

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 482 468**

21 Número de solicitud: 201201292

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

**21.12.2012**

43 Fecha de publicación de la solicitud:

**01.08.2014**

Fecha de la concesión:

**05.05.2015**

45 Fecha de publicación de la concesión:

**12.05.2015**

56 Se remite a la solicitud internacional:

**PCT/ES2013/070910**

73 Titular/es:

**UNIVERSIDAD DE GRANADA (50.0%)  
Hospital Real C/ Cuesta del Hospicio s/n  
18071 Granada (Granada) ES y  
SERVICIO ANDALUZ DE SALUD (50.0%)**

72 Inventor/es:

**FÁREZ VIDAL, M. Esther;  
SÁNCHEZ-PALENCIA RAMOS, Abel;  
GÓMEZ MORALES, Mercedes y  
GÓMEZ CAPILLA, José Antonio**

54 Título: **Método de obtención de datos útiles para el diagnóstico diferencial del cáncer de pulmón**

57 Resumen:

Método de obtención de datos, útiles para el diagnóstico diferencial del cáncer de pulmón. Uso de los genes PKP1, KRT15, DSG3, o de las proteínas PKP1, KRT15, DSG3, para el diagnóstico diferencial del cáncer de pulmón, kits de diagnóstico y usos.

**ES 2 482 468 B1**

## DESCRIPCIÓN

### Método de obtención de datos útiles para el diagnóstico diferencial del cáncer de pulmón.

#### 5 CAMPO DE LA INVENCION

La presente invención se encuentra dentro de la medicina y la biología molecular, y se refiere a un método de obtención de datos útiles para el diagnóstico diferencial de los distintos tipos de cáncer de pulmón, permitiendo el establecimiento de grupos de pacientes.

#### ANTECEDENTES DE LA INVENCION

El cáncer de pulmón es la principal causa de muerte por cáncer en todo el mundo.

Los dos tipos principales de cáncer de pulmón son el **cáncer de pulmón de células no pequeñas** (de inglés *Non-small cell lung cancer* (NSCLC)) ( $\approx$  85% de todos los cánceres de pulmón) y el **cáncer de pulmón de células pequeñas** (del inglés *Small cell lung cancer* (SCLC)) ( $\approx$  15%). El cáncer de pulmón de células no pequeñas se puede dividir en tres grandes subtipos histológicos:

- **Carcinoma de células escamosas**, que supone del 25 al 30% de NSCLC. Este tipo es generalmente encontrado cerca de los bronquios, hacia el centro de la cavidad torácica (del pecho). Es también conocido como carcinoma epidermoide y está usualmente asociado a la exposición al humo de tabaco.
- **Adenocarcinoma**: conforma el 40% de todos los NSCLC. Este tipo de cáncer se encuentra generalmente en las regiones más externas

del pulmón. También existe una forma rara de adenocarcinoma, llamado carcinoma broncoalveolar (Bronchioalveolar Carcinoma (BAC)) que se está viendo con mayor frecuencia a nivel mundial. BAC se disemina a lo largo de todo el pulmón a diferencia de otros tipos de cáncer de pulmón más comunes que forman tumores únicos. Se desconoce la causa del BAC. A pesar de presentarse en persona que fuman, generalmente se da en aquellas que nunca han fumado.

- **Carcinoma de Células Grandes:** conforma del 10% al 15% de NSCLC. Es de crecimiento rápido y puede aparecer en cualquier parte del pulmón., el cáncer de pulmón.

CÁNCER DE PULMÓN DE CÉLULAS NO PEQUEÑAS (NSCLC)	CÁNCER DE PULMÓN DE CÉLULAS PEQUEÑAS (SCLC)
Carcinoma de células escamosas	
Adenocarcinoma	
Carcinoma de Células Grandes	

Tabla 1. Tipos de cáncer de pulmón.

La capacidad de responder a este problema de salud depende de la continua investigación sobre los mecanismos fundamentales celulares y moleculares que controlan la tumorigénesis y metástasis.

La inmunohistoquímica es una herramienta muy valiosa y de uso frecuente en el diagnóstico diferencial de los carcinomas pulmonares. Disponibilidad de marcadores específicos de tumor de pulmón sería importante para el diagnóstico diferencial de pulmón o de sus tipos histológicos. La importancia de las proteínas de unión de célula a célula-(incluyendo proteínas armadillo) en la biología del tumor se conoce, pero el conocimiento es limitado con respecto a las proteínas específicas de las adhesiones intercelulares.

El contacto entre las células epiteliales está mediado por varios tipos de uniones de células. Estas uniones se componen de complejas agregaciones de proteínas transmembrana y de placa, y están típicamente conectados a los

componentes del citoesqueleto. Los desmosomas son complejos célula-célula que se encuentran principalmente en los tejidos epiteliales. Además de las proteínas constitutivas de la placa desmosomal, al menos uno de los tres miembros clásicos de la familia de placofilinas (PKP1 a PKP3) es necesaria para la formación de desmosomas funcionales. La placofilina 1 (PKP1) es un componente principal de placa desmosomal que funciona para reclutar filamentos intermedios a los sitios de contacto célula-célula a través de interacciones con desmoplaquina. Las cadherinas desmosómicas son posibles moléculas de adhesión celular del tipo desmosoma de unión celular en virtud de su homología con la clase cadherina de moléculas de adhesión celular. Dos clases de cadherinas desmosómicas son conocidas, a saber, los desmogleínas y la desmocollinas. El dominio citoplasmático de las cadherinas desmosómicas interactúan con las proteínas de placa que a su vez interactúan con filamentos intermedios de queratina del citoesqueleto.

En los últimos años, se ha acumulado evidencia de que varias proteínas de unión desempeñan papeles importantes en la carcinogénesis, invasión del tumor, y la metástasis. Múltiples estudios han implicado a miembros de la familia de proteínas armadillo, incluyendo placoglobina y  $\beta$ -catenina en la regulación aberrante de células de adhesión celular que promueve la progresión del tumor. Sin embargo, carecemos de la evidencia de un papel para PKPs y otras proteínas desmosomales en la patología tumoral.

Recientemente (Sanchez-Palencia *et al.*, 2010. *Int J Cancer* 129(2):355-364) se han establecido perfiles de expresión génica en NSCLC, en carcinomas primarios de células escamosas o epidermoides y adenocarcinomas. Después del análisis de microarrays de muestras tumorales y no tumorales, el nivel de expresión de 92 genes seleccionados fue validado por qPCR utilizando el test robusto test de Bonferroni en un conjunto independiente de las muestras. En este primer estudio, que mostró resultados acerca de las secuencias de genes expresados diferencialmente en función del tipo de tumor, el estadio y el grado de diferenciación en NSCLC. También puso de manifiesto los datos relacionados con las secuencias de genes expresados diferencialmente correspondientes a las proteínas desmosómicas placofilina 1, queratina 15 y desmogleína 3 en el cáncer no microcítico de pulmón.

Después de que el tipo de cáncer de pulmón es identificado y el estadio determinado, el paciente y su familia pueden discutir opciones de tratamiento con el equipo médico. El tratamiento para cáncer de pulmón se basa en el tipo y estadio del cáncer. Los tratamientos pueden incluir: cirugía para remover el tumor, quimioterapia (medicamentos que matan o reducen el tamaño del tumor) o radiación (rayos X que destruyen o dañan las células cancerosas).

Es por tanto, necesario, desarrollar un método de diagnóstico diferencial que permita conocer el tipo específico de cáncer de pulmón, así como el estadio en el que se encuentra.

## **DESCRIPCIÓN DE LA INVENCION**

En los ejemplos de la presente invención se muestra la localización inmunohistoquímica de las proteínas PKP1, KRT15 y DSG3 involucradas en las adhesiones intercelulares, en setenta y cinco muestras de tumores de NSCLC primarios de pulmón de pacientes no tratados. El patrón de tinción de estas proteínas fue diferente entre los carcinomas escamosos y adenocarcinomas. Los adenocarcinomas no mostraron tinción de membrana. La tinción de membrana es característica de los carcinomas escamosos de pulmón para las tres proteínas analizadas. En nuestro estudio, en los carcinomas de células escamosas, se observó una relación entre la presencia o ausencia de estas proteínas en la membrana y el grado de diferenciación con una tinción más intensa en las áreas mejor diferenciadas en cada neoplasia. La expresión de estas proteínas dibujaron las uniones intercelulares que son características del estrato escamoso del epitelio plano monoestratificado y de neoplasias con este tipo de diferenciación (carcinomas de células escamosas) y se puede utilizar en el diagnóstico de los pacientes afectados por carcinoma de células escamosas del pulmón.

Los autores de la presente invención han desarrollado un método que permite el diagnóstico diferencial del cáncer de pulmón mediante la detección de tres

biomarcadores, así como un kit de diagnóstico diferencial que permite el establecimiento de grupos de pacientes.

Por tanto, un primer aspecto de la invención se refiere al uso de cualquiera de los genes *PKP1*, *KRT15*, *DSG3*, o cualquiera de sus combinaciones, o de  
5 cualquiera de las proteínas PKP1, KRT15, DSG3, o cualquiera de sus combinaciones, para el diagnóstico diferencial del cáncer de pulmón. En una realización preferida, el cáncer de pulmón es el cáncer de pulmón de células no pequeñas, y más preferiblemente es el carcinoma de células escamosas.

Otro aspecto de la invención se refiere al uso simultáneo de los genes *PKP1*,  
10 *KRT15*, *DSG3*, o de cualquiera de las proteínas PKP1, KRT15, DSG3, para el diagnóstico diferencial del cáncer de pulmón. Sin embargo, el uso independiente de cualquiera de ellos o de cualquiera de sus combinaciones podrían ser suficientes para el diagnóstico, pronóstico o predicción de la respuesta al tratamiento de dicha enfermedad. En una realización preferida, el  
15 cáncer de pulmón es el cáncer de pulmón de células no pequeñas, y más preferiblemente es el carcinoma de células escamosas.

#### *MÉTODO DE OBTENCIÓN DE DATOS ÚTILES Y MÉTODO DE DIAGNÓSTICO DIFERENCIAL*

20

Otro aspecto de la invención se refiere a un método de obtención de datos útiles para el diagnóstico diferencial del cáncer de pulmón, que comprende:

- a) obtener una muestra biológica aislada que comprende células de un individuo, y
- 25 b) detectar el nivel de expresión de los genes *PKP1*, *KRT15*, *DSG3*, o cualquiera de sus combinaciones, y/o detectar la cantidad de cualquiera de las proteínas PKP1, KRT15, DSG3, o cualquiera de sus combinaciones, en la muestra aislada de (a).

c) Comparar la expresión del gen o los genes del paso (b) con una cantidad de referencia.

Otro aspecto de la invención se refiere a un método de obtención de datos útiles para el diagnóstico diferencial del cáncer de pulmón, que comprende:

5 a) obtener una muestra biológica aislada que comprende células de un individuo, y

b) detectar simultáneamente el nivel de expresión de los genes *PKP1*, *KRT15*, *DSG3*, y/o detectar la cantidad de las proteínas PKP1, KRT15, DSG3, en la muestra aislada de (a).

10 c) Comparar la expresión de los genes del paso (b) con una cantidad de referencia.

Los pasos (b) y/o (c) de los métodos descritos anteriormente pueden ser total o parcialmente automatizados, por ejemplo, por medio de un equipo robótico sensor para la detección de la cantidad en el paso (b) o la comparación  
15 computerizada en el paso (c).

Otro aspecto de la invención se refiere a un método para el diagnóstico diferencial del cáncer de pulmón en un individuo, de ahora en adelante segundo método de la invención, que comprende los pasos (a) – (c) según el primer método de la invención, y además comprende:

20 (d) diagnosticar al individuo del paso (a) como un individuo con carcinomas escamosos, cuando presente una expresión aumentada del gen *PKP1*, *KRT15* y/o *DSG3* o una cantidad mayor de la proteína PKP1, KRT15 y/o DSG3 en la muestra obtenida en (a), en relación a la cantidad de expresión detectada para dicho gen o dicha proteína en una población de pacientes de referencia. En  
25 una realización preferida de este aspecto de la invención, la expresión aumentada del gen *PKP1*, o una cantidad mayor de la proteína PKP1 se detecta en el citoplasma y la membrana de las células.

