

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 662 846**

51 Int. Cl.:

**G01N 27/00** (2006.01)

**B01D 59/44** (2006.01)

**G01N 33/68** (2006.01)

**G01N 33/574** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.09.2008 E 14161288 (7)**

97 Fecha y número de publicación de la concesión europea: **13.12.2017 EP 2781913**

54 Título: **Identificación de proteínas en suero humano indicativo de patologías de los tejidos pulmonares humanos**

30 Prioridad:

**11.09.2007 US 971422 P**

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

**10.04.2018**

73 Titular/es:

**CANCER PREVENTION AND CURE, LTD.  
(100.0%)  
105 Washington Street  
Michigan City, IN 46360, US**

72 Inventor/es:

**STREEPER, ROBERT T.;  
IZBICKA, ELZBIETA y  
BAEK, SUNG H.**

74 Agente/Representante:

**UNGRÍA LÓPEZ, Javier**

**ES 2 662 846 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

Identificación de proteínas en suero humano indicativo de patologías de los tejidos pulmonares humanos

### Antecedentes de la invención

#### Campo de la invención

5 La presente invención, definida en las reivindicaciones, se refiere en general a métodos que facilitan el diagnóstico de patologías del tejido pulmonar humano. Más concretamente, la presente invención se refiere a métodos que facilitan el diagnóstico del cáncer pulmonar no microcítico y el asma, utilizando la cromatografía líquida-espectrometría de masas para identificar proteínas presentes en sueros humanos que, cuando se alteran en términos de intensidad relativa de expresión en el suero humano de las mismas proteínas que se encuentran en una población normal, son indicativas de patologías asociadas con el tejido pulmonar humano y el sistema respiratorio humano.

10 Al identificar las proteínas asociadas con estas patologías, determinar las identidades de expresión representativas y comparar estas intensidades de expresión con las intensidades de expresión presentes en el suero de un paciente, resulta posible detectar la presencia de patologías en una fase temprana de su progresión a través de un simple análisis de sangre y diferenciar entre las diferentes patologías.

#### 2. Descripción de la técnica relacionada

15 Las patologías del sistema respiratorio, como el asma y el cáncer de pulmón, afectan a millones de estadounidenses. En efecto, la American Lung Association afirma que casi 20 millones de estadounidenses padecen asma. Según estimaciones de la American Cancer Society, solo en el año 2007 se detectaron 229.400 nuevos casos de cáncer del sistema respiratorio y 164.840 muertes producidas por esta enfermedad. Mientras que la tasa de supervivencia a los cinco años en los casos de cáncer detectados mientras se encuentra localizado es del 46%, esta tasa de supervivencia en los pacientes con cáncer de pulmón es tan solo del 13%. Por otra parte, solo el 16% de los cánceres de pulmón se descubren antes de que la enfermedad se haya propagado. En general, los cánceres de pulmón se clasifican en dos categorías principales en función de la patología de las células cancerosas. Cada una de estas categorías se designa por los tipos de células que han sufrido una transformación para convertirse en cancerosas. El cáncer microcítico procede de las células pequeñas del tejido pulmonar humano, mientras que, por lo general, el cáncer no microcítico abarca todos los cánceres de pulmón que no son de células pequeñas. Los cánceres de pulmón no microcíticos se incluyen en un mismo grupo porque, por lo general, se aplica el mismo tratamiento a todos los cánceres de este tipo. En conjunto, los cánceres de pulmón no microcíticos (o NSCLC, por sus siglas en inglés) representan alrededor del 75% de todos los cánceres de pulmón.

20 Un importante factor para la reducción de la tasa de supervivencia de los pacientes con cáncer de pulmón es el hecho de que resulta difícil diagnosticarlo en una fase temprana. Los métodos actuales para diagnosticar el cáncer de pulmón o identificar su existencia en el ser humano se limitan a los rayos X, el TAC y otras pruebas similares de los pulmones, para determinar físicamente la presencia o ausencia de un tumor. Por tanto, a menudo el diagnóstico del cáncer de pulmón se realiza únicamente en respuesta a unos síntomas que se han manifestado durante un periodo de tiempo significativo y después de que la enfermedad haya estado presente en el ser humano el tiempo suficiente para producir una masa físicamente detectable.

25 De un modo similar, los métodos actuales para detectar el asma se realizan típicamente mucho después de la aparición de síntomas como sibilancias recurrentes, tos y opresión en el pecho. 10 métodos actuales para detectar el asma se limitan típicamente a pruebas pulmonares funcionales como la espirometría o las pruebas de provocación. Por otra parte, a menudo el médico encarga estas pruebas junto con muchas otras a fin de descartar otras patologías o enfermedades tales como la enfermedad pulmonar obstructiva crónica (EPOC), bronquitis, neumonía e insuficiencia cardíaca congestiva.

30 Hasta la fecha no existe en la técnica un método sencillo y fiable para diagnosticar patologías del tejido pulmonar humano en una fase temprana de su evolución. Asimismo, en la actualidad no existe ningún análisis de sangre disponible que sea capaz de indicar la presencia de una patología concreta del tejido pulmonar. Así pues, resulta recomendable desarrollar un método para determinar la existencia de cánceres de pulmón en una fase temprana de su progresión. Asimismo, resulta recomendable desarrollar un método para diagnosticar el asma y el cáncer de pulmón microcítico y para diferenciar estas patologías entre sí, así como de otras enfermedades pulmonares, tales como infecciones, tan pronto como se manifiesten los síntomas. Resulta asimismo recomendable identificar proteínas específicas presentes en la sangre humana, que, cuando sufren alteraciones en términos de las intensidades de expresión relativas, sean indicativas de la presencia de cáncer de pulmón microcítico y/o asma.

35 WO01/60857 es un documento que revela la misma proteína (llamada aquí PAP1) tal como se describe en la SEC. ID. N°:8 (ver reivindicación 8, abstracto) pero no vinculada con la enfermedad pulmonar descrita aquí. Dai Songwei et al. (06-2007), Science in China, Series C Life Sciences, vol. 50, n°. 3, p.305-311 revela los marcadores para el asma y el cancer de pulmón pero sin vínculo con la proteína de la SEC. ID N°:8 descrita.

## Resumen de la invención

La invención, en su forma más amplia, se define por las reivindicaciones independientes 1 y 3:

Un método que facilita el diagnóstico de las patologías del asma y/o el cáncer de pulmón en pacientes que se encuentran en una etapa temprana de la enfermedad, que consiste en determinar la concentración de proteína específica de CAC69571 (SEC. ID. N°: 8), y opcionalmente una o más de las proteínas que contienen la proteína 4 del dominio FERM (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 (SEC. ID. N°: 9), sintaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), AAK13049 (SEC. ID. N°: 11), BACO4615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°: 2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°: 4), y/o Q8N7P1 (SEC. ID. N°: 7) en el suero de un paciente y comparar los datos de concentración con los obtenidos de las poblaciones con cáncer de pulmón y/o las poblaciones con asma, a fin de verificar o descartar la existencia de las patologías en cuestión.

Un método para evaluar la respuesta de un paciente a intervenciones terapéuticas que consisten en determinar la concentración de proteína específica CAC69571 (SEC. ID. N°: 8), y opcionalmente una o más de las proteínas de la proteína 4 que contienen un dominio (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma JCC1445 (SEC. ID. N°: 9), sintaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), AAK13049 (SEC. ID. N°: 11), BACO4615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°: 2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°: 4), y/o Q8N7P1 (SEC. ID. N°: 7) en el suero de un paciente, donde el paciente padece asma o cáncer de pulmón.

Las realizaciones preferibles se definen en las reivindicaciones dependientes 2, 4 y 5.

## Otros aspectos de la presente divulgación

La presente divulgación ofrece un método novedoso para identificar proteínas presentes en el suero humano que se expresan de forma diferenciada entre los individuos sanos y los pacientes que se sabe que padecen cáncer no microcítico y asma, diagnosticado por un médico, utilizando una espectrometría de masas con ionización por electronebulización acoplada a cromatografía líquida (LC-ESIMS, por sus siglas en inglés). Se realizó una selección de 30 proteínas indicativas de cáncer de pulmón no microcítico y/o asma comparando los datos de la espectrometría de masas, en concreto la masa de los péptidos y las indicaciones gráficas de las intensidades de las proteínas expresadas en el tiempo en una única dimensión. Se compararon miles de proteínas y esto dio como resultado la selección de once proteínas que se expresaron con unas intensidades sustancialmente diferenciadas entre las poblaciones de individuos que no padecen ninguna patología del tejido pulmonar, las poblaciones de individuos que padecen asma, diagnosticado por un médico, y las poblaciones de individuos que padecen cáncer de pulmón no microcítico, diagnosticado por un médico.

Concretamente, los sueros se obtuvieron de una "población normal", una "población con asma" y una "población con cáncer de pulmón". Para los fines del presente, el término "población normal" define de los individuos que se sabe que no padecen asma ni cáncer de pulmón. Para los fines del presente, el término "población con asma" define a los individuos que se sabe que padecen asma y en los que esta patología ha sido diagnosticada por un médico. Para los fines del presente, el término "población con cáncer de pulmón" define a los individuos que se sabe que padecen cáncer de pulmón no microcítico y en los que esta patología ha sido diagnosticada por un médico.

Después de haber obtenido los sueros de la población normal, la población con asma y la población con cáncer de pulmón, cada una de las muestras de suero se dividió en partes alícuotas y se expuso a un agente digestivo o proteasa, concretamente tripsina, para digerir las proteínas presentes en las muestras de suero y convertirlas en péptidos o clivajes predecibles y definidos. Los péptidos creados por la acción enzimática de la tripsina, normalmente conocidos como péptidos trípticos, se separaron de la materia insoluble digerida por la tripsina, sometiendo las muestras a centrifugación para precipitar la materia insoluble. A continuación se sometió la solución supernatante que contenía los péptidos trípticos a cromatografía líquida capilar para realizar la separación temporoespacial de los péptidos trípticos.

Después estos péptidos trípticos se sometieron a LC-ESIMS. Cada péptido se fue separando haciéndolo pasar por una columna de fluido hidrófobo, concretamente agua, acetonitrilo con un contenido del 0,1% en volumen de ácido fórmico en una columna cromatográfica que contenía una fase estacionaria de material de empaquetamiento Supelcosil ABZ + 5 urn, con una longitud de lecho de 18 cm y un diámetro interior de 0,375 mm. Los péptidos separados son transportados por un efluente de la columna. La columna tiene un extremo desde el que los péptidos separados se someten a electronebulización mediante aplicación de alto voltaje al extremo de la columna que tiene un sesgo positivo con respecto a la tierra, formando un haz de pequeñas gotas cargadas que se aceleran hacia la entrada de la LC-ESIMS por la fuerza del campo eléctrico aplicado. La nebulización resultante formada consistía en pequeñas gotas de solvente que contenían los péptidos trípticos disueltos. Las pequeñas gotas se separaron del solvente haciendo pasar por una región de presión atmosférica la fuente de electronebulización y, a continuación, por una entrada capilar calentada de la LC-ESIMS.

La separación del solvente de las pequeñas gotas provocó la deposición de iones de carga positiva, más típicamente hidrógeno ( $H^+$ ) sobre los péptidos, dotando de carga a los péptidos. Estos péptidos cargados en la fase gaseosa se describen en la técnica como "iones pseudomoleculares". Los iones pseudomoleculares se atraen a

través de diversos potenciales eléctricos hacia el analizador de masas de la LC-ESIMS, donde son separados en el espacio y el tiempo en función del ratio de masa:carga. Una vez separados en función del ratio masa:carga, los iones pseudomoleculares se dirigen mediante gradientes del campo eléctrico adicionales hacia un detector de la LC-ESIMS, donde el haz de iones pseudomoleculares se convierte en impulsos eléctricos que son registrados por los dispositivos de registro de datos.

De este modo, los péptidos presentes en los trípticos digeridos pasaron al analizador de masas de la LC-ESIMS, donde se midieron los pesos moleculares de cada péptido, produciendo espectros de masas de tiempo incrementado que se adquieren repetidamente durante la totalidad del tiempo que los péptidos de la muestra salen por la columna. Por lo general, los resultados de los espectros de masas son ilustraciones gráficas de los péptidos encontrados por la LC-ESIMS, donde el eje X es la medida del ratio masa:carga y el eje Y es la intensidad de la señal del péptido. Estos espectros de masas se pueden combinar después en el tiempo, en una visualización tridimensional en la que el eje X es el tiempo de la separación cromatográfica, el eje Z es el eje de la masa del espectro de masas y el eje Y es la intensidad de las señales de los espectros de masas, que es proporcional a la cantidad de un ion pseudomolecular dado detectado por la LC-ESIMS.

A continuación, se realizó un análisis comparativo para contrastar los resultados de los espectros de masas de cada muestra analizada de la población con asma y la población con cáncer de pulmón con los de cada muestra analizada de la población normal. Cada señal ("máxima") de un ion pseudomolecular de un péptido tríptico asociada con una proteína identificada putativa detectada en la LC-ESIMS se comparó en la población con asma, la población con cáncer de pulmón y la población normal. Se consideraron no significativos y, por tanto, se descartaron aquellos péptidos con unas intensidades máximas de espectros de masas que indicaban que las cantidades de los péptidos no presentaban una alteración sustancial al comparar la población con asma o la población con cáncer de pulmón con la población normal. Por lo general, los criterios de exclusión utilizados implicaban la comparación de las intensidades máximas de los péptidos de al menos la mitad de los péptidos característicos identificados para una proteína dada en al menos diez conjuntos de datos obtenidos del análisis de sueros de pacientes individuales de cada patología. Si la intensidad de la mayoría de los máximos de los péptidos obtenidos de una proteína dada presentaban una intensidad al menos diez veces superior en el 80% de los conjuntos de datos de los sueros, la proteína se clasificaba como regulada de forma diferenciada entre las dos clases patológicas.

Como resultado del análisis comparativo, se determinó que once proteínas se expresan constantemente de forma diferenciada entre la población con asma, la población con cáncer de pulmón y la población normal. Las once proteínas se identificaron por referencia a bases de datos o bibliotecas de proteínas y péptidos conocidas. Algunos ejemplos de estas bases de datos incluyen Entrez Protein mantenida por el National Center for Biotechnology Information ("NCBIInr"), ExPASy mantenida por el Swiss Bioinformatics Institute ("SwissProt"), y la Mass Spectral Database ("MSDB") del Medical Research Council Clinical Science Center del Imperial College of London.

