y la producción de anticuerpos usando un vector de expresión de mamífero; y e) someter a ensayo los anticuerpos, identificando de este modo un anticuerpo que se une a una determinada molécula o tejido diana o la identificación de una molécula diana que se une a un determinado anticuerpo.

30 Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente que comprenden: a) proporcionar una muestra biológica procedente de un mamífero tal como un ser humano, murido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo que se inmuniza o infecta con un agente o una molécula diana; b) recoger los linfocitos B de la muestra biológica; c) obtener ARNm que codifican inmunoglobulinas (por ejemplo, IgGs) en los linfocitos B recogidos; d) generar los ADNc de las inmunoglobulinas (por ejemplo, mediante transcriptasa inversa-PCR) y emplear cebadores específicos de IgG para la amplificación, obteniendo de este modo una mezcla de los ADNc; e) secuenciar cada una de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas (por ejemplo, mediante la obtención de secuencias independientes de VH y VL), obteniendo de este modo datos de las secuencias de las cadenas VH y VL de las inmunoglobulinas presentes en la muestra; f) analizar los datos de las secuencias de las cadenas VH y VL para determinar las cadenas VH y VL más dominantes presentes en la muestra; g) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo predeterminado implementado por ordenador; h) sintetizar polinucleótidos de las cadenas VH y VL más abundantes; i) integrar los polinucleótidos de VH y VL sintetizados en un vector de expresión de mamífero; j) permitir que los vectores con los polinucleótidos de VH y VL integrados se expresen en un medio de cultivo, produciendo de este modo anticuerpos o fragmentos de los mismos; y k) someter a ensayo los anticuerpos mediante el uso de un inmunoensayo, identificando de este modo el anticuerpo, la molécula diana o el agente de la molécula diana. En ciertas realizaciones el mamífero es un mamífero enfermo.

45 En ciertas realizaciones la inmunoglobulina es una cualquiera entre las clases de inmunoglobulinas IgG, IgM, IgA, IgD o IgE, o un fragmento de las mismas. En algunas realizaciones, la inmunoglobulina es de la clase IgG.

En ciertas realizaciones la inmunoglobulina es una cualquiera entre las subclases de inmunoglobulinas G, IgG1, IgG2, IgG3 o IgG4.

50 Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente, en los que los ADNc de las inmunoglobulinas son generados mediante transcriptasa inversa-PCR.

Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente, en los que se utilizan cebadores específicos de IgG en la amplificación con transcriptasa inversa-PCR.

55 Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente, en los que se utilizan cebadores en la amplificación con transcriptasa inversa-PCR que son específicos de una cualquiera de las subclases de inmunoglobulinas G, IgG1, IgG2, IgG3 o IgG4.

Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente, en los que se obtienen datos de secuencias independientes para las cadenas VH y VL de las inmunoglobulinas. Los datos de las secuencias de las cadenas VH y VL se almacenan en una base de datos y las cadenas VH y VL más abundantes se

determinan a través de un algoritmo implementado por ordenador.

Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente, en los que el mamífero es un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo.

5 Otra realización proporciona métodos de identificación de un anticuerpo, una diana o un agente en una muestra que comprenden las etapas de: a) analizar los datos de las secuencias de las cadenas VH y VL presentes en una muestra; y b) determinar la abundancia de las cadenas VH y VL presentes en la muestra. Los métodos comprenden además preparar una o varias inmunoglobulinas que contienen cadenas VH y VL abundantes, identificadas en la etapa b).

10 En ciertas realizaciones el ser un humano es un paciente con cáncer, o está inmunizado o infectado con un agente o una molécula diana.

En todavía otra realización de esta descripción, se proporcionan métodos para identificar un anticuerpo, una diana o un agente, en donde el mamífero es un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo

15 **DESCRIPCIÓN DETALLADA**

Esta descripción proporciona métodos para identificar un anticuerpo, una molécula diana o un agente mediante el análisis de los datos de secuencias de inmunoglobulinas de las cadenas pesadas variables (VH) y las cadenas ligeras variables (VL) presentes en una muestra biológica, así como materiales utilizados en los métodos descritos en esta memoria. Esta descripción también se refiere a métodos para determinar las cadenas VH y VL más abundantes presentes en una muestra, utilizando un algoritmo predeterminado implementado por ordenador, sintetizando polinucleótidos de las cadenas VH y VL más abundantes para la expresión en un vector, y sometiendo a ensayo los anticuerpos expresados para identificar la molécula diana.

De acuerdo con un aspecto, esta descripción proporciona métodos para identificar o detectar un anticuerpo, una molécula diana o un agente mediante la secuenciación del "inmunonoma", que se refiere al ARNm completo (o ADNc después de la transcripción inversa-PCR) que codifica inmunoglobulinas (tales como IgGs), en una muestra de mamífero (tal como ser humano, mívrido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos), en donde el mamífero se ha inmunizado con una molécula diana, es un paciente con cáncer o está infectado con una determinada especie vírica, bacteriana o protozoaria. En ciertas realizaciones el mamífero (por ejemplo, ser humano, mívrido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos) está inmunizado con una molécula diana, tal como una vacuna. En otras realizaciones el mamífero es un paciente con cáncer.

De acuerdo con otro aspecto, los linfocitos B se aíslan a partir de un mamífero inmunizado o un paciente con cáncer, el ARNm se extrae de la muestra respectiva, y las inmunoglobulinas codificadas por dicho ARNm se transcriben de forma inversa y se secuencian. En algunas realizaciones, al menos  $10^4$  linfocitos B se aíslan a partir del mamífero inmunizado o el paciente con cáncer, en otras realizaciones, al menos  $10^5$  linfocitos B;  $10^6$  linfocitos B; o incluso  $10^7$  linfocitos B se aíslan a partir del mamífero inmunizado o el paciente con cáncer. Los genes que codifican las cadenas variables de las inmunoglobulinas tienen generalmente una longitud de aproximadamente 1000 nucleótidos. Por lo tanto, la secuenciación de aproximadamente  $2 \times 10^9$  bases de información de secuencias, abarca todo el inmunonoma al menos dos veces. Se identifican las secuencias de CDRs dominantes, así como (hiper)mutaciones de las mismas. Estas secuencias, son específicas, o al menos representativas del estímulo inmunológico. Las secuencias de la cadena pesada y ligera (que no se pueden recuperar juntas) se agrupan según la frecuencia de dominancia, es decir, por su abundancia. Las secuencias más abundantes de VH y VL se sintetizan y se transfieren (una a una o en masa) a un vector de expresión de mamífero. Los vectores se expresan en un medio adecuado y se escrutan para obtener inmunoglobulinas, o fragmentos de las mismas, en busca de especificidad hacia la diana de interés. Los aglutinantes obtenidos se someten a ensayo para estudiar su eficacia terapéutica y se pueden producir como fármacos o emplearlos para identificar nuevos anticuerpos o dianas moleculares.

