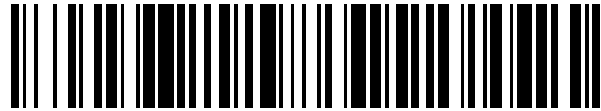


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 486 340**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.10.2001 E 01989310 (6)**

97 Fecha y número de publicación de la concesión europea: **18.06.2014 EP 1404712**

54 Título: **Ácidos nucleicos y proteínas Nod2**

30 Prioridad:

30.10.2000 US 244266 P
25.04.2001 US 286316 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.08.2014

73 Titular/es:

THE REGENTS OF THE UNIVERSITY OF MICHIGAN (50.0%)
3003 South State Street
Ann Arbor, Michigan 48109-1274, US y
THE UNIVERSITY OF CHICAGO (50.0%)

72 Inventor/es:

NUNEZ, GABRIEL;
INOHARA, NAHIRO;
OGURA, YASUNORI;
CHO, JUDY;
NICOLAE, DAN L. y
BONEN, DENISE

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 486 340 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

Ácidos nucleicos y proteínas Nod2

Campo de la invención

5 La presente invención está definida en las reivindicaciones y se refiere a moléculas de señalización intracelular, en concreto a variantes de la proteína Nod2 y a los ácidos nucleicos que codifican las variantes de la proteína Nod2. La presente divulgación proporciona ensayos para la detección de Nod2 y polimorfismos de Nod2 asociados con estados de enfermedad. La presente divulgación proporciona además inhibidores de la señalización de Nod2 y procedimientos de identificación de los componentes de la vía de Nod2. La presente invención proporciona
10 adicionalmente procedimientos de determinación del riesgo que tiene un individuo de desarrollar estados de enfermedad.

Antecedentes de la invención

15 Las enfermedades inflamatorias intestinales (EII) se definen por una inflamación intestinal crónica y recidivante de origen oculto. La EII se refiere a dos trastornos distintos, la enfermedad de Crohn y la colitis ulcerosa (CU). En ambas enfermedades parece que participa una respuesta inmunitaria mal regulada a los antígenos del tracto GI, una brecha en la barrera mucosa y/o una reacción inflamatoria adversa a una infección intestinal persistente. Los contenidos luminales del tracto GI y las bacterias estimulan de forma constante el sistema inmunitario de la mucosa y un delicado equilibrio de las células y moléculas proinflamatorias y antiinflamatorias mantiene la integridad del tracto GI sin provocar una inflamación grave y dañina. Se desconoce cómo comienza la cascada inflamatoria de la EII, pero la estimulación constante dependiente de antígenos GI de los sistemas inmunitarios de la mucosa y
20 sistémicos perpetúa la cascada inflamatoria y dirige la formación de lesiones.

No existe una cura conocida para la EII, que afecta a 2 millones de americanos. Se ha estimado que los costes de los procedimientos actuales del tratamiento de los síntomas de EII son de 1,2 miles de millones de \$ anuales solo en EE.UU.

25 En pacientes con EII, las úlceras y la inflamación del revestimiento interno del intestino conducen a síntomas de dolor abdominal, diarrea y sangrado rectal. La colitis ulcerosa se produce en el intestino grueso, mientras que en la enfermedad de Crohn, la enfermedad puede afectar a todo el tracto GI así como a los intestinos delgado y grueso. Para la mayoría de los pacientes la EII es una afección crónica con síntomas que duran de meses a años. Es más habitual en adultos jóvenes, pero se puede producir a cualquier edad. Se encuentra en todo el mundo, pero es más frecuente en los países industrializados como EE.UU., Inglaterra y el norte de Europa. Es especialmente frecuente
30 en personas de ascendencia judía y también existen diferencias por razas en cuanto a la incidencia. Los síntomas clínicos de la EII son sangrado rectal intermitente, dolor abdominal de tipo calambres, pérdida de peso y diarrea. El diagnóstico de la EII se basa en los síntomas clínicos, el uso de un enema de bario, pero la prueba más precisa es la visualización directa (sigmoidoscopia o colonoscopia). La protracción de EII es un factor de riesgo del cáncer de colon. El riesgo de cáncer comienza a elevarse significativamente tras de ocho a diez años de EII.

35 Algunos pacientes con CU solo sufren la enfermedad en el recto (proctitis). Otros con CU tienen la enfermedad limitada al recto y al colon adyacente izquierdo (proctosigmoiditis). Aunque otros tienen CU en todo el colon (EII universal). En general, los síntomas de CU son más graves cuanto más extensa es la enfermedad (una mayor parte del colon está afectada por la enfermedad).

40 El pronóstico de los pacientes con la enfermedad limitada al recto (proctitis) o con CU limitada al final del colon izquierdo (proctosigmoiditis) es mejor que cuando la CU afecta a todo el colon. Puede ser suficiente con tratamientos periódicos breves con medicamentos orales o enemas. En aquellos con la enfermedad más extensa, la pérdida de sangre de los intestinos inflamados puede producir anemia y puede requerir tratamiento con suplementos de hierro o incluso con transfusiones de sangre. Rara vez el colon se dilata de forma aguda a un tamaño grande cuando la transformación pasa a ser muy intensa. Esta afección se llama megacolon tóxico. Los pacientes con
45 megacolon tóxico están extremadamente enfermos con fiebre, dolor y distensión abdominal, deshidratación y malnutrición. A menos que el paciente mejore rápidamente con la medicación, habitualmente se necesita cirugía para evitar la rotura del colon.

50 La enfermedad de Crohn se puede producir en todas las regiones del tracto gastrointestinal. Con esta enfermedad se produce obstrucción intestinal por inflamación y fibrosis en un gran número de pacientes. Los granulomas y la formación de fístulas son complicaciones frecuentes de la enfermedad de Crohn. Las consecuencias de la progresión de la enfermedad incluyen alimentación intravenosa, cirugía y colostomía.

55 Los medicamentos más usados para tratar la EII son fármacos antiinflamatorios como los salicilatos. Las preparaciones con salicilato han sido eficaces en el tratamiento de la enfermedad leve a moderada. También pueden disminuir la frecuencia de los brotes de la enfermedad cuando los medicamentos se toman de forma prolongada. Ejemplos de salicilatos incluyen sulfasalazina, olsalazina y mesalamina. Todos estos medicamentos se administran por vía oral a dosis elevadas para obtener un beneficio terapéutico máximo. Estos medicamentos no carecen de efectos secundarios. La azulfidina puede producir malestar de estómago cuando se toma a dosis elevadas y, en

casos raros de inflamación renal leve, se han notificado con algunas preparaciones de salicilatos.

Los corticosteroides son más potentes y de acción más rápida que los salicilatos en el tratamiento de la EII, pero los efectos secundarios potencialmente graves limitan el uso de corticosteroides a los pacientes con enfermedad más grave. Los efectos secundarios de los corticosteroides suelen producirse con el uso prolongado. Incluyen adelgazamiento del hueso y la piel, infecciones, diabetes, emaciación muscular, redondeado de la cara, alteraciones psiquiátricas y, rara vez, destrucción de las articulaciones de las caderas.