Los resultados de los espectros de masas de cada muestra de la población normal, la población con cáncer de pulmón y la población con asma se introdujeron en un motor de búsqueda conocido llamado Mascot. Mascot es un motor de búsqueda conocido en la técnica que emplea los datos de la espectrometría de masas para identificar proteínas de cuatro importantes bases de datos de secuencias, concretamente las bases de datos MSDB, NCBIInr, SwissProt y dbEST. Se introdujeron los parámetros y criterios de búsqueda en el programa Mascot y se procesó cada una de las muestras con este programa. A continuación, el programa Mascot procesó los péptidos introducidos con las bases de datos de secuencias, comparando las masas e intensidades máximas de cada péptido con las masas e intensidades máximas de las proteínas y péptidos conocidos. El programa produjo una lista de candidatas de posibles coincidencias, normalmente denominadas "coincidencias significativas" para cada péptido procesado.

El programa Mascot determina las coincidencias significativas asignando una puntuación denominada "puntuación Mowse" para cada muestra analizada. La puntuación Mowse es un algoritmo donde la puntuación es  $-10 \cdot \text{LOG}_{10}(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio, que se correlaciona con un valor p de significación donde p es inferior a 0,05, que es el estándar generalmente aceptado en la comunidad científica. Por lo general, unas puntuaciones Mowse de aproximadamente 55 a 66 o superiores se consideran significativas. El nivel de significación varía en cierta medida debido a consideraciones de búsqueda y a parámetros de las bases de datos específicos. Se obtuvieron coincidencias significativas para cada péptido procesado, lo que dio como resultado una lista de proteínas candidatas.

A continuación se contrastan los péptidos con las proteínas de las coincidencias significativas para determinar la identidad de los péptidos procesados con el programa Mascot. Se realizó un análisis manual de cada péptido identificado con el programa Mascot y de cada proteína de las coincidencias significativas. Se excluyeron las coincidencias de intensidad máxima que se determinó que eran resultado del "ruido" químico o electrónico. Los datos de los resultados de los espectros de masas se contrastaron con las coincidencias significativas para confirmar los datos brutos, las identidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos.

A continuación se realizó una búsqueda inversa para añadir péptidos a la lista de candidatos que podrían haber sido omitidos por la búsqueda automática a través del programa Mascot. Los péptidos adicionales se identificaron seleccionando la "mejor coincidencia", es decir la proteína individual que coincidía de forma sustancial con cada

parámetro del péptido comparado, realizando una digestión con silicio donde los péptidos trípticos y sus respectivas masas moleculares se calculan basándose en la secuencia genética o de aminoácidos conocida de la proteína. Estas masas estimadas de los péptidos se comparan entonces con los datos de espectros de masas en bruto y cualesquiera máximos identificados se examinan y cualifican tal como se ha descrito anteriormente. A continuación, todos los péptidos, incluyendo aquellos identificados automáticamente por Mascot y aquellos identificados mediante examen manual, se introducen en la lista de masas utilizada por Mascot. La coincidencia refinada se utiliza después para obtener la puntuación Mowse refinada, tal como se debate a continuación.

Como resultado del proceso de identificación, las once proteínas que se determinó que se expresaban de forma significativamente diferenciada entre la población con asma, la población con cáncer de pulmón y la población normal se identificaron como BAC04615, Q6NSC8, CAF17350, Q6ZUD4, Q8N7P1, CAC69571, proteína 4 que contiene el dominio FERM, forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445, sintaxina 11, AAK13083, y AAK130490. BAC04615, Q6NSC8, CAF 17350, Q6ZUD4, Q8N7P1 son proteínas identificadas resultantes de esfuerzos de secuenciación genética. La proteína 4 que contiene el dominio FERM se sabe que participa en el anclaje a la membrana de la proteína intracitoplásmica. La forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 es un proteasoma conocido. La sintaxina 11 es activa en la respuesta inmune celular. BAC04615, AAK13083, y AAK130490 son proteínas asociadas con el complejo mayor de histocompatibilidad (MHC).

#### Breve descripción de las ilustraciones

La Fig. 1 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína BAC04615;

La Fig. 2 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína Q6NSC8;

La Fig. 3 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína CAF17350;

La Fig. 4 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína Q6ZUD4;

La Fig. 5 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína Q8N7P1;

La Fig. 6 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína CAC69571;

La Fig. 7 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína 4 que contiene el dominio FERM;

La Fig. 8 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445;

La Fig. 9 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para la proteína sintaxina 11;

La Fig. 10 divulga una tabla que muestra los resultados del algoritmo Mowse y coincidencias significativas para las proteínas de AAK13083 y AAK13049;

#### Descripción detallada de la invención

La presente invención, en su forma más amplia, se define en las reivindicaciones independientes 1 y 3:

Un método que facilita el diagnóstico de las patologías del asma y/o el cáncer de pulmón en pacientes que se encuentran en una etapa temprana de la enfermedad, que consiste en determinar la concentración de proteína específica de CAC69571 (SEC. ID. N°: 8), y opcionalmente una o más de las proteínas de proteína 4 que contiene el dominio FERM (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 (SEC. ID. N°: 9), sintaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), y AAK13049 (SEC. ID. N°: 11) BAC04615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°:2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°:4), y/o Q8N7P1 (SEC. ID. N°:7) en el suero de un paciente y comparar los datos de concentración con los obtenidos de las poblaciones con cáncer de pulmón y/o las poblaciones con asma, a fin de verificar o descartar la existencia de las patologías en cuestión.

Un método para evaluar la respuesta de un paciente a intervenciones terapéuticas que consisten en determinar la concentración de proteína específica de CAC69571 (SEC. ID. N°:8), y opcionalmente una o más de las proteínas de la proteína 4 que contiene el dominio FERM (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 (SEC. ID. N°: 9), sintaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), y AAK13049 (SEC. ID. N°: 11) BAC04615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°:2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°:4), y/o Q8N7P1 (SEC. ID. N°:7) en el suero de un paciente, donde el paciente padece asma o cáncer de pulmón.

Las realizaciones preferibles se definen en las reivindicaciones dependientes 2, 4 y 5.

La presente invención proporciona un método para identificar e identifica proteínas presentes en el suero humano, que se expresan de forma diferenciada entre los individuos sanos y en los individuos que se sabe que padecen cáncer de pulmón no microcítico y asma, diagnosticados por un médico, utilizando espectrometría de masas con ionización por electronebulización acoplada a cromatografía líquida. Al determinar las proteínas que se expresan de forma sustancial y constantemente diferenciada entre las poblaciones que no padecen ninguna patología de tejido pulmonar humano, las poblaciones a las que se les ha diagnosticado asma y las poblaciones a las que se les ha diagnosticado cáncer de pulmón no microcítico, y al obtener la identidad de estas proteínas, resulta posible identificar la presencia de la patología en un paciente a través de análisis de sangre en los que se identifiquen las mismas proteínas y se cuantifiquen los niveles de expresión de estas proteínas, a fin de identificar y diagnosticar el asma o cáncer de pulmón no microcítico en una fase mucho más temprana de la progresión de las respectivas enfermedades.

Se tomaron muestras de sangre humana a los voluntarios. Se recogieron treinta muestras de individuos que se sabe que no padecen cáncer de pulmón no microcítico ni asma. Los individuos que se sabe que no padecen cáncer de pulmón no microcítico ni asma comprenden y se denominan en el presente "población normal". Por otra parte, para los fines del presente, el término "cáncer de pulmón" describe los cánceres de pulmón no microcíticos. Se recogieron veintiocho muestras de individuos que se sabe que padecen asma, con el diagnóstico confirmado por un médico. Los individuos que se sabe que padecen asma comprenden y se denominan en el presente la "población con asma". Se recogieron treinta muestras de sangre de individuos que se sabe que padecen cáncer de pulmón no microcítico, con el diagnóstico confirmado por un médico. Los individuos que se sabe que padecen asma comprenden y se denominan en el presente la "población con cáncer de pulmón". Por lo general, para los fines del presente, los términos "cáncer de pulmón" o "cánceres de pulmón" se refieren a los cánceres de pulmón no microcíticos. Por último, se recogieron setenta y una muestras de individuos que se sabe que corren el riesgo de padecer cáncer de pulmón debido a su historial de tabaquismo, confirmado por un médico. Estas setenta y una muestras están siendo objeto de la investigación y experimentación en la actualidad y, por consiguiente, no se comentan en el presente.

Se tomaron muestras de sangre a los voluntarios de conformidad con un protocolo aprobado por el Comité Institucional de Revisión (IRB, por sus siglas en inglés), tras la firma de un consentimiento informado, utilizando técnicas estándar de venopunción con tubos de vidrio estériles de tapón rojo para suero de 10 ml BD Vacutainer®. Las muestras de sangre se dejaron reposar a temperatura ambiente durante treinta minutos para permitir la coagulación de la sangre. A continuación se centrifugaron en una centrifugadora de sobremesa estándar a 2.000 rpm a temperatura ambiente durante diez minutos, para separar el suero de las muestras de sangre. Se retiró el suero de cada muestra vertiéndolo en tubos secundarios. Los tubos secundarios se enfriaron previamente con hielo para garantizar la integridad de cada muestra de suero, limitando la degradación resultante de la proteólisis y la desnaturalización. Se recogieron las muestras de suero de cada muestra recogida y se dividieron en partes alícuotas de 1,0 ml en tubos Cryovial previamente enfriados con hielo. Las partes alícuotas de las muestras de suero se almacenaron a una temperatura no superior a ochenta grados Celsius bajo cero (-80°C). El tiempo de procesamiento no fue superior a una hora desde el momento de la flebotomía hasta su almacenamiento a -80°C.

Se seleccionaron aleatoriamente entre ocho y diez muestras de suero de cada una de las poblaciones (normal, con asma y con cáncer de pulmón) para su análisis. Cada una de las muestras de suero de cada población se trató con un agente digestivo o proteasa, en este caso tripsina. La tripsina fue la proteasa utilizada y resulta recomendable su utilización debido a su capacidad para realizar clivajes altamente específicos y altamente predecibles, debido al hecho de que se sabe que la tripsina cliva las cadenas peptídicas en el extremo carboxilo de la lisina y arginina, salvo cuando hay una prolina presente inmediatamente después de la lisina o arginina. A pesar de que se utilizó tripsina, resulta posible utilizar otras proteasas o agentes digestivos. Resulta recomendable utilizar una proteasa, o mezcla de proteasas, que permitan un clivaje al menos tan específico como la tripsina.

Los péptidos tripticos, que son los péptidos que deja la tripsina tras el clivaje, se separan después de la materia insoluble sometiendo las muestras a centrifugado y a una cromatografía líquida capilar, con un gradiente de acetonitrilo acuoso con un 0,1% de ácido fórmico utilizando una columna 0.375 X 180 mm Supelcosil ABZ+ en un sistema de HPLC capilar Eksigent 2D para conseguir la resolución cromatográfica de los péptidos tripticos generados. Esta separación de los péptidos es necesaria porque el proceso de ionización por electronebulización está sujeto a una co-supresión de iones, donde los iones de un tipo que tienen una mayor afinidad por el protón suprimirá la formación de iones que tienen afinidades inferiores por el protón si se extraen simultáneamente del emisor de electronebulización, que en este caso es coterminal con el extremo de la columna de HPLC.

Esta metodología permite la separación del mayor número de péptidos producidos en las digestiones de tripticos y contribuye a minimizar los problemas de co-supresión, maximizando así las oportunidades de formación de co-supresión de iones pseudomoleculares y, por tanto, el muestreo de iones. Los péptidos tripticos de cada muestra se sometieron a continuación a una LC-ESIMS. La LC-ESIMS separó cada péptido de cada muestra con el tiempo, pasando los péptidos de cada muestra a través de una columna de un sistema de solventes compuesto por agua, acetonitrilo y ácido fórmico, tal y como se ha descrito anteriormente.

A continuación, los péptidos fueron nebulizados con una fuente de ionización por electronebulización para ionizar los péptidos y producir los iones pseudomoleculares de los péptidos, tal como se ha descrito anteriormente. A

continuación, los péptidos pasaron por un analizador de masas de la LC-ESIMS, donde se midieron las masas moleculares de cada uno de los iones pseudomoleculares de los péptidos. Después de someterse a la LC-ESIMS, se produjeron resultados de los espectros de masas de los péptidos presentes en cada muestra a partir de los datos de los espectros de masas, concretamente de las intensidades de los pesos moleculares y del tiempo de elución de los péptidos de una columna cromatográfica. Por lo general, los resultados de los espectros de masas son ilustraciones gráficas de las señales de los iones pseudomoleculares de los péptidos registradas por la LC-ESIMS, donde el eje X es la medida del ratio masa:carga, mientras que el eje Y es la intensidad de la señal de los iones pseudomoleculares. A continuación, estos datos se procesan con un sistema de software que controla la LC-ESIMS y captura y almacena los datos resultantes.

Después de obtener los datos de los espectros de masas y de generar los correspondientes resultados, se realizó un análisis comparativo en el que se generaron los resultados de los espectros de masas de cada muestra de suero testada con la LC-ESIMS para cada población, tanto dentro de las poblaciones con la misma patología como entre las poblaciones con distintas patologías (en adelante, intrapatológica e interpatológicamente). Se compararon los picos de los espectros de masas entre cada muestra testada de la población normal. A continuación, se compararon los picos de los espectros de masas entre cada muestra testada de la población con asma y la población con cáncer de pulmón. Una vez realizadas las comparaciones intrapatológicas, se realizaron comparaciones interpatológicas, en las que los resultados de los espectros de masas para cada muestra testada en la LC-ESIMS para la población con asma se compararon con cada muestra testada de la población normal. De igual modo, los resultados de los espectros de masas de cada muestra testada en la LC-ESIMS para la población con cáncer de pulmón se compararon con cada muestra testada de la población normal.

Se consideraron no significativos y, por tanto, se descartaron aquellos péptidos con unos resultados de espectros de masas que indicaban que las intensidades de los péptidos se expresaban de forma diferenciada de forma inconstante intrapatológicamente o que no presentaban una alteración sustancial (una variación de la intensidad inferior a 10 x) cuando se comparaba la población con asma o la población con cáncer con la población normal. Por lo general, los criterios de exclusión utilizados implicaban la comparación las intensidades peptídicas máximas para al menos la mitad de los péptidos característicos identificados para una proteína dada en al menos diez conjuntos de datos obtenidos del análisis de los sueros de pacientes individuales de cada patología. Si la intensidad de la mayoría de los máximos peptídicos obtenidos de una proteína dada era al menos 10 veces superior para el 80% de los conjuntos de los datos de los sueros, la proteína se clasificaba como regulada de forma diferenciada entre las dos clases patológicas.