Las realizaciones de esta descripción proporcionan ventajas que incluyen:

- 50 - Las realizaciones descritas en esta memoria pueden reemplazar la biología húmeda de ultra alto rendimiento por la bioinformática y la tecnología de secuenciación masiva en paralelo de moléculas, lo que permite un paralelismo del análisis de linfocitos B procedentes de muchos pacientes en plazos reducidos significativamente (por lo general solo se lleva a cabo una secuenciación aleatoria, que produce tramos comparativamente cortos de 75-500 pares de bases, y el ensamblado se realiza mediante alineación de los fragmentos obtenidos).
- 55 - La homología elevada en las regiones constantes de las inmunoglobulinas no interfiere con los algoritmos de la alineación.
- Las realizaciones descritas en esta memoria permiten evitar las etapas de clonación.
- Las realizaciones descritas en esta memoria pueden superar el problema del aislamiento de IgMs de baja

calidad empleando cebadores selectivos de IgGs, cebadores específicos de subclases de IgGs.

- Los métodos en esta memoria permiten emplear el inmunonoma para métodos de diagnóstico. Por ejemplo, pacientes que padecen alergias contra alérgenos idénticos pueden mostrar patrones de secuencias similares en su respuesta inmune. En principio, por ello, el uso de un perfil de inmunonoma como biomarcador también es factible si se obtiene un conjunto de datos de tamaño suficiente.
- El conocimiento del inmunonoma también puede facilitar el diseño de genotecas sintéticas de anticuerpos totalmente humanos, de múrido o de roedor.

De acuerdo con una realización de esta descripción, un anticuerpo, una molécula diana, un marcador de enfermedad, un agente o un agente para una molécula diana en una muestra, por ejemplo, en una muestra biológica procedente de un mamífero (por ejemplo, ser humano, múrido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos) que se inmuniza o se infecta con un agente o una molécula diana, se identifica, se determina, se detecta o se diagnostica a través de un método que comprende las etapas de: a) obtener los ADNc de los ARNm que codifican inmunoglobulinas en la muestra (por ejemplo, mediante la recogida de linfocitos B a partir de la muestra), obteniendo de este modo una mezcla de los ADNc; b) secuenciar cada una de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas, obteniendo con ello datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas (tales como IgGs) presentes en la muestra; c) determinar las cadenas VH y VL más abundantes presentes en la muestra (p. ej., mediante el análisis de los datos de la secuencia de las cadenas VH y VL presentes en la muestra usando un algoritmo predeterminado implementado por ordenador); d) sintetizar polinucleótidos de las cadenas VH y VL más abundantes y producir anticuerpos utilizando un vector de expresión de mamífero (por ejemplo, mediante la integración de los polinucleótidos de VH y VL sintetizados en el vector de expresión de mamífero); y e) someter a ensayo los anticuerpos (por ejemplo, mediante el empleo de un inmunoensayo), identificando de este modo el anticuerpo, la diana o el agente para la molécula diana en la muestra.

Según otra realización de esta descripción, un anticuerpo, una molécula diana, un marcador de enfermedad, un agente o un agente para una molécula diana se identifica, se determina, se detecta o se diagnostica a través de un método que comprende las etapas de: a) proporcionar una muestra biológica procedente de un mamífero (por ejemplo, ser humano, múrido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos) que se inmuniza o se infecta con un agente o una molécula diana o la muestra que contiene el marcador de enfermedad o anticuerpos para una molécula diana; b) recoger los linfocitos B a partir de la muestra biológica; c) obtener los ARNm que codifican inmunoglobulinas en los linfocitos B recogidos; d) generar los ADNc de las inmunoglobulinas, obteniendo de este modo una mezcla de los ADNc; e) secuenciar cada una de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas (preferibles las secuencias de VH y VL y las obtenidas de forma independiente), obteniendo de este modo datos de las secuencias de las cadenas VH y VL de las inmunoglobulinas (tales como IgGs) presentes en la muestra; f) analizar los datos de las secuencias de las cadenas VH y VL para determinar las cadenas VH y VL más dominantes presentes en la muestra; g) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo predeterminado implementado por ordenador; h) sintetizar polinucleótidos de las cadenas VH y VL más abundantes; i) integrar los polinucleótidos de VH y VL sintetizados en un vector de expresión de mamífero; j) permitir que los vectores con los polinucleótidos de VH y VL integrados se expresen en un medio de cultivo, produciendo de este modo anticuerpos; y k) someter a ensayo los anticuerpos mediante el uso de un inmunoensayo, identificando de este modo el anticuerpo, la diana, la molécula diana, el marcador de enfermedad, el agente o el agente para la molécula diana.

En realizaciones de esta descripción, los ARNm que codifican la inmunoglobulina son de la clase IgG.

En otras realizaciones de esta descripción, los ADNc de las inmunoglobulinas son generados mediante transcriptasa inversa-PCR.

Según otras realizaciones de esta descripción, los cebadores específicos de IgG se utilizan en la amplificación con transcriptasa inversa-PCR.

En otras realizaciones de esta descripción, los datos de secuencias independientes para las cadenas VH y VL de las inmunoglobulinas se obtienen para la determinación del algoritmo implementado por ordenador.

En otra realización de esta descripción, el mamífero es un ser humano, múrido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo.

Según otra realización de esta descripción, un anticuerpo, una molécula diana, un marcador de enfermedad, un agente o un agente para una molécula diana se identifica, se determina, se detecta o se diagnostica a través de un método que comprende las etapas de:

- a) analizar los datos de las secuencias de las cadenas VH y VL presentes en la muestra; y
- b) determinar la abundancia de las cadenas VH y VL presentes en la muestra.

En otra realización, el método comprende adicionalmente preparar una o varias inmunoglobulinas que contienen

5 cadenas VH y VL abundantes identificadas en la etapa b). En otras realizaciones de esta descripción, la muestra es una muestra biológica. En otras realizaciones, la muestra biológica es procedente de un mamífero. En aún otras realizaciones, dicho mamífero es un ser humano, múnido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. En otras realizaciones, dicho ser humano es un paciente con cáncer o está inmunizado o infectado con un agente o una molécula diana. En otras realizaciones, las inmunoglobulinas son de la clase IgG.

#### Definiciones y otras realizaciones de la descripción:

10 El término "inmunoglobulina (Ig)" se refiere a una proteína que consiste en uno o varios polipéptidos codificados sustancialmente por genes de inmunoglobulinas. Las inmunoglobulinas incluyen pero no se limitan a los anticuerpos. Las inmunoglobulinas pueden tener una variedad de formas estructurales, incluyendo, pero no limitadas a anticuerpos de longitud completa, fragmentos de anticuerpos y dominios de inmunoglobulinas individuales. Por "dominio de inmunoglobulina (Ig)" en este documento se entiende una región de una inmunoglobulina que existe como una entidad estructural diferente tal y como lo comprueba un experto en la técnica de estructuras proteicas. El término "IgG" tal y como se utiliza en esta memoria significa una proteína que pertenece a la clase de anticuerpos que están codificados sustancialmente por un gen gamma de inmunoglobulina reconocido.

15 El término "anticuerpos" incluye anticuerpos policlonales, anticuerpos policlonales purificados por afinidad, anticuerpos monoclonales, anticuerpos completamente humanos, de múnido o de roedor, y fragmentos que se unen a antígenos, tales como F(ab')<sub>2</sub> y fragmentos proteolíticos Fab. Anticuerpos o fragmentos intactos modificados genéticamente, tales como anticuerpos quiméricos, fragmentos Fv, anticuerpos de cadena sencilla y similares, así como péptidos y polipéptidos sintéticos que se unen a antígeno, también se incluyen. Un anticuerpo de acuerdo con esta descripción puede estar codificado sustancialmente por genes de inmunoglobulinas que pertenecen a cualquiera de las clases de anticuerpos. El anticuerpo comprende secuencias que pertenecen a la clase de anticuerpos IgG, incluyendo las subclases humanas IgG1, IgG2, IgG3 e IgG4. El anticuerpo presente comprende secuencias que pertenecen a la clase de anticuerpos IgA (incluyendo las subclases humanas IgA1 e IgA2), IgD, IgE, IgG o IgM.