En otra realización preferida, el individuo del paso (a) se diagnostica como un individuo con adenocarcinoma de pulmón, cuando presenta la proteína PKP1  
30 y/o KRT15 en el núcleo de la célula tumoral. Aún más preferiblemente, el

individuo del paso (a) se diagnostica como un individuo con adenocarcinoma, cuando no presenta la proteína DSG3, PKP1 o KRT15 en la membrana, ni de la célula.

5 En el presente estudio, la expresión de PKP1 y KRT15 estuvo restringida generalmente a carcinoma epidermoide y se localizó en citoplasma y principalmente en membrana. La expresión de estas proteínas fue más extensa y más intensa especialmente en membranas de carcinomas bien-moderadamente diferenciados.

10 El patrón de tinción de DSG3 también fue distinto entre carcinomas epidermoides y adenocarcinomas. Así, la tinción de membrana fue característica de carcinomas epidermoides para las tres proteínas.

15 Nosotros hemos observado una relación en carcinomas epidermoides entre la presencia/ausencia de estas proteínas en la membrana y el grado de diferenciación, encontrando una tinción más intensa en las áreas mejor diferenciadas de los tumores.

Una "muestra biológica aislada" incluye, pero sin limitarnos a, células, tejidos y/o fluidos biológicos de un organismo, obtenidos mediante cualquier método conocido por un experto en la materia.

20 En una realización preferida de este aspecto de la invención la muestra biológica aislada de un individuo del paso (a) es una muestra de tejido, preferiblemente que comprende células tumorales.

La detección de la expresión de los genes, o la detección de la cantidad de proteína, puede realizarse por cualquier medio conocido en el estado de la técnica.

25 La medida de la concentración, preferiblemente de manera cuantitativa, puede ser llevada a cabo de manera directa o indirecta. La medida directa se refiere a la medida de la expresión de los genes, basada en una señal que se obtiene directamente de los transcritos de dichos genes, o de las proteínas a las que se traducen, y que está correlacionada directamente con el número de moléculas  
30 de RNA o de proteínas producidas por los genes. Dicha señal – a la que



también podemos referirnos como señal de intensidad – puede obtenerse, por ejemplo, midiendo un valor de intensidad de una propiedad química o física de dichos productos. La medida indirecta incluye la medida obtenida de un componente secundario o un sistema de medida biológica (por ejemplo la  
5 medida de respuestas celulares, ligandos, “etiquetas” o productos de reacción enzimática).

El término “comparación”, tal y como se utiliza en la descripción, se refiere pero no se limita, a la comparación de los niveles de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* en la muestra biológica a analizar, también llamada muestra  
10 biológica problema, con los niveles de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* de una o varias muestras de referencia deseable descrita en otra parte de la presente descripción. La muestra de referencia puede ser analizada, por ejemplo, simultánea o consecutivamente, junto con la muestra biológica problema. La comparación descrita en el apartado (c) del método de  
15 la presente invención puede ser realizada manualmente o asistida por ordenador.

Los niveles de expresión de los genes van a dar un determinado perfil de expresión génica. El término “nivel de expresión”, también denominado “cantidad producto génico” se refiere al material bioquímico, ya sea ARN o  
20 proteína, resultado de la expresión de un gen. Algunas veces se usa una medida de la cantidad de producto génico para inferir qué tan activo es un gen. Se entiende por “perfil de expresión génica” el perfil génico obtenido tras la cuantificación del ARNm y/o de proteína producida por los genes de interés o biomarcadores, es decir, por los genes *PKP1*, *KRT15* y/o *DSG3*, en una  
25 muestra biológica aislada. El perfil de expresión de los genes se realiza, preferiblemente, determinando el nivel de ARNm derivado de su transcripción, previa extracción del ARN total presente en la muestra biológica aislada, lo cual puede realizarse mediante protocolos conocidos en el estado de la técnica. La determinación del nivel de ARNm derivado de la transcripción de los genes  
30 *PKP1*, *KRT15* y/o *DSG3*, puede realizarse, por ejemplo, aunque sin limitarnos, mediante amplificación por reacción en cadena de la polimerasa (PCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa (RT-PCR), RT-PCR cuantitativa, retrotranscripción en combinación con la

reacción en cadena de la ligasa (RT-LCR), o cualquier otro método de amplificación de ácidos nucleicos; análisis en serie de la expresión génica (SAGE, SuperSAGE); chips de ADN elaborados con oligonucleótidos depositados por cualquier mecanismo; microarrays de ADN elaborados con oligonucleótidos sintetizados *in situ* mediante fotolitografía o por cualquier otro mecanismo; hibridación *in situ* utilizando sondas específicas marcadas con cualquier método de marcaje; mediante geles de electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante resonancia magnética nuclear o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio. El perfil de expresión génica también podría obtenerse mediante la detección y/o cuantificación de las proteínas producto de la traducción del ARNm derivado de la transcripción de los genes *PKP1*, *KRT15* y/o *DSG3*, mediante por ejemplo, pero sin limitarnos, inmunodetección por western blot. La detección cuantitativa de la expresión de los genes *PKP1*, *KRT15* y/o *DSG3* puede realizarse más preferiblemente mediante PCR en tiempo real (RT-PCR ó RTqPCR). La detección en tiempo real de los productos amplificados puede llevarse a cabo mediante la utilización de moléculas fluorescentes que se intercalan en el ADN de cadena doble o mediante hibridación con diferentes tipos de sondas.

Más preferentemente, la detección de los niveles de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* se realiza mediante Q-RT-PCR.

La PCR cuantitativa en tiempo real (Q-RT-PCR) es una técnica de cuantificación de la expresión génica sensible y reproducible que se puede usar de manera particular para la expresión del perfil génico en células y tejidos. Se puede usar cualquier procedimiento para la evaluación de los resultados de la RT-PCR y se puede preferir el procedimiento  $\Delta\Delta Ct$ . El procedimiento  $\Delta\Delta Ct$  se describe en detalle en Livak y col. (*Methods* 2001, 25:402-408). (Ct = Valores umbral del ciclo). Cuando se lleva la presente invención a la práctica, se deberá usar preferiblemente el procedimiento  $\Delta\Delta Ct$  tal como describen Livak y col. (*Methods* 2001, 25:402-408). El procedimiento  $\Delta\Delta Ct$  implicará una "muestra del control" y una "muestra del sujeto". La "muestra del sujeto" es una muestra

procedente del sujeto que se va a analizar. Por cada muestra, se incluyen un gen diana (aquí: el gen de interés) y un gen endógeno del control (tal como se describe a continuación) para la amplificación de la PCR a partir de alícuotas (normalmente diluciones en serie). Normalmente se usan varias réplicas de cada concentración diluida para derivar la eficacia de la amplificación. La eficacia de la amplificación de la PCR se puede definir como el porcentaje de amplificación (de 0 a 1). Durante la reacción de la qPCR, un software mide normalmente el número de ciclos de cada muestra en el cual la fluorescencia cruza una línea arbitraria (indicadora de la amplificación de la PCR), el umbral. Este punto de cruce es el valor Ct. Muestras más diluidas cruzarán a valores Ct posteriores. Para cuantificar la expresión génica de un gen particular, se divide el Ct de un ácido nucleico procedente del gen de interés por el Ct del ácido nucleico procedente del control endógeno en la misma muestra para normalizar la variación en la cantidad y calidad del ARN entre diferentes muestras y obtener la expresión relativa (con respecto al control endógeno) de cada una de la "muestra del sujeto" y de la "muestra del control". Opcionalmente, esto se lleva a cabo por duplicado, triplicado, cuadruplicado y de manera similar, respectivamente. Se puede obtener de manera adecuada un valor  $\Delta Ct$  del control calculando el promedio de los valores  $\Delta Ct$  obtenidos a partir de muestras de un grupo del control de varios individuos con los cuales se van a comparar los valores de la "muestra del sujeto". El grupo del control (del cual se calcula el valor promedio) consiste en los individuos adecuados a los respectivos fines (de comparación). La persona experta aprenderá de esta divulgación que un grupo de control adecuado es para un fin concreto. En una realización particular, la presente invención se puede llevar a la práctica omitiendo la determinación del valor  $\Delta Ct$  del grupo del control, es decir, determinar (solo) el valor  $\Delta Ct$  de la "muestra del sujeto" y a continuación comparando posteriormente este con el respectivo valor  $\Delta Ct$  promedio del control indicado en los ejemplos.

Como se ha dicho, otros métodos están también disponibles en el estado de la técnica, como por ejemplo la Transferencia Northern Blot, o los microarrays.

Por tanto, en otra realización preferida de este aspecto de la invención la detección del producto de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* se

realiza mediante Transferencia Northern Blot. En otro aspecto de la invención, la detección del producto de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* se realiza mediante microarrays.

El gen *PKP1*, o *plakophilin 1* (*ectodermal dysplasia/skin fragility syndrome*)  
5 codifica un miembro de la familia de genes placofilina. Las proteínas placofilina contienen numerosas repeticiones armadillo, localiza a los desmosomas y a los núcleos celulares, y participar en la vinculación de las cadherinas a los filamentos intermedios del citoesqueleto. Esta proteína puede estar implicada en el reclutamiento molecular y la estabilización durante la formación de  
10 desmosoma. Las mutaciones en este gen se han asociado con la displasia ectodérmica / síndrome de fragilidad de la piel. Dos variantes de la transcripción que codifican diferentes isoformas se han encontrado para este gen. se encuentra en el cromosoma 1 (1q32).

En el contexto de la presente invención, *PKP1* se define también por una  
15 secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína recogida en la SEQ ID NO: 1, y que comprendería diversas variantes procedentes de:

- a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 1,
- 20 b) moléculas de ácido nucleico cuya cadena complementaria híbrida con la secuencia polinucleotídica de a),
- c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,
- d) moléculas de ácido nucleico que codifican un polipéptido que comprende la  
25 secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 1, y en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *PKP1*. Entre dichas moléculas de ácido nucleico se encuentra la recogida en la secuencia SEQ ID NO: 2.

El gen *KRT15* (keratina 15; K15; CK15; K1CO) codifica una proteína que es un miembro de la familia de las queratina. Las queratinas son proteínas de filamentos intermedios responsables de la integridad estructural de las células epiteliales y se subdividen en citoqueratinas y las queratinas del pelo. La mayoría de las citoqueratinas de tipo I consisten en proteínas ácidas que están dispuestas en pares de cadenas de queratina heterotípicas y se agrupan en una región en el cromosoma 17 (17q21.2). En el contexto de la presente invención, *KRT15* se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína recogida en la SEQ ID NO: 3, y que comprendería diversas variantes procedentes de:

- a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 3,
- b) moléculas de ácido nucleico cuya cadena complementaria híbrida con la secuencia polinucleotídica de a),
- 15 c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,
- d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 3, y en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *KRT15*. Entre dichas moléculas de ácido nucleico se encuentra la recogida en la secuencia SEQ ID NO: 4.

El gen *DSG3* o Desmogleína 3 (PVA; CDHF6) es una glicoproteína componente del calcio transmembrana de unión de desmosomas en las células epiteliales de vertebrados. Actualmente, tres miembros de la subfamilia antidesmogleína han sido identificados y todos son miembros de la superfamilia de la molécula de adhesión celular cadherina. Estos miembros de la familia de genes antidesmogleína se encuentran en un clúster en el cromosoma 18. Esta proteína se ha identificado como el autoantígeno de la enfermedad pénfigo vulgaris.

En el contexto de la presente invención, *DSG3* se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína recogida en la SEQ ID NO: 5, y que comprendería diversas variantes procedentes de:

- 5 a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 5,
- b) moléculas de ácido nucleico cuya cadena complementaria híbrida con la secuencia polinucleotídica de a),
- c) moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la  
10 degeneración del código genético,
- d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 5, y en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características  
15 estructurales de la proteína *DSG3*. Entre dichas moléculas de ácido nucleico se encuentra la recogida en la secuencia SEQ ID NO: 6.

El término "diagnóstico", tal y como se utiliza en la presente invención, se refiere a la capacidad de discriminar entre individuos afectados o no por carcinoma de células escamosas, o afectados o no por adenocarcinoma de  
20 pulmón.