Sin embargo, se desconocía la identidad de las proteínas que daban lugar a los péptidos que se observó que estaban regulados de forma diferenciada, por lo que fue necesario identificarlas. Para realizar la identificación de las proteínas, se compararon las intensidades de la señal de los iones pseudomoleculares de los péptidos entre las bases de datos conocidas que contienen bibliotecas de proteínas y péptidos conocidos y de proteínas y péptidos sospechados.

Los resultados de los espectros de masas de los trípticos digeridos de cada muestra de la población normal, la población con cáncer y la población con asma se introdujeron en un motor de búsqueda conocido denominado Mascot. Mascot es un motor de búsqueda conocido en la técnica, que utiliza los datos de la espectrometría de masas para identificar proteínas de cuatro importantes bases de datos de secuencias, concretamente las bases de datos MSDB, NCBI, SwissProt y dbEST. Estas bases de datos contienen información sobre todas las secuencias conocidas y todas las proteínas putativas, basándose en la observación de las regiones de inicio de la transcripción de proteínas características obtenidas de las secuencias genéticas. La precisión y redundancia de estas bases de datos se comprueba continuamente y están sujetas a una adición constante, a medida que se identifican y publican nuevas secuencias genéticas y proteínas en la literatura de patentes y científica.

Como resultado del análisis comparativo, se determinó que once proteínas se expresan constantemente de forma diferenciada entre la población con asma, la población con cáncer de pulmón y la población normal. Se introdujeron parámetros y criterios de búsqueda en el programa Mascot y los datos de los espectros de masas de los resultados de los espectros de masas de cada población se procesaron en el mencionado programa. Los datos de los espectros de masas introducidos en el programa Mascot fueron los de todas las muestras de cada patología. A continuación, el programa Mascot procesó los datos espectrales de los péptidos introducidos con las bases de datos de secuencias, comparando las masas e intensidades máximas de cada péptido con las masas e intensidades máximas de las proteínas y péptidos conocidos. Mascot produjo un resultado de la búsqueda que proporcionó una lista de candidatos de posibles coincidencias de identificación de proteínas, normalmente conocidas como "coincidencias significativas" para cada muestra analizada.

El programa Mascot determina las coincidencias significativas asignando una puntuación denominada "puntuación Mowse" para cada muestra analizada. La puntuación Mowse es un algoritmo donde la puntuación es  $-10 \cdot \text{LOG}_{10}(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio, que se correlaciona con un valor p de significación donde p es inferior a 0,05, que es el estándar generalmente aceptado en la comunidad científica. Por lo general, unas puntuaciones Mowse de aproximadamente 55 a 66 o superiores se consideran significativas. El nivel de significación varía en cierta medida debido a consideraciones de búsqueda y a parámetros de las bases de datos específicos. Se obtuvieron coincidencias significativas para cada péptido procesado, lo que dio como resultado una lista de proteínas candidatas.

A continuación, se realizó un análisis comparativo, comparando los resultados de los espectros de masas de cada muestra analizada de la población con asma y la población con cáncer de pulmón con los de cada muestra analizada de la población normal. La señal (máxima) de los iones pseudomoleculares de cada péptido tríptico asociada con una proteína identificada putativa que se detectó en la LC-ESIMS se comparó en la población con asma, la población con cáncer y la población normal. Los péptidos con intensidades máximas de los espectros de masas que indicaban que las cantidades de péptido no se encontraban sustancialmente alteradas al comparar la población con asma o la población con cáncer de pulmón con la población normal se consideraron no significativos y se excluyeron. Por lo general, los criterios de exclusión utilizados implicaban la comparación de las intensidades peptídicas máximas para al menos la mitad de los péptidos característicos identificados para una proteína dada en al menos diez conjuntos de datos obtenidos del análisis de los sueros de pacientes individuales de cada patología. Si la intensidad de la mayoría de los máximos de los péptidos obtenidos de una proteína determinada presentaban una intensidad al menos 10 veces superior en el 80% de los conjuntos de datos de los sueros, la proteína se clasificó como regulada de forma diferenciada entre las dos clases patológicas.

Los datos de los resultados de los espectros de masas se contrastaron con las coincidencias significativas para confirmar los datos brutos, las identidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. A continuación se realizó una búsqueda inversa para añadir péptidos a la lista de candidatos que podrían haber sido omitidos por la búsqueda automática a través del programa Mascot. Los péptidos adicionales se identificaron seleccionando la "mejor coincidencia", es decir la proteína individual que coincidía de forma sustancial con cada parámetro del péptido comparado, realizando una digestión con silicio donde los péptidos trípticos y sus respectivas masas moleculares se calculan basándose en la secuencia genética o de aminoácidos conocida de la proteína. Estas masas estimadas de los péptidos se comparan entonces con los datos de espectros de masas en bruto y cualesquiera máximos identificados se examinan y cualifican tal como se ha descrito anteriormente. A continuación, todos los péptidos, incluyendo aquellos identificados automáticamente por Mascot y aquellos identificados mediante examen manual, se introducen en la lista de masas utilizada por Mascot. La coincidencia refinada se utiliza después para obtener la puntuación Mowse refinada, tal como se presenta a continuación.

Con respecto a las Fig. 1 a 10, los resultados de la búsqueda de Mascot se muestran para cada proteína identificada y expresada de forma diferenciada entre la población con cáncer de pulmón o la población con asma en comparación con la población normal. En cada caso, se introdujeron los parámetros y criterios de búsqueda, y se estableció un máximo para la puntuación Mowse en términos de aceptabilidad y significación. Con respecto a la Fig. 1, se muestra el resultado de una búsqueda con Mascot para la proteína BAC04615. La base de datos seleccionada para la búsqueda fue NCBI nr 10 y la taxonomía de las muestras introducida en el programa Mascot se estableció como Homo sapiens 12. El máximo de la puntuación Mowse en términos de significación se estableció en un valor Mowse de 66 o superior 14. Como resultado de la búsqueda de Mascot, se obtuvo una puntuación máxima de 121, tal como indica el gráfico de la puntuación Mowse 18, donde el eje Y del gráfico indica el número de proteínas identificadas que tienen una puntuación Mowse concreta.

Con respecto también a la Fig. 1, la puntuación Mowse más alta de ciento veintiuno se otorgó a gi/21755032, tal como indica la fila 20. Una puntuación Mowse de 121 es altamente significativa, lo que quiere decir que existe una probabilidad muy baja de que la coincidencia sea aleatoria. De hecho, tal como se indica en la columna 28, la expectativa de que esta coincidencia se pueda producir de forma aleatoria se indica por el programa Mascot como  $1,7 \times 10^{-07}$ . Sin embargo, las proteínas indicadas en las filas 22, 24 y 26 también obtuvieron unas puntuaciones Mowse muy elevadas, lo que indica que estas tres proteínas también son coincidencias significativas. A continuación se realizó el análisis manual, donde se excluyeron los datos no significativos o ruido, y se contrastaron los datos en bruto, las intensidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. Como resultado del análisis manual, la probabilidad de que las proteínas indicadas en las filas 22, 24 y 26 sean coincidencias significativas se redujo de forma notable y, por tanto, las proteínas indicadas en estas filas se excluyeron de las coincidencias. La proteína indicada en la fila 20, gi/21 755032, se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 1. El número de proteína indicado en la fila 20, gi/21755032, donde el número gi (en ocasiones escrito "GI") es simplemente una serie de dígitos que se asignan consecutivamente a cada registro de secuencia procesado por NCBI. gi/21755032 corresponde a la proteína BAC04615.

Con respecto a la Fig. 2, se divulga el resultado de la búsqueda de Mascot para la proteína Q6NSC8. El umbral de la puntuación Mowse de significación 29 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse máxima de ciento diecisiete, tal como indica la barra de puntuación Mowse 36 del gráfico de la puntuación Mowse 30. La proteína identificada que se correlacionaba con la barra de puntuación Mowse 36 es Q6NSC8, tal como se indica en la fila 32. Tal como se muestra en la Fig. 2, la parte sombreada 34 del gráfico de la puntuación Mowse 30 indica proteínas que fueron registradas, pero que se situaron por debajo del umbral de significación y que, por tanto, no fueron consideradas.

Con respecto a la Fig. 3, se divulga el resultado de la búsqueda de Mascot para la proteína CAF17350. El umbral de la puntuación Mowse de significación 38 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse de ciento cincuenta y dos, tal como indica la barra de puntuación Mowse 42 del gráfico de la puntuación Mowse 40. La proteína identificada que se correlacionaba con la barra de puntuación Mowse 42 es

CAF17350, tal como se indica en la fila 46. Tal como se muestra en la Fig. 3, la parte sombreada 44 del gráfico de la puntuación Mowse 40 indica proteínas que fueron registradas, pero que se situaron por debajo del umbral de significación y que, por tanto, no fueron consideradas.

Con respecto a la Fig. 4, se divulga el resultado de la búsqueda de Mascot para la proteína Q6ZUD4. El umbral de la puntuación Mowse de significación 48 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse de doscientos veinte, tal como indica la barra de puntuación Mowse 52 del gráfico de la puntuación Mowse 50. La proteína identificada que se correlacionaba con la barra de puntuación Mowse 52 es Q6ZUD4, tal como se indica en la fila 56. Tal como se muestra en la Fig. 4, la parte sombreada 54 del gráfico de la puntuación Mowse 50 indica proteínas que fueron registradas, pero que se situaron por debajo del umbral de significación y que, por tanto, no fueron consideradas.

Con respecto a la Fig. 5, se divulga el resultado de la búsqueda de Mascot para la proteína Q8N7P1. El umbral de la puntuación Mowse de significación 58 se estableció como el valor Mowse de sesenta y seis, y se obtuvo una puntuación Mowse de setenta y cuatro, tal como indica la barra de puntuación Mowse 62 del gráfico de la puntuación Mowse 60. La proteína identificada que se correlacionaba con la barra de puntuación Mowse 62 es gi/71682143, tal como se indica en la fila 64. De forma similar a lo que ocurre en la Fig. 1, gi/71682143 se corresponde con la proteína Q8N7P1. Las proteínas indicadas en las filas 66 y 68 también presentaban unas puntuaciones Mowse muy elevadas, lo que indica que también son coincidencias significativas. A continuación se realizó el análisis manual, donde se excluyeron los datos no significativos o ruido, y se contrastaron los datos en bruto, las intensidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. Como resultado del análisis manual, la probabilidad de que las proteínas indicadas en las filas 66 y 68 sean coincidencias significativas se redujo de forma notable y, por tanto, las proteínas indicadas en estas filas se excluyeron de las coincidencias. Q8N7P1 se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 5. La indicación en 70 de la proteína Q8NB22 se debe a que es la misma proteína que Q8N7P1.

Con respecto a la Fig. 6, se divulga el resultado de la búsqueda de Mascot para la proteína CAC69571. El umbral de la puntuación Mowse de significación 72 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse de ciento setenta y uno, tal como indica la barra de puntuación Mowse 76 del gráfico de la puntuación Mowse 74. La proteína indicada que se correlacionaba con la barra de puntuación Mowse 76 es CAC69571, tal como se indica en la fila 78. Las proteínas indicadas en las filas 80, 82, 84 y 86 también presentaban unas puntuaciones Mowse muy elevadas, lo que indica que estas cuatro proteínas también son coincidencias significativas. A continuación se realizó el análisis manual, donde se excluyeron los datos no significativos o ruido, y se contrastaron los datos en bruto, las intensidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. Como resultado del análisis manual, la probabilidad de que las proteínas indicadas en las filas 80, 82, 84 y 86 sean coincidencias significativas se redujo de forma notable y, por tanto, las proteínas indicadas en estas filas se excluyeron de las coincidencias. CAC69571 se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 6.

Con respecto a la Fig. 7, se divulga el resultado de la búsqueda de Mascot para la proteína 4 que contiene el dominio FERM 4. El umbral de la puntuación Mowse de significación 88 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse de trescientos treinta y cinco, tal como indica la barra de puntuación Mowse 92 del gráfico de la puntuación Mowse 90. La proteína indicada que se correlacionaba con la barra de puntuación Mowse 92 es la proteína 4 que contiene el dominio FERM 4, tal como se indica en la fila 98. Las proteínas indicadas en las filas 100, 102, 104, 106 y 108 también presentaban unas puntuaciones Mowse muy elevadas, lo que indica que estas cinco proteínas también son coincidencias significativas. A continuación se realizó el análisis manual, donde se excluyeron los datos no significativos o ruido, y se contrastaron los datos en bruto, las intensidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. Como resultado del análisis manual, la probabilidad de que las proteínas indicadas en las filas 100, 102, 104, 106 y 108 sean coincidencias significativas se redujo de forma notable y, por tanto, las proteínas indicadas en estas filas se excluyeron de las coincidencias. La proteína 4 que contiene el dominio FERM 4 se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 7.

Con respecto a la Fig. 8, se divulga el resultado de la búsqueda de Mascot para la forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 ("JCC1445"). El umbral de la puntuación Mowse de significación 110 se estableció como el valor Mowse de sesenta y seis, y se obtuvo una puntuación Mowse de ciento veintitrés, tal como indica la barra de puntuación Mowse 114 del gráfico de la puntuación Mowse 112. La proteína identificada que se correlacionaba con la barra de puntuación Mowse 114 es gi/4506179, tal como se indica en la fila 116. gi/4506179 se corresponde con la proteína JCC1445. Las proteínas indicadas en las filas 118, 120, 122, 124, 126 y 128 también presentaban unas puntuaciones Mowse muy elevadas, lo que indica que estas seis proteínas también son coincidencias significativas. A continuación se realizó el análisis manual, donde se excluyeron los datos no significativos o ruido, y se contrastaron los datos en bruto, las intensidades máximas, las multiplicidades de carga, la distribución de isótopos y los estados de carga de los flancos. Como resultado del análisis manual, la probabilidad de que las proteínas indicadas en las filas 118, 120, 122, 124, 126 y 128 sean coincidencias significativas se redujo de forma notable y, por tanto, las proteínas indicadas en estas filas se

excluyeron de las coincidencias. JCC1445 se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 8.