20 La expresión "cadena variable" o la "región variable" tal y como se emplea en esta memoria, significa la región de una inmunoglobulina que comprende uno o varios dominios de Ig codificados sustancialmente por cualquiera de los genes VL (incluyendo V<sub>k</sub> y V<sub>λ</sub>), VH, JL (incluyendo J<sub>k</sub> y J<sub>λ</sub>) y JH que forman los loci genéticos de la cadena ligera (incluyendo kappa y lambda) y la cadena pesada de inmunoglobulinas, respectivamente. Una región variable de la cadena ligera o pesada (VL y VH) consiste en una región "estructural" o región "FR" interrumpida por tres regiones hipervariables denominadas "regiones determinantes de complementariedad" o "CDRs". La extensión de la región estructural y las CDRs se ha definido con precisión (véase Kabat, 1991, J. Immunol., 147, 915-920; Chothia y Lesk, 1987, J. Mol. Biol. 196: 901-917; Chothia et al., 1989, Nature 342: 877-883; Al-Lazikani et al., 1997, J. Mol. Biol. 273: 927-948). Las regiones estructurales de un anticuerpo, es decir, las regiones estructurales combinadas de las cadenas ligeras y pesadas constituyentes, sirven para situar y alinear las CDRs, que son las principales responsables de la unión a un antígeno.

30 La expresión "muestra biológica" se puede referir a cualquier muestra de origen biológico, incluyendo tejido o líquido, sangre completa, muestras de suero o plasma, o una muestra que contiene cualquier componente, tal como proteínas, polipéptidos, ácidos nucleicos o polinucleótidos. En ciertas realizaciones la muestra biológica es de un mamífero, tal como de un ser humano, múnido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos. En otras realizaciones la muestra biológica es de un mamífero, en donde el mamífero (por ejemplo, ser humano, múnido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas o caballos) está inmunizado con una molécula diana, es un paciente con cáncer o está infectado con una especie determinada de virus, bacteria o protozoo.

35 Una "diana" o "molécula diana" se refiere a una molécula que es reactiva o que se une a una inmunoglobulina, tal como un anticuerpo. La diana puede ser conocida, desconocida y/o se puede identificar por métodos conocidos en la técnica, por ejemplo, inmunoprecipitación, seguido de un análisis de espectrometría de masas, secuenciación de proteínas N-terminal, etc.

40 El término "mamífero" se refiere a cualquier organismo clasificado como mamífero, por ejemplo, ser humano, múnido o roedor, ratones, ratas, ardillas, ardillas rayadas, taltuzas, puercoespines, castores, hámsteres, jerbos, cobayas, conejos, perros, gatos, vacas y caballos. En unas realizaciones, el mamífero es un ratón. En otras realizaciones de esta descripción, el mamífero es un ser humano o una rata.

45 La información de la secuencia se puede almacenar en cualquier formato y por lo tanto la expresión "base de datos" tal como se utiliza en esta memoria, se refiere a cualquier recopilación de información, en particular información de la secuencia, tal como un archivo de una base de datos, una tabla de búsqueda, una hoja de cálculo Excel o similares. En ciertas realizaciones la base de datos está almacenada en forma electrónica, tal como un dispositivo de memoria legible por ordenador. Esto incluye medios tales como un servidor, un cliente, un disco duro, un CD, un DVD, un asistente digital personal tal como una Palm Pilot, una cinta, un disco zip, la memoria ROM interna del ordenador (del inglés, read-only-memory) o Internet o la red de extensión mundial. Otros medios para el almacena-

miento de archivos accesibles a través de un ordenador serán obvios para un experto en la técnica.

“Algoritmo implementado por ordenador” tal y como se usa en este contexto, se refiere a cualquier medio estadístico que se puede utilizar para determinar qué cadenas VH y VL son las más abundantes en una determinada muestra biológica. Tal algoritmo implementado por ordenador puede formar parte de un paquete de programas informáticos más grande o un parche o una aplicación de un programa informático independiente. El algoritmo implementado por ordenador actúa normalmente sobre una base de datos que contiene datos de la secuencia de variantes de las cadenas VH y VL. El producto del algoritmo implementado por ordenador es normalmente una lista de las secuencias más dominantes o abundantes, por ejemplo, cadenas VH o VL *v*, presentes en una muestra biológica.

De acuerdo con realizaciones de esta descripción, las secuencias más dominantes o abundantes identificadas a través del algoritmo implementado por ordenador se sintetizan y se integran en un vector de expresión para una expresión tal y como se describe en esta memoria.

El término "(poli) péptido" se refiere a moléculas que consisten en una o varias cadenas de aminoácidos múltiples, es decir, dos o más, enlazadas a través de enlaces peptídicos.

El término "proteína" se refiere a (poli) péptidos en los que al menos una parte del (poli) péptido tiene o es capaz de adquirir una disposición tridimensional definida formando estructuras secundarias, terciarias o cuaternarias dentro y/o entre su(s) cadena(s) (poli) peptídica(s). Esta definición comprende proteínas tales como proteínas de origen natural o proteínas al menos parcialmente artificiales, así como fragmentos o dominios de proteínas completas, siempre que estos fragmentos o dominios sean capaces de adquirir una disposición tridimensional definida tal y como se ha descrito anteriormente. Ejemplos de (poli) péptidos/proteínas que consisten en una cadena son fragmentos de anticuerpos Fv de cadena sencilla, y ejemplos de (poli) péptidos/proteínas que consisten en varias cadenas son fragmentos de anticuerpos Fab.

En ciertas realizaciones, la descripción proporciona genotecas de (poli) péptidos que comprenden al menos partes de miembros o derivados de la superfamilia de las inmunoglobulinas, preferiblemente de inmunoglobulinas tales como las inmunoglobulinas de la clase IgG. Algunas realizaciones proporcionan genotecas de anticuerpos humanos. Otras realizaciones proporcionan genotecas de anticuerpos de mamífero o de roedor. Las regiones de las cadenas pesadas y ligeras variables comprenden preferiblemente regiones estructurales (FR) 1, 2, 3 y 4 y regiones determinantes de la complementariedad (CDR) 1, 2 y 3.

Estos genes artificiales que codifican las cadenas VH y VL más abundantes se construyen a continuación, por ejemplo, mediante síntesis génica total o mediante el uso de subunidades genéticas sintéticas. Estas subunidades genéticas se pueden corresponder a subelementos estructurales a nivel de (poli) péptido, por ejemplo, a una o más regiones estructurales y/o a una o más regiones determinantes de la complementariedad. A nivel del ADN, estas subunidades genéticas pueden estar definidas por sitios de escisión al inicio y al final de cada uno de los subelementos, que son únicos en el sistema del vector. En ciertas realizaciones, los subelementos son compatibles con la genoteca HuCAL (Human Combinatorial Antibody), como se describe en el documento US 7.264.963).