En los pacientes de EII que no responden a los salicilatos o a los corticosteroides se usan medicamentos que suprimen el sistema inmunitario. Ejemplos de inmunosupresores incluyen azatioprina y 6-mercaptopurina. Los inmunosupresores usados en esta situación ayudan a controlar la EII y permiten la reducción gradual o la eliminación de los corticosteroides. No obstante, los inmunosupresores aumentan el riesgo de infección, insuficiencia renal y la necesidad de hospitalización.

Claramente existe una gran necesidad de identificar la base molecular de la EII o de sus trastornos asociados, enfermedad de Crohn y colitis ulcerosa. El documento WO-A-01/72822 divulga la asociación entre polimorfismos en el gen IBD1 y la enfermedad de Crohn.

15 **Sumario de la invención**

La presente invención está definida en las reivindicaciones y se refiere a moléculas de señalización intracelular, en concreto a variantes de la proteína Nod2 y a los ácidos nucleicos que codifican las variantes de la proteína Nod2. La presente divulgación proporciona ensayos para la detección de Nod2 y polimorfismos de Nod2 asociados con estados de enfermedad. La presente divulgación proporciona además inhibidores de la señalización de Nod2 y procedimientos de identificación de los componentes de la vía de Nod2.

La presente divulgación proporciona un ácido nucleico aislado y purificado que comprende una secuencia que codifica una proteína seleccionada del grupo que consiste en las SEC ID N°: 2, 3 y 34. En algunas formas de realización, la secuencia de ácidos nucleicos está unida operablemente a un promotor heterólogo. En algunas formas de realización, la secuencia de ácido nucleico está contenida dentro de un vector. En algunas formas de realización, el vector está dentro de una célula huésped.

En otras realizaciones, la presente invención proporciona una secuencia de ácido nucleico aislado y purificado tal como se reivindica, que hibrida en condiciones de baja rigurosidad con un ácido nucleico seleccionado del grupo que consiste en la SEC ID N°: 33. En algunas formas de realización, la secuencia de ácido nucleico codifica una proteína que activa NF- κ B. En otras formas de realización, la presente invención proporciona un vector que comprende la secuencia de ácido nucleico reivindicada. En otras formas de realización más, el vector está dentro de una célula huésped. En algunas formas de realización, la célula huésped está localizada en un organismo seleccionado del grupo que consiste en una planta y un animal no humano.

La presente divulgación proporciona una proteína codificada por un ácido nucleico seleccionado del grupo que consiste en las SEC ID N° 1, 33 y variantes de la misma que tienen una identidad de al menos el 80 % con las SEC ID N° 1 y 33, y en la que la proteína tiene al menos una actividad de Nod2. La actividad puede ser activación de NF- κ B. La actividad puede ser unión a RICK. La proteína puede tener una identidad de al menos 90 % con las SEC ID N° 1 y 33. La proteína puede tener una identidad de al menos 95% con las SEC ID N° 1 y 33.

La presente divulgación proporciona un procedimiento para producir variantes de Nod2 que comprenden: proporcionar una secuencia de aminoácidos seleccionada del grupo que consiste en las SEC ID N° 1 y 33, mutageneizar la secuencia de ácido nucleico y seleccionar la variante según la actividad Nod2.

La presente divulgación proporciona un ácido nucleico que codifica Nod2, en el que Nod2 compete por la unión a RICK con una proteína codificada por una secuencia de ácido nucleico seleccionada del grupo que consiste en las SEC ID N° 1 y 33.

La presente divulgación proporciona una composición que comprende un ácido nucleico que inhibe la unión de al menos una porción de un ácido nucleico seleccionado del grupo que consiste en las SEC ID N° 1 y 33 a sus secuencias complementarias. La presente divulgación proporciona una secuencia de polinucleótidos que comprende al menos quince nucleótidos capaces de hibridar en condiciones rigurosas con la secuencia nucleotídica aislada seleccionada del grupo que consiste en las SEC ID N° 1 y 33.

La presente invención es útil un procedimiento para la detección de un polinucleótido que codifica la proteína Nod2 en una muestra biológica que se sospecha que contiene un polinucleótido que codifica Nod2. El procedimiento incluye hibridar la secuencia polinucleotídica seleccionada del grupo que consiste en las SEC ID N° 1 y 33 y variantes de la misma que tienen una identidad de al menos el 80 % con las SEC ID N° 1 y 33 (y en la que la proteína tiene al menos una actividad de Nod2) con el ácido nucleico de la muestra biológica para producir un complejo de hibridación. El procedimiento puede incluir adicionalmente la etapa de detectar el complejo de hibridación, en el que la presencia del complejo de hibridación indica la presencia de un polinucleótido que codifica Nod2 en la muestra biológica. Antes de la hibridación, el ácido nucleico de la muestra biológica se puede amplificar.

La presente invención es útil en un procedimiento para seleccionar compuestos por su capacidad para alterar la actividad Nod2, que comprende: proporcionar: una primera secuencia polipeptídica que comprende al menos una porción de Nod2; ii) una segunda secuencia polipeptídica que comprende al menos una porción de una proteína que se sabe que interacciona con Nod2; y iii) uno o más compuestos de ensayo, combinando en cualquier orden, la primera secuencia polipeptídica que comprende al menos una porción de Nod2, la segunda secuencia polipeptídica que comprende al menos una porción de una proteína que se sabe que interacciona con Nod2, y uno o más compuestos de ensayo en condiciones tales que la primera secuencia polipeptídica, la segunda secuencia polipeptídica y el compuesto de ensayo interaccionan; y detectar la presencia o ausencia de una interacción entre la secuencia polipeptídica que comprende al menos una porción de Nod2 y la secuencia polipeptídica que comprende al menos una porción de una proteína que se sabe que interacciona con Nod2. La primera secuencia polipeptídica se puede seleccionar del grupo que consiste en las SEC ID N° 2-17 y 34. La segunda secuencia polipeptídica puede comprender RICK.

La presente divulgación proporciona un polipéptido purificado seleccionado del grupo que consiste en las SEC ID N° 2, 3 y 34.

La presente divulgación proporciona adicionalmente un compuesto capaz de inhibir la unión de un Nod2 a un polipéptido RICK.

La presente divulgación proporciona un procedimiento de identificación de sujetos en riesgo de desarrollar enfermedad de Crohn, que comprende: proporcionar ácido nucleico de un sujeto, en el que dicho ácido nucleico comprende un gen de Nod2 y detectar la presencia o ausencia de una o más variaciones en el gen de Nod2. El procedimiento se refiere adicionalmente a la determinación de si el sujeto está en riesgo de desarrollar enfermedad de Crohn en base a la presencia o ausencia de la una o más variaciones. Por ejemplo, la determinación comprende determinar un riesgo relativo al genotipo para dicho sujeto o la determinación comprende un riesgo atribuible a la población para dicho sujeto. La variación puede ser una mutación. La variación puede ser un polimorfismo. La mutación puede ser una inserción de un residuo de citosina. La mutación puede producir una delección de al menos una repetición LRR de Nod2. La presente invención es útil en la detección de una variación específica. Cualquier variación se puede detectar, incluidas, entre otras, las seleccionadas del grupo que consiste en las secuencias de ácido nucleico descritas por las SEC ID N° 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86, and 88. En algunas formas de realización, la etapa de detección se consigue mediante análisis de hibridación. En algunas formas de realización, la etapa de detección comprende comparar la secuencia del ácido nucleico con la secuencia del ácido nucleico de Nod2 de tipo silvestre.