Con respecto a la Fig. 9, se divulga el resultado de la búsqueda de Mascot para la proteína syntaxina 11. El umbral de la puntuación Mowse de significación 130 se estableció como el valor Mowse de sesenta y seis, y se obtuvo en dos ocasiones una puntuación Mowse de ciento veintisiete, tal como indican las barras de puntuación Mowse 134 y las filas 136 y 138. Se obtuvo una tercera puntuación Mowse de 95 para la syntaxina 11, tal como se indica en la fila 140. La syntaxina 11 se identificó como la proteína indicada por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 9.

Con respecto a la Fig. 10, se divulgan los resultados de la búsqueda de Mascot para las proteínas AAK13083 y AAK13049. El umbral de la puntuación Mowse de significación 142 se estableció como el valor Mowse de sesenta y cuatro, y se obtuvo una puntuación Mowse de doscientos setenta y tres con la proteína Q5VY82, tal como indica la fila 148 y la barra de puntuación Mowse 146. Las proteínas indicadas en las filas 150, 152 y 154 también presentaban unas puntuaciones Mowse muy elevadas, lo que indica que estas tres proteínas también son coincidencias significativas. Sin embargo, como resultado del análisis manual realizado, las proteínas indicadas en las filas 150 y 154 se eliminaron como coincidencias probables. Q5VY82 se está sometiendo a nuevos trabajos de investigación y experimentación para determinar si se expresa de forma significativamente diferenciada. AAK13049, indicada en la fila 152, y AAK13083 fueron ambas identificadas como las proteínas indicadas por los datos de espectros de masas introducidos en el programa Mascot en la Fig. 10.

Las Fig. 1 a 10 divulgan análisis de datos que se realizaron para identificar las once proteínas que se expresan de forma diferenciada en la población con asma y/o la población con cáncer de pulmón en comparación con la población normal. El proceso descrito en el presente, tal como se indica en las Fig. 1 a 10, se realizó para cada una de las once proteínas, para la población con asma, la población normal y la población con cáncer de pulmón.

Como resultado del proceso de identificación, las once proteínas que se determinó que se expresaban de forma significativamente diferenciada entre la población con asma, la población con cáncer de pulmón y la población normal se identificaron como BAC04615, Q6NSC8, CAF17350, Q6ZUD4, Q8N7P1, CAC69571, proteína 4 que contiene el dominio FERM, forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445, syntaxina 11, AAK13083, y AAK130490. BAC04615, Q6NSC8, CAF 17350, Q6ZUD4, Q8N7P1 son proteínas identificadas resultantes de esfuerzos de secuenciación genética. La proteína 4 que contiene el dominio FERM se sabe que participa en el anclaje a la membrana de la proteína intracitoplásmica. La forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 es un proteasoma conocido. La syntaxina 11 es activa en la respuesta inmune celular. BAC04615, AAK13083, y AAK130490 son proteínas asociadas con el complejo mayor de histocompatibilidad (MHC).

Habiendo identificado once proteínas específicas que se expresan de forma constantemente diferenciada en los pacientes con asma y cáncer de pulmón, resulta posible diagnosticar estas patologías en una fase de su progresión temprana, sometiendo las proteínas BAC04615, Q6NSC8, CAF17350, Q6ZUD4, Q8N7P1, CAC69571, proteína 4 que contiene el dominio FERM, forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445, syntaxina 11, AAK13083, y AAK130490 del suero de un paciente a la LC-ESIMS, obteniendo los datos de espectros de masas de estas proteínas, y comparando estos datos con los datos de espectros de masas de la población normal. Se puede realizar otro análisis comparando los datos de espectros de masas con los de la población con cáncer de pulmón y/o asma, para verificar o descartar la presencia de las patologías dadas.

Por supuesto, el análisis se podría extrapolar a muchas otras técnicas en las que las que se pueden determinar las concentraciones de proteínas específicas, incluyendo, a título meramente enunciativo, las siguientes: radioinmunoensayo, inmunoensayo absorbente vinculado a enzimas, cromatografía líquida de alta resolución con radiométrica, detección espectrométrica mediante absorbancia de luz ultravioleta o visible, análisis cuantitativo y cualitativo por espectrometría de masas, inmunoensayo western-blot, electroforesis en gel unidimensional o bidimensional con visualización cuantitativa mediante detección de núcleos o sondas radiactivas, detección basada en anticuerpos con fotometría de absorción o fluorescencia, cuantificación por luminiscencia de cualesquiera de una serie de sistemas reporteros quimioluminiscentes, ensayos enzimáticos, inmunoprecipitación o inmunoensayos de captura, o cualesquiera de una serie de inmunoensayos en fase sólida y líquida.

Además de determinar la existencia de cáncer de pulmón o asma en una fase temprana de la enfermedad, las proteínas identificadas en el presente como indicativas de estas patologías se podrían utilizar y aplicar de formas relacionadas para el objetivo del tratamiento del cáncer de pulmón y/o del asma. Por ejemplo, se pueden desarrollar anticuerpos que se unan a estas proteínas. Los anticuerpos se podrían juntar en un grupo de biomarcadores donde cualquiera de los anticuerpos o todos ellos se unirían en un grupo o kit basado en un único gránulo para un inmunoensayo de los gránulos. A continuación, las proteínas se podrían someter a un inmunoensayo multiplexado utilizando tecnologías basadas en gránulos, como las tecnologías xMAP de Luminex, y cuantificarse. Asimismo, se podrían emplear ensayos no basados en gránulos para cuantificar los niveles de expresión de las proteínas. Al cuantificar los niveles de expresión de las proteínas, los resultados cuantificables se pueden comparar con los niveles de expresión de la población normal, la población con asma y/o la población con cáncer de pulmón para verificar o descartar la presencia de cáncer de pulmón o asma en el paciente.

Las proteínas se podrían utilizar y aplicar también en el campo farmacológico, para evaluar la respuesta de un paciente a intervenciones terapéuticas tales como el tratamiento con fármacos, la radiación/quimioterapia o el tratamiento quirúrgico. Por otra parte, se podrían utilizar kits para medir proteínas individuales o un grupo de proteínas para el análisis rutinario de un paciente diseñado para controlar el estado de salud de un paciente que corre un mayor riesgo de padecer las patologías, como fumadores, o aquellos con un historial familiar de estas patologías.

Finalmente, a continuación se proporciona una lista de las secuencias de aminoácidos para cada una de las once proteínas identificadas en el presente documento. En el Listado de secuencias, la secuencia de aminoácidos divulgada en la SEC. ID. N°: 1 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína BAC04615. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 2 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína Q6NSC8. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 3 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína CAF17350. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 4 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína Q6ZUD4. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 5 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína 4 que contiene el dominio PERM. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 6 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína AKK13083. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 7 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína Q8N7P1. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 8 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína CAC69571. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 9 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma JCC1445. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 10 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína syntaxina 11. La secuencia de aminoácidos divulgada en la SEC. ID. N°: 11 es la secuencia de aminoácidos primaria conocida en la fecha de presentación de la presente solicitud para la proteína AKK13049.

<110> Streeper, Robert T.; Izbicka, Elzbieta; Baek, Sung H.

<120> Método para identificar proteínas en el suero humano que son indicativas de patologías del tejido pulmonar humano

<130> P9286.2

5 <140>

<141>

<150> US 60/971,422

<151> 2007-09-11

<160> 11

10 <210> 1

<211> 319

<212> PRT

<213> Homo sapiens

<400> 1

Met	Val	Leu	Ser	Glu	Leu	Ala	Ala	Arg	Leu	Asn	Cys	Ala	Glu	Tyr
				5					10					15
Lys	Asn	Trp	Val	Lys	Ala	Gly	His	Cys	Leu	Leu	Leu	Leu	Arg	Ser
				20					25					30
Cys	Leu	Gln	Gly	Phe	Val	Gly	Arg	Glu	Val	Leu	Ser	Phe	His	Arg
				35					40					45
Gly	Leu	Leu	Ala	Ala	Ala	Pro	Gly	Leu	Gly	Pro	Arg	Ala	Val	Cys
				50					55					60
Arg	Gly	Gly	Ser	Arg	Cys	Ser	Pro	Arg	Ala	Arg	Gln	Phe	Gln	Pro
				65					70					75
Gln	Cys	Gln	Val	Cys	Ala	Glu	Trp	Lys	Arg	Glu	Ile	Leu	Arg	His
				80					85					90
His	Val	Asn	Arg	Asn	Gly	Asp	Val	His	Trp	Gly	Asn	Cys	Arg	Pro
				95					100					105
Gly	Arg	Trp	Pro	Val	Asp	Ala	Trp	Glu	Val	Ala	Lys	Ala	Phe	Met
				110					115					120
Pro	Arg	Gly	Leu	Ala	Asp	Lys	Gln	Gly	Pro	Glu	Glu	Cys	Asp	Ala
				125					130					135
Val	Ala	Leu	Leu	Ser	Leu	Ile	Asn	Ser	Cys	Asp	His	Phe	Val	Val
				140					145					150
Asp	Arg	Lys	Lys	Val	Thr	Glu	Val	Ile	Lys	Cys	Arg	Asn	Glu	Ile
				155					160					165
Met	His	Ser	Ser	Glu	Met	Lys	Val	Ser	Ser	Thr	Trp	Leu	Arg	Asp
				170					175					180
Phe	Gln	Met	Lys	Ile	Gln	Asn	Phe	Leu	Asn	Glu	Phe	Lys	Asn	Ile
				185					190					195
Pro	Glu	Ile	Val	Ala	Val	Tyr	Ser	Arg	Ile	Glu	Gln	Leu	Leu	Thr
				200					205					210
Ser	Asp	Trp	Ala	Val	His	Ile	Pro	Glu	Glu	Asp	Gln	Arg	Asp	Gly
				215					220					225
Cys	Glu	Cys	Glu	Met	Gly	Thr	Tyr	Leu	Ser	Glu	Ser	Gln	Val	Asn
				230					235					240
Glu	Ile	Glu	Met	Gln	Leu	Leu	Lys	Glu	Lys	Leu	Gln	Glu	Ile	Tyr
				245					250					255
Leu	Gln	Ala	Glu	Glu	Gln	Glu	Val	Leu	Pro	Glu	Glu	Leu	Ser	Asn
				260					265					270
Arg	Leu	Glu	Val	Val	Lys	Glu	Phe	Leu	Arg	Asn	Asn	Glu	Asp	Leu
				275					280					285
Arg	Asn	Gly	Leu	Thr	Glu	Asp	Met	Gln	Lys	Leu	Asp	Ser	Leu	Cys
				290					295					300

Leu	His	Gln	Lys	Leu	Asp	Ser	Gln	Glu	Pro	Gly	Arg	Gln	Thr	Pro
				305					310					315
Asp	Arg	Lys	Ala											

<210> 2

<211> 57

5 <212> PRT

<213> Homo sapiens

<400> 2

Met	Ser	Cys	Leu	Met	Val	Glu	Arg	Cys	Gly	Glu	Ile	Leu	Phe	Glu
				5					10					15
Asn	Pro	Asp	Gln	Asn	Ala	Lys	Cys	Val	Cys	Met	Leu	Gly	Asp	Ile
				20					25					30
Arg	Leu	Arg	Gly	Gln	Thr	Gly	Val	Arg	Ala	Glu	Arg	Arg	Gly	Ser
				35					40					45
Tyr	Pro	Phe	Ile	Asp	Phe	Arg	Leu	Leu	Asn	Ser	Glu			
				50					55					

<210> 3

<211> 62

<212> PRT

5 <213> Homo sapiens

<400> 3

Met	Ile	Arg	Ser	Lys	Phe	Arg	Val	Pro	Arg	Ile	Leu	His	Val	Leu
				5					10					15
Ser	Ala	His	Ser	Gln	Ala	Ser	Asp	Lys	Asn	Phe	Thr	Ala	Glu	Asn
				20					25					30
Ser	Glu	Val	Val	Val	Ser	Ser	Arg	Thr	Asp	Val	Ser	Pro	Met	Lys
				35					40					45
Ser	Asp	Leu	Leu	Leu	Pro	Pro	Ser	Lys	Pro	Gly	Cys	Asn	Asn	Val
				50					55					60
Leu	Asn													

<210> 4

<211> 146

<212> PRT

10 <213> Homo sapiens

<400> 4

Met	Val	Gln	Gly	Met	Cys	Ser	Pro	Ser	Pro	Phe	Gly	Thr	Ser	Arg
				5					10					15
Ala	Cys	Thr	Val	Gly	Thr	Gln	Val	Asp	Ser	Arg	Ser	Leu	Pro	Trp
				20					25					30
Ala	Leu	Gly	Ala	Ser	Ala	Gln	Arg	Gly	Asn	Ile	Pro	Thr	Ala	Thr
				35					40					45
Cys	Ala	Arg	Thr	Ala	Gly	Thr	Leu	Arg	Arg	Gly	Leu	Gln	Pro	Gly
				50					55					60
Trp	Gly	Trp	Glu	Asp	Phe	Leu	Asp	Glu	Gly	Gln	Pro	Gly	Phe	Ser
				65					70					75
Ser	Arg	Met	Ser	Trp	Ser	Arg	Pro	Pro	Ala	Gln	Glu	Gln	Gly	Ala
				80					85					90
Gly	Arg	Gly	Pro	Ser	Trp	Val	Arg	Gly	Leu	Gly	Gln	Pro	Thr	Ala
				95					100					105
Ala	Phe	Glu	Gln	Gly	Pro	Arg	Ser	Ser	Val	Ser	Pro	Gln	Trp	Glu
				110					115					120
Gly	Gly	Gly	Gln	Gly	Pro	Gly	Glu	Leu	Gly	Arg	Lys	His	Leu	Leu
				125					130					135