Esta colección de moléculas de ADN se puede utilizar a continuación para crear genotecas de anticuerpos o de fragmentos de anticuerpos, tales como Fv, Fv enlazado a disulfuro, Fv de cadena sencilla (scFv), o fragmentos Fab, que se pueden utilizar como fuentes de especificidades contra nuevos antígenos diana. Por otra parte, la afinidad de los anticuerpos se puede optimizar utilizando casetes de genotecas constituidas previamente y un procedimiento de maduración conocido.

La descripción proporciona un método para identificar uno o varios genes que codifican uno o varios anticuerpos o fragmentos de anticuerpos que se unen a una diana, que comprende las etapas de expresar los anticuerpos o los fragmentos de anticuerpos, y después escrutarlos para aislar uno o varios anticuerpos o fragmentos de anticuerpos que se unen a una molécula diana dada.

Expresión génica: La locución "expresión génica" se refiere a procedimientos *in vivo* o *in vitro*, a través de los cuales la información de un gen se transcribe en ARNm y luego se traduce en una proteína/(poli) péptido. Por lo tanto, la locución expresión génica se refiere a un procedimiento que se produce dentro de las células, a través del cual la información de un gen se transcribe en ARNm y luego en una proteína. El término expresión también incluye todos los eventos de modificaciones postraduccionales y el transporte, que son necesarios para que el (poli) péptido sea funcional. Análisis de genes homólogos: Las secuencias de aminoácidos correspondientes a dos o más genes se alinean entre sí de una manera que se maximiza la correspondencia entre los residuos de aminoácidos idénticos o similares en todas las posiciones. Estas secuencias alineadas se denominan homólogos si el porcentaje de la suma de residuos idénticos y/o similares supera un umbral definido.

El término "vector" se refiere a una molécula de ácido nucleico capaz de transportar entre diferentes entornos genéticos otro ácido nucleico al que se ha ligado funcionalmente. Los vectores preferidos son aquellos capaces de una replicación autónoma y/o una expresión de los ácidos nucleicos a los que están enlazados. Los vectores capaces de dirigir la expresión de genes a los que están ligados funcionalmente, se denominan en esta memoria "vectores de expresión". La elección del vector depende de los requisitos específicos y de las propiedades funcionales de un vector dado.

- En una realización de esta descripción, el vector incluye un replicón procariota, es decir, una secuencia de ADN que tiene la capacidad de dirigir la replicación autónoma y el mantenimiento de la molécula de ADN recombinante de forma extracromosómica en una célula hospedadora procariota, tal como una célula hospedadora bacteriana, transformada con el mismo. Tales replicones son bien conocidos en la técnica. Además, aquellas realizaciones que incluyen un replicón procariota también incluyen un gen cuya expresión confiere una ventaja selectiva, tal como resistencia a los fármacos, frente a un hospedador bacteriano transformado con el mismo.
- Los vectores que incluyen un replicón procariota también pueden incluir un promotor procariota capaz de dirigir la expresión (transcripción y traducción) de los homólogos que codifican  $V_H$  y/o  $V_L$  en una célula hospedadora bacteriana, tales como *Escherichia coli* transformada con el mismo. Un promotor es un elemento para el control de la expresión formado por una secuencia de ADN que permite la unión de una ARN polimerasa y que se produzca la transcripción. Las secuencias promotoras compatibles con hospedadores bacterianos se proporcionan típicamente en vectores plasmídicos que contienen sitios de restricción de conveniencia para la inserción de un segmento de ADN. Ejemplos de tales plásmidos de vectores incluyen pUC8, pUC9, pBR322 y pBR329, pPL y pKK223, disponibles comercialmente. Tales vectores se denominan "vectores de expresión procariota".
- Los vectores de expresión eucariota preferidos incluyen los que son compatibles con células de vertebrados. La expresión "vectores de expresión eucariota" se refiere a cualquier vector de expresión, útil en la expresión de ácidos nucleicos en células hospedadoras eucariotas. En una realización particular de esta descripción, los vectores eucariotas son vectores de expresión de mamífero. La expresión "vectores de expresión de mamífero" se refiere a cualquier vector de expresión, útil en la expresión de ácidos nucleicos en células hospedadoras de mamífero.
- Los vectores de expresión eucariotas son bien conocidos en la técnica y también están disponibles comercialmente. Típicamente, se proporcionan tales vectores conteniendo sitios de restricción convenientes para la inserción del homólogo de ADN deseado. Ejemplos de tales vectores incluyen pSV<sub>L</sub> y pKSV-10, pBPV-1/PML2d y pTDT1 (ATCC, nº 31255).
- En otra realización de esta descripción, los vectores de expresión eucariotas incluyen un marcador de selección que es eficaz en una célula eucariota, preferiblemente un marcador de selección resistente a fármacos. Un marcador de resistencia a fármacos preferido es el gen cuya expresión da como resultado la resistencia a la neomicina, es decir, el gen de la neomicina fosfotransferasa (neo). Southern et al., J. Mol. Appl. Genet., 1: 327-341 (1982).
- Vectores de expresión retrovíricos para expresar los genes de los homólogos de ADN que codifican  $V_H$  y/o  $V_L$ , también se contemplan. La expresión "vector de expresión retrovírico" se refiere a una molécula de ADN que incluye secuencias de un promotor obtenidas a partir de la región de repetición terminal larga (LTR) de un genoma de retrovirus.
- Estructura(s) artificial(es) de ADN que codifica(n)  $V_H$  y/o  $V_L$  es/son introducida(s) en un hospedador apropiado para proporcionar la amplificación y/o la expresión de los homólogos de ADN que codifican  $V_H$  y/o  $V_L$ , ya sea de forma separada o en combinación. Cuando los polipéptidos de  $V_H$  y  $V_L$  se expresan en diferentes organismos, los polipéptidos respectivos se aíslan y luego se combinan en un medio apropiado para formar un anticuerpo o un fragmento del mismo. Los hospedadores celulares en los que se ha introducido una estructura artificial que contiene homólogos de ADN que codifican  $V_H$  y/o  $V_L$ , se hace referencia a los mismos en esta memoria diciendo que se han "transformado" o que son "transformantes".
- Las células hospedadoras pueden ser procariotas o eucariotas. Las células bacterianas se prefieren como células hospedadora procariotas y, normalmente, son una cepa de *Escherichia coli* (*E. coli*), tal como, por ejemplo, la cepa de *E. coli* DH5 disponible en Bethesda Research Laboratories, Inc., Bethesda, Md. Las células hospedadoras eucariotas preferidas incluyen levaduras y células de mamífero incluyendo muridos y roedores, preferiblemente células de vertebrado tales como las de una línea celular de ratón, rata, mono o ser humano.
- La transformación de los hospedadores celulares apropiados con una molécula de ADN recombinante se lleva a cabo por métodos que dependen típicamente del tipo de vector usado. Con respecto a la transformación de células hospedadoras procariotas, véase, por ejemplo, Cohen et al., Proceedings National Academy of Science, USA, vol. 69, pág. 2110 (1972); y Maniatis et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982). Con respecto a la transformación de células de vertebrados con vectores retrovíricos que contienen ADN<sub>r</sub>, véase, por ejemplo, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); y Wigler et al., Proceedings National Academy of Sciences, USA, vol. 76, págs. 1373-1376 (1979).
- Escrutinio de la expresión de polipéptidos de  $V_H$  y/o  $V_L$ : células transformadas con éxito, es decir, células que contienen un homólogo de ADN que codifica  $V_H$  y/o  $V_L$  ligado funcionalmente a un vector, se pueden identificar mediante cualquier técnica conocida, adecuada para detectar la unión de un receptor a un ligando o la presencia de un polinucleótido que codifica el receptor, preferiblemente su sitio activo. En una realización, los ensayos de escrutinio se llevan a cabo de tal manera que la unión del ligando a través del receptor produce una señal detectable, ya sea directa o indirectamente. Tales señales incluyen, por ejemplo, la producción de un complejo, la formación de un producto de reacción catalítica, la liberación o la absorción de energía y similares. Las células procedentes de una población sometida a transformación con un ADN recombinante objeto, se pueden clonar para producir colonias



monoclonales, por ejemplo. Las células que forman esas colonias se pueden recoger, lisar y examinar su contenido en ADN para estudiar la presencia del ADN recombinante, utilizando un método conocido en la técnica, por ejemplo, tal y como se describe en Southern, *J. Mol. Biol.*, 98: 503 (1975) o Berent et al, *Biotech.* 3: 208 (1985).