La presente divulgación se refiere adicionalmente a un kit para determinar si un sujeto está en riesgo de desarrollar la enfermedad de Crohn, que comprende: Un ensayo de detección capaz de detectar específicamente un alelo de Nod2 variante; e instrucciones para determinar si el sujeto está en mayor riesgo de desarrollar la enfermedad de Crohn, en el que el ensayo de detección puede comprender una sonda de ácido nucleico que hibrida en condiciones rigurosas con una secuencia de ácido nucleico seleccionada del grupo que consiste en las SEC ID N° 70-83.

La presente divulgación proporciona un ácido nucleico aislado que comprende una secuencia que codifica un polipéptido seleccionado del grupo que consiste en las SEC ID N°: 55, 57, 59, 61, 63, 65, 67, 69, 85, 87 y 89. En algunas formas de realización, la secuencia está unida operablemente a un promotor heterólogo. En algunas formas de realización, la secuencia está contenida dentro de un vector. En algunas formas de realización, el vector está contenido en una célula huésped. En algunas formas de realización, la célula huésped está localizada en un organismo seleccionado del grupo que consiste en una planta y un animal.

En otras formas de realización más, la presente divulgación proporciona una secuencia de ácido nucleico aislado seleccionado del grupo que consiste en las SEC ID N° 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88. Un medio legible por ordenador puede codificar una representación de la secuencia de ácido nucleico.

En otras formas de realización, la presente divulgación proporciona un polipéptido seleccionado del grupo que consiste en las SEC ID N° 55, 57, 59, 61, 63, 65, 67, 69, 85, 87 y 89.

Un medio legible por ordenador puede codificar una representación del polipéptido.

La presente invención también proporciona un procedimiento implementado por ordenador de determinar el riesgo de un paciente de desarrollar enfermedad de Crohn que comprende proporcionar ácido nucleico de un paciente, en el que el ácido nucleico comprende un gen Nod2; y un ordenador que comprende software para la predicción del riesgo de un paciente de desarrollar enfermedad de Crohn; y detectar la presencia de una o más variaciones en el gen Nod2 del paciente para generar información sobre la variación genética,

en el que dichas variaciones se seleccionan del grupo que consiste en Nod2 3020Insc, Nod2 c802t + 3020Insc, Nod2 c2104t, Nod2 g2722c, Nod2 c802t + g2722c, and Nod2 c802t + c2104t; introducir la información sobre la variación genética en dicho ordenador y calcular el riesgo del paciente con el software,

en el que dicho cálculo es indicativo de que dicho sujeto puede estar en riesgo de desarrollar enfermedad de Crohn. En algunas formas de realización, el procedimiento comprende además la etapa de mostrar el riesgo del paciente.

En algunas formas de realización, el riesgo comprende un riesgo relativo al genotipo. En otras formas de realización, el riesgo comprende un riesgo atribuible a la población. En otras formas de realización, la determinación comprende determinar un riesgo atribuible a la población para dicho sujeto. En algunas formas de realización, la variación es una mutación. En otras formas de realización, la variación es un polimorfismo. En algunas formas de realización, la mutación es una inserción de un residuo de citosina. En algunas formas de realización, la mutación produce una delección en al menos una repetición LRR de Nod2.

En algunas formas de realización, la etapa de detección comprende comparar la secuencia del ácido nucleico variante con la secuencia del ácido nucleico de Nod2 de tipo silvestre.

Descripción de las figuras

La Figura 1 muestra la secuencia de aminoácidos deducida y la estructura del dominio de Nod2 humana. La Figura 1A muestra la secuencia de aminoácidos de Nod2 (SEC ID N° 4). Los dominios de reclutamiento de la caspasa (CARD 1 Y 2; SEC ID N° 5 y 6), el dominio de unión al nucleótido Nod2. La Figura 1A muestra la secuencia de aminoácidos de Nod2 (SEC ID N° 4). Los dominios de reclutamiento de la caspasa (CARD 1 Y 2; SEC ID N° 5 y 6), el dominio de unión al nucleótido (NBD, SEC ID N° 7) y repeticiones ricas en leucina (LRRS; SEC ID N° 8-17) se indican mediante resaltado inverso, subrayado y flechas, respectivamente. La secuencia consenso del bucle P (caja Walker A; SEC ID N° 18) y el sitio de unión a Mg²⁺ (caja Walker B; SEC ID N° 19) se indican con recuadros. La Figura 1B muestra la estructura del dominio de Nod2. Los números corresponden a los residuos de aminoácidos del panel A. la región homóloga a los CARD, NBD y LRR se indican con recuadros cerrados negros, cerrados oscuros y tramados, respectivamente.

La Figura 2 muestra una alineación de la Nod2 humana y proteínas relacionadas. La Figura 2A muestra una alineación de CARD de Nod2 (SEC ID N° 5 y 6), Nod1 (número de acceso en GeneBank AF113925; SEC ID N° 20), RICK (AF027706; SEC ID N°: 21), ARC (AF043244; SEC ID N° 22), RAIDD (U79115; SEC ID N° 23), Caspasa-2 (U13021; SEC ID N° 24), Ced-3 (L29052; SEC ID N° 25), Ced-4 (X69016; SEC ID N° 26), Caspasa-9 (U56390; SEC ID N° 27), Apaf-I (AF013263; SEC ID N° 28) y c-IAP-1 (L49431; SEC ID N° 29). Los residuos hidrófobos se muestran en resaltado inverso. Los residuos con carga positiva y negativa se señalan en color gris claro y oscuro, respectivamente. Los residuos de prolina y glicina ((fragmentador αβ) están en negrita. Las supuestas α hélices, H1 a H5, se muestran de acuerdo con la estructura tridimensional de los CARD de RAIDD (Chou y col., Cell, 94:171 [1998]). La Figura 2B muestra una alineación de los NBD de Nod2 (SEC ID N° 7), Nod1 (SEC ID N° 30), Apaf-I (SEC ID N° 31) y Ced-4 (SEC ID N°: 32). Los residuos idénticos y similares a los de Nod2 se muestran mediante resaltado inverso y oscuro, respectivamente. La secuencia consenso del bucle P (Walker A box) y el sitio de unión a Mg²⁺ (caja Walker B) se indican con recuadros. Los residuos idénticos y similares a los de Nod2 se muestran mediante resaltado inverso y oscuro, respectivamente. La Figura 2CA muestra una alineación de las LRR de Nod2 (SEC ID N° 8-17). Las posiciones conservadas con leucina y otros residuos hidrofóbicos se indican en resaltado de color gris oscuro y claro, respectivamente. Las supuestas α hélice y lámina β se muestran de acuerdo con la estructura tridimensional del inhibidor de ribonucleasa (Kobe y Deisenhofer, Curr. Opin. Struct Biol., 5:409-416 [1995]).