**Gly Pro Ser Gln His His Pro Thr Asp Arg His**  
**140 145**

<210> 5

<211> 1039

<212> PRT

5 <213> Homo sapiens

<400> 5

Met	Ala	Val	Gln	Leu	Val	Pro	Asp	Ser	Ala	Leu	Gly	Leu	Leu	Met
				5					10					15
Met	Thr	Glu	Gly	Arg	Arg	Cys	Gln	Val	His	Leu	Leu	Asp	Asp	Arg
				20					25					30
Lys	Leu	Glu	Leu	Leu	Val	Gln	Pro	Lys	Leu	Leu	Ala	Lys	Glu	Leu
				35					40					45
Leu	Asp	Leu	Val	Ala	Ser	His	Phe	Asn	Leu	Lys	Glu	Lys	Glu	Tyr
				50					55					60
Phe	Gly	Ile	Ala	Phe	Thr	Asp	Glu	Thr	Gly	His	Leu	Asn	Trp	Leu
				65					70					75
Gln	Leu	Asp	Arg	Arg	Val	Leu	Glu	His	Asp	Phe	Pro	Lys	Lys	Ser
				80					85					90
Gly	Pro	Val	Val	Leu	Tyr	Phe	Cys	Val	Arg	Phe	Tyr	Ile	Glu	Ser
				95					100					105
Ile	Ser	Tyr	Leu	Lys	Asp	Asn	Ala	Thr	Ile	Glu	Leu	Phe	Phe	Leu
				110					115					120
Asn	Ala	Lys	Ser	Cys	Ile	Tyr	Lys	Glu	Leu	Ile	Asp	Val	Asp	Ser
				125					130					135
Glu	Val	Val	Phe	Glu	Leu	Ala	Ser	Tyr	Ile	Leu	Gln	Glu	Ala	Lys
				140					145					150
Gly	Asp	Phe	Ser	Ser	Asn	Glu	Val	Val	Arg	Ser	Asp	Leu	Lys	Lys
				155					160					165
Leu	Pro	Ala	Leu	Pro	Thr	Gln	Ala	Leu	Lys	Glu	His	Pro	Ser	Leu
				170					175					180
Ala	Tyr	Cys	Glu	Asp	Arg	Val	Ile	Glu	His	Tyr	Lys	Lys	Leu	Asn
				185					190					195
Gly	Gln	Thr	Arg	Gly	Gln	Ala	Ile	Val	Asn	Tyr	Met	Ser	Ile	Val
				200					205					210
Glu	Ser	Leu	Pro	Thr	Tyr	Gly	Val	His	Tyr	Tyr	Ala	Val	Lys	Asp
				215					220					225
Lys	Gln	Gly	Ile	Pro	Trp	Trp	Leu	Gly	Leu	Ser	Tyr	Lys	Gly	Ile
				230					235					240
Phe	Gln	Tyr	Asp	Tyr	His	Asp	Lys	Val	Lys	Pro	Arg	Lys	Ile	Phe
				245					250					255
Gln	Trp	Arg	Gln	Leu	Glu	Asn	Leu	Tyr	Phe	Arg	Glu	Lys	Lys	Phe
				260					265					270
Ser	Val	Glu	Val	His	Asp	Pro	Arg	Arg	Ala	Ser	Val	Thr	Arg	Arg
				275					280					285
Thr	Phe	Gly	His	Ser	Gly	Ile	Ala	Val	His	Thr	Trp	Tyr	Ala	Cys
				290					295					300
Pro	Ala	Leu	Ile	Lys	Ser	Ile	Trp	Ala	Met	Ala	Ile	Ser	Gln	His
				305					310					315
Gln	Phe	Tyr	Leu	Asp	Arg	Lys	Gln	Ser	Lys	Ser	Lys	Ile	His	Ala
				320					325					330
Ala	Arg	Ser	Leu	Ser	Glu	Ile	Ala	Ile	Asp	Leu	Thr	Glu	Thr	Gly
				335					340					345
Thr	Leu	Lys	Thr	Ser	Lys	Leu	Ala	Asn	Met	Gly	Ser	Lys	Gly	Lys
				350					355					360
Ile	Ile	Ser	Gly	Ser	Ser	Gly	Ser	Leu	Leu	Ser	Ser	Gly	Ser	Gln
				365					370					375
Glu	Ser	Asp	Ser	Ser	Gln	Ser	Ala	Lys	Lys	Asp	Met	Leu	Ala	Ala
				380					385					390
Leu	Lys	Ser	Arg	Gln	Glu	Ala	Leu	Glu	Glu	Thr	Leu	Arg	Gln	Arg
				395					400					405

Leu Glu Glu Leu Lys Lys Leu Cys Leu Arg Glu Ala Glu Leu Thr  
 410 415 420  
 Gly Lys Leu Pro Val Glu Tyr Pro Leu Asp Pro Gly Glu Glu Pro  
 425 430 435  
 Pro Ile Val Arg Arg Arg Ile Gly Thr Ala Phe Lys Leu Asp Glu  
 440 445 450  
 Gln Lys Ile Leu Pro Lys Gly Glu Glu Ala Glu Leu Glu Arg Leu  
 455 460 465  
 Glu Arg Glu Phe Ala Ile Gln Ser Gln Ile Thr Glu Ala Ala Arg  
 470 475 480  
 Arg Leu Ala Ser Asp Pro Asn Val Ser Lys Lys Leu Lys Lys Gln  
 485 490 495  
 Arg Lys Thr Ser Tyr Leu Asn Ala Leu Lys Lys Leu Gln Glu Ile  
 500 505 510  
 Glu Asn Ala Ile Asn Glu Asn Arg Ile Lys Ser Gly Lys Lys Pro  
 515 520 525  
 Thr Gln Arg Ala Ser Leu Ile Ile Asp Asp Gly Asn Ile Ala Ser  
 530 535 540  
 Glu Asp Ser Ser Leu Ser Asp Ala Leu Val Leu Glu Asp Glu Asp  
 545 550 555  
 Ser Gln Val Thr Ser Thr Ile Ser Pro Leu His Ser Pro His Lys  
 560 565 570  
 Gly Leu Pro Pro Arg Pro Pro Ser His Asn Arg Pro Pro Pro Pro  
 575 580 585  
 Gln Ser Leu Glu Gly Leu Arg Gln Met His Tyr His Arg Asn Asp  
 590 595 600  
 Tyr Asp Lys Ser Pro Ile Lys Pro Lys Met Trp Ser Glu Ser Ser  
 605 610 615  
 Leu Asp Glu Pro Tyr Glu Lys Val Lys Lys Arg Ser Ser His Ser  
 620 625 630  
 His Ser Ser Ser His Lys Arg Phe Pro Ser Thr Gly Ser Cys Ala  
 635 640 645  
 Glu Ala Gly Gly Gly Ser Asn Ser Leu Gln Asn Ser Pro Ile Arg  
 650 655 660  
 Gly Leu Pro His Trp Asn Ser Gln Ser Ser Met Pro Ser Thr Pro  
 665 670 675  
 Asp Leu Arg Val Arg Ser Pro His Tyr Val His Ser Thr Arg Ser  
 680 685 690  
 Val Asp Ile Ser Pro Thr Arg Leu His Ser Leu Ala Leu His Phe  
 695 700 705  
 Arg His Arg Ser Ser Ser Leu Glu Ser Gln Gly Lys Leu Leu Gly  
 710 715 720  
 Ser Glu Asn Asp Thr Gly Ser Pro Asp Phe Tyr Thr Pro Arg Thr  
 725 730 735  
 Arg Ser Ser Asn Gly Ser Asp Pro Met Asp Asp Cys Ser Ser Cys  
 740 745 750  
 Thr Ser His Ser Ser Ser Glu His Tyr Tyr Pro Ala Gln Met Asn  
 755 760 765  
 Ala Asn Tyr Ser Thr Leu Ala Glu Asp Ser Pro Ser Lys Ala Arg  
 770 775 780  
 Gln Arg Gln Arg Gln Arg Gln Arg Ala Ala Gly Ala Leu Gly Ser  
 785 790 795  
 Ala Ser Ser Gly Ser Met Pro Asn Leu Ala Ala Arg Gly Gly Ala  
 800 805 810  
 Gly Gly Ala Gly Gly Ala Gly Gly Gly Val Tyr Leu His Ser Gln  
 815 820 825  
 Ser Gln Pro Ser Ser Gln Tyr Arg Ile Lys Glu Tyr Pro Leu Tyr  
 830 835 840  
 Ile Glu Gly Gly Ala Thr Pro Val Val Val Arg Ser Leu Glu Ser  
 845 850 855  
 Asp Gln Glu Gly His Tyr Ser Val Lys Ala Gln Phe Lys Thr Ser  
 860 865 870  
 Asn Ser Tyr Thr Ala Gly Gly Leu Phe Lys Glu Ser Trp Arg Gly  
 875 880 885  
 Gly Gly Gly Asp Glu Gly Asp Thr Gly Arg Leu Thr Pro Ser Arg

				890						895				900
Ser	Gln	Ile	Leu	Arg	Thr	Pro	Ser	Leu	Gly	Arg	Glu	Gly	Ala	His
				905						910				915
Asp	Lys	Gly	Ala	Gly	Arg	Ala	Ala	Val	Ser	Asp	Glu	Leu	Arg	Gln
				920						925				930
Trp	Tyr	Gln	Arg	Ser	Thr	Ala	Ser	His	Lys	Glu	His	Ser	Arg	Leu
				935						940				945
Ser	His	Thr	Ser	Ser	Thr	Ser	Ser	Asp	Ser	Gly	Ser	Gln	Tyr	Ser
				950						955				960
Thr	Ser	Ser	Gln	Ser	Thr	Phe	Val	Ala	His	Ser	Arg	Val	Thr	Arg
				965						970				975
Met	Pro	Gln	Met	Cys	Lys	Ala	Thr	Ser	Ala	Ala	Leu	Pro	Gln	Ser
				980						985				990
Gln	Arg	Ser	Ser	Thr	Pro	Ser	Ser	Glu	Ile	Gly	Ala	Thr	Pro	Pro
				995						1000				1005
Ser	Ser	Pro	His	His	Ile	Leu	Thr	Trp	Gln	Thr	Gly	Glu	Ala	Thr
				1010						1015				1020
Glu	Asn	Ser	Pro	Ile	Leu	Asp	Gly	Ser	Glu	Ser	Pro	Pro	His	Gln
				1025						1030				1035
Ser	Thr	Asp	Glu											

<210> 6

<211> 244

<212> PRT

5 <213> Homo sapiens

<400> 6

Met	Ala	Ala	Ala	Ala	Ser	Pro	Ala	Ile	Leu	Pro	Arg	Leu	Ala	Ile
				5					10					15
Leu	Pro	Tyr	Leu	Leu	Phe	Asp	Trp	Ser	Gly	Thr	Gly	Arg	Ala	Asp
				20					25					30
Ala	His	Ser	Leu	Trp	Tyr	Asn	Phe	Thr	Ile	Ile	His	Leu	Pro	Arg
				35					40					45
His	Gly	Gln	Gln	Trp	Cys	Glu	Val	Gln	Ser	Gln	Val	Asp	Gln	Lys
				50					55					60
Asn	Phe	Leu	Ser	Tyr	Asp	Cys	Gly	Ser	Asp	Lys	Val	Leu	Ser	Met
				65					70					75
Gly	His	Leu	Glu	Glu	Gln	Leu	Tyr	Ala	Thr	Asp	Ala	Trp	Gly	Lys
				80					85					90
Gln	Leu	Glu	Met	Leu	Arg	Glu	Val	Gly	Gln	Arg	Leu	Arg	Leu	Glu
				95					100					105
Leu	Ala	Asp	Thr	Glu	Leu	Glu	Asp	Phe	Thr	Pro	Ser	Gly	Pro	Leu
				110					115					120
Thr	Leu	Gln	Val	Arg	Met	Ser	Cys	Glu	Cys	Glu	Ala	Asp	Gly	Tyr
				125					130					135
Ile	Arg	Gly	Ser	Trp	Gln	Phe	Ser	Phe	Asp	Gly	Arg	Lys	Phe	Leu
				140					145					150
Leu	Phe	Asp	Ser	Asn	Asn	Arg	Lys	Trp	Thr	Val	Val	His	Ala	Gly
				155					160					165
Ala	Arg	Arg	Met	Lys	Glu	Lys	Trp	Glu	Lys	Asp	Ser	Gly	Leu	Thr
				170					175					180
Thr	Phe	Phe	Lys	Met	Val	Ser	Met	Arg	Asp	Cys	Lys	Ser	Trp	Leu
				185					190					195
Arg	Asp	Phe	Leu	Met	His	Arg	Lys	Lys	Arg	Leu	Glu	Pro	Thr	Ala
				200					205					210
Pro	Pro	Thr	Met	Ala	Pro	Gly	Leu	Ala	Gln	Pro	Lys	Ala	Ile	Ala
				215					220					225
Thr	Thr	Leu	Ser	Pro	Trp	Ser	Phe	Leu	Ile	Ile	Leu	Cys	Phe	Ile
				230					235					240
Leu	Pro	Gly	Ile											