5 Además de someter a ensayo directamente la presencia de un ADN que codifica  $V_H$  y/o  $V_L$ , una transformación exitosa también puede ser confirmada por métodos inmunológicos bien conocidos, especialmente cuando los polipéptidos de  $V_H$  y/o  $V_L$  producidos contienen un epítipo preseleccionado. Muestras de células de las cuales se sospecha que están transformadas, se sometieron a ensayo para determinar la presencia del epítipo preseleccionado usando, por ejemplo, un anticuerpo contra el epítipo.

10 Un "inmunoensayo" tal y como se emplea en esta memoria, se refiere a cualquier medición de la reacción de unión específica entre un antígeno y una inmunoglobulina, tal como un anticuerpo. Típicamente, el antígeno es un (poli) péptido o proteína, pero cualquier otra sustancia, tal como un ácido nucleico, un lípido, un ácido graso o una molécula orgánica pequeña, puede servir como antígeno. El experto en la materia comprenderá fácilmente y determinará qué inmunoensayo es el más adecuado para medir la unión específica entre un antígeno y una inmunoglobulina.

15 La expresión "una muestra representativa" tal y como se usa en el contexto de la presente invención, se refiere a una muestra representativa de cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas de una muestra que se tiene que secuenciar, con el fin de obtener una visión general de las cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas presentes en dicha muestra. El número total de las secuencias necesarias para obtener una visión general de este tipo puede depender de la naturaleza de la muestra respectiva, pero debe ser por lo menos tan alto, que se pueda realizar una estimación razonable de qué cadenas ligeras variables (VL) y qué cadenas pesadas variables (VH) de las inmunoglobulinas son más abundantes en dicha muestra. Preferiblemente, al menos 20%, al menos 30%, al menos 40%, al menos 50%, al menos 60%, al menos 70%, al menos 80%, al menos 90% o al menos 95% de las cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas presentes en dicha muestra se secuencian. También preferiblemente al menos 1.000, al menos 10.000, al menos 50.000 o al menos 100.000 de las cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas presentes en dicha muestra se secuencian.

La expresión "más abundantes" tal y como se emplea en el contexto de la presente invención, se refiere a aquellas cadenas ligeras variables (VL) y pesadas variables (VH) que se identifican con mayor frecuencia cuando se secuencian una muestra. Las cadenas ligeras variables (VL) y pesadas variables (VH) más abundantes son normalmente las que están presentes en mayor número en una determinada muestra.

30 En ciertas realizaciones, la presente invención proporciona un método para la identificación de un anticuerpo, una diana o un agente en una muestra, que comprende:

- a) obtener los ADNc a partir de los ARNm que codifican inmunoglobulinas en la muestra, obteniendo de este modo una mezcla de los ADNc;
- 35 b) secuenciar al menos una muestra representativa de las cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas, obteniendo de este modo datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas (IgGs) presentes en la muestra;
- c) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo implementado por ordenado;
- 40 d) sintetizar polinucleótidos de las cadenas VH y VL más abundantes y producir anticuerpos que comprenden dichas cadenas VH y VL utilizando un vector de expresión de mamífero; y
- e) someter a ensayo los anticuerpos, identificando de este modo el anticuerpo o la diana. Preferiblemente, dicha muestra es una muestra biológica. Preferiblemente, dicha muestra biológica es de un mamífero. Preferiblemente dicho mamífero es un ser humano, murido, roedor, ratón, rata, ardilla, ardilla, tuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. Preferiblemente, dicho humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana. Preferiblemente, los ARNm que codifican dichas inmunoglobulinas en la muestra se obtienen recogiendo o aislando los linfocitos B de la muestra. Preferiblemente, dichas inmunoglobulinas son de la clase IgG. Preferiblemente, los ADNc obtenidos a partir de los ARNm que codifican inmunoglobulinas en la muestra son generados por medio de transcriptasa inversa-PCR. Preferiblemente se emplean cebadores específicos de IgG para la generación de los ADNc a partir de los ARNm que codifican las inmunoglobulinas. Preferiblemente, las cadenas VH y VL presentes en la muestra se determinan mediante el análisis de datos de las secuencias de las cadenas VH y VL presentes en la muestra. Lo más preferiblemente, los métodos de la presente invención no requieren ninguna etapa de clonación. En particular, todas las etapas del método realizado hasta la determinación de las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo implementado por ordenado, no requieren ninguna etapa de clonación. Preferiblemente, los polinucleótidos de VH y VL de las cadenas VH y VL más abundantes en la etapa (d) se integran en un vector de expresión de mamífero. Preferiblemente, los anticuerpos producidos en dicha etapa se liberan en el medio de cultivo. Preferiblemente, el ensayo de los anticuerpos en la etapa (e) del método se lleva a cabo mediante el

empleo de un inmunoensayo. Preferiblemente, en la etapa (e) del método se identifica un anticuerpo que se une a una determinada molécula o tejido diana. Preferiblemente en la etapa (e) del método se identifica una molécula diana que se une a un determinado anticuerpo.

5 En ciertas realizaciones, la presente invención proporciona un método para la identificación de un anticuerpo, una diana o un agente que comprende:

a) proporcionar una muestra biológica a partir de un mamífero que está inmunizado, infectado con un agente o una molécula diana o que padece cáncer;

b) recoger los linfocitos B de la muestra biológica;

c) obtener ARNm a partir de los linfocitos B recogidos;

10 d) generar ADNc de las inmunoglobulinas codificadas por el ARNm, obteniendo de este modo una mezcla de los ADNc;

e) secuenciar al menos una muestra representativa de las cadenas ligeras variables (VL) y pesadas variables (VH) de las inmunoglobulinas, obteniendo con ello datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas presentes en la muestra;

15 f) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo implementado por ordenador;

h) sintetizar polinucleótidos de las cadenas VH y VL más abundantes;

i) integrar los polinucleótidos de VH y VL sintetizados en un vector de expresión de mamífero;

20 j) permitir que los vectores con los polinucleótidos de VH y VL integrados expresen los polinucleótidos de VH y VL, produciendo de este modo anticuerpos; y

25 k) someter a ensayo los anticuerpos mediante el uso de un inmunoensayo, identificando de este modo el anticuerpo, la diana o el agente. Preferiblemente, dicha muestra es una muestra biológica. Preferiblemente, dicha muestra biológica es de un mamífero. Preferiblemente dicho mamífero es un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla, tuza, puerco espín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. Preferiblemente, dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana. Preferiblemente dichas inmunoglobulinas son de la clase IgG.