La Figura 3 muestra un análisis de la expresión de Nod2. La Figure 3A muestra un análisis de transferencia de tipo Northern de la expresión de nod2 en tejidos humanos; PBL (leucocitos de sangre periférica). La Figure 3B muestra un análisis de RT-PCR de la expresión de nod2 en poblaciones enriquecidas de granulocitos, monocitos y linfocitos. Para amplificar las secuencias de nucleótidos de los CARD y las LRR se usaron dos conjuntos de cebadores oligonucleotídicos de Nod2 (P1-P2 y P3-P4), respectivamente. La Figura 3C muestra la secuencia de nucleótidos de la región 5' de nod2. Dos potenciales sitios de inicio de la traducción dentro del marco separados por 81 nucleótidos se indican con flechas. La Figura 3D muestra inmunotransferencia de los productos del gen *nod2* expresados en células HEK293T. Las células se transfectaron con el plásmido control (carril 1) o construcciones que contienen potenciales sitios iniciadores de la traducción de Nod2 (carril 2) o como control el segundo sitio de inicio de la traducción correspondiente al de Nod2b (carril 3) o el sitio de inicio de la traducción más en el extremo NH₂ (carril 4) en el contexto del motivo de inicio de la traducción canónico. En todos los casos se expresó una proteína Nod2 que carece de los residuos 302-1040 y marcada con HA en su extremo COOH, para facilitar la detección de productos génicos de *nod2*. Las proteínas Nod2 se detectaron mediante inmunotransferencia con el anticuerpo anti-HA y se indican como a y b.

La Figura 4 muestra un análisis mutacional de Nod2. La Figura 4A muestra las proteínas Nod2 silvestre y mutante. Los dominios CARD, NBD y las LRR se indican con cuadros cerrados negros, oscuros y tramados, respectivamente. Los números representan el residuo de aminoácido en la proteína Nod2. La Figura 4B muestra el análisis de expresión de las proteínas Nod2 silvestre y mutante. Las células HEK293T se transfectaron con el plásmido control (-) o 5 µg de los plásmidos que producen las proteínas Nod2 marcadas con HA indicadas. Los extractos de igual número de células se inmunoprecipitaron con el anticuerpo anti-HA de conejo y se sometieron a inmunotransferencia con el anticuerpo monoclonal anti-HA de ratón. El tamaño previsto de las proteínas mutantes en CARD, CARD1 y las LRR se indica mediante puntas de flecha negras. La Figura 4C muestra la activación de NF-κB por las proteínas Nod2. la inducción de la activación de NF-κB se determinó a partir de cultivos por triplicado de células HEK293T cotransfectadas con la cantidad indicada de plásmidos de expresión de Nod2 silvestre o mutante en presencia de pBVIx-Luc y pEF-BOS-p-gal como se describe más adelante. Los

valores representan la media \pm SD de los cultivos por triplicado.

La Figura 5 muestra que la Nod2 actúa a través del complejo IKK para activar NF- κ B. La Figura 5A muestra inhibición de Nod2 y activación de NF- κ B inducida por TNF- α por las proteínas mutantes negativas dominantes de la vía de NF- κ B. La inducción de la activación de NF- κ B se determinó en cultivos por triplicado de células HEK293T transfectadas con 30 ng de plásmido Nod2 (barras abiertas) o estimuladas con 10 ng/ml de TNF- α durante 4 horas (barras cerradas) y 70 ng del plásmido de expresión I- κ B α S32A/S36A, IKK α K44A, IKK β K44A, RICK (406-540) o RIP (558-671) en presencia de pBVluc-Luc y pEF-BOS-p-gal. Los resultados se presentan como un porcentaje de valores obtenidos con Nod2 y el plásmido control. En los experimentos mostrados, Nod2 y TNF- α indujeron 58 ± 8 veces y 14 ± 1 la activación de NF- κ B, respectivamente. Los valores representan la media \pm SD de los cultivos por triplicado. La Figura 5B muestra la inducción de NF- κ B en células Rat-1 y 5R derivadas. La inducción de la activación de NF- κ B se determinó a partir de cultivos por triplicado de 1×10^5 células HEK293T cotransfectadas con los plásmidos indicados y pBVluc-Luc en presencia del plásmido control pEF-BOS-p-gal. Los valores representan la media \pm SD de cultivos por triplicado.

La Figura 6 muestra la interacción de Nod2 con RICK. Las Figuras 6A y B muestran la interacción entre Nod2 silvestre y mutante con RICK. Las células HEK293T se cotransfectaron con el plásmido de expresión de RICK y Nod2 silvestre y mutante. El RICK co-inmunoprecipitado se detectó mediante inmunotransferencia con anticuerpo anti-Flag (panel superior). Los inmunoprecipitados de Nod2 se muestran en el panel inferior. Los lisados totales se transfirieron con el anticuerpo anti-Flag (panel del medio). La Figura 6C muestra la interacción entre Nod2 y RICK silvestre y mutante. Las células HEK293T se cotransfectaron con el plásmido de expresión de Nod2 silvestre y RICK- Δ CARD (residuos 1-3-74) o RICK-CARD (residuos 374-540) el plásmido de expresión de RICK y Nod2 silvestre y mutante. La Nod2 co-inmunoprecipitada se detectó mediante inmunotransferencia con anticuerpo anti-HA (panel superior). Los lisados totales se transfirieron con el anticuerpo anti-Flag (panel del medio) o anti-HA (panel inferior). Una banda de fondo se muestra con un asterisco.

La Figura 7 muestra que la oligomerización forzada de Nod2 induce la activación de NF- κ B. La Figura 7A muestra un análisis de expresión de las proteínas quiméricas Fpk3-Nod2 silvestre y mutante. Las células HEK293T se transfectaron con el plásmido control (-) o los plásmidos que producen las proteínas Fpk3Nod2 marcadas con Myc indicadas. Los extractos de igual número de células se inmunoprecipitaron e inmunotransfirieron con anticuerpo anti-Myc de conejo. La Figura 7B muestra que la oligomerización forzada de Nod2 induce la activación de NF- κ B. 2×10^5 células HEK293T se transfectaron con 1 ng de los plásmidos indicados en presencia de pBVluc-luc and pEF-BOS- β -gal. 8 horas después de la transfección se trató a las células con AP1510 500 nM (barras negras) o se dejaron sin tratar (barras blancas). 24 horas después de la transfección se determinó la transcripción dependiente de B como se describe más adelante. Los valores representan la media \pm SD de los cultivos por triplicado.