<210> 7

<211> 536

<212> PRT

5 <213> Homo sapiens

&lt;400&gt; 7

Met	Glu	Ile	Arg	Gln	His	Glu	Trp	Leu	Ser	Ala	Ser	Pro	His	Glu
				5					10					15
Gly	Phe	Glu	Gln	Met	Arg	Leu	Lys	Ser	Arg	Pro	Lys	Glu	Pro	Ser
				20					25					30
Pro	Ser	Leu	Thr	Arg	Val	Gly	Ala	Asn	Phe	Tyr	Ser	Ser	Val	Lys
				35					40					45
Gln	Gln	Asp	Tyr	Ser	Ala	Ser	Val	Trp	Leu	Arg	Arg	Lys	Asp	Lys
				50					55					60
Leu	Glu	His	Ser	Gln	Gln	Lys	Cys	Ile	Val	Ile	Phe	Ala	Leu	Val
				65					70					75
Cys	Cys	Phe	Ala	Ile	Leu	Val	Ala	Leu	Ile	Phe	Ser	Ala	Val	Asp
				80					85					90
Ile	Met	Gly	Glu	Asp	Glu	Asp	Gly	Leu	Ser	Glu	Lys	Asn	Cys	Gln
				95					100					105
Asn	Lys	Cys	Arg	Ile	Ala	Leu	Val	Glu	Asn	Ile	Pro	Glu	Gly	Leu
				110					115					120
Asn	Tyr	Ser	Glu	Asn	Ala	Pro	Phe	His	Leu	Ser	Leu	Phe	Gln	Gly
				125					130					135
Trp	Met	Asn	Leu	Leu	Asn	Met	Ala	Lys	Lys	Ser	Val	Asp	Ile	Val
				140					145					150
Ser	Ser	His	Trp	Asp	Leu	Asn	His	Thr	His	Pro	Ser	Ala	Cys	Gln
				155					160					165
Gly	Gln	Arg	Leu	Phe	Glu	Lys	Leu	Leu	Gln	Leu	Thr	Ser	Gln	Asn
				170					175					180
Ile	Glu	Ile	Lys	Leu	Val	Ser	Asp	Val	Thr	Ala	Asp	Ser	Lys	Val
				185					190					195
Leu	Glu	Ala	Leu	Lys	Leu	Lys	Gly	Ala	Glu	Val	Thr	Tyr	Met	Asn
				200					205					210
Met	Thr	Ala	Tyr	Asn	Lys	Gly	Arg	Leu	Gln	Ser	Ser	Phe	Trp	Ile
				215					220					225
Val	Asp	Lys	Gln	His	Val	Tyr	Ile	Gly	Ser	Ala	Gly	Leu	Asp	Trp
				230					235					240
Gln	Ser	Leu	Gly	Gln	Met	Lys	Glu	Leu	Gly	Val	Ile	Phe	Tyr	Asn
				245					250					255
Cys	Ser	Cys	Leu	Val	Leu	Asp	Leu	Gln	Arg	Ile	Phe	Ala	Leu	Tyr
				260					265					270
Ser	Ser	Leu	Lys	Phe	Lys	Ser	Arg	Val	Pro	Gln	Thr	Trp	Ser	Lys
				275					280					285
Arg	Leu	Tyr	Gly	Val	Tyr	Asp	Asn	Glu	Lys	Lys	Leu	Gln	Leu	Gln
				290					295					300
Leu	Asn	Glu	Thr	Lys	Ser	Gln	Ala	Phe	Val	Ser	Asn	Ser	Pro	Lys
				305					310					315
Leu	Phe	Cys	Pro	Lys	Asn	Arg	Ser	Phe	Asp	Ile	Asp	Ala	Ile	Tyr
				320					325					330
Ser	Val	Ile	Asp	Asp	Ala	Lys	Gln	Tyr	Val	Tyr	Ile	Ala	Val	Met
				335					340					345
Asp	Tyr	Leu	Pro	Ile	Ser	Ser	Thr	Ser	Thr	Lys	Arg	Thr	Tyr	Trp
				350					355					360
Pro	Asp	Leu	Asp	Ala	Lys	Ile	Arg	Glu	Ala	Leu	Val	Leu	Arg	Ser
				365					370					375
Val	Arg	Val	Arg	Leu	Leu	Leu	Ser	Phe	Trp	Lys	Glu	Thr	Asp	Pro
				380					385					390
Leu	Thr	Phe	Asn	Phe	Ile	Ser	Ser	Leu	Lys	Ala	Ile	Cys	Thr	Glu
				395					400					405
Ile	Ala	Asn	Cys	Ser	Leu	Lys	Val	Lys	Phe	Phe	Asp	Leu	Glu	Arg
				410					415					420
Glu	Asn	Ala	Cys	Ala	Thr	Lys	Glu	Gln	Lys	Asn	His	Thr	Phe	Pro
				425					430					435
Arg	Leu	Asn	Arg	Asn	Lys	Tyr	Met	Val	Thr	Asp	Gly	Ala	Ala	Tyr

ES 2 662 846 T3

				440					445					450
Ile	Gly	Asn	Phe	Asp	Trp	Val	Gly	Asn	Asp	Phe	Thr	Gln	Asn	Ala
				455					460					465
Gly	Thr	Gly	Leu	Val	Ile	Asn	Gln	Ala	Asp	Val	Arg	Asn	Asn	Arg
				470					475					480
Ser	Ile	Ile	Lys	Gln	Leu	Lys	Asp	Val	Phe	Glu	Arg	Asp	Trp	Tyr
				485					490					495
Ser	Pro	Tyr	Ala	Lys	Thr	Leu	Gln	Pro	Thr	Lys	Gln	Pro	Asn	Cys
				500					505					510
Ser	Ser	Leu	Phe	Lys	Leu	Lys	Pro	Leu	Ser	Asn	Lys	Thr	Ala	Thr
				515					520					525
Asp	Asp	Thr	Gly	Gly	Lys	Asp	Pro	Arg	Asn	Val				
				530					535					

<210> 8

<211> 344

<212> PRT

5 <213> Homo sapiens

<400> 8

Gln	Asn	Leu	Pro	Ser	Ser	Pro	Ala	Pro	Ser	Thr	Ile	Phe	Ser	Gly
				5					10					15
Gly	Phe	Arg	His	Gly	Ser	Leu	Ile	Ser	Ile	Asp	Ser	Thr	Cys	Thr
				20					25					30
Glu	Met	Gly	Asn	Phe	Asp	Asn	Ala	Asn	Val	Thr	Gly	Glu	Ile	Glu
				35					40					45
Phe	Ala	Ile	His	Tyr	Cys	Phe	Lys	Thr	His	Ser	Leu	Glu	Ile	Cys
				50					55					60
Ile	Lys	Ala	Cys	Lys	Asn	Leu	Ala	Tyr	Gly	Glu	Glu	Lys	Lys	Lys
				65					70					75
Lys	Cys	Asn	Pro	Tyr	Val	Lys	Thr	Tyr	Leu	Leu	Pro	Asp	Arg	Ser
				80					85					90
Ser	Gln	Gly	Lys	Arg	Lys	Thr	Gly	Val	Gln	Arg	Asn	Thr	Val	Asp
				95					100					105
Pro	Thr	Phe	Gln	Glu	Thr	Leu	Lys	Tyr	Gln	Val	Ala	Pro	Ala	Gln
				110					115					120
Leu	Val	Thr	Arg	Gln	Leu	Gln	Val	Ser	Val	Trp	His	Leu	Gly	Thr
				125					130					135
Leu	Ala	Arg	Arg	Val	Phe	Leu	Gly	Glu	Val	Ile	Ile	Ser	Leu	Ala
				140					145					150
Thr	Trp	Asp	Phe	Glu	Asp	Ser	Thr	Thr	Gln	Ser	Phe	Arg	Trp	His
				155					160					165
Pro	Leu	Arg	Ala	Lys	Ala	Glu	Lys	Tyr	Glu	Asp	Ser	Val	Pro	Gln
				170					175					180
Ser	Asn	Gly	Glu	Leu	Thr	Val	Arg	Ala	Lys	Leu	Val	Leu	Pro	Ser
				185					190					195
Arg	Pro	Arg	Lys	Leu	Gln	Glu	Ala	Gln	Glu	Gly	Thr	Asp	Gln	Pro
				200					205					210
Ser	Leu	His	Gly	Gln	Leu	Cys	Leu	Val	Val	Leu	Gly	Ala	Lys	Asn
				215					220					225
Leu	Pro	Val	Arg	Pro	Asp	Gly	Thr	Leu	Asn	Ser	Phe	Val	Lys	Gly
				230					235					240
Cys	Leu	Thr	Leu	Pro	Asp	Gln	Gln	Lys	Leu	Arg	Leu	Lys	Ser	Pro
				245					250					255
Val	Leu	Arg	Lys	Gln	Ala	Cys	Pro	Gln	Trp	Lys	His	Ser	Phe	Val
				260					265					270
Phe	Ser	Gly	Val	Thr	Pro	Ala	Gln	Leu	Arg	Gln	Ser	Ser	Leu	Glu
				275					280					285
Leu	Thr	Val	Trp	Asp	Gln	Ala	Leu	Phe	Gly	Met	Asn	Asp	Arg	Leu
				290					295					300
Leu	Gly	Gly	Thr	Arg	Leu	Gly	Ser	Lys	Gly	Asp	Thr	Ala	Val	Gly
				305					310					315
Gly	Asp	Ala	Cys	Ser	Leu	Ser	Lys	Leu	Gln	Trp	Gln	Lys	Val	Leu

320

325

330

Ser Ser Pro Asn Leu Trp Thr Asp Met Thr Leu Val Leu His

<210> 9

<211> 263

5 <212> PRT

<213> Homo sapiens

<400> 9

Met	Phe	Arg	Asn	Gln	Tyr	Asp	Asn	Asp	Val	Thr	Val	Trp	Ser	Pro
				5					10					15
Gln	Gly	Arg	Ile	His	Gln	Ile	Glu	Tyr	Ala	Met	Glu	Ala	Val	Lys
				20					25					30
Gln	Gly	Ser	Ala	Thr	Val	Gly	Leu	Lys	Ser	Lys	Thr	His	Ala	Val
				35					40					45
Leu	Val	Ala	Leu	Lys	Arg	Ala	Gln	Ser	Glu	Leu	Ala	Ala	His	Gln
				50					55					60
Lys	Lys	Ile	Leu	His	Val	Asp	Asn	His	Ile	Gly	Ile	Ser	Ile	Ala
				65					70					75
Gly	Leu	Thr	Ala	Asp	Ala	Arg	Leu	Leu	Cys	Asn	Phe	Met	Arg	Gln
				80					85					90
Glu	Cys	Leu	Asp	Ser	Arg	Phe	Val	Phe	Asp	Arg	Pro	Leu	Pro	Val
				95					100					105
Ser	Arg	Leu	Val	Ser	Leu	Ile	Gly	Ser	Lys	Thr	Gln	Ile	Pro	Thr
				110					115					120
Gln	Arg	Tyr	Gly	Arg	Arg	Pro	Tyr	Gly	Val	Gly	Leu	Leu	Ile	Ala
				125					130					135
Gly	Tyr	Asp	Asp	Met	Gly	Pro	His	Ile	Phe	Gln	Thr	Cys	Pro	Ser
				140					145					150
Ala	Asn	Tyr	Phe	Asp	Cys	Arg	Ala	Met	Ser	Ile	Gly	Ala	Arg	Ser
				155					160					165
Gln	Ser	Ala	Arg	Thr	Tyr	Leu	Glu	Arg	His	Met	Ser	Glu	Phe	Met
				170					175					180
Glu	Cys	Asn	Leu	Asn	Glu	Leu	Val	Lys	His	Gly	Leu	Arg	Ala	Leu
				185					190					195
Arg	Glu	Thr	Leu	Pro	Ala	Glu	Gln	Asp	Leu	Thr	Thr	Lys	Asn	Val
				200					205					210
Ser	Ile	Gly	Ile	Val	Gly	Lys	Asp	Leu	Glu	Phe	Thr	Ile	Tyr	Asp
				215					220					225
Asp	Asp	Asp	Val	Ser	Pro	Phe	Leu	Glu	Gly	Leu	Glu	Glu	Arg	Pro
				230					235					240
Gln	Arg	Lys	Ala	Gln	Pro	Ala	Gln	Pro	Ala	Asp	Glu	Pro	Ala	Glu
				245					250					255
Lys	Ala	Asp	Glu	Pro	Met	Glu	His							
				260										

<210> 10

<211> 287

5 <212> PRT

<213> Homo sapiens

<400> 10

Met Lys Asp Arg Leu Ala Glu Leu Leu Asp Leu Ser Lys Gln Tyr  
 5 10 15  
 Asp Gln Gln Phe Pro Asp Gly Asp Asp Glu Phe Asp Ser Pro His  
 20 25 30  
 Glu Asp Ile Val Phe Glu Thr Asp His Ile Leu Glu Ser Leu Tyr  
 35 40 45  
 Arg Asp Ile Arg Asp Ile Gln Asp Glu Asn Gln Leu Leu Val Ala  
 50 55 60  
 Asp Val Lys Arg Leu Gly Lys Gln Asn Ala Arg Phe Leu Thr Ser

65 70 75  
 Met Arg Arg Leu Ser Ser Ile Lys Arg Asp Thr Asn Ser Ile Ala  
 80 85 90  
 Lys Ala Ile Lys Ala Arg Gly Glu Val Ile His Cys Lys Leu Arg  
 95 100 105  
 Ala Met Lys Glu Leu Ser Glu Ala Ala Glu Ala Gln His Gly Pro  
 110 115 120  
 His Ser Ala Val Ala Arg Ile Ser Arg Ala Gln Tyr Asn Ala Leu  
 125 130 135  
 Thr Leu Thr Phe Gln Arg Ala Met His Asp Tyr Asn Gln Ala Glu  
 140 145 150  
 Met Lys Gln Arg Asp Asn Cys Lys Ile Arg Ile Gln Arg Gln Leu  
 155 160 165  
 Glu Ile Met Gly Lys Glu Val Ser Gly Asp Gln Ile Glu Asp Met  
 170 175 180  
 Phe Glu Gln Gly Lys Trp Asp Val Phe Ser Glu Asn Leu Leu Ala  
 185 190 195  
 Asp Val Lys Gly Ala Arg Ala Ala Leu Asn Glu Ile Glu Ser Arg  
 200 205 210  
 His Arg Glu Leu Leu Arg Leu Glu Ser Arg Ile Arg Asp Val His  
 215 220 225  
 Glu Leu Phe Leu Gln Met Ala Val Leu Val Glu Lys Gln Ala Asp  
 230 235 240  
 Thr Leu Asn Val Ile Glu Leu Asn Val Gln Lys Thr Val Asp Tyr  
 245 250 255  
 Thr Gly Gln Ala Lys Ala Gln Val Arg Lys Ala Val Gln Tyr Glu  
 260 265 270  
 Glu Lys Asn Pro Cys Arg Thr Leu Cys Cys Phe Cys Cys Pro Cys  
 275 280 285  
 Leu Lys

<210> 11

<211> 244

5 <212> PRT

<213> Homo sapiens

<400> 11

Met	Ala	Ala	Ala	Ala	Ser	Pro	Ala	Ile	Leu	Pro	Arg	Leu	Ala	Ile
				5					10					15
Leu	Pro	Tyr	Leu	Leu	Phe	Asp	Trp	Ser	Gly	Thr	Gly	Arg	Ala	Asp
				20					25					30
Ala	His	Ser	Leu	Trp	Tyr	Asn	Phe	Thr	Ile	Ile	His	Leu	Pro	Arg
				35					40					45
His	Gly	Gln	Gln	Trp	Cys	Glu	Val	Gln	Ser	Gln	Val	Asp	Gln	Lys
				50					55					60
Asn	Phe	Leu	Ser	Tyr	Asp	Cys	Gly	Ser	Asp	Lys	Val	Leu	Ser	Met
				65					70					75
Gly	His	Leu	Glu	Glu	Gln	Leu	Tyr	Ala	Thr	Asp	Ala	Trp	Gly	Lys
				80					85					90
Gln	Leu	Glu	Met	Leu	Arg	Glu	Val	Gly	Gln	Arg	Leu	Arg	Leu	Glu
				95					100					105
Leu	Ala	Asp	Thr	Glu	Leu	Glu	Asp	Phe	Thr	Pro	Ser	Gly	Pro	Leu
				110					115					120
Thr	Leu	Gln	Val	Arg	Met	Ser	Cys	Glu	Cys	Glu	Ala	Asp	Gly	Tyr
				125					130					135
Ile	Arg	Gly	Ser	Trp	Gln	Phe	Ser	Phe	Asp	Gly	Arg	Lys	Phe	Leu
				140					145					150
Leu	Phe	Asp	Ser	Asn	Asn	Arg	Lys	Trp	Thr	Val	Val	His	Ala	Gly
				155					160					165
Ala	Arg	Arg	Met	Lys	Glu	Lys	Trp	Glu	Lys	Asp	Ser	Gly	Leu	Thr
				170					175					180
Thr	Phe	Phe	Lys	Met	Val	Ser	Met	Arg	Asp	Cys	Lys	Ser	Trp	Leu
				185					190					195