En ciertas realizaciones, la presente invención proporciona un método para la identificación de un anticuerpo, una diana o un agente en una muestra que comprende las etapas de:

30 a) analizar los datos de la secuencia de las cadenas VH y VL presentes en la muestra; y

b) determinar la abundancia de las cadenas VH y VL presentes en la muestra.

35 Preferiblemente dicho método comprende además la preparación de una o varias inmunoglobulinas que contienen las cadenas VH y VL abundantes, identificadas en la etapa b). Preferiblemente, dicha muestra es una muestra biológica. Preferiblemente, dicha muestra biológica es de un mamífero. Preferiblemente dicho mamífero es un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla, tuza, puerco espín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. Preferiblemente, dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana. Preferiblemente dichas inmunoglobulinas son de la clase IgG.

40 En ciertas realizaciones, la presente invención proporciona un método para identificar un anticuerpo, una diana o un agente en una muestra que comprende las etapas de:

a) obtener los ADNc de los ARNm que codifican inmunoglobulinas en la muestra, obteniendo de este modo una mezcla de los ADNc;

45 b) secuenciar directamente al menos una muestra representativa de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas, obteniendo con ello datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas presentes en la muestra;

c) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo implementado por ordenador;

d) sintetizar polinucleótidos que codifican las cadenas VH y VL más abundantes presentes en la muestra y producir anticuerpos que comprenden dichas cadenas VH y VL utilizando un vector de expresión; y

e) someter a ensayo los anticuerpos, identificando de este modo el anticuerpo o la diana.

La expresión "secuenciación directa" tal y como se usa en el contexto de la etapa (b) del método indicado anteriormente, se refiere a una situación en la que los ADNc obtenidos en la etapa (a) se secuencian sin etapas de modificación biológica molecular adicionales. En particular, no se requieren etapas de clonación.

5 En realizaciones preferidas, la muestra es una muestra biológica. En realizaciones más preferidas, la muestra biológica es de un mamífero. Dicho mamífero puede ser un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla, tuza, puerco espín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. Dicha muestra biológica puede ser también de un ser humano, en donde dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana.

10 En realizaciones preferidas, los ARNm que codifican inmunoglobulinas en la etapa (a) del método indicado anteriormente, se obtienen mediante la recogida o el aislamiento de los linfocitos B de la muestra. En otras realizaciones preferidas las inmunoglobulinas son de la clase IgG.

15 En realizaciones preferidas los ADNc de los ARNm que codifican inmunoglobulinas en la etapa (a) del método indicado anteriormente, son generados mediante transcriptasa inversa-PCR. En realizaciones más preferidas los cebadores específicos de IgG se utilizan en dicha transcriptasa inversa-PCR.

En realizaciones preferidas, las cadenas VH y VL más abundantes presentes en la muestra en la etapa (c) del método indicado anteriormente, se determinan mediante el análisis de datos de la secuencia de las cadenas VH y VL presentes en la muestra.

20 En realizaciones preferidas, las etapas (a) - (c) del método indicado anteriormente no requieren ninguna etapa de clonación.

En realizaciones preferidas los polinucleótidos de VH y VL sintetizados en la etapa (d) del método indicado anteriormente, se integran en un vector de expresión. Más preferiblemente dicho vector de expresión es un vector de expresión de mamífero.

25 En realizaciones preferidas los anticuerpos producidos en la etapa (d) del método indicado anteriormente, se liberan en el medio de cultivo.

En realizaciones preferidas, el ensayo de los anticuerpos en la etapa (e) del método indicado anteriormente, se lleva a cabo mediante el empleo de un inmunoensayo.

30 En realizaciones preferidas en la etapa (e) del método indicado anteriormente, se identifica un anticuerpo que se une a una cierta molécula o tejido diana. En otras realizaciones preferidas en la etapa (e) del método citado anteriormente, se identifica una molécula diana que se une a un determinado anticuerpo.

En ciertas realizaciones, la presente invención proporciona un método para identificar un anticuerpo, una diana o un agente que comprende las etapas de:

35 a) proporcionar una muestra biológica procedente de un mamífero que está inmunizado, infectado con un agente o una molécula diana o que padece cáncer;

b) recoger los linfocitos B de la muestra biológica;

c) obtener ARNm a partir de los linfocitos B recogidos;

d) generar ADNc de las inmunoglobulinas codificadas por el ARNm, obteniendo de este modo una mezcla de los ADNc;

40 e) secuenciar directamente al menos una muestra representativa de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas, obteniendo con ello datos de la secuencia de las cadenas VH y VL de las inmunoglobulinas presentes en la muestra;

f) determinar las cadenas VH y VL más abundantes presentes en la muestra utilizando un algoritmo implementado por ordenador;

h) sintetizar polinucleótidos que codifican las cadenas VH y VL más abundantes presentes en la muestra;

45 i) integrar los polinucleótidos de VH y VL sintetizados en un vector de expresión de mamífero;

j) permitir que los vectores con los polinucleótidos de VH y VL integrados expresen los polinucleótidos de VH y VL, produciendo de este modo anticuerpos; y

k) someter a ensayo los anticuerpos mediante el uso de un inmunoensayo, identificando de este modo el anticuerpo, la diana o el agente. En realizaciones preferidas, dicha muestra es una muestra biológica. En reali-

zaciones más preferidas, dicha muestra biológica es de un mamífero. Dicho mamífero puede ser un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla, tuza, puerco espín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. En realizaciones más preferidas, dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana. En realizaciones más preferidas, las inmunoglobulinas son de la clase IgG.

En ciertas realizaciones, la presente invención proporciona un método para identificar un anticuerpo, una diana o un agente en una muestra que comprende las etapas de:

a) analizar los datos de la secuencia de las cadenas VH y VL presentes en la muestra; y

b) determinar la abundancia de las cadenas VH y VL presentes en la muestra. En otras realizaciones dicho método comprende además la preparación de una o varias inmunoglobulinas que contienen cadenas VH y VL abundantes identificadas en la etapa b). En realizaciones preferidas, dicha muestra es una muestra biológica. En realizaciones más preferidas, dicha muestra biológica es de un mamífero. Dicho mamífero puede ser un ser humano, mívrido, roedor, ratón, rata, ardilla, ardilla, tuza, puerco espín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo. En realizaciones más preferidas, dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana. En realizaciones más preferidas las inmunoglobulinas son de la clase IgG.

Esta descripción se describe adicionalmente mediante los siguientes ejemplos, que no limitan de ningún modo el alcance de la descripción.