En la presente invención "pronóstico" se entiende como la evolución esperada de una enfermedad y se refiere a la valoración de la probabilidad según la cual un sujeto padece una enfermedad así como a la valoración de su inicio, estado de desarrollo, evolución, o de su regresión, y/o el pronóstico del curso de la  
25 enfermedad en el futuro. Como entenderán los expertos en la materia, tal valoración, aunque se prefiere que sea, normalmente puede no ser correcta para el 100% de los sujetos que se va a diagnosticar. El término, sin embargo, requiere que una parte estadísticamente significativa de los sujetos se pueda identificar como que padecen la enfermedad o que tienen predisposición a la  
30 misma. Si una parte es estadísticamente significativa se puede determinar sin más por el experto en la materia usando varias herramientas de evaluación

estadística bien conocidas, por ejemplo, determinación de intervalos de confianza, determinación de valores p, prueba t de Student, prueba de Mann-Whitney, etc. Los intervalos de confianza preferidos son al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%. Los valores de p son, preferiblemente, 0,2, 0,1, 0,05.

Por "predicción de la respuesta" se entiende, en el contexto de la presente invención, la determinación de la probabilidad de que el paciente responda de forma favorable o desfavorable a una terapia o a un tratamiento determinado, incluyendo el tratamiento quirúrgico. Especialmente, el término "predicción", como se usa aquí, se refiere a una evaluación individual de cualquier parámetro que pueda ser útil en determinar la evolución de un paciente. Como entenderán los expertos en la materia, la predicción de la respuesta clínica al tratamiento, aunque se prefiere que sea, no necesita ser correcta para el 100% de los sujetos a ser diagnosticados o evaluados. El término, sin embargo, requiere que se pueda identificar una parte estadísticamente significativa de los sujetos como que tienen una probabilidad aumentada de tener una respuesta positiva. El experto en la materia puede determinar fácilmente si un sujeto es estadísticamente significativo usando varias herramientas de evaluación estadística bien conocidas, por ejemplo, determinación de intervalos de confianza, determinación de los valores de p, prueba t de Student, prueba de Mann Whitney, etc. Los intervalos de confianza preferidos son al menos del 50%>, al menos del 60%>, al menos del 70%>, al menos del 80%>, al menos del 90%), al menos del 95%>. Los valores de p son, preferiblemente, 0,2, 0,1 ó 0,05. La predicción de la respuesta clínica se puede hacer utilizando cualquier criterio de valoración usado en oncología y conocido por el experto en la materia.

A su vez, atendiendo al método de la presente invención, se podrían establecer otras subclasificaciones dentro de esta principal, facilitando, por tanto, la elección y el establecimiento de regímenes terapéuticos o tratamiento adecuados. Esta discriminación tal y como es entendida por un experto en la materia no pretende ser correcta en un 100% de las muestras analizadas. Sin embargo, requiere que una cantidad estadísticamente significativa de las muestras analizadas sean clasificadas correctamente. La cantidad que es

estadísticamente significativa puede ser establecida por un experto en la materia mediante el uso de diferentes herramientas estadísticas, por ejemplo, pero sin limitarse, mediante la determinación de intervalos de confianza, determinación del valor significación P, test de Student o funciones discriminantes de Fisher, medidas no paramétricas de Mann Whitney, correlación de Spearman, regresión logística, regresión lineal, área bajo la curva de ROC (AUC). Preferiblemente, los intervalos de confianza son al menos del 90%, al menos del 95%, al menos del 97%, al menos del 98% o al menos del 99%. Preferiblemente, el valor de p es menor de 0,1, de 0,05, de 0,01, de 0,005 o de 0,0001. Preferiblemente, la presente invención permite detectar correctamente la enfermedad de forma diferencial en al menos el 60%, más preferiblemente en al menos el 70%, mucho más preferiblemente en al menos el 80%, o aún mucho más preferiblemente en al menos el 90% de los sujetos de un determinado grupo o población analizada.

El término "individuo", tal y como se utiliza en la descripción, se refiere a animales, preferiblemente mamíferos, y más preferiblemente, humanos. El término "individuo" no pretende ser limitativo en ningún aspecto, pudiendo ser éste de cualquier edad, sexo y condición física.

La cantidad de referencia se obtiene a partir de los valores de expresión constitutiva de los genes, en un grupo de individuos sanos, o de la expresión de los genes en el grupo de individuos antes de ser sometidos al tratamiento.

La cantidad de referencia será, por ejemplo, en el caso de la diferenciación entre los pacientes afectados por cáncer de pulmón de los individuos sanos, la expresión constitutiva del gen o la cantidad de proteína detectada en un grupo control de individuos sanos. Sin embargo, en el caso de la subclasificación de los pacientes afectados por adenocarcinoma de pulmón o de los afectados por carcinoma de células escamosas, el grupo control estará formado por un grupo de enfermos con adenocarcinoma de pulmón que no tuvieron esa manifestación clínica. En otra realización preferida de este aspecto de la presente invención, la cantidad de referencia se obtiene a partir de una muestra de referencia. La cantidad de referencia puede obtenerse también, por ejemplo, de los límites de distribución normal de una cantidad encontrada en muestras



obtenidas de una población de individuos con adenocarcinoma de pulmón, o carcinoma escamoso en distintas fases, mediante técnicas estadísticas bien conocidas. En otra realización preferida, la muestra de referencia se obtiene de los pacientes antes y después del tratamiento.

- 5 En una realización preferida, el cáncer de pulmón es el cáncer de pulmón de células no pequeñas, y más preferiblemente es el carcinoma de células escamosas y adenocarcinomas.

Otro aspecto de la invención se refiere a una composición farmacéutica que comprende un agente modulador de al menos uno de los genes que se  
10 seleccionan de entre *PKP1*, *KRT15* y/o *DSG3* para tratar un individuo diagnosticado de carcinoma escamoso según los métodos de la invención. Preferiblemente, el agente modulador es un inhibidor de la expresión de dichos genes.

Otro aspecto de la invención se refiere a un anticuerpo para tratar un individuo  
15 diagnosticado de carcinoma escamoso, identificable por un método de la invención, donde el anticuerpo se selecciona de entre anti- *PKP1*, anti- *KRT15*, y/o anti- *DSG3*.

#### *Cuantificación de la proteína*

La cuantificación de la cantidad de proteínas *PKP1*, *KRT15* y/o *DSG3* puede  
20 hacerse por cualquiera de las técnicas conocidas por el experto en la materia, preferiblemente mediante técnicas inmunológicas.

En una realización más preferida, las técnicas inmunológicas están basadas en reacciones de precipitación, basadas en reacciones de aglutinación, inmunomarcación, radioinmunoanálisis y técnicas radioinmunométricas, ELISA  
25 (*Enzyme Linked Immunoabsorbent Assay*), o en cualquiera de sus combinaciones. En otra realización más preferida, las técnicas inmunológicas comprenden el inmunomarcaje. En otra realización aún más preferida, el inmunomarcaje se selecciona de entre inmunomarcaje con anticuerpos conjugados a enzimas, inmunomarcaje con anticuerpos conjugados a  
30 fluorocromos, o citometría. Aún más preferiblemente, la citometría es citometría de flujo.

*KIT O DISPOSITIVO DE DIAGNÓSTICO, MICROARRAY o MICROARRAY DE PROTEÍNAS Y USOS*

5 Otro aspecto de la presente invención se refiere a un kit o dispositivo, de aquí en adelante kit o dispositivo de la invención, que comprende los elementos necesarios para analizar el nivel de expresión de los genes *PKP1*, *KRT15* y/o *DSG3*. En otra realización preferida el kit puede contener oligonucleótido  
10 capaces de hibridar con la secuencia de los genes *PKP1*, *KRT15* y/o *DSG3*, para la posterior amplificación por PCR. Más preferiblemente, la secuencia de los genes *PKP1*, *KRT15* y/o *DSG3* y es la secuencia nucleotídica SEQ ID NO: 2, SEQ ID NO: 4, y SEQ ID NO: 6, respectivamente.

En otra realización preferida, el kit o dispositivo de la invención comprende al  
15 menos un anticuerpo anti- *PKP1*, un anticuerpo anti- *KRT15*, y/o un anticuerpo anti- *DSG3*. En una realización preferida de este aspecto de la invención, el anticuerpo es humano, humanizado o sintético. En otra realización más preferida, el anticuerpo es monoclonal. En otra realización más preferida, el anticuerpo se encuentra marcado con un fluorocromo. Más preferiblemente el  
20 fluorocromo se selecciona de la lista que comprende Fluoresceína (FITC), Tetrametilrodamina y derivados, Ficoeritrina (PE), PerCP, Cy5, Texas, alofococianina, o cualquiera de sus combinaciones.

Más preferiblemente comprende los medios necesarios para comparar la cantidad detectada en el paso (b) con una cantidad de referencia

25 Dicho kit puede contener todos aquellos reactivos necesarios para analizar el nivel de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* por medio de cualquiera de los métodos descritos anteriormente en este documento. El kit además puede incluir, sin ningún tipo de limitación, tampones, agentes para prevenir la contaminación, inhibidores de la degradación de las proteínas, etc.  
30 Por otro lado el kit puede incluir todos los soportes y recipientes necesarios para su puesta en marcha y optimización. Preferiblemente, el kit comprende

además las instrucciones para llevar a cabo cualquiera de los métodos de la invención.

En el caso de (a) un kit adecuado para la RQ-PCR, una técnica de cuantificación de la expresión génica sensible y reproducible, se desea que el  
5 kit comprenda adicionalmente un cebador del oligonucleótido poliT además del (de los) oligonucleótido(s) del kit. Estos reactivos pueden estar comprendidos opcionalmente en el kit.

Una Transferencia Northern implica el uso de electroforesis para separar las muestras de ARN por tamaño y la posterior detección con un(os)  
10 oligonucleótido(s) (sonda de hibridación) complementaria con (parte de) la secuencia diana del ARN de interés.

Es también posible que el(los) oligonucleótido(s) estén inmovilizados en manchas sobre una superficie (preferiblemente sólida). En una de sus realizaciones, el kit comprende una micromatriz, o micromatriz de la invención.  
15 Una micromatriz de ARN es una matriz sobre un sustrato sólido (normalmente un porta de vidrio o una celda de una película fina de silicio) que evalúa grandes cantidades de diferentes ARN que son detectables mediante sondas específicas inmovilizadas sobre manchas sobre un sustrato sólido. Cada mancha contiene una secuencia específica de ácido nucleico, normalmente una secuencia de ADN, como sondas (o indicadores). Aunque el número de  
20 manchas no está limitado de manera alguna, existe una realización preferida en la que la micromatriz se personaliza para los procedimientos de la invención. En una realización, dicha matriz personalizada comprende cincuenta manchas o menos, tal como treinta manchas o menos, incluyendo veinte  
25 manchas o menos. Por tanto, otro aspecto de la invención se refiere a una micromatriz que comprende oligonucleótidos diseñados a partir de una secuencia conocida o un ARNm de los genes, y/o capaces de hibridar con la secuencia de los genes *PKP1*, *KRT15* y/o *DSG3*. Más preferiblemente, la secuencia de los genes *PKP1*, *KRT15* y/o *DSG3* es la secuencia nucleotídica  
30 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, respectivamente.

Otro aspecto de la invención se refiere a un microarray, de ahora en adelante microarray de la invención, que comprende oligonucleótidos o microarreglos de

canal único diseñados a partir de una secuencia conocida o un ARNm de los genes *PKP1*, *KRT15* y/o *DSG3*. Más preferiblemente, la secuencia de los genes *PKP1*, *KRT15* y/o *DSG3* es la secuencia nucleotídica SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16 y/o SEQ ID NO: 18, respectivamente.

Así, por ejemplo, las secuencias de oligonucleótidos son construidas en la superficie de un chip mediante el elongamiento secuencial de una cadena en crecimiento con un sólo nucleótido utilizando fotolitografía. Así, los oligonucleótidos son anclados por el extremo 3' mediante un método de activación selectiva de nucleótidos, protegidos por un reactivo fotolábil, mediante la incidencia selectiva de luz a través de una fotomáscara. La fotomáscara puede ser física o virtual.