Arg	Asp	Phe	Leu	Met	His	Arg	Lys	Lys	Arg	Leu	Glu	Pro	Thr	Ala
				200					205					210
Pro	Pro	Thr	Met	Ala	Pro	Gly	Leu	Ala	Gln	Pro	Lys	Ala	Ile	Ala
				215					220					225
Thr	Thr	Leu	Ser	Pro	Trp	Ser	Phe	Leu	Ile	Ile	Leu	Cys	Phe	Ile
				230					235					240
Leu	Pro	Gly	Ile											

**REIVINDICACIONES**

1. Un método que facilita el diagnóstico de las patologías del asma y/o el cáncer de pulmón en pacientes que se encuentran en una etapa temprana de la enfermedad, que consiste en determinar la concentración de proteína específica de CAC69571 (SEC. ID. N°: 8) y opcionalmente una o más de las proteínas de la proteína 4 que contiene el dominio FERM (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 (SEC. ID. N°: 9), syntaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), AAK13049 (SEC. ID. N°: 11) BACO4615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°:2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°:4), y/o Q8N7P1 (SEC. ID. N°:7) en el suero de un paciente y comparar los datos de concentración con los obtenidos de las poblaciones con cáncer de pulmón y/o las poblaciones con asma, a fin de verificar o descartar la existencia de las patologías en cuestión.
2. El método de la reivindicación 1 donde las concentraciones de proteínas específicas se determinan obteniendo los datos de los espectros de masas, radioinmunoensayo, inmunoensayo absorbente vinculado a enzimas, cromatografía líquida de alta resolución con radiométrica, detección espectrométrica mediante absorbancia de luz ultravioleta o visible, análisis cuantitativo y cualitativo por espectrometría de masas, inmunoensayo western-blot, electroforesis en gel unidimensional o bidimensional con visualización cuantitativa mediante detección de núcleos o sondas radiactivas, detección basada en anticuerpos con fotometría de absorción o fluorescencia, cuantificación por luminiscencia de cualesquiera de una serie de sistemas reporteros quimioluminiscentes, ensayos enzimáticos, inmunoprecipitación o inmunoensayos de captura, o cualesquiera de una serie de inmunoensayos en fase sólida y líquida.
3. Un método para evaluar la respuesta de un paciente a intervenciones terapéuticas que consisten en determinar la concentración de proteína específica de CAC69571 (SEC. ID. N°: 8), y opcionalmente una o más de las proteínas de la proteína 4 que contiene el dominio FERM (SEC. ID. N°: 5), forma de unión larga de C2 de la cadena del complejo de endopeptidasa del proteasoma de JCC1445 (SEC. ID. N°: 9), syntaxina 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), AAK13049 (SEC. ID. N°: 11) BACO4615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°:2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N° :4), y/o Q8N7P1 (SEC. ID. N°:7) en el suero de un paciente, donde el paciente padece asma o cáncer de pulmón.
4. El método de la reivindicación 3, donde la intervención terapéutica es un tratamiento con fármacos, radiación/quimioterapia o tratamiento quirúrgico.
5. El método de cualquier reivindicación precedente, que comprenda la determinación de las concentraciones de proteínas específicas de CAC69571 (SEC. ID. N°: 8), y al menos una de las proteínas de la proteína 4 que contiene el dominio FERM (SEQ ID NO: 5), forma de unión larga de C2 de la cadena del complejo de endopetidasa del proteasoma JC1445 (SEC. ID. N°: 9), Syntaxin 11 (SEC. ID. N°: 10), AAK13083 (SEC. ID. N°: 6), AAK13049 (SEC. ID. N°: 11), BAC04615 (SEC. ID. N°: 1), Q6NSC8 (SEC. ID. N°: 2), CAF17350 (SEC. ID. N°: 3), Q6ZUD4 (SEC. ID. N°: 4), y/o Q8N7P1 (SEC. ID. N°: 7) en el suero de un paciente, donde el paciente padece asma o cáncer de pulmón.

40

45

50

**FIG. 1**

**10 Resultado de la búsqueda de Mascot**

Título de la búsqueda: BAC04615

Base de datos: NCBI nr 20080903 (693453 secuencias; 2394607247 residuos)

**12** Taxonomía: Homo sapiens (humano) (216752 secuencias)

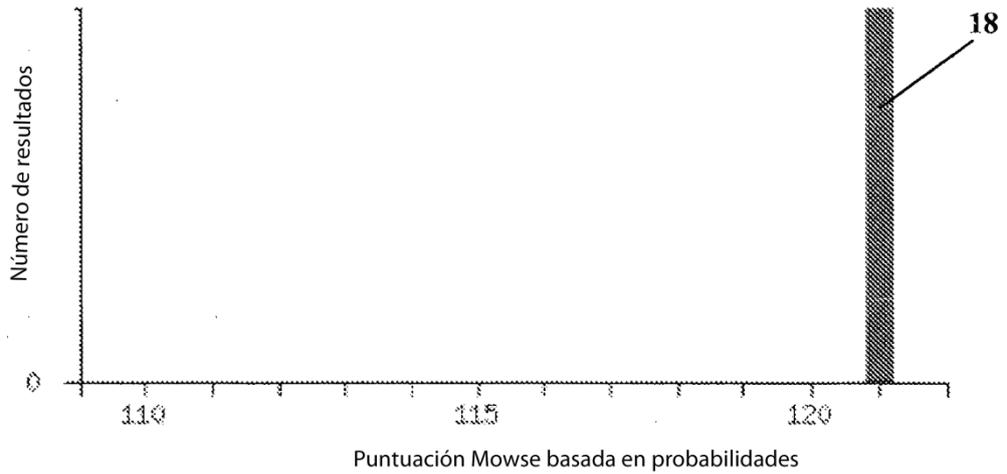
Puntuación superior: 121 para gi/21755032, producto de proteína sin nombre designado (Homo sapiens)

**16 Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 66 son significativas ( $p < 0,05$ ).

**14**



**28**

Informe conciso del resumen de proteínas

	Proteína	Masa	Puntuación	Valor previsto
<b>20</b>	gi/21755032, producto de proteína sin nombre designado (Homo sapiens)	23270	121	1.7e-07
<b>22</b>	gi/194385926, producto de proteína sin nombre designado (Homo sapiens)	26844	113	1.1e-06
<b>24</b>	Gi/21450663; proteína hipotética LOC159013 [Homo sapiens]	36646	110	2.2e-06
<b>26</b>	gi/189067542, producto de proteína sin nombre designado (Homo sapiens)	36645	109	1.7e-06

**FIG. 2**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Búsqueda refinada de Q6NSC8

Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)

Taxonomía: Homo sapiens (humano) (148148 secuencias)

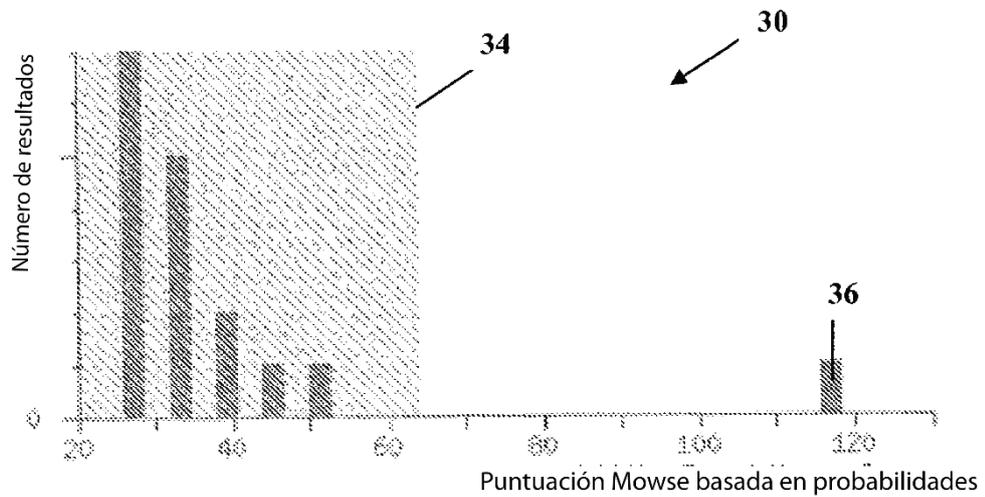
Puntuación superior: 117 para Q6NSC8\_HUMAN, proteína PDE7B. - Homo sapiens (Humano)

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).

29



Informe conciso del resumen de proteínas

32

Proteína	Masa	Puntuación	Valor previsto
Q6NSC8 HUMAN, proteína PDE7B. - Homo sapiens (Humano)	6535	117	3e-07

**FIG. 3**

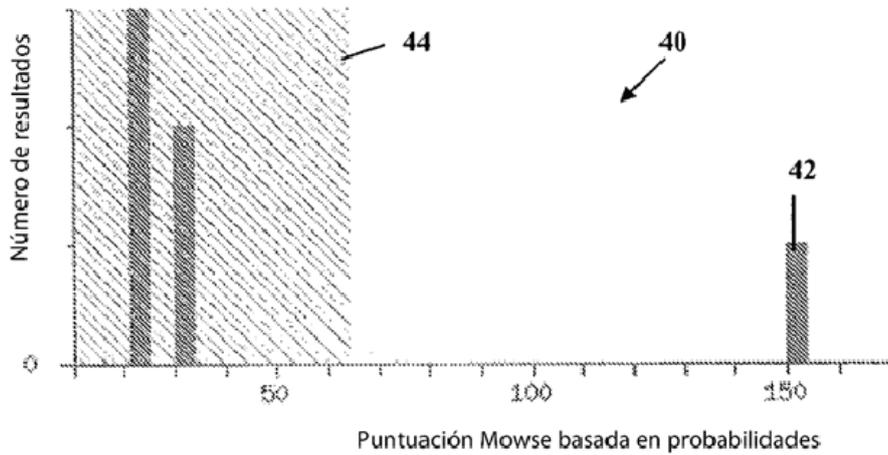
**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Coincidencia de AX972242 depurada.  
 Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)  
 Taxonomía: Homo sapiens (humano) (148148 secuencias)  
 Puntuación superior: 152 para CAF17350, AX972242 NID - Homo Sapiens

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

38 — Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



Informe conciso del resumen de proteínas

Proteína	Masa	Puntuación	Valor previsto
CAF17350; AX972242 NID: - Homo sapiens	6816	152	9.3e-11

46 —

**FIG. 4**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Q6ZUD4

Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)

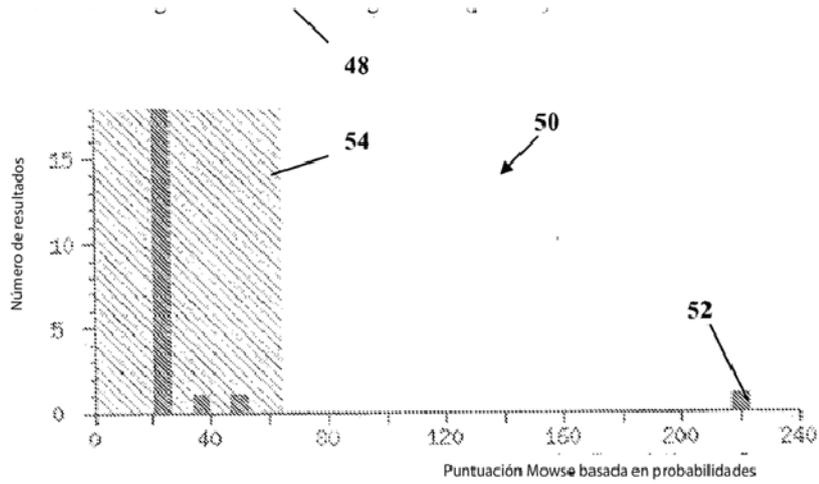
Taxonomía: Homo sapiens (humano) (148148 secuencias)

Puntuación superior: 220 para Q6ZUD4\_HUMAN, CDNA FLJ43798 fis, clon TEST14000250. Homo sapiens

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



**Informe conciso del resumen de proteínas**

Proteína	Masa	Puntuación	Valor previsto
Q6ZUD4_HUMAN, CDNA FLJ43798 fis, clon TEST14000250. Homo sapiens (Humano)	15462	220	1.5e-17

**FIG. 5**

**Resultado de la búsqueda de Mascot**

70

Título de la búsqueda: Q8N7P1/Q8NB22

Base de datos: NCBI nr 20080903 (6934543 secuencias; 2394607247 residuos)

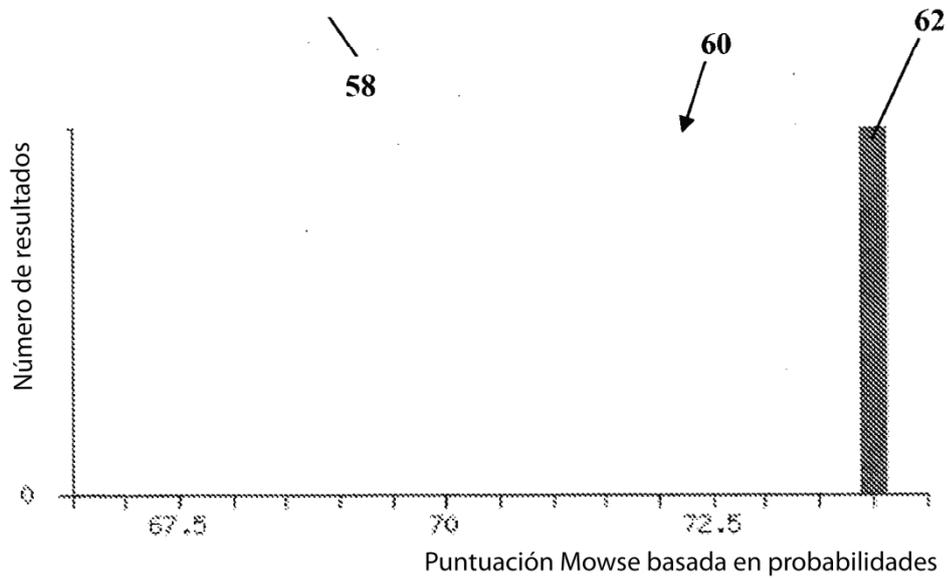
Taxonomía: Homo sapiens (humano) (216752 secuencias)