### **EJEMPLOS**

#### ***Recogida de linfocitos B a partir de un paciente infectado/inmunizado:***

Los linfocitos B se pueden aislar a partir de pacientes inmunizados o infectados de varias maneras diferentes, tales técnicas son conocidas por el experto en la materia. En muchas de estas técnicas, linfocitos B en fase de reposo (células B) se aíslan a partir de bazo utilizando selección negativa con anticuerpos monoclonales anti-CD43 y anti-Mac-1/CD11b, por ejemplo, a través de microperlas magnéticas. Esta estrategia agota las células no B a partir de una población mixta de esplenocitos y se basa en el hecho de que la mayoría de los leucocitos maduros, con la excepción de los linfocitos B esplénicos en reposo, expresan CD43 (de hecho, la expresión de CD43 se ha demostrado en linfocitos B inmaduros, células plasmáticas y algunos linfocitos B1 maduros, además de granulocitos, monocitos, macrófagos, plaquetas, células citotóxicas naturales (NK), timocitos y linfocitos T CD8pos periféricos y la mayoría de los linfocitos T CD4pos). Las microperlas anti-Mac-1/CD11b se incluyen en la selección negativa para mejorar la eliminación de las células mieloides. El aislamiento de linfocitos B se puede automatizar mediante el uso de un clasificador automático de células con perlas magnéticas AutoMACS (Miltenyi Biotec). Como se determina mediante un análisis con fluorescencia de linfocitos B220+, tal aislamiento produce rutinariamente aproximadamente  $4 \times 10^7$  linfocitos B por bazo que tienen una pureza >95%. Véase también Miltenyi S, Muller W, Weichel W y Radbruch A. (1990) Cytometry 11(2), 231-238.

#### ***Extracción de ARNm y transcripción inversa:***

Las inmunoglobulinas, preferiblemente inmunoglobulinas de tipo IgG, se pueden amplificar selectivamente a partir de linfocitos B mediante la extracción del ARNm, seguida de transcripción inversa.

La extracción del ARNm a partir de células eucariotas, tales como linfocitos B, es un procedimiento tecnológico bien conocido. Existen numerosos protocolos y kits comercialmente disponibles. Tales como el sistema de aislamiento de ARNm PolyATtract® (Promega, Madison, WI, EE.UU.) o varios kits de RNeasy y Oligotex DirectmRNA (ambos de Qiagen, Hilden, Alemania). Muchas de estas técnicas utilizan la cola poliA del ARNm eucariota, por ejemplo, a través de la purificación por afinidad de matrices de oligo (dT), tales como celulosa oligo (dT).

Las inmunoglobulinas se pueden amplificar de forma selectiva a partir del ARNm aislado mediante transcripción inversa, usando cebadores específicos, seguida por PCR convencional. Los cebadores específicos pueden ser específicos de inmunoglobulinas, de una determinada clase de inmunoglobulinas, es decir, ya sea IgG, IgM, IgA, IgD o IgE, o incluso de una cierta subclase de inmunoglobulinas, tal como IgG1, IgG2, IgG3 o IgG4. Los cebadores que se pueden usar para amplificar genes de la cadena pesada y ligera de las inmunoglobulinas, se describen por ejemplo en Cancer Surv 1997;30:21-44, J Clin Pathol 1994;47:493-6, J Clin Pathol 1990;43:888-90 o Mol Pathol. Abr 2002; 55(2): 98-101.

#### ***Secuenciación genómica de los ADNc:***

Las secuencias completas de las inmunoglobulinas se secuencian. Existen varias compañías que son capaces de secuenciar genomas completos, tales como Helicos BioSciences Corporation (Cambridge, MA, EE.UU.). Con su tecnología True Individual Molecule Sequencing®, Helicos es capaz de secuenciar directamente moléculas individuales de ADN o ARN con una velocidad y eficacia elevadas. Otras compañías capaces de realizar iniciativas similares con secuencias incluyen Illumina (San Diego, CA, EE.UU., sistema Solexa) y Roche (Basilea, CH, sistema 454). No

son necesarias etapas de clonación antes de la secuenciación.

Las secuencias de las cadenas VH y VL de las inmunoglobulinas se determinan por separado. Se determinan más de  $10^3$  secuencias independientes de las cadenas VH y VL, preferiblemente más de  $10^4$  secuencias independientes, más preferiblemente más de  $10^5$  secuencias independientes, e incluso más preferiblemente más de  $10^6$  secuencias independientes.

Las secuencias determinadas se pueden almacenar en cualquier sistema de base de datos. Tales sistemas de bases de datos pueden formar parte del sistema de secuenciación utilizado. Alternativamente, la información de la secuencia también se puede almacenar en cualquier otro formato, tal como en forma de una hoja de cálculo Excel o en formato delimitado por tabulaciones.

**Determinación de las secuencias de VH y de VL dominantes/más abundantes basada en un algoritmo predefinido:**

La abundancia de las cadenas VH y VL de las inmunoglobulinas secuenciadas se puede determinar mediante diversos algoritmos. Las cadenas VH y las cadenas VL se analizan preferiblemente por separado.

Una primera etapa puede ser la identificación de secuencias de las cadenas VH y VL que se obtienen a partir de las mismas inmunoglobulinas. Las secuencias que se obtienen a partir de las mismas inmunoglobulinas no son necesariamente completamente idénticas. Estas pequeñas diferencias pueden surgir debido a que las secuencias no comienzan o terminan por los mismos nucleótidos en cada secuencia, o porque un nucleótido se leyó erróneamente en el proceso de secuenciación, un evento que puede ocurrir durante este tipo de proyectos de secuenciación a gran escala. Se sabe que ciertos nucleótidos y, en particular, ciertas secuencias de nucleótidos, son más propensos a ser leídos incorrectamente que otros (por ejemplo, tramos de nucleótidos ricos en GC). Herramientas y algoritmos bioinformáticos, muchos de los cuales forman parte de los sistemas de secuenciación respectivos, son capaces de determinar tales acontecimientos, o al menos señalar ejemplos en los que podrían haber tenido lugar tales errores.

La abundancia de las cadenas VH y VL se puede determinar mediante diversas pruebas estadísticas. Lo más sencillo es contar simplemente las cadenas VH y VL individuales. Pruebas estadísticas más sofisticadas pueden tener en cuenta varios otros parámetros. A modo de ejemplos no limitativos, las siguientes pruebas estadísticas y referencias pueden orientar como ejemplos de los numerosos enfoques que se han realizado en un análisis de este tipo o similar: Estimación Bayesiana de Encogimiento (véase, p. ej., *Biometrics* 59 (2003): 476-486), DADA (Digital Analysis of cDNA Abundance, véase, p. ej., *BMC Genomics* 2002, 3:7), modelado lineal (Pacific Symposium on Biocomputing, 1999, 4:41-52) ) y diversos métodos de agrupamiento (*BMC Bioinformatics* 2006, 7:397, Fourth IEEE International Conference on Data Mining (ICDM'04), págs. 403-406).

**Síntesis de VH y VL:**

Los genes de las cadenas VH y VL más abundantes se sintetizan por medios habituales. Tal síntesis es una tecnología convencional y muchas compañías ofrecen servicios respectivos, por ejemplo, Entelechon (Regensburg, Alemania), Geneart (Regensburg, Alemania) o Sloning Biotecnología (Puchheim, Alemania), por nombrar unas pocas. Idealmente los genes respectivos ya son portadores de sitios de restricción apropiados para la clonación en vectores adecuados.

**Clonación y expresión de las cadenas VH y VL dominantes:**

Los genes sintetizados de las cadenas VH y VL se clonan en vectores de expresión respectivos. Para ello el vector de expresión respectivo se digiere con enzimas de restricción apropiadas que sean compatibles con los genes sintetizados. Como se ha indicado anteriormente, los genes sintetizados preferiblemente ya son compatibles con el vector, es decir, los sitios de restricción respectivos ya están presentes en los genes sintetizados. Vectores ejemplares incluyen pcDNA, pMORPH, pUC, pBR, pBAD y otros. La expresión de dichos vectores conduce a la producción de inmunoglobulinas de longitud completa que comprenden las cadenas VH y VL sintetizadas, que se pueden caracterizar o modificar adicionalmente en etapas posteriores.