La Figura 8 muestra la respuesta de las células HEK293T que expresan Nod1 a los componentes patógenos bacterianos y fúngicos. La Figura 8A muestra datos de 1×10^5 células HEK293T que se transfeccionaron con 0,3 ng de pcDNA3-Flag (barras blancas) o pcDNA3Nod1-Flag (barras negras) en presencia de 600 ng de pcDNA3, 73 ng de pEFIBOS- β gal y 7,3 ng de pBXIV-luc. 8 horas después de la transfección se trató a las células con 10 μ g/ml de cada producto patógeno, ácido lipoteicoico (LTA) o peptidoglicano (PGN) de *Staphylococcus aureus*, lipopolisacárido (LPS) de *Escherichia coli* 055:B5, manano de *Candida albicans* 20A, lipoproteína bacteriana soluble sintética (SBLP) o se dejaron sin tratar (control). 24 horas después de la transfección, la transcripción dependiente de B se determinó mediante la actividad luciferasa relativa y los valores se normalizaron frente a β -galactosidasa en cultivos por triplicado. Como control, el inserto mostró proteínas Nod1 inmunodetectadas con Ac anti-FLAG en lisados de células transfectadas con 1 ng de pcDNA3-Nod1 en presencia (derecha) y ausencia (izquierda) de 10 μ g/ml de LPS. La Figura 8B muestra datos de 1×10^5 células HEK293T que se transfeccionaron con 0,3 ng de pcDNA3-Flag (-), pcDNA3-Nod1-Flag (Nod1) pcDNA3-Nod1 (1-648)-Flag (Nod1 Δ LRR), 300 ng de pcDNA3-FLAG-TLR4, 3 ng de pCMV-IL1R1 más 100 ng de pcDNA3-IL1 β -HA (IL1) o 1 ng de pcDNA3-RIP-Flag (RIP) en presencia de 600 ng de pcDNA3, 73 ng de pEF 1BOS- β gal y 7,3 ng de pBXIV-luc. Ocho horas después de la transfección se trató a las células con 10 μ g/n-A LPS (barras negras) o se dejaron sin tratar (barras blancas). Veinticuatro horas después de la transfección se determinó la transcripción dependiente de κ B como se ha descrito anteriormente.

La Figura 9 muestra la capacidad de respuesta diferencial de Nod1 y Nod2 a LPS de varias fuentes. 1×10^5 células HEK293T se transfeccionaron con 0,3 ng de pcDNA3-Flag (-), pcDNA3-Nod1-Flag (Nod1) o pcDNA3-Nod1 (1-648)-Flag (Nod1 Δ LRR), 0,3 ng de pcDNA3-Flag (-), pcDNA3-Nod1-Flag (Nod1) o pcDNA3-Nod1(1-648)-Flag (Nod1 Δ LRR), 0,03ng de pcDNA3-Nod2 o pcDNA3-Nod2(1-744)-Flag (Nod2 Δ LRR) en presencia de 600 ng de pcDNA3, 73 ng de pEF1BOS- β gal y 7,3 ng de pBXIV-luc. 8 horas después de la transfección se trató a las células con 10 μ g/ml de cada patógeno, LTA de *S. aureus* o *S. sanguis*, PGN de *S. aureus*, LPS de *Pseudomonas aeruginosa*, *Shigella flexneri* 1A, *Serratia marcescens*, *Salmonella typhimurium*, *Klebsiella pneumoniae* o *E. coli* 055:B5, o se dejaron sin tratamiento. Para la estimulación de TNF- α , 22 horas después de la transfección se incubaron las células con 10 ng/ml de TNF- α durante 2 horas.

La Figura 10 muestra la interacción física entre Nod1 y LPS. 1×10^8 células HEK293T se transfectaron con 30 μ g

- de pcDNA3-Flag-Nod1, pRK7-FLAG-IKK α , pcDNA3-FLAG-TKKi, pcDNA3-FLAG-IKK γ o pcDNA3-CIPER-FLAG (Takeuchi y col., Immunity, 4:443 [1999]). 24 horas después de la transfección se prepararon fracciones S100 de células transfectadas como se describe más adelante. La radiactividad de [³H] LPS co-inmunoprecipitada con el Ab anti-FLAG se determinó como se describe más adelante. La Figura 10A muestra el lisado S100 de células transfectadas incubado con [³H] LPS, Ac M2 anti-FLAG, proteína A-Sefarosa y Proteína G-Sefarosa. La Figura 10B muestra los datos de proteínas inmunoprecipitadas primero a partir de 20 mg del lisado S100 e incubadas con [³H] LPS en presencia de 10 mg de BSA. La radiactividad co-inmunoprecipitada se determinó como se describe con detalle más adelante. La expresión de cada proteína en 50 μ g del lisado S100 se inmunodetectó con Ac anti-FLAG.
- 5
- 10 La Figura 11 muestra la secuencia de ácido nucleico de SEC ID N° 33.
La Figura 12 muestra la secuencia de ácido nucleico de SEC ID N° 1.
La Figura 13 muestra la secuencia polipeptídica de SEC ID N° 2.
La Figura 14 muestra la secuencia polipeptídica de SEC ID N° 3.
La Figura 15 muestra la secuencia polipeptídica de SEC ID N° 34.
- 15 La Figura 16 muestra el ácido nucleico (SEC ID N° 35 (tipo silvestre) y 36 (mutante) y el polipéptido (SEC ID N° 98 (tipo silvestre) y la SEC ID N° 99 (mutante)) del exón 11 de Nod2.
- La Figura 17 muestra la identificación de una mutación de cambio de marco de Nod2 en individuos afectados de familias de EC. La Figura 17a muestra un mapa físico de la región de interés en 16q12. Las posiciones aproximadas de los marcadores cromosómicos y genéticos se muestran en base a la ref. 23. El clon de BAC genómico humano RP 11-32722 contiene el gen *Nod2* y los marcadores stSG46415 y SGC32374. La organización exón-intrón del gen *Nod2* humano se muestra debajo. La figura 17b muestra electroferogramas de la secuencia de ADN (exón 11) del control y de tres individuos afectados de familias de EC. Los pacientes de las familias 1 y 6 son heterocigotos, mientras que el paciente de la familia 7 es homocigoto para 3020InsC. La inserción C se muestra con una flecha. La Figura 17c muestra la secuencia de nucleótidos y de aminoácidos predicha del exón 11 y los intrones flanqueantes de controles normales y pacientes con 3020InsC. La secuencia del exón se muestra en negrita. El sitio de la inserción C se indica con una flecha. El residuo (W) indica que un nucleótido del exón 12 contribuye al codón. La Figura 17d muestra un esquema de la estructura de dominio de Nod2 para ilustrar el sitio de truncamiento proteico. Se muestran los dominios de reclutamiento de caspasa (CARD), dominio de unión al nucleótido (NBS) y diez LRR. Los residuos de diez LRR están subrayados. Los números indican las posiciones de los residuos.
- 20
- 25
- 30
- La Figura 18 muestra la determinación de la transmisión de la mutación 3020InsC en una familia CD mediante PCR específica de alelos. Se usó PCR multiplex para generar un producto inespecífico de 533 pb junto con amplicones específicos de alelos: un fragmento de 319 pb (de tipo silvestre) y un fragmento de 214 pb (3020InsC). En esta familia, ambos padres (calles 1 y 4) son heterocigotos para 3020InsC, mientras que los hijos (calles 2 y 3) tienen CD y son homocigotos para 3020 InsC. Calle 5, control de tipo silvestre. Calle 6, marcadores de ADN-Mspl de pBR322. Los números en el gel indican el tamaño de los fragmentos.
- 35
- La Figura 19 muestra la capacidad de respuesta diferencial de Nod2 de tipo silvestre y mutante al LPS. La Figure 19a muestra las células HEK293T contranfectadas con las cantidades indicadas de los plásmidos indicadores pcDNA3-Nod2, pcDNA3-Nod2 3020InsC o pcDNA3 (vector) y pEF-BOS- β -gal y pBVI-luc por triplicado. Los valores representan las medias \pm s. d. La expresión de las proteínas Nod2 de tipo silvestre y mutante en extractos celulares se muestra en la parte superior. La Figura 19b muestra las células HEK293T cotransfectadas con 0,3 ng de pcDNA3-Nod2, 3 ng de pcDNA3-Nod2 3020InsC, 3 ng de pcDNA3-TLR4 más 3 ng de pcDNA3-MD-2 (indicados por TLR4) o pcDNA3 (vector) y pEF1BOS- β -gal y pBMV-luc por triplicado. En estas condiciones, las construcciones de Nod2 de tipo silvestre y mutante indujeron niveles similares de actividad de NF- κ B basal. 8 horas después de la transfección, las células se trataron con 10 μ g/ml de LPS de las bacterias idnicadasa. Los valores representan las medias \pm s.d. Los resultados son representativos de al menos 3 experimentos independientes.
- 40
- 45
- La Figura 20 muestra la secuencia de ácido nucleico de SEC ID N° 53.
La Figura 21 muestra la secuencia de ácido nucleico de SEC ID N° 54.
- 50
- La Figura 22 muestra la secuencia de aminoácidos SEC ID N° 55.
La Figura 23 muestra la secuencia de ácido nucleico de SEC ID N° 56.
La Figura 24 muestra la secuencia de aminoácidos SEC ID N° 57.
La Figura 25 muestra la secuencia de ácido nucleico de SEC ID N° 58.