Puntuación superior: 74 para gi/71682143, familia fosfolipasa D, miembro 5 [homo sapiens]

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



Informe conciso del resumen de proteínas

	Proteína	Masa	Puntuación	Valor previsto
64	gi/71682143, familia fosfolipasa D, miembro 5 [homo sapiens]	50816	74	0.0086
66	gi/22749353, familia fosfolipasa D, miembro 5 [homo sapiens]	50763	73	0.011
68	gi/46392519; proteína PLDc [Homo sapiens]	53956	69	0.0026

**FIG. 6**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: CAC69571

Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)

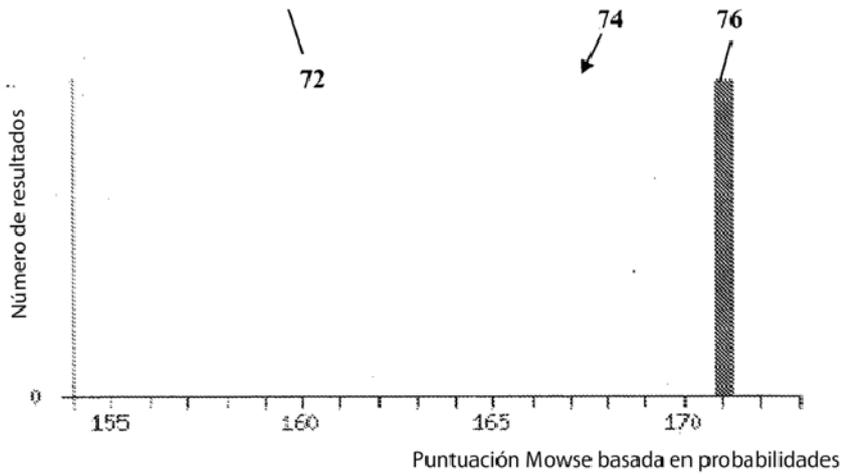
Taxonomía: Homo sapiens (humano) (148148 secuencias)

Puntuación superior: 171 para CAC69571, secuencia 1 de la Patente WO0160857 (Fragmento). – Homo sapiens (Humano)

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



**Informe conciso del resumen de proteínas**

	Proteína	Masa	Puntuación	Valor previsto
78	CAC69571; secuencia 1 de la Patente WO0160857 (Fragmento). – Homo sapiens (Humano)	38244	171	1.2e-12
80	Q4VX76 HUMAN; similar a sinaptotagmina-III. – Homo sapiens (Humano)	68516	145	4.7e-10
82	Q496J4 HUMAN; proteína SYTL3. – Homo sapiens (Humano)	68531	101	1.2e-05
84	Q496J6 HUMAN ; proteína SYTL3. – Homo sapiens (Humano)	68531	101	1.2e-05
86	Q5U3B9 HUMAN ; dsimilar a sinaptotagmina-III (proteína SYTL3) – Homo sapiens (Humano)	60901	95	4.3e-05

**FIG. 7**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Búsqueda refinada de FRMD-4

Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)

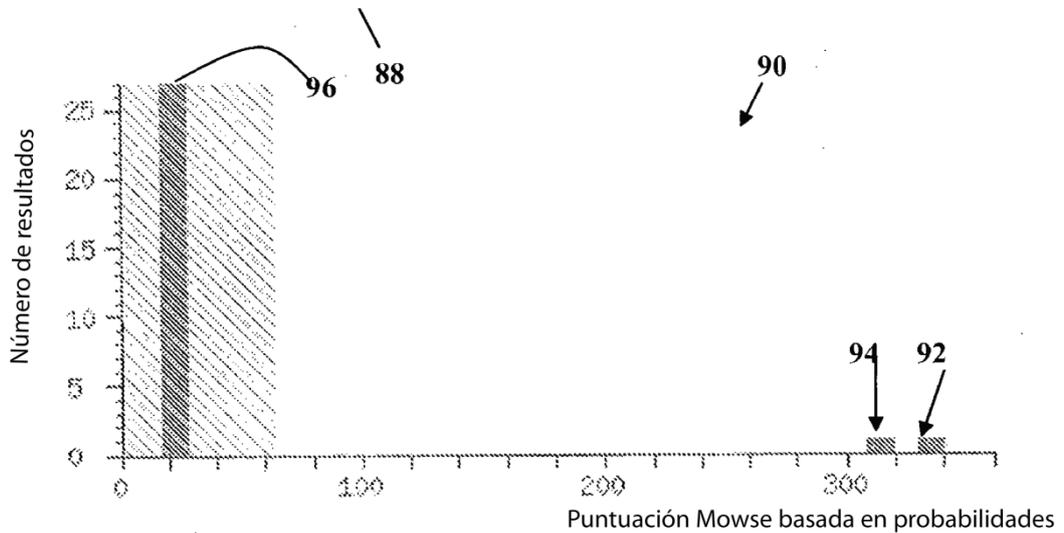
Taxonomía: Homo sapiens (humano) (148148 secuencias)

Puntuación superior: 335 para FRM4A\_HIMAN, proteína 4A que contiene el dominio FERM. - Homo sapiens

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



**Informe conciso del resumen de proteínas**

	Proteína	Masa	Puntuación	Valor previsto
98	FRM4A_HIMAN, proteína 4A que contiene el dominio FERM. - Homo sapiens (Humano)	113894	335	4.7E-29
100	BAA92532; AB037715 NID: - Homo sapiens	116819	309	1.9e-26
102	Q5T376_HUMAN; proteína nueva (FLJ10210) (Fragmento). - Homo sapiens (Humano)	60164	111	1.2e-06
104	Q5T378_HUMAN ; proteína nueva (FLJ10210) (Fragmento). - Homo sapiens (Humano)	51532	80	0.0016
106	Q9NW91 HUMAN: CDNA FLJ10210 fis, clon HEMBA1006344, ligeramente similar a RADIX FN. - Homo sapiens	51460	71	0.011
108	Q5T377_HUMAN ; proteína nueva (FLJ10210) (Fragmento). - Homo sapiens (Humano)	115387	314	5.9e-27

**FIG. 8**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: JCC 1445

Base de datos: NCBI nr 20080903 (6934543 secuencias; 2394607247 residuos)

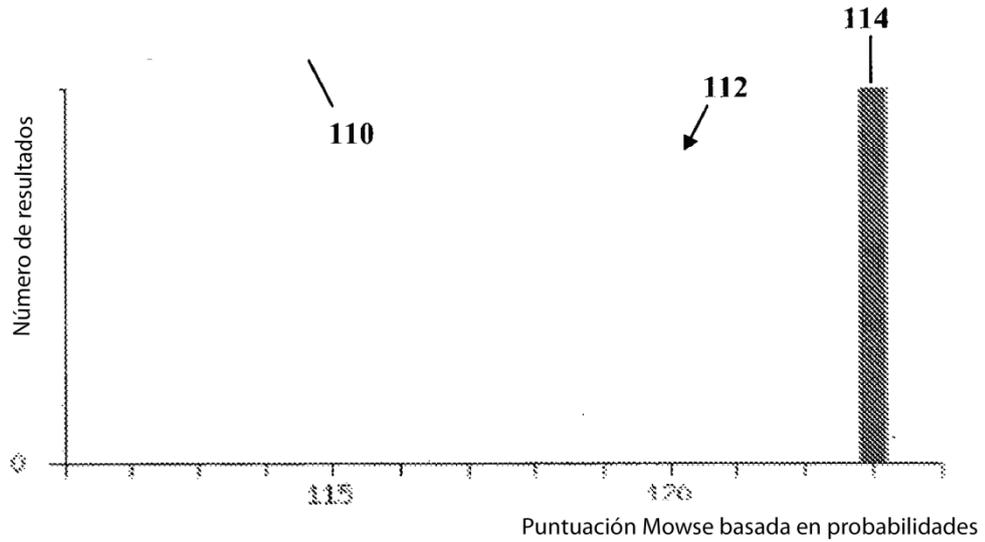
Taxonomía: Homo sapiens (humano) (216752 secuencias)

Puntuación superior: 123 para gi/4506179, isoforma 2 subunidad alfa 1 del proteasoma [Homo sapiens]

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 66 son significativas ( $p < 0,05$ ).



**Informe conciso del resumen de proteínas**

	Proteína	Masa	Puntuación	Valor previsto
116	gi/4506179, isoforma 2 subunidad alfa 1 del proteasoma [Homo sapiens]	29537	123	1.1e-07
118	gi/13543551; subunidad del proteasoma (prosoma, macropain), tipo alfa, 1 [Homo sapiens]	29579	123	1.1e-07
120	gi/23110935; isoforma 1 subunidad alfa 1 del proteasoma 1 [Homo sapiens]	30220	122	1.4e-07
122	gi/119588883; isoforma CRAb, tipo alfa, subunidad 1 del proteasoma (prosoma, macropain) [Homo sapiens]	33933	120	2.2e-07
124	gi/190447; proteína prosómica P30-33K	30208	92	0.00012
126	gi/194382836; producto de proteína sin nombre designado [Homo sapiens]	14615	82	0.0013
128	gi/194375750; producto de proteína sin nombre designado [Homo sapiens]	18446	78	0.0031

**FIG. 9**

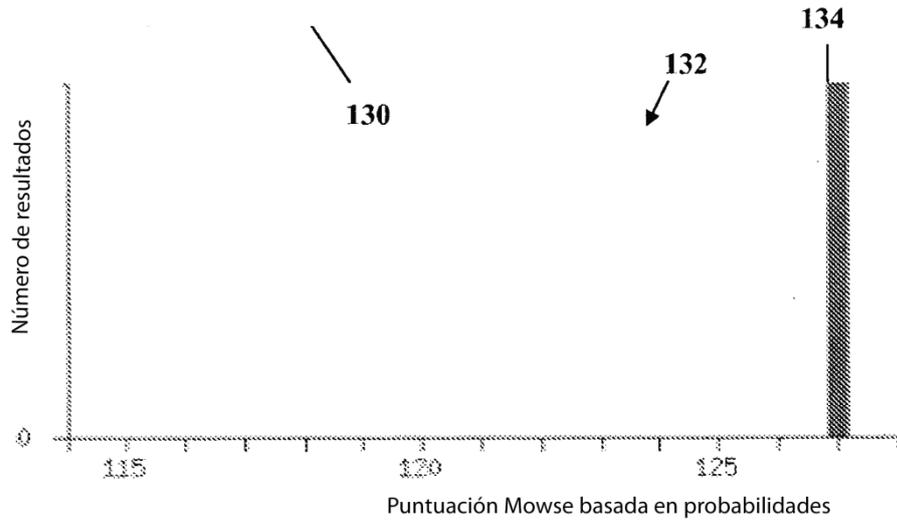
**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Sintaxina 11  
 Base de datos: NCBI nr 20080903 (6934543 secuencias; 2394607247 residuos)  
 Taxonomía: Homo sapiens (humano) (216752 secuencias)  
 Puntuación superior: 127 para gi/33667038, syntaxina 11 [Homo sapiens]

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 66 son significativas ( $p < 0,05$ ).



Informe conciso del resumen de proteínas

	Proteína	Masa	Puntuación	Valor previsto
136	gi/33667038, syntaxina 11 [Homo sapiens]	33175	127	4.3e-08
138	gi/3248918, syntaxina 11 [Homo sapiens]	32934	127	4.3e-08
140	gi/4104685, syntaxina 11 [Homo sapiens]	32877	95	6.7e-05

**FIG. 10**

**Resultado de la búsqueda de Mascot**

Título de la búsqueda: Búsqueda refinada de AAK13083

Base de datos: MSDB 20060831 (3239079 secuencias; 1079594700 residuos)

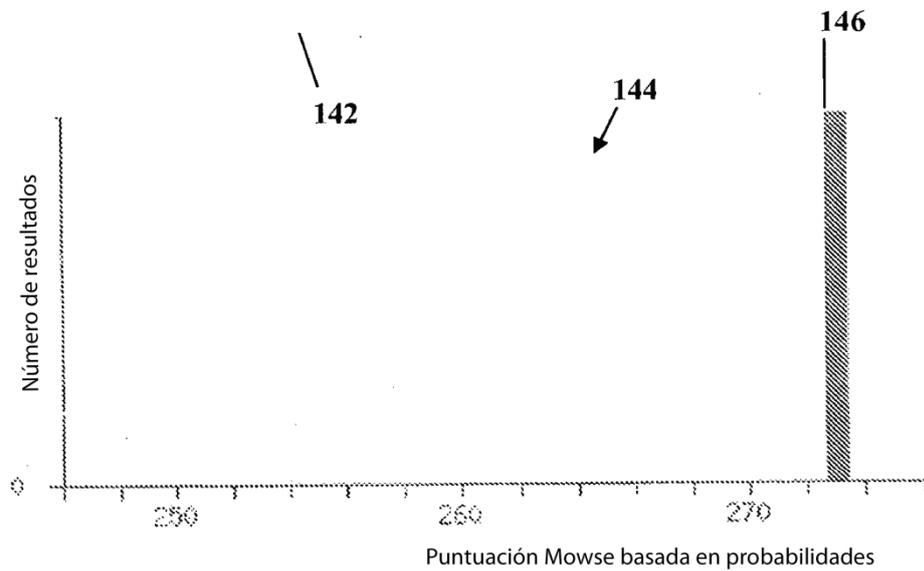
Taxonomía: Homo sapiens (humano) (148148 secuencias)

Puntuación superior: 273 para Q5VY82\_HUMAN, proteína 3 de unión a UL16. - Homo sapiens (Humano)

**Puntuación Mowse basada en probabilidades**

La puntuación de proteína es  $-10 \cdot \log(P)$ , donde P es la probabilidad de que la coincidencia observada sea un evento aleatorio.

Las puntuaciones de proteína superiores a 64 son significativas ( $p < 0,05$ ).



Informe conciso del resumen de proteínas

	Proteína	Masa	Puntuación	Valor previsto
148	Q5VY82_HUMAN, proteína 3 de unión a UL16. - Homo sapiens (Humano)	27931	273	7.4E-23
150	AAN40839; AF3346596 NID - Homo sapiens	27261	260	1.4e-21
152	AAK13049; AY027538 NID - Homo sapiens	27035	257	3e-21
154	CAD28547; HSM803089 NID; - Homo sapiens	7688	40	15