Así, las sondas oligonucleótidos pueden ser de entre 10 y 100 nucleótidos, más preferiblemente, de entre 20 y 70 nucleótidos, y aún más preferiblemente, de entre 24 y 30 nucleótidos. Para la cuantificación de la expresión génica, preferiblemente se emplean aproximadamente unos 40 oligonucleótidos por gen.

La síntesis *in situ* sobre un soporte sólido (por ejemplo, vidrio), podría hacerse mediante tecnología chorro de tinta (ink-jet), lo que requiere sondas más largas. Los soportes podrían ser, pero sin limitarse, filtros o membranas de NC o nylon (cargadas), silicio, o Portas de vidrio para microscopios cubiertos con aminosilanos, polilisina, aldehídos o epoxy. La sonda es cada una de las muestras del chip. El target es la muestra a analizar: RNA mensajero, RNA total, un fragmento de PCR, etc.

Otro aspecto de la invención se refiere a un microarray de proteínas, de ahora en adelante microarray de proteínas de la invención, que comprende anticuerpos anti- *PKP1*, anticuerpos anti- *KRT15*, y/o anticuerpos anti- *DSG3*. Las sondas son anticuerpos fijados a portaobjetos de vidrio y los blancos son muestras de suero o tejido.

Otro aspecto de la invención se refiere al uso del kit o dispositivo, la micromatriz, o el microarray de la invención, de la para la obtención de datos

escamoso.

Otro aspecto de la invención se refiere a un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer  
5 que un ordenador lleve a cabo los pasos de cualquiera de los métodos de la invención (del primer o del segundo método de la invención).

Otro aspecto de la invención se refiere a una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos de cualquiera de los métodos de la invención.

10 Los términos "polinucleótido" y "ácido nucleico" se usan aquí de manera intercambiable, refiriéndose a formas poliméricas de nucleótidos de cualquier longitud, tanto ribonucleótidos (ARN ó RNA) como desoxiribonucleótidos (ADN ó DNA). Los términos "secuencia aminoacídica", "péptido", "oligopéptido", "polipéptido" y "proteína" se usan aquí de manera intercambiable, y se refieren  
15 a una forma polimérica de aminoácidos de cualquier longitud, que pueden ser codificantes o no codificantes, química o bioquímicamente modificados.

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas  
20 y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

## 25 **BREVE DESCRIPCIÓN DE LAS FIGURAS**

**Fig. 1.** Localización inmunohistoquímica de la proteína pKP1 en el carcinoma de pulmón de células escamosas o epidermoide. A) Tinción de la membrana. B) Tinción del citoplasma. C) Tinción del citoplasma con tinción intensa en el globo córneo (flecha). D) Intensa tinción nuclear en células con apariencia más  
30 inmadura. (A, B, E y F 40x; C 10x).

**Fig. 2.** Localización inmunohistoquímica de la proteína pKP1 en el adenocarcinoma de pulmón. A) Tinción negativa B, C) Tinción en citoplasma. D) Tinción nuclear. (A y B, 40x; C y D 20x).

**Fig. 3.** Localización inmunohistoquímica de la proteína KRT15 en membranas del carcinoma de pulmón de células escamosas o epidermoides (40x).

## EJEMPLOS DE LA INVENCION

A continuación se ilustrará la invención mediante unos ensayos realizados por los inventores, que pone de manifiesto la especificidad y efectividad de los métodos de la invención para obtener datos útiles en el diagnóstico diferencial del adenocarcinoma de pulmón y del carcinoma escamoso.

La localización inmunohistoquímica de las proteínas pKP1, KRT15 y DSG3 se examinó en carcinoma de células escamosas y adenocarcinomas de pulmón. Setenta y cinco tumores de NSCLC se incluyeron en el estudio, que comprende 47 carcinomas escamosos y 28 adenocarcinomas. Los carcinomas de células escamosas fueron bien moderadamente diferenciado en 26 muestras y pobremente diferenciados en 21 muestras. Los adenocarcinomas fueron bien diferenciados moderadamente en 17 muestras y pobremente diferenciados en 11 muestras.

Los resultados de inmunohistoquímica para la proteína pKP1 se resumen en la Tabla 1. La localización inmunohistoquímica para pKP1 proteína se restringió principalmente a carcinomas escamosos, con una distribución heterogénea y la intensidad de la tinción entre los carcinomas diferentes y en diferentes áreas de un carcinoma. En carcinomas epidermoides la expresión de PKP 1 se localiza principalmente en el citoplasma y la membrana (Cohran,  $p = 0,001$ ) dibujando las uniones intercelulares, (Fig. 1). La expresión membranosa de la proteína PKP1 se observó principalmente en los tumores bien-moderadamente diferenciados de carcinomas epidermoides (chi-cuadrado  $p = 0,176$ ; Fisher  $p = 0,135$ ), y considerando una neoplasia específica, la positividad se encontró

predominantemente en las mejores zonas diferenciadas. La tinción nuclear de PKP1 fue observada en células con un aspecto más inmaduro (Fig. 1D).

La tinción para PKP 1 en adenocarcinomas fue generalmente focal y se observó con una intensidad variable en el núcleo y en el citoplasma, y nunca en la membrana celular (Cohran,  $p = 0,039$ ) (Fig. 2).

Los resultados de inmunohistoquímica para la proteína KRT15 se resumen en la Tabla 2. La localización inmunohistoquímica para la proteína KRT15 se restringió principalmente a carcinomas escamosos, con una distribución heterogénea y la intensidad de la tinción entre los carcinomas diferentes y en diferentes áreas de un carcinoma. La expresión de KRT15 se localizó en el citoplasma y en su mayoría en la membrana celular (Cohran,  $p = 0,000$ ) dibujando las uniones intercelulares, (Fig. 3). En los carcinomas de células escamosas, tanto en los tumores bien- moderadamente como pobremente diferenciados (chi-cuadrado  $p = 1,000$ ) mostraron tinción de la membrana celular para KRT15 y cuando se consideró una neoplasia específica, la positividad se observó principalmente en las mejores zonas diferenciadas.

La tinción de la membrana celular para KRT15 fue generalmente ausente en los adenocarcinomas (Cohran,  $p = 0,005$ ). La tinción positiva en este tipo de neoplasia fue generalmente focal y con una intensidad variable y se observaron en el núcleo y el citoplasma, pero nunca en la membrana celular. Se observó tinción nuclear para KRT15 en células con un aspecto más inmaduro.

Los resultados de inmunohistoquímica para la proteína DSG3 se resumen en la Tabla 3. La tinción inmunohistoquímica para proteína DSG3 fue principalmente ausente en adenocarcinomas, en cambio, carcinomas de células escamosas mostraron positividad en las membranas celulares (Cohran,  $p = 0,000$ ). La expresión de DSG3 se observó principalmente en las membranas de carcinomas escamosos moderadamente bien diferenciados. (chi-cuadrado  $p = 0,792$ ). Los adenocarcinomas mostraron ausencia de tinción para la proteína DSG3 ni en el núcleo, ni en el citoplasma ni en la membrana (Cohran,  $p =$

0,223) y no hubo tinción en adenocarcinomas moderados ni pobremente diferenciados (Fisher,  $p = 1,000$ ).

5

10

15

20



Tabla 1. Análisis de la localización de PKP1.

Tipo of tumor (N)	Núcleo PKP1		Citoplasma PKP1		Membrana PKP1	
	+ (N)	- (N)	+ (N)	- (N)	+ (N)	- (N)
Adenocarcinoma (28)	3.6% (1)	96.4% (27)	14.3% (4)	85.7% (24)	0% (0)	100% (28)
	(Cochran p=0.039)		(Cochran p=0.039)		(Cochran p=0.039)	
Carcinomas de células escamosas (47)	36.1% (17)	63.8% (30)	59.6% (28)	40.4% (19)	57.4%(27)	42.5% (20)
	(Cochran p= 0.001)		(Cochran p= 0.001)		(Cochran p= 0.001)	
Carcinoma de células escamosas Bien- Moderadamente diferenciados (26)	38.5% (10)	61.5% (16)	65.4% (17)	34.6% (9)	69.2% (18)	30.8% (8)
Carcinoma de células escamosas Pobremente diferenciados (21)	35% (7)	65% (13)	55% (11)	45% (9)	45% (9)	55% (11)
	(chi-square p= 1.000)		(chi-square p= 0.681)		(chi-square p= 0.176; Fisher p= 0.135)	

ES 2 482 468 B1

<p><b>Adenocarcinoma: Bien- Moderadamente diferenciados  (17)</b></p>	<p><b>5.9% (1) 94.1% (16)</b></p>	<p><b>11.8% (2) 88.2% (15)</b></p>	<p><b>0% (0) 100% (17)</b></p>
<p><b>Adenocarcinoma: Pobrementemente diferenciados  (11)</b></p>	<p><b>0% (0) 100% (11) (Fisher p= 1.000)</b></p>	<p><b>18.2% (2) 81.8% (9) (Fisher p= 1.000)</b></p>	<p><b>0% (0) 100% (11)</b></p>

N= Número de casos

Tabla 2. Análisis de la localización de KRT15.

Tipo de tumor (N)	Núcleo KRT15 + (N)      - (N)	Citoplasma KRT15 + (N)      - (N)	Membrana KRT15 + (N)      - (N)
Adenocarcinoma (26)	26.9% (7)      73.1% (19) (Cochran p=0.005)	3.8% (1)      96.2% (25) (Cochran p=0.005)	0% (0)      100% (26) (Cochran p=0.005)
Carcinoma de células escamosas (47)	0% (0)      100% (47) (Cochran p= 0.000)	46.8% (22)      53.2% (25) (Cochran p= 0.000)	70.2%(33)      29.8% (14) (Cochran p= 0.000)
Carcinoma de células escamosas: Bien- Moderadamente diferenciados (26)	0% (0)      100% (26)	46.2% (12)      53.8% (14)	73.1% (19)      26.9% (7)
Carcinoma de células escamosas: Pobremente diferenciados (21)	0% (0)      100% (21)	41.7% (10)      50% (11) (chi-square p= 1.000)	70% (15)      30% (6) (chi-square p= 1.000)
Adenocarcinoma: Bien- Moderadamente diferenciados (16)	25% (4)      75% (12)	0% (0)      100% (16)	0% (0)      100% (16)
Adenocarcinoma: Pobremente diferenciados (10)	30% (3)      70% (7) (Fisher p= 1.000)	10% (1)      90% (9) (Fisher p= 0.385)	0% (0)      100% (10)

N= Número de casos

Tabla 3. Análisis de la localización de DSG3.