La Figurs 26 describe polimorfismos en el gen de Nod2. La Tabla 1 describe los alelos y sus correspondientes SEC ID N°

La Figura 27 describe las frecuencias de alelos para los polimorfismos descritos en la Figura 26.

La Figura 28 muestra la secuencia de aminoácidos SEC ID N° 59.

5 La Figura 29 muestra la secuencia de ácido nucleico de SEC ID N° 60.

La Figura 30 muestra la secuencia de aminoácidos SEC ID N° 61.

La Figura 31 muestra la secuencia de ácido nucleico de SEC ID N° 62.

La Figura 32 muestra la secuencia de aminoácidos SEC ID N° 63.

La Figura 33 muestra la secuencia de ácido nucleico de SEC ID N° 64.

10 La Figura 34 muestra la secuencia de aminoácidos SEC ID N° 65.

La Figura 35 muestra la secuencia de ácido nucleico de SEC ID N° 66.

La Figura 36 muestra la secuencia de aminoácidos SEC ID N° 67.

La Figura 37 muestra la secuencia de ácido nucleico de SEC ID N° 68.

La Figura 38 muestra la secuencia de aminoácidos SEC ID N° 69.

15 La Figura 39 muestra la secuencia de ácido nucleico de SEC ID N° 84.

La Figura 40 muestra la secuencia de aminoácidos SEC ID N° 85.

La Figura 41 muestra la secuencia de ácido nucleico de SEC ID N° 86.

La Figura 42 muestra la secuencia de aminoácidos SEC ID N° 87.

La Figura 43 muestra la secuencia de ácido nucleico de SEC ID N° 88.

20 La Figura 44 muestra la secuencia de aminoácidos SEC ID N° 89.

Descripción general de la invención

La presente invención se define en las reivindicaciones y se refiere a moléculas de señalización intracelular, en concreto a variantes de la proteína y a los ácidos nucleicos que codifican las variantes de la proteína Nod2. Se encontró que la proteína Nod2 tenía homología estructural con la proteína Nod1. Apaf-1 y Nod1 (también denominado CARD4) son miembros de una familia de proteínas intracelulares que están compuestas por un dominio de reclutamiento de la caspasa en NH₂-terminal (CARD), un dominio de unión al nucleótido (NBD) localizado centralmente y un dominio regulador en COOH-terminal (Bertin y col., J. Biol. Chem. 274: 12955-12958 [1999], Inohara y col., J. Biol. Chem. 274: 14560-14568 [1999]). Mientras que Apaf-1 posee repeticiones de WD40, Nod1 contiene repeticiones ricas en leucina (LRR) en su extremo C. Las similitudes estructurales y funcionales entre Apaf-1 y Nod1 sugieren que estas proteínas comparten un mecanismo molecular común para la activación y la función efectora. En el caso de Apaf-1, las repeticiones WD-40 actúan como dominio de reconocimiento para el daño mitocondrial a través de la unión al citocromo c, permitiendo que Apaf-1 oligomerice e interaccione con la procaspasa-9 a través de una interacción hemofílica CARD-CARD (Li y col., Cell 91: 479-489 [1997], Zou et al., J. Bio. Chem. 274: 11549-11556 [1999]). La oligomerización de Apaf-1 está mediada por el NBD y se piensa que induce la proximidad y la activación proteolítica de las moléculas de procaspasa-9 en el complejo del apoptosoma (Srinivasula y col., Mol. Cell 1: 949-957 [1998], Hu et al., J. Bio. Chem. 273: 33489-34494 [1998]).

Estudios previos mostraron que Nod1 estimula la apoptosis cuando se sobreexpresa en las células, pero al contrario que Apaf-1, induce la activación de NF-κB (Bertin y col., citado anteriormente, Inohara y col., citado anteriormente). La activación de NF-κB inducida por Nod1 está mediada por la asociación del CARD de Nod1 con el correspondiente CARD de RICK (también denominado RIP2 and CARDIAK), una proteína quinasa que activa NF-κB (Bertin y col., citado anteriormente, Inohara y col., citado anteriormente, Inohara y col., J. Biol. Chem. 273: 12296-12300 [1998], McCarthy et al., J. Bio. Chem. 273, 16968-16975 [1998], Thome et al., Curr. Biol. 8: 885-888 [1998]). Los análisis con formas silvestre (sv) y mutante de Nod1 y RICK han sugerido que Nod1 y RICK actúan en la misma vía de activación de NF-κB, en la que RICK funciona como mediador cadena abajo de la señalización de Nod1 (Bertin y col., citado anteriormente, Inohara y col., [1999] citado anteriormente, Inohara et al., J. Biol. Chem. 275: 27823-27831 [2000]). Nod1 se autoasocia a través de su NBD y la oligomerización de Nod1 estimula la proximidad de las moléculas de RICK y la activación de NF-κB (Inohara y col., [2000], citado anteriormente). Nod1 también muestra similitud con una clase de proteínas de resistencia (R) a la enfermedad encontrada en plantas (Parniske y col., Cell 91: 821-832 [1997], Dixon et al., Proc. Natl. Acad. Sci. U. S. A. 97: 8807-8814 [2000]). Como Nod1, estas proteínas R intracelulares contienen dominios efectores en N-terminal unidos a un NBD y comparten con Nod1 la