**Inmunoensayos para el escrutinio de los polipéptidos expresados y selección de parejas de VH y VL**

Inmunoglobulinas de longitud completa producidas después de la expresión a partir de vectores respectivos, tales como pcDNA, pMORPH, pUC, pBR, pBAD y otros, se pueden utilizar para diversos tipos de ensayos. Por ejemplo, se pueden realizar inmunoensayos.

Por ejemplo, la unión de las inmunoglobulinas a una determinada molécula diana, tal como un antígeno, se puede someter a ensayo. Esto se puede llevar a cabo mediante procedimientos de laboratorio convencionales, tales como pruebas de ELISA, transferencia Western o cualquier otro medio equivalente. Tales experimentos pueden conducir a la identificación de aquellas inmunoglobulinas que se unen a una determinada molécula diana. Tales ensayos también se pueden realizar de una manera más cuantitativa, es decir, no se determina solamente si una inmunoglobulina respectiva se une o no a una cierta molécula diana, sino también cómo es de fuerte la interacción producida. Esto se puede lograr a través de la determinación de la afinidad de la unión, la constante de disociación o cualquier otro

parámetro equivalente, de una inmunoglobulina frente a una molécula diana dada. Técnicas respectivas incluyen resonancia de plasmón superficial, titulación en solución de equilibrio, soporte, biosensor acústico y otros métodos conocidos en la técnica.

- 5 También es posible identificar la molécula diana a la que se une una inmunoglobulina dada. Para ello se elige una cierta inmunoglobulina y se somete a una mezcla de proteínas de unión potenciales en condiciones que permiten la unión de la inmunoglobulina con al menos una molécula diana de la mezcla. Las condiciones de unión respectivas se pueden ajustar mediante la selección apropiada de parámetros, tales como la composición del tampón y el rigor.

***Identificación de la diana/anticuerpo:***

- 10 La identificación de la inmunoglobulina que se une a una molécula diana dada, o la identificación de la molécula diana que se une a una inmunoglobulina dada, se puede lograr mediante cualquier metodología conocida. Muchos de estos métodos son conocidos por el experto en la materia y como referencias ejemplares se proporcionan las siguientes: Valle RP, Curr Opin Drug Discov Devel. Mar 2003; 6(2):197-203; Ackermann BL Expert Rev Proteomics. Abr 2007; 4(2):175-86; y Anderson KS J Proteome Res. Jul-Ago 2005; 4(4):1123-33

- 15 Se debe entender que la descripción, los ejemplos y los datos específicos, aunque indican realizaciones ejemplares, se proporcionan a modo de ilustración y no se pretende limitar la descripción. Diversos cambios y modificaciones dentro de esta descripción serán evidentes para el experto en la materia a partir de la exposición, la descripción y los datos contenidos en el presente documento.

**REIVINDICACIONES**

1. Un método para identificar un anticuerpo, una diana o un agente en una muestra, que comprende las etapas de:
  - 5 a) obtener los ADNc a partir de los ARNm que codifican inmunoglobulinas en la muestra, obteniendo de este modo una mezcla de los ADNc;
  - b) secuenciar sin ninguna etapa de clonación previa al menos 1000 de las cadenas pesadas variables (VH) y cadenas ligeras variables (VL) de las inmunoglobulinas presentes en dicha muestra, obteniendo con ello datos de la secuencia;
  - 10 c) determinar las cadenas VH y VL más abundantes presentes en la muestra usando un algoritmo implementado por ordenador;
  - d) sintetizar polinucleótidos que codifican las cadenas VH y VL más abundantes presentes en la muestra y producir anticuerpos que comprenden dichas cadenas VH y VL usando un vector de expresión; y
  - e) someter a ensayo los anticuerpos, identificando de este modo el anticuerpo o la diana.
2. El método según la reivindicación 1, en el que la muestra es una muestra biológica.
- 15 3. El método según la reivindicación 2, en el que dicha muestra es una muestra biológica procedente de un mamífero.
4. El método según la reivindicación 3, en el que dicho mamífero es un ser humano, múnido, roedor, ratón, rata, ardilla, ardilla rayada, taltuza, puercoespín, castor, hámster, jerbo, cobaya, conejo, perro, gato, vaca o caballo.
5. El método según la reivindicación 4, en el que dicho ser humano es un paciente con cáncer, dicho ser humano está inmunizado o dicho ser humano está infectado con un agente o una molécula diana.
- 20 6. El método según la reivindicación 1, en el que los ARNm que codifican inmunoglobulinas se obtienen mediante la recogida o el aislamiento de linfocitos B de la muestra.
7. El método según la reivindicación 1, en el que las inmunoglobulinas son de la clase IgG.
8. El método según la reivindicación 1, en el que los ADNc de los ARNm que codifican inmunoglobulinas en la etapa (a) se generan mediante transcriptasa inversa-PCR.
- 25 9. El método según la reivindicación 8, en el que se utilizan cebadores específicos de IgG.
10. El método según la reivindicación 1, en el que las cadenas VH y VL más abundantes presentes en la muestra se determinan mediante el análisis de datos de las secuencias de las cadenas VH y VL presentes en la muestra.
11. El método según la reivindicación 1, en el que los polinucleótidos de VH y VL sintetizados en la etapa (d) se integran en un vector de expresión.
- 30 12. El método según la reivindicación 11, en el que dicho vector de expresión es un vector de expresión útil en la expresión de ácidos nucleicos en células hospedadoras de mamífero.
13. El método según la reivindicación 1, en el que los anticuerpos producidos en la etapa (d) se liberan en el medio de cultivo.
- 35 14. El método según la reivindicación 1, en el que el ensayo de los anticuerpos en la etapa (e) se lleva a cabo empleando un inmunoensayo.
15. Un método para identificar un anticuerpo, una diana o un agente que comprende las etapas de:
  - a) proporcionar una muestra biológica procedente de un mamífero que está inmunizado, infectado con un agente o una molécula diana o que padece cáncer;
  - 40 b) recoger los linfocitos B de la muestra biológica;
  - c) obtener los ARNm a partir de los linfocitos B recogidos;
  - d) generar los ADNc de las inmunoglobulinas codificadas por el ARNm, obteniendo de este modo una mezcla de los ADNc;
  - 45 e) secuenciar sin ninguna etapa previa de clonación al menos 1000 de las cadenas pesadas variables (VH) y ligeras variables (VL) de las inmunoglobulinas presentes en dicha muestra, obteniendo con ello datos de la

secuencia;

f) determinar las cadenas VH y VL más abundantes presentes en la muestra usando un algoritmo implementado por ordenador;

g) sintetizar los polinucleótidos que codifican las cadenas VH y VL más abundantes presentes en la muestra;

5 h) integrar los polinucleótidos de VH y VL sintetizados en un vector de expresión útil en la expresión de ácidos nucleicos en células hospedadoras de mamífero;

i) permitir que los vectores con los polinucleótidos de VH y VL integrados expresen los polinucleótidos de VH y VL, produciendo de este modo anticuerpos; y

10 j) someter a ensayo los anticuerpos mediante el uso de un inmunoensayo, identificando de este modo el anticuerpo, la diana o el agente.