Tipo de tumor (N)	Núcleo DSG3		Citoplasma DSG3		Membrana DSG3	
	+(N)	-(N)	+(N)	-(N)	+(N)	-(N)
Adenocarcinoma (28)	7.1% (2) (26) (Cochran p=0.223)	92.9%	3.6% (1) (27) (Cochran p=0.223)	96.4%	0% (0) (28) (Cochran p=0.223)	100%
Carcinoma de células escamosas (45)	2.2% (1) (44) (Cochran p= 0.000)	97.8%	15.6% (7) (38) (Cochran p= 0.000)	84.4%	53.3%(24) (21) (Cochran p= 0.000)	46.7%
Carcinoma de células escamosas : Bien- Moderadamente diferenciados (25)	0% (0) (25)	100%	16% (4) (21)	84%	56% (14) (11)	44%
Carcinoma de células escamosas : Pobrememente diferenciados (20)	5.3% (1) (19) (Fisher p= 0.432)	94.7%	15.8% (3) (17) (Fisher p= 1.000)	84.2%	47.4% (9) (11) (chi-square p= 0.792)	52.6%
Adenocarcinoma: Bien- Moderadamente diferenciados (17)	5.9% (1) (16)	94.1%	5.9% (1) (16)	94.1%	0% (0) (17)	100%
Adenocarcinoma: Pobrememente diferenciados (11)	9.1% (1) (10) (Fisher p= 1.000)	90.9%	0% (0) (11) (Fisher p= 1.000)	100%	0% (0) (11)	100%

N= Número de casos

## REIVINDICACIONES

- 1.- El uso de los genes *PKP1*, *KRT15*, *DSG3*, o de las proteínas PKP1,  
5 KRT15, DSG3, para el diagnóstico diferencial del cáncer de pulmón.
- 2.- El uso de los genes *PKP1*, *KRT15*, *DSG3*, o de las proteínas PKP1,  
KRT15, DSG3 según la reivindicación anterior, donde el cáncer de pulmón es  
cáncer de pulmón de células no pequeñas.
- 3.- El uso de los genes *PKP1*, *KRT15*, *DSG3*, o de las proteínas PKP1,  
10 KRT15, DSG3 según cualquiera de las reivindicaciones 1-2, donde el cáncer de  
pulmón es el carcinoma de células escamosas.
- 4.- Un método de obtención de datos útiles para el diagnóstico diferencial del  
cáncer de pulmón, que comprende:
- 15 a. obtener una muestra biológica aislada que comprende células de un  
individuo, y
- b. detectar simultáneamente el nivel de expresión de los genes *PKP1*,  
*KRT15*, *DSG3*, o la cantidad de las proteínas PKP1, KRT15, DSG3, en  
la muestra aislada de (a), y
- 20 c. comparar la expresión de los genes del paso (b) con una cantidad de  
referencia.
- 5.- Un método para el diagnóstico diferencial del cáncer de pulmón, que  
comprende los pasos (a) – (c) según la reivindicación 4, y además comprende
- 25 (d) diagnosticar al individuo del paso (a) como un individuo con  
carcinomas escamoso, cuando presente una expresión aumentada del  
gen *PKP1*, *KRT15* y/o *DSG3* o una cantidad mayor de la proteína PKP1,  
KRT15 y/o DSG3 en la muestra obtenida en (a), en relación a la cantidad  
de expresión detectada para dicho gen o dicha proteína en una  
población de pacientes de referencia.

- 6.- El método según cualquiera de las reivindicaciones 4-5, donde la expresión aumentada del gen *PKP1*, o una cantidad mayor de la proteína PKP1 se detecta en el citoplasma y la membrana de las células.
- 7.- El método según cualquiera de las reivindicaciones 4-6, donde el individuo del paso (a) se diagnostica como un individuo con adenocarcinoma de pulmón, cuando presenta la proteína PKP1 y/o KRT15 en el núcleo de la célula y está ausente en membrana.
- 8.- El método según cualquiera de las reivindicaciones 4-7, donde el individuo del paso (a) se diagnostica como un individuo con adenocarcinoma, cuando no presenta la proteína DSG3 ni en la membrana, ni en el citoplasma, ni en el núcleo de la célula.
- 9.- El método según cualquiera de las reivindicaciones 5-8, donde la muestra biológica aislada de un individuo del paso (a) es una muestra de tejido, preferiblemente tumoral.
- 10.- El método según cualquiera de las reivindicaciones 5-9, donde la detección de los niveles de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* se realiza mediante Q-RT-PCR.
- 11.- El método según cualquiera de las reivindicaciones 5-10, donde la identificación de la cantidad de proteínas *PKP1*, *KRT15* y/o *DSG3*, se realiza mediante técnicas inmunológicas.
- 12.- El método según la reivindicación 11, donde las técnicas inmunológicas están basadas en reacciones de precipitación, basadas en reacciones de aglutinación, inmunomarcación, radioinmunoanálisis y técnicas radioinmunométricas, ELISA (*Enzyme Linked Immunoabsorbent Assay*), o en cualquiera de sus combinaciones.
- 13.- El método según cualquiera de las reivindicaciones 11-12, donde las técnicas inmunológicas comprenden el inmunomarcaje.
- 14.- El método según la reivindicación 13, donde el inmunomarcaje se selecciona de entre inmunomarcaje con anticuerpos conjugados a enzimas, inmunomarcaje con anticuerpos conjugados a fluorocromos, o citometría.

15.- El método según cualquiera de las reivindicaciones 4-14, donde el cáncer de pulmón es cáncer de pulmón de células no pequeñas.

16.- El método según cualquiera de las reivindicaciones 4-14, donde el cáncer de pulmón es el carcinoma de células escamosas.

5 17.- Un kit o dispositivo que comprende los elementos necesarios para analizar el nivel de expresión de los genes *PKP1*, *KRT15* y/o *DSG3* o la cantidad de proteína PKP1, KRT15 y/o DSG3.

10 18.- El kit o dispositivo según la reivindicación anterior, que comprende un anticuerpo anti- PKP1, un anticuerpo anti- KRT15, y/o un anticuerpo anti- DSG3.

19.- El kit o dispositivo según cualquiera de las reivindicaciones 17-18, donde el anticuerpo es monoclonal

20.- El kit o dispositivo según cualquiera de las reivindicaciones 17-19, donde el anticuerpo se encuentra marcado con un fluorocromo.

15 21.- El kit o dispositivo según la reivindicación anterior, donde el fluorocromo se selecciona de la lista que comprende Fluoresceína (FITC), Tetrametilrodamina y derivados, Ficoeritrina (PE), PerCP, Cy5, Texas, alofococianina, o cualquiera de sus combinaciones.

20 22.- El uso del kit o dispositivo según cualquiera de las reivindicaciones 17-22, para el diagnóstico diferencial del cáncer de pulmón.

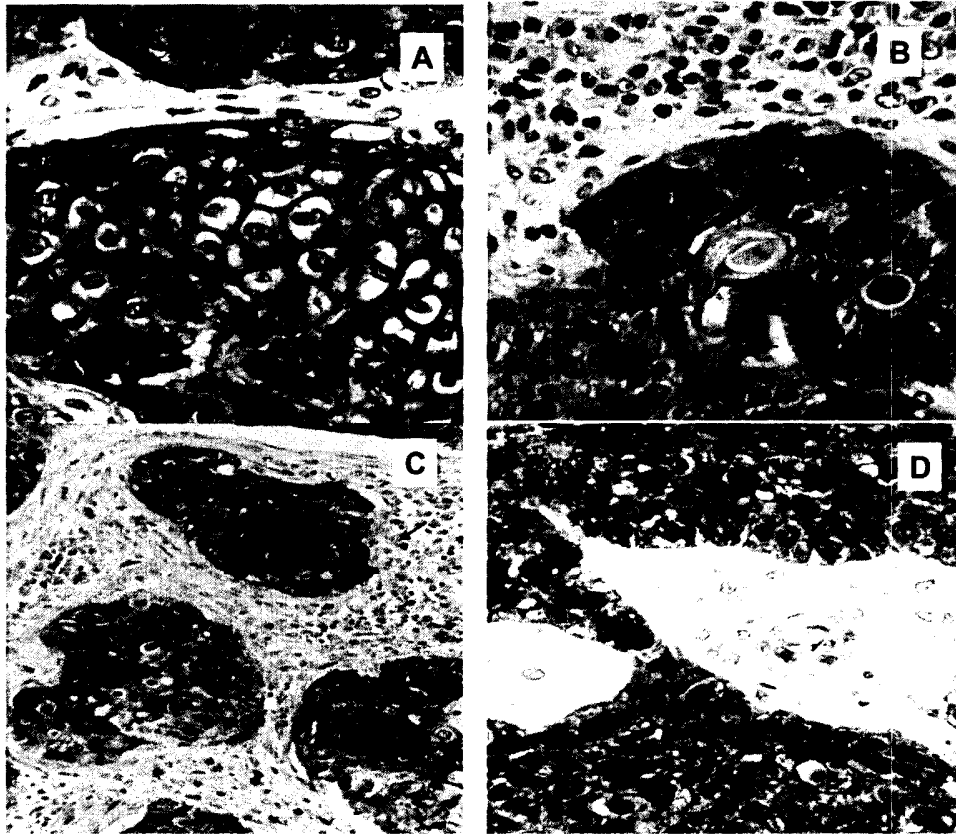


Fig. 1



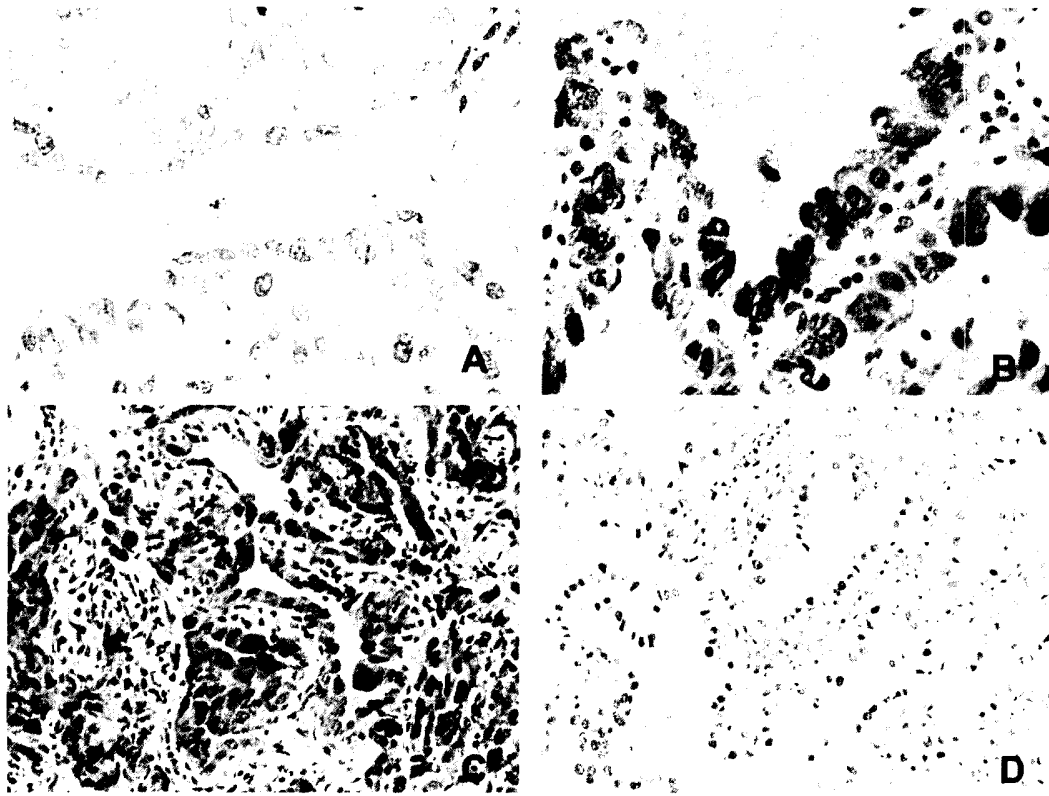


Fig. 2

ES 2 482 468 B1



**Fig. 3**

Lista de Secuencias

- <110> SERVICIO ANDALUZ DE SALUD  
UNIVERSIDAD DE GRANADA
- <120> Método de obtención de datos útiles para el diagnóstico diferencial del cáncer de pulmón.
- <130> FIBAO-1411
- <160> 6
- <170> PatentIn version 3.5
- <210> 1
- <211> 747
- <212> PRT
- <213> Homo sapiens
- <400> 1

Met Asn His Ser Pro Leu Lys Thr Ala Leu Ala Tyr Glu Cys Phe Gln  
 1 5 10 15

Asp Gln Asp Asn Ser Thr Leu Ala Leu Pro Ser Asp Gln Lys Met Lys  
 20 25 30

Thr Gly Thr Ser Gly Arg Gln Arg Val Gln Glu Gln Val Met Met Thr  
 35 40 45

Val Lys Arg Gln Lys Ser Lys Ser Ser Gln Ser Ser Thr Leu Ser His  
 50 55 60

Ser Asn Arg Gly Ser Met Tyr Asp Gly Leu Ala Asp Asn Tyr Asn Tyr  
 65 70 75 80

Gly Thr Thr Ser Arg Ser Ser Tyr Tyr Ser Lys Phe Gln Ala Gly Asn  
 85 90 95

Gly Ser Trp Gly Tyr Pro Ile Tyr Asn Gly Thr Leu Lys Arg Glu Pro  
 100 105 110

Asp Asn Arg Arg Phe Ser Ser Tyr Ser Gln Met Glu Asn Trp Ser Arg  
 115 120 125

His Tyr Pro Arg Gly Ser Cys Asn Thr Thr Gly Ala Gly Ser Asp Ile  
 130 135 140

Cys Phe Met Gln Lys Ile Lys Ala Ser Arg Ser Glu Pro Asp Leu Tyr  
 145 150 155 160