presencia de múltiples LRR localizados en C terminal del NBD (Bertin y *col.*, citado anteriormente, Dixon y *col.*, citado anteriormente). Tras el reconocimiento específico de los productos patogénicos, estas proteínas R participan en una respuesta de defensa asociada con alteraciones metabólicas y localizan la muerte celular en el lugar de la invasión del patógeno (Dixon y *col.*, citado anteriormente). Las LRR de las proteínas R son muy diversas y parecen estar implicadas en el reconocimiento de una amplia gama de componentes patogénicos (Pamiske y *col.*, citado anteriormente, Dixon y *col.*, citado anteriormente). La pareja de unión de las LRR de Nod1 sigue sin conocerse. La homología estructural de Nod1 con proteínas R de planta sugiere que otras moléculas de tipo Nod1 que contienen LRR pueden existir en el genoma humana para permitir la activación de estas moléculas mediante diferentes conjuntos de estímulos intracelulares.

La identificación y caracterización de Nod2, una proteína que contiene LRR con similitud estructural y funciona con Nod1 se divulgan en el presente documento. Estos estudios indican que Nod2 activa NF- κ B, pero al contrario que Nod1, este nuevo homólogo se expresa principalmente en los monocitos. La presente divulgación no está limitada a ningún mecanismo de acción concreto. De hecho, no es necesario conocer el mecanismo de acción para la práctica de la presente invención. No obstante, Nod2 es un miembro de la familia Nod1/Apaf-1 que activa NF- κ B mediante interacciones con su CARDS en NH₂ terminal, ya que se descubrió que estos dominios eran necesarios y suficientes para la activación de NF- κ B. Nod2 asociado con RICK a través de una interacción CARD-CARD homófila. La actividad de inducción de NF- κ B de Nod2 se correlacionó con su capacidad para asociarse con RICK y es inhibida por un mutante de RICK, lo que sugiere que RICK es una diana cadena abajo directa de Nod2. Por tanto, las vías de señalización de Nod1 y Nod2 parecen usar RICK como un mediador cadena debajo de activación de NF- κ B. En contraste con Nod1, dos CARD en tándem están presentes en el extremo NH₂ de Nod2 y ambos eran necesarios para la asociación con RICK y la activación de NF- κ B.

Nod2 es la primera molécula que se sabe que contiene dos CARD. La base molecular subyacente al requisito de ambos CARD de Nod2 para a unión a RICK sigue sin estar clara. La presente divulgación no está limitada a ningún mecanismo de acción concreto. De hecho, no es necesario conocer el mecanismo de acción para la práctica de la presente invención. No obstante, se contempla que la presencia de ambos CARD puede potenciar la afinidad por el CARD de RICK, Otra posibilidad es que tras una interacción inicial que implica un CARD de Nod2 y el CARD de RICK, Nod2 puede sufrir un cambio conformacional que permita al segundo CARD asociarse a RICK con afinidad elevada. La región intermedia de RICK se asocia con IKK γ (Inohara y *col.*, [2000], citada anteriormente), lo que proporciona un vínculo directo entre Nod1/Nod2 y el complejo IKK. Consistente con este modelo, la activación de NF- κ B por Nod2 así como la inducida por Nod1 requería IKK γ y estaba inhibida por las formas negativas dominantes de IKK γ , IKK α e IKK β . El papel funcional para las LRR de Nod1 y Nod2 sigue sin estar claro. La LRR es un módulo de interacción proteína-proteína repetida que probablemente esté implicado en la activación de Nod1 y Nod2 por señales cadena arriba. En el caso de las proteínas R que contienen NBD/LRR de planta, sus LRR parecen ser importantes para el reconocimiento de componentes patógenos y sus dominios en N-terminal parecen mediar en una cascada de señalización que regula la expresión génica (Pamiske y *col.*, citado anteriormente, Dixon y *col.*, citado anteriormente). Dado que Nod1 y Nod2 activan NF- κ B, sus LRR pueden actuar reconociendo un conjunto diferente de estímulos intracelulares que median en la oligomerización de Nod1 y Nod2 y la asociación con RICK. Dado que Nod2 se expresa principalmente en monocitos, Nod2 podría servir como receptor intracelular que transduce las señales en el monocito/macrófago que conducen a la activación y transcripción de genes reguladores.

Las proteínas Nod2 de la presente invención también están implicados en el reconocimiento de patógenos microbianos. El sistema inmunitario innato regula la respuesta inmediata a los patógenos microbianos en múltiples organismos, incluidos los seres humanos. La respuesta inmunitaria innata se inicia mediante el reconocimiento de componentes patogénicos específicos por las células inmunitarias huésped. Las células de mamífero tienen receptores de superficie celular y mecanismos intracelulares que inician la respuesta de defensa contra patógenos microbianos (Aderem y Ulevitch, *Nature*, 406:785-787 [2000]; Philpott et al., *J. Immunol.*, 165:903-914 [2000]). Los receptores de tipo toll (TLR) comprenden una familia de receptores de superficie celular que están relacionados con la proteína Toll de *Drosophila*, una molécula implicada en la defensa contra la infección fúngica en la mosca (Aderem y Ulevitch, *citado anteriormente*). Se han identificado diez TLR de mamífero (Aderem y Ulevitch, *citado anteriormente*). Dos miembros de la familia, TLR2 y TLR4, se han caracterizado y se ha demostrado que median en la respuesta a múltiples componentes de la pared celular bacteriana, incluidos el lipopolisacárido (LPS), lipopéptidos, peptidoglicanos (PGN) y ácido lipoteicoico (LTA) (Yang y *col.*, *Nature*, 395:284-288 [1998]; Poltorak y *col.*, *Science*, 282:2085-2088 [1998]; Aliprantis et al., *Science*, 285:736-739 [1999]; Chow et al., *J. Biol. Chem.*, 274:10689-10692 [2000]; and Schwandner et al., *J. Biol. Chem.*, 274: 17406-17409 [2000]). Los TLR de mamífero tienen múltiples repeticiones ricas en leucina en el ectodominio y un dominio del receptor de IL-1 de tipo Toll intracelular (TIR) que media en una cascada de señalización al núcleo (Aderem y Ulevitch, *citado anteriormente*). La estimulación de TLR2 y TLR4 conduce al reclutamiento de la molécula adaptadora MyD88 y la quinasa asociada con la serina quinasa IL-1R (IRAK), dos componentes de señalización que junto con TRAF-6 median en la activación de NF- κ B (Aderem y Ulevitch, *citado anteriormente*).