Cys Asp Pro Arg Gly Thr Leu Arg Lys Gly Thr Leu Gly Ser Lys Gly  
 165 170 175

Gln Lys Thr Thr Gln Asn Arg Tyr Ser Phe Tyr Ser Thr Cys Ser Gly  
 180 185 190

ES 2 482 468 B1

Gln Lys Ala Ile Lys Lys Cys Pro Val Arg Pro Pro Ser Cys Ala Ser  
 195 200 205

Lys Gln Asp Pro Val Tyr Ile Pro Pro Ile Ser Cys Asn Lys Asp Leu  
 210 215 220

Ser Phe Gly His Ser Arg Ala Ser Ser Lys Ile Cys Ser Glu Asp Ile  
 225 230 235 240

Glu Cys Ser Gly Leu Thr Ile Pro Lys Ala Val Gln Tyr Leu Ser Ser  
 245 250 255

Gln Asp Glu Lys Tyr Gln Ala Ile Gly Ala Tyr Tyr Ile Gln His Thr  
 260 265 270

Cys Phe Gln Asp Glu Ser Ala Lys Gln Gln Val Tyr Gln Leu Gly Gly  
 275 280 285

Ile Cys Lys Leu Val Asp Leu Leu Arg Ser Pro Asn Gln Asn Val Gln  
 290 295 300

Gln Ala Ala Ala Gly Ala Leu Arg Asn Leu Val Phe Arg Ser Thr Thr  
 305 310 315 320

Asn Lys Leu Glu Thr Arg Arg Gln Asn Gly Ile Arg Glu Ala Val Ser  
 325 330 335

Leu Leu Arg Arg Thr Gly Asn Ala Glu Ile Gln Lys Gln Leu Thr Gly  
 340 345 350

Leu Leu Trp Asn Leu Ser Ser Thr Asp Glu Leu Lys Glu Glu Leu Ile  
 355 360 365

Ala Asp Ala Leu Pro Val Leu Ala Asp Arg Val Ile Ile Pro Phe Ser  
 370 375 380

Gly Trp Cys Asp Gly Asn Ser Asn Met Ser Arg Glu Val Val Asp Pro  
 385 390 395 400

Glu Val Phe Phe Asn Ala Thr Gly Cys Leu Arg Lys Arg Leu Gly Met  
 405 410 415

Arg Glu Leu Leu Ala Leu Val Pro Gln Arg Ala Thr Ser Ser Arg Val  
 420 425 430

Asn Leu Ser Ser Ala Asp Ala Gly Arg Gln Thr Met Arg Asn Tyr Ser  
 435 440 445

Gly Leu Ile Asp Ser Leu Met Ala Tyr Val Gln Asn Cys Val Ala Ala  
 450 455 460

ES 2 482 468 B1

Ser Arg Cys Asp Asp Lys Ser Val Glu Asn Cys Met Cys Val Leu His  
 465 470 475 480  
 Asn Leu Ser Tyr Arg Leu Asp Ala Glu Val Pro Thr Arg Tyr Arg Gln  
 485 490 495  
 Leu Glu Tyr Asn Ala Arg Asn Ala Tyr Thr Glu Lys Ser Ser Thr Gly  
 500 505 510  
 Cys Phe Ser Asn Lys Ser Asp Lys Met Met Asn Asn Asn Tyr Asp Cys  
 515 520 525  
 Pro Leu Pro Glu Glu Glu Thr Asn Pro Lys Gly Ser Gly Trp Leu Tyr  
 530 535 540  
 His Ser Asp Ala Ile Arg Thr Tyr Leu Asn Leu Met Gly Lys Ser Lys  
 545 550 555 560  
 Lys Asp Ala Thr Leu Glu Ala Cys Ala Gly Ala Leu Gln Asn Leu Thr  
 565 570 575  
 Ala Ser Lys Gly Leu Met Ser Ser Gly Met Ser Gln Leu Ile Gly Leu  
 580 585 590  
 Lys Glu Lys Gly Leu Pro Gln Ile Ala Arg Leu Leu Gln Ser Gly Asn  
 595 600 605  
 Ser Asp Val Val Arg Ser Gly Ala Ser Leu Leu Ser Asn Met Ser Arg  
 610 615 620  
 His Pro Leu Leu His Arg Val Met Gly Asn Gln Val Phe Pro Glu Val  
 625 630 635 640  
 Thr Arg Leu Leu Thr Ser His Thr Gly Asn Thr Ser Asn Ser Glu Asp  
 645 650 655  
 Ile Leu Ser Ser Ala Cys Tyr Thr Val Arg Asn Leu Met Ala Ser Gln  
 660 665 670  
 Pro Gln Leu Ala Lys Gln Tyr Phe Ser Ser Ser Met Leu Asn Asn Ile  
 675 680 685  
 Ile Asn Leu Cys Arg Ser Ser Ala Ser Pro Lys Ala Ala Glu Ala Ala  
 690 695 700  
 Arg Leu Leu Leu Ser Asp Met Trp Ser Ser Lys Glu Leu Gln Gly Val  
 705 710 715 720  
 Leu Arg Gln Gln Gly Phe Asp Arg Asn Met Leu Gly Thr Leu Ala Gly  
 725 730 735

ES 2 482 468 B1

Ala Asn Ser Leu Arg Asn Phe Thr Ser Arg Phe  
 740 745

<210> 2  
 <211> 5447  
 <212> DNA  
 <213> Homo sapiens

<400> 2  
 ggggtggtgca gggcaggggt ggtatatacct gtctgacgga gggcgggcct cgccagtgcc 60  
 agagagggac gaaccaggtt ggaagcgcca ggagcagctg caggagagccc tcacgcggac 120  
 cacgcactct atggccgtag ggagccgctg agagcgagaa gaggcagctc ctgcccgcc 180  
 gctgcaccgc acctgcctc gcctctctgc tctcctaggc cccggccgcg cgccaccgc 240  
 ctcccgccac catgaaccac tcgccgctca agaccgcctt ggcgtacgaa tgcttccagg 300  
 accaggacaa ctccacgttg gctttgccgt cggaccaaaa gatgaaaaca ggcacgtctg 360  
 gcaggcagcg cgtgcaggag caggtgatga tgaccgtcaa gcggcagaag tccaagtctt 420  
 cccagtcgtc caccctgagc cactccaatc gaggttccat gtatgatggc ttggctgaca 480  
 attacaacta tgggaccacc agcaggagca gctactactc caagttccag gcagggaaatg 540  
 gctcatgggg atatccgatc tacaatggaa ccctcaagcg ggagcctgac aacaggcgct 600  
 tcagctccta cagccagatg gagaactgga gccggcacta cccccggggc agctgtaaca 660  
 ccaccggcgc aggcagcgac atctgcttca tgcagaaaat caaggcgagc cgcagtgagc 720  
 ccgacctcta ctgtgaccca cggggcacc tgcgcaaggg cagcgtgggc agcaagggcc 780  
 agaagaccac ccagaaccgc tacagctttt acagcacctg cagtggtcag aaggccataa 840  
 agaagtgcc tgtgcgccg ccctcttggt cctccaagca ggaccctgtg tatatcccgc 900  
 ccattctctg caacaaggac ctgtcctttg gccactctag ggccagctcc aagatctgca 960  
 gtgaggacat cgagtgcagt gggctgacca tccccaggc tgtgcagtac ctgagctccc 1020  
 aggatgagaa gtaccaggcc attggggcct attacatcca gcatacctgc ttccaggatg 1080  
 aatctgccaa gcaacaggtc tatcagctgg gaggcattctg caagctgggtg gacctctcc 1140  
 gcagcccaaa ccagaacgtc cagcaggccg cggcaggggc cctgcgcaac ctggtgttca 1200  
 ggagcaccac caacaagctg gagacccgga ggcagaatgg gatccgagag gcagtcagcc 1260  
 tcctgaggag aaccgggaac gccgagatcc agaagcagct gactgggctg ctctggaacc 1320  
 tgtcttccac tgacgagctg aaggaggaac tcattgccga cgccctgcct gttctggccg 1380  
 accgcgtcat cattccctc tctggctggt gcgatggcaa tagcaaatg tcccgggaag 1440  
 tgggtgaccc tgaggtcttc ttcaatgcca caggctgctt gagaaagaga ctgggcatgc 1500  
 gggagcttct ggctcttggt ccgcaaaggg cactagtag cagggtgaac ctgagctcgg 1560  
 ccgatgcagg ccgccagacc atgcgtaact actcagggtc cattgattcc ctcatggcct 1620  
 atgtccagaa ctgtgtagcg gccagccgct gtgacgacaa gtctgtggaa aactgcatgt 1680  
 gtgttctgca caacctctcc taccgctgg acgccagggt gccaccgcg taccgccagc 1740

ES 2 482 468 B1

tgaggtataa	cgcccgaac	gcctacaccg	agaagtcctc	cactggctgc	ttcagcaaca	1800
agagcgacia	gatgatgaac	aacaactatg	actgccccct	gcctgaggaa	gagaccaacc	1860
ccaagggcag	cggctggtg	taccattcag	atgccatccg	cacctacctg	aacctcatgg	1920
gcaagagcaa	gaaagatgct	accctggagg	cctgtgctgg	tgccctgcag	aacctgacag	1980
ccagcaaggg	gctgatgtcc	agtggcatga	gccagttgat	tgggctgaag	gaaaagggcc	2040
tgccacaaat	tgcccgcctc	ctgcaatctg	gcaactctga	tgtggtgcgg	tccggagcct	2100
ccctcctgag	caacatgtcc	cgccaccctc	tgctgcacag	agtgatgggg	aaccaggtgt	2160
tcccgagggt	gaccaggctc	ctcaccagcc	acactggcaa	taccagcaac	tccgaagaca	2220
tcttgtcctc	ggcctgctac	actgtgagga	acctgatggc	ctcgcagcca	caactggcca	2280
agcagtactt	ctccagcagc	atgctcaaca	acatcatcaa	cctgtgccga	agcagtgctt	2340
caccaaggc	cgcagaagct	gcccggcttc	tcctgtctga	catgtggtcc	agcaaggaac	2400
tgcaaggtgt	cctcagacag	caaggtttcg	ataggaacat	gctgggaacc	ttagctgggg	2460
ccaacagcct	caggaacttc	acctcccgat	tctaagaaga	gactgtccaa	gcaagttagg	2520
cttgcaggaa	gatatgacc	agctgagaag	ccctcaggcc	tcgctggatg	gggttttctg	2580
tccatcctgt	gcagtatttg	ggaaagtca	caagaaactg	agaagaaacc	taaaaactgt	2640
ggatagtggg	aagattttta	gatttttttt	ttccttgggg	aaactggcag	gcaatggggg	2700
ttagggaggt	tggggcgggg	ggggctttct	tgagttaaag	gggcttatat	gtgatgtcaa	2760
tatttcttcc	tctgagaaat	ggtatatata	tgtgtataat	gtaagtgtgt	gcatgcatgt	2820
gcgctgcat	gtgtgtgtgt	gtgagtgtct	taaagcataa	ccacaaactg	caaaaagcta	2880
ggtaagctat	tttgttgag	ctcataaggt	ggtgaaaagg	actctcctgt	gtttcttact	2940
cataggcaag	gacaacatgt	gctttttggt	gagctgctca	taattcctga	aatgtgtggt	3000
gccagggcaa	ggggccatc	actgcagtca	ggccctcaga	ggagtcctgc	aggcttcccta	3060
ccagtggctt	ccaggggtgc	aggagtaact	ggggctgggc	cagcctcccc	acttacaagg	3120
ctgctttcca	ggaagggagg	tctggtgtat	ctcatgggag	aatctggggg	gtctgtaatg	3180
tcaccctcc	agcagcgcca	caaggactga	ggttgggtag	gtgtgggggt	ccagaggaca	3240
gcaggacact	ctgcatact	ttgccaaatg	aggcctgctc	agaggagtag	gagctgaaag	3300
atggtgcctt	ccaccctctt	gggctgtgtg	cccatcagag	caggctcagc	ctgcaaaggc	3360
cctgcattca	gaggtcttgt	aatctacttg	ttgcaggaga	aagaaggtaa	aaaatgattt	3420
ttttaagaaa	agctatttta	ttgcagctct	ttccaagag	ctgttctggg	aatggctggt	3480
cttcatattc	ccagtggaga	ggggaacaag	tggggctggg	catataccta	ttccggcttc	3540
tagtgggatg	gagttgggg	atagaaatta	accaggaaga	tgtttccacc	aagcctgctg	3600
tgagtcaatt	gagggagtgt	ttggggctcc	aggagacttg	gacgggggga	gtttgggtag	3660
actaggaaa	gaaagtgcc	tatcagggtg	ccggtaccgg	caagctcaca	tctcagccag	3720
gggcatgcc	ccacttcccc	tgaccccagc	tgtcttgtct	ccactctgtg	aaaccacag	3780

ES 2 482 468 B1

gggatgtgat aaacagggct attaggggta tcagccacgt cgagcccca gactctgtgc 3840  
 acttcagacc agcagcagca ggagggctcc cgagggcctt atgagaaaac ctgtgtggac 3900  
 atcccttgggt gtacactaag acagagcaga gccagcgct cccaagcctt cctccttcca 3960  
 gcttctacct ccatgctagc attgctgggtg ttagagagga attaacttcc tggctctgtgc 4020  
 ctttctctag aagaatataa gatgctcctc ctctcacc cttctcagcc tcctccaag 4080  
 tcttctctt ctgcaccacc cccgagtcca aaccacctc ttgccccagc attcaggctg 4140  
 gaaaacactg atgtggactc agtatgataa ctgagatggg ggacgccaga catgtgagga 4200  
 cgctgtcctc cgagaggtgt ccccggctgt tagccagctg tgctgtgggtg ctgtgggtct 4260  
 gtcataccct cccttgcttc tgttcacact gggaggcca ctctggctc acctctcct 4320  
 ctcagggacc cacgtgggag cctggatccc tggactgtcc tgggcatagg tttcagggg 4380  
 ctctttgtt gtcatcagaa cccagaggaa ttcttctct aaaaaatag tatggcatac 4440  
 caatctgtgc ggggcagtgt cctaagcact tagactacat caggaagaa cacagaccac 4500  
 atccctgtcc tcatgcggct tatgttttct ggaggaaagt ggagacaaa gtccttggct 4560  
 ttagggctcc cccggctggg ggctgtgcag tccggctcagg gcgggagggg aatgcaccg 4620  
 ctgcatgtga acctaccag cccaggcggg tgccttcc ccttagcact acctggcct 4680  
 cctgcatccc ctgcctcat gttcctcca cttcaaaga atgaagagcc catggggccc 4740  
 agcccctgcc ctgggaacca ggcagccttc cagacctcag gggctgaggc agactattag 4800  
 ggcagggctg actttgggta cactgcccac tccctctcag gccagctcag gtcaccggg 4860  
 cctctgacct aggcctgtca ctttgagagg ggcaaaactg agaggggctt ttcctagaga 4920  
 aagagaaca ggagcttgcc aggcctcatg tagccgacac acgtctcagg attttaagtc 4980  
 cacattggcc tcacactacc agggccaatg ccaaaaataa ggagttcaa tttggggcca 5040  
 aatgaggaag gacacagact ctgccctggg atctcctgtg ctagcggcca atgacaaatc 5100  
 cagtcattgg ccaccagcca cctctgcagt ggggaccaca ctagcagccc tgactccaca 5160  
 ctctcctgg ggacccaaga ggcagtgttg ctgtctgcat gtccacctg gaatctggct 5220  
 gaactggctg gcaggaccaa gactgcggct ggggtgggca ggaaggaa gccgggggct 5280  
 gctgtgaggg atcttgagc ttccctgtag cccaccttcc ccttgcttca tgtttgtaga 5340  
 ggaacctgt gccggcagg cccagtttcc ttgtgtgata cactaatgta tttgctttt 5400  
 ttgaaatag agaaaatcaa taaattgcta gtgttcttt gaacttt 5447

<210> 3  
 <211> 456  
 <212> PRT  
 <213> Homo sapiens

<400> 3

Met Thr Thr Thr Phe Leu Gln Thr Ser Ser Ser Thr Phe Gly Gly Gly

1

5

10

15



ES 2 482 468 B1

Ser Thr Arg Gly Gly Ser Leu Leu Ala Gly Gly Gly Gly Phe Gly Gly  
 20 25 30

Gly Ser Leu Ser Gly Gly Gly Gly Ser Arg Ser Ile Ser Ala Ser Ser  
 35 40 45

Ala Arg Phe Val Ser Ser Gly Ser Gly Gly Gly Tyr Gly Gly Gly Met  
 50 55 60

Arg Val Cys Gly Phe Gly Gly Gly Ala Gly Ser Val Phe Gly Gly Gly  
 65 70 75 80

Phe Gly Gly Gly Val Gly Gly Gly Phe Gly Gly Gly Phe Gly Gly Gly  
 85 90 95

Asp Gly Gly Leu Leu Ser Gly Asn Glu Lys Ile Thr Met Gln Asn Leu  
 100 105 110

Asn Asp Arg Leu Ala Ser Tyr Leu Asp Lys Val Arg Ala Leu Glu Glu  
 115 120 125

Ala Asn Ala Asp Leu Glu Val Lys Ile His Asp Trp Tyr Gln Lys Gln  
 130 135 140

Thr Pro Thr Ser Pro Glu Cys Asp Tyr Ser Gln Tyr Phe Lys Thr Ile  
 145 150 155 160

Glu Glu Leu Arg Asp Lys Ile Met Ala Thr Thr Ile Asp Asn Ser Arg  
 165 170 175

Val Ile Leu Glu Ile Asp Asn Ala Arg Leu Ala Ala Asp Asp Phe Arg  
 180 185 190

Leu Lys Tyr Glu Asn Glu Leu Ala Leu Arg Gln Gly Val Glu Ala Asp  
 195 200 205

Ile Asn Gly Leu Arg Arg Val Leu Asp Glu Leu Thr Leu Ala Arg Thr  
 210 215 220

Asp Leu Glu Met Gln Ile Glu Gly Leu Asn Glu Glu Leu Ala Tyr Leu  
 225 230 235 240

Lys Lys Asn His Glu Glu Glu Met Lys Glu Phe Ser Ser Gln Leu Ala  
 245 250 255

Gly Gln Val Asn Val Glu Met Asp Ala Ala Pro Gly Val Asp Leu Thr  
 260 265 270

Arg Val Leu Ala Glu Met Arg Glu Gln Tyr Glu Ala Met Ala Glu Lys  
 275 280 285

ES 2 482 468 B1

Asn Arg Arg Asp Val Glu Ala Trp Phe Phe Ser Lys Thr Glu Glu Leu  
 290 295 300

Asn Lys Glu Val Ala Ser Asn Thr Glu Met Ile Gln Thr Ser Lys Thr  
 305 310 315 320

Glu Ile Thr Asp Leu Arg Arg Thr Met Gln Glu Leu Glu Ile Glu Leu  
 325 330 335

Gln Ser Gln Leu Ser Met Lys Ala Gly Leu Glu Asn Ser Leu Ala Glu  
 340 345 350

Thr Glu Cys Arg Tyr Ala Thr Gln Leu Gln Gln Ile Gln Gly Leu Ile  
 355 360 365

Gly Gly Leu Glu Ala Gln Leu Ser Glu Leu Arg Cys Glu Met Glu Ala  
 370 375 380

Gln Asn Gln Glu Tyr Lys Met Leu Leu Asp Ile Lys Thr Arg Leu Glu  
 385 390 395 400

Gln Glu Ile Ala Thr Tyr Arg Ser Leu Leu Glu Gly Gln Asp Ala Lys  
 405 410 415

Met Ala Gly Ile Gly Ile Arg Glu Ala Ser Ser Gly Gly Gly Gly Ser  
 420 425 430

Ser Ser Asn Phe His Ile Asn Val Glu Glu Ser Val Asp Gly Gln Val  
 435 440 445

Val Ser Ser His Lys Arg Glu Ile  
 450 455

<210> 4  
 <211> 1861  
 <212> DNA  
 <213> Homo sapiens

<400> 4  
 cactcaaggt gtgcaggcag ctgtgtttgt caggaaggca gaaggagttg gctttgcttt 60  
 aggggaggag acgaggtccc acaacaccct ctgaagggta tataaggagc cccagcgtgc 120  
 agcctggcct ggtacctcct gccagcatct cttgggtttg ctgagaactc acgggctcca 180  
 gctacctggc catgaccacc acatttctgc aaacttcttc ctccaccttt ggggggtggc 240  
 caaccgagg gggttccctc ctggctgggg gaggtggcct tgggtggggg agtctctctg 300  
 ggggaggtgg aagccgaagt atctcagctt cttctgctag gtttgtctct tcagggtcag 360  
 gaggaggata tgggggtggc atgagggctt gtggcctttg tggaggggct ggtagtgttt 420  
 tcggtggagg ctttggaggg ggcgttggtg ggggttttgg tgggtggcttt ggtggtggcg 480  
 atggtggtct cctctctggc aatgagaaaa ttaccatoca aaacctcaat gaccgcctgg 540

ES 2 482 468 B1

cctcctacct ggacaaggta cgtgccctgg aggaggccaa tgctgacctg gaggtgaaga 600  
 tccatgactg gtaccagaag cagaccccaa ccagcccaga atgcgactac agccaatact 660  
 tcaagacat tgaagagctc cgggacaaga tcatggccac caccatcgac aactcccggg 720  
 tcatcctgga gatcgacaat gccaggctgg ctgctggacga cttcaggctc aagtatgaga 780  
 atgagctggc cctgcgccag ggcgttgagg ctgacatcaa cggcttgccg cgagtcctgg 840  
 atgagctgac cctggccagg actgacctgg agatgcagat cgagggcctg aatgaggagc 900  
 tagcctacct gaagaagaac cacgaagagg agatgaagga gttcagcagc cagctggccg 960  
 gccagggtcaa tgtggagatg gacgcagcac cgggtgtgga cctgaccctg gtgctggcag 1020  
 agatgagggg gcagtacgag gccatggcgg agaagaaccg ccgggatgtc gaggcctggt 1080  
 tcttcagcaa gactgaggag ctgaacaaag aggtggcctc caacacagaa atgatccaga 1140  
 ccagcaagac ggagatcaca gacctgagac gcacgatgca ggagctggag atcgagctgc 1200  
 agtcccagct cagcatgaaa gctgggctgg agaactcact ggccgagaca gagtgccgct 1260  
 atgccacgca gctgcagcag atccaggggc tcattggtgg cctggaggcc cagctgagtg 1320  
 agtcccgatg cgagatggag gctcagaacc aggagtacaa gatgctgctt gacataaaga 1380  
 cacggctgga gcaggagatc gctacttacc gcagcctgct cgagggccag gatgccaaga 1440  
 tggctggcat tggcatcagg gaagcctctt caggaggtgg tggtagcagc agcaatttcc 1500  
 acatcaatgt agaagagtca gtggatggac aggtggtttc ttcccacaag agagaaatct 1560  
 aagtgtctat tgcaggagaa acgtcccttg ccaactcccca ctctcatcag gccaagtgga 1620  
 ggactggcca gagggcctgc acatgcaaac tccagtcctt gccttcagag agctgaaaag 1680  
 ggtccctcgg tcttttattt cagggtttg catgcgctct attccccctc tgcctctccc 1740  
 caccttcttt ggagcaagga gatgcagctg tattgtgtaa caagctcatt tgtacagtg 1800  
 ctgttcatgt aataaagaat tacttttctt ttgcaaata aaaaaaaaaa aaaaaaaaaa 1860  
 a 1861

<210> 5  
 <211> 999  
 <212> PRT  
 <213> Homo sapiens

<400> 5

Met Met Gly Leu Phe Pro Arg Thr Thr Gly Ala Leu Ala Ile Phe Val  
 1 5 10 15

Val Val Ile Leu Val His Gly Glu Leu Arg Ile Glu Thr Lys Gly Gln  
 20 25 30

Tyr Asp Glu Glu Glu Met Thr Met Gln Gln Ala Lys Arg Arg Gln Lys  
 35 40 45

Arg Glu Trp Val Lys Phe Ala Lys Pro Cys Arg Glu Gly Glu Asp Asn





ES 2 482 468 B1

595					600					605					
Gly	Arg	Pro	His	Ser	Gly	Arg	Leu	Gly	Pro	Ala	Ala	Ile	Gly	Leu	Leu
	610					615					620				
Leu	Leu	Gly	Leu	Leu	Leu	Leu	Leu	Leu	Ala	Pro	Leu	Leu	Leu	Leu	Thr
625					630					635					640
Cys	Asp	Cys	Gly	Ala	Gly	Ser	Thr	Gly	Gly	Val	Thr	Gly	Gly	Phe	Ile
				645					650					655	
Pro	Val	Pro	Asp	Gly	Ser	Glu	Gly	Thr	Ile	His	Gln	Trp	Gly	Ile	Glu
			660					665					670		
Gly	Ala	His	Pro	Glu	Asp	Lys	Glu	Ile	Thr	Asn	Ile	Cys	Val	Pro	Pro
		675					680					685			
Val	Thr	Ala	Asn	Gly	Ala	Asp	Phe	Met	Glu	Ser	Ser	Glu	Val	Cys	Thr
	690					695					700				
Asn	Thr	Tyr	Ala	Arg	Gly	Thr	Ala	Val	Glu	Gly	Thr	Ser	Gly	Met	Glu
705					710					715					720
Met	Thr	Thr	Lys	Leu	Gly	Ala	Ala	Thr	Glu	Ser	Gly	Gly	Ala	Ala	Gly
				725					730					735	
Phe	Ala	Thr	Gly	Thr	Val	Ser	Gly	Ala	Ala	Ser	Gly	Phe	Gly	Ala	Ala
			740					745					750		
Thr	Gly	Val	Gly	Ile	Cys	Ser	Ser	Gly	Gln	Ser	Gly	Thr	Met	Arg	Thr
		755					760					765			
Arg	His	Ser	Thr	Gly	Gly	Thr	Asn	Lys	Asp	Tyr	Ala	Asp	Gly	Ala	Ile
	770					775					780				
Ser	Met	Asn	Phe	Leu	Asp	Ser	Tyr	Phe	Ser	Gln	Lys	Ala	Phe	Ala	Cys
785					790					795					800
Ala	Glu	Glu	Asp	Asp	Gly	Gln	Glu	Ala	Asn	Asp	Cys	Leu	Leu	Ile	Tyr
				805					810					815	
Asp	Asn	Glu	Gly	Ala	Asp	Ala	Thr	Gly	Ser	Pro	Val	Gly	Ser	Val	Gly
			820					825					830		
Cys	Cys	Ser	Phe	Ile	Ala	Asp	Asp	Leu	Asp	Asp	Ser	Phe	Leu	Asp	Ser
		835					840					845			
Leu	Gly	Pro	Lys	Phe	Lys	Lys	Leu	Ala	Glu	Ile	Ser	Leu	Gly	Val	Asp
	850					855					860				
Gly	Glu	Gly	Lys	Glu	Val	Gln	Pro	Pro	Ser	Lys	Asp	Ser	Gly	Tyr	Gly



ES 2 482 468 B1

agaaggacta tcaactcaat gtgaatgtaa tattaagtg aaagatgtca acgataactt 900  
 cccaatgttt agagactctc agtattcagc acgtattgaa gaaaatattt taagttctga 960  
 attacttcga tttcaagtaa cagatttgga tgaagagtac acagataatt ggcttgca 1020  
 atatttcttt acctctggga atgaaggaaa ttggtttgaa atacaaactg atcctagaac 1080  
 taatgaaggc atcctgaaag tggagaaggc tctagattat gaacaactac aaagcgtgaa 1140  
 acttagtatt gctgtcaaaa acaaagctga atttcaccaa tcagttatct ctcgataccg 1200  
 agttcagtca accccagtca caattcaggt aataaatgta agagaaggaa ttgcattccg 1260  
 tcctgcttcc aagacattta ctgtgcaaaa aggcataagt agcaaaaaat tgggtgatta 1320  
 ttcctggga acatatcaag ccatcgaatg ggacactaac aaagctgcct caaatgtcaa 1380  
 atatgtcatg ggacgtaacg atgggtgata cctaattgatt gattcaaaaa ctgctgaaat 1440  
 caaatttgtc aaaaatatga accgagattc tactttcata gttaacaaaa caatcacagc 1500  
 tgaggttctg gccatagatg aatacacggg taaaacttct acaggcacgg tatatgttag 1560  
 agtaccgat ttcaatgaca attgtccaac agctgtcctc gaaaaagatg cagtttgca 1620  
 ttcttcacct tccgtggtg tctccgctag aacctgaat aatagataca ctggccccta 1680  
 tacatttgca ctggaagatc aacctgtaaa gttgcctgcc gtatggagta tcacaaccct 1740  
 caatgctacc tcggccctcc tcagagccca ggaacagata cctcctggag tataccacat 1800  
 ctccctggta cttacagaca gtcagaacaa tcggtgtgag atgccacgca gcttgacact 1860  
 ggaagtctgt cagtgtgaca acaggggcat ctgtggaact tcttaccaa ccacaagccc 1920  
 tgggaccagg tatggcaggc cgcactcagg gaggtgggg cctgccgcca tcggcctgct 1980  
 gctccttggg ctctgctgc tgctgttggc cccccttctg ctgttgacct gtgactgtgg 2040  
 ggcaggttct actgggggag tgacaggtgg ttttatcca gttcctgatg gtcagaagg 2100  
 aacaattcat cagtgggaa ttgaaggagc ccatcctgaa gacaaggaaa tcacaaatat 2160  
 ttgtgtgcct cctgtaacag ccaatggagc cgatttcatg gaaagtctg aagtttgtac 2220  
 aaatacgtat gccagaggca cagcgggtgga aggcacttca ggaatggaaa tgaccactaa 2280  
 gcttgagca gccactgaat ctggaggtgc tgcaggcttt gcaacagga cagtgtcagg 2340  
 agctgcttca ggattcggag cagccactgg agttggcatc tgttcctcag ggcagtctgg 2400  
 aacctgaga acaaggcatt cactggagg aaccaataag gactacgctg atggggcgat 2460  
 aagcatgaat tttctggact cctacttttc tcagaaagca tttgcctgtg cggaggaaga 2520  
 cgatggccag gaagcaaatg actgcttgtt gatctatgat aatgaaggcg cagatgccac 2580  
 tggttctcct gtgggctccg tgggttgtg cagttttatt gctgatgacc tggatgacag 2640  
 cttcttgac tcaactggac ccaaatttaa aaaacttgca gagataagcc ttggtgttga 2700  
 tggagaaggc aaagaagttc agccaccctc taaagacagc ggttatggga ttgaatcctg 2760  
 tggccatccc atagaagtcc agcagacagg atttgttaag tgccagactt tgcaggaag 2820  
 tcaaggagct tctgcttgt ccacctctgg gtctgtccag ccagctgttt ccatccctga 2880



ES 2 482 468 B1

ccctctgcag catggtaact atttagtaac ggagacttac tcggcttctg gttccctcgt 2940  
 gcaaccttcc actgcaggct ttgatccact tctcacacaa aatgtgatag tgacagaaag 3000  
 ggtgatctgt cccatttcca gtgttcctgg caacctagct ggcccaacgc agctacgagg 3060  
 gtcacatact atgctctgta cagaggatcc ttgctcccgt ctaatatgac cagaatgagc 3120  
 tggaatacca cactgaccaa atctggatct ttggactaaa gtattcaaaa tagcatagca 3180  
 aagctcactg tattgggcta ataatttggc acttatttagc ttctctcata aactgatcac 3240  
 gattataaat taaatgtttg ggttcatacc ccaaaagcaa tatgtttgca ctcttaattc 3300  
 tcaagtacta ttcaaattgt agtaaactctt aaagtttttc aaaaccctaa aatcatattc 3360  
 gccaggaaat tttcctaaac attcttaagc ttctatTTTT cccctgccaa aggaagggtg 3420  
 ttatcatttt aaaatgcaat gtgatttagt ggattaagca ggagcgctgg ttcttgtctc 3480  
 cattgccttt tcttatatca ttgataatga tgtaagaatc acaaggggccc gggcgcggtg 3540  
 gctcacgcct gtaatcccag cactttggga ggccgaggca ggtggatcat gaggtcagga 3600  
 gatcgagacc atcctggcta acaagggtgaa accccgtctc tactaaaaat acaaaaaatt 3660  
 agccggggcg agtggcgggc gcctgtagtc ccagctactc gggaggctga ggcaggagaa 3720  
 tggcatgaac ccggaagcg gagcttgcaag tgagccgaga ttgcgccact gcagtccgca 3780  
 gtccggcctg ggcgacagag cgagactccg tctcaaaaaa aaaaaaaaaa aaagaatcac 3840  
 aaggatattg ctaaagcatt ttgagctgct tggaaaaagg gaagtagttg cagtagagtt 3900  
 tcttccatct tcttgggtgct gggaaagccat atatgtgtct tttactcaag ctaaggggta 3960  
 taagcttatg tgttgaattt gctacatcta ttttccatc attctcaca taagagaatt 4020  
 ttgaaataga aatatcatag aacatttaag aaagtttagt ataaataata ttttgtgtgt 4080  
 tttaatccct ttgaagggat ctatccaaag aaaatatTTT acactgagct ccttccata 4140  
 cgtctcagta acagatcctg tgtagtctt tgaaaatagc tcatttttta aatgtcagtg 4200  
 agtagatgta gcatacatat gatgtataat gacgtgtatt atgttaacaa tgtctgcaga 4260  
 ttttgtagga atacaaaaca tggcctTTTT tataagcaa acgggccaat gactagaata 4320  
 acacataggg caatctgtga atatgtatta taagcagcat tccagaaaag tagttgggtga 4380  
 aataattttc aagtcaaaaa gggatatgga aagggaatta tgagtaacct ctatttttta 4440  
 agccttgctt ttaaattaaa cagctacagc catttaagcc ttgaggataa taaagcttga 4500  
 gagtaataat gttaggttag caaaggTTT gatgtatcac ttcatgcatg ctacatgat 4560  
 agtaatgcag ctcttcgagt catttctggc cattcaagat attcaccctt ttgccatag 4620  
 aaagcacct acctcacctg cttactgaca ttgtcttagc tgatcacaag atcattatca 4680  
 gcctcatta ttcttactg tatataaaat acagagTTTT atattttcct ttcttcgttt 4740  
 ttcacatat tcaaaaccta aatttgtttt tgcagatgga atgcaaagta atcaagtgtt 4800  
 tgtgctttca cctagaaggg tgtggctcctg aaggaaagag gtcccctaaa tatccccac 4860  
 cctgggtgctc ctccctctcc ctggtaccct gactaccagg aagtcagggtg ctagagcagc 4920

ES 2 482 468 B1

tggagaagtg caggcagcct gtgcttccac agatgggggt gctgctgcaa caaggctttc	4980
aatgtgccca tcttaggtgg gagaagctag atcctgtgca gcagcctggt aagtcctgag	5040
gaggttccat tgctcttcct gctgctgtcc tttgcttctc aacggtggct cgctctacag	5100
tctagagcac atgcagctaa cttgtgcctc tgcttatgca tgaggggtaa attaacaacc	5160
ataaccttca tttgaagttc aaaggtgtat tcaggatcct caaagcattt taaccttgcc	5220
gcttaaaacc caatttaccg tgaaatggga attttgctgc attgttaaac tgtagtggaa	5280
accatgctat agtaataaag gttatataag agagaaattg aaattaaatg tgtttttaaa	5340
tttcaaaaaa aatcaatct ttaggatgac ttaaaaattg atttgccatg taaaatgtat	5400
ctgcattttt tacacaaaac ttgttttaag cataaaattt taaaactgta ctacttgatg	5460
tattatacat tttgaacat atgtattaaa ccataaacag tataatgttg ttataataaa	5520
acaggcaata aatttataaa taaaagctga aaaaaaaaaa a	5561