Las plantas tienen varias clases de genes que regulan la defensa contra los patógenos invasores. Una clase importante de estas moléculas se denominan proteínas de resistencia (R) a la enfermedad sus miembros incluyen proteínas unidas a la membrana y citosólicas. Estas son esenciales para la defensa contra múltiples patógenos, incluidas bacterias, hongos y virus (Dixon y *col.*, *PNAS*, 97:8807-8814 [2000]). El tipo citosólico de las proteínas R que incluyen el producto génico de tabaco N y hasta 200 productos génicos en *Arabidopsis thaliana* están

compuestos por un TIR en N-terminal o dominio efector en dedo de cinc, un dominio de unión al nucleótido localizado centralmente (NBD) y repeticiones ticas en leucina en C-terminal (LRR) (Dixon y col., *citado anteriormente*). Las LRR de las proteínas R citosólicas son muy diversas y parecen estar implicadas en el reconocimiento de una amplia gama de componentes microbianos (Dixon y col., *citado anteriormente*). Esta clase de proteínas resistentes a la enfermedad participa en la respuesta de hipersensibilidad (HS) en plantas que incluye alteraciones metabólicas y muerte celular localizada en el lugar de la invasión del patógeno (Dixon y col., *citado anteriormente*). Las proteínas R citosólicas de plantas tienen una considerable homología estructural con Nod1/CARD4, una proteína recientemente descrita relacionada con el regulador de la apoptosis Apaf-1 (Zou y col., *Cell*, 90:405-413 [1997]; Bertin et al., *J. Biol. Chem.*, 274:12955-12958; and Inohara et al., *J. Biol. Chem.*, 274:14560-14568 [1999]). Como las proteínas R de plantas, Nod1 está compuesto por un dominio efector en N-terminal, un NBD localizado centralmente y múltiples LRR en el extremo C (Bertin y col., *citado anteriormente*, Inohara y col., *citado anteriormente*). Nod1 induce la activación de NF- κ B, que está mediada por la asociación de su dominio de reclutamiento de caspasa en N-terminal (CARD) con la de RICK, una proteína cinasa que también activa NF- κ B (Bertin y col., *citado anteriormente*, Inohara y col., *citado anteriormente*, *J. Biol. Chem.*, 273:12296 [1998]; McCarthy et al., *J. Biol. Chem.*, 273:16968; Thome et al., *Curr. Biol.*, 8:885 [1998]; Inohara et al., *J. Biol. Chem.*, 275:27823 [2000]). No obstante, la(s) molécula(s) desencadenante(s) que activa(n) Nod1 para mediar en la activación de NF- κ B siguen sin conocerse.

La presente invención también demuestra que el lipopolisacárido (LPS) induce activación de NF- κ B en células HEK293T que expresan Nod1, mientras que las células HEK293T parentales son insensibles al LPS. La presente invención no está limitada a ningún mecanismo de acción concreto. De hecho, no es necesario conocer el mecanismo de acción para la práctica de la presente invención. No obstante, en el sistema humano, se ha demostrado que el complejo TLR4/MD2/CD14 sirve como receptor de superficie para el LPS (Aderem y Ulevitch, *citado anteriormente*). Además del complejo TLR4 de superficie celular existen abundante evidencia de que las células de mamífero tienen un receptor intracelular que detecta el LPS en el citoplasma de células infectadas por bacterias (Philpott y col., *citado anteriormente*). Por ejemplo, las células epiteliales no responden al LPS extracelular purificado o presentado en el contexto de cepas bacterianas gramnegativas no invasivas (Philpott y col., *citado anteriormente*). No obstante, el LPS introducido dentro de las células epiteliales activa NF- κ B (Philpott y col., *citado anteriormente*). Sin embargo, hasta la fecha, la identificación de un sistema de reconocimiento intracelular para LPS y/u otros productos microbianos sigue sin estar clara. Dado que Nod1 puede conferir capacidad de respuesta a LPS, Nod1 puede actuar como receptor intracelular para el LPS. La función de Nod1 podría ser importante en la respuesta intracelular de las células epiteliales contra las bacterias invasoras, ya que Nod1 se expresa en superficies epiteliales intestinales, pulmonares y nasales en el embrión de ratón tardío (Inohara y col., *citado anteriormente*). La presencia de un sistema de detección intracelular para LPS bacteriano se esperaría en las superficies epiteliales como las del intestino que están muy expuestas a bacterias y a productos bacterianos. En dichos órganos, el desencadenamiento de una respuesta inflamatoria a productos bacterianos a través de receptores de superficie tales como TLR4 sería perjudicial para el organismo. Las células HEK293T que expresan Nod2, otro miembro de la familia de Nod2, responden a LPS pero Nod1 y Nod2 parecen tener diferentes preferencias por las preparaciones de LPS de diferentes bacterias. Estas observaciones sugieren que, además de los TLR, los miembros de la familia Nod pueden representar otro sistema inmunitario innato para el reconocimiento de una amplia gama de productos patogénicos. Por ejemplo, el genoma de la planta *Arabidopsis thaliana* contiene aproximadamente 200 genes de resistencia a la enfermedad que codifican proteínas NBD-LRR intracelulares relacionadas con Nod1 y Nod2 (Dixon y col., *citado anteriormente*).

El sistema inmunitario innato regula la respuesta inmediata a los patógenos microbianos y se inicia mediante el reconocimiento de componentes patogénicos específicos por las células inmunitarias huésped. Las células de mamífero tienen receptores de superficie celular (p. ej., TLRs) y mecanismos intracelulares que inician la respuesta de defensa contra patógenos microbianos (Aderem y Ulevitch, *citado anteriormente*). Nod1 y Nod2 parecen funcionar como receptores intracelulares para el LPS con las LRR requeridos para la capacidad de respuesta (véase el Ejemplo 8). Los resultados descritos en el presente documento sugiere un truncamiento prematuro de la décima LRR en Nod2 se asocia con el desarrollo de EC. Consistente con los datos de análisis de enlace publicados (Ohmen et al., *Hum Mol Genet.* 5:1679 [1996]; Cho et al., *Proc Natl Acad Sci U S A.* 95:7502 [1998]) esta variante genética se asoció únicamente con la EC y no con la CU. El análisis funcional indica que el mutante LRR Nod2 truncado es menos activo que la proteína de tipo silvestre al conferir capacidad de respuesta al LPS bacteriano. En homólogos de Nod2 vegetal, las LRR parecen determinar la especificidad por productos patógenos y las alteraciones en sus LRR pueden tener como resultado una falta de respuesta a patógenos y enfermedades concretas (Pamiske et al., *Cell* 91:821 [1997]; Ellis et al., *Plant Cell* 11:495 [1999]; Dixon et al., *Proc. Natl. Acad. Sci. U.S.A.* 97:8807 [2000]). De un modo similar, la variación genética en las LRR de TLR4 representan las diferencias entre individuos en la capacidad de respuesta bronquial a LPS aerosolizados (Arbour et al., *Nat Genet.* 25:187 [2000]). La presente invención no está limitada a ningún mecanismo concreto. De hecho, no es necesario conocer el mecanismo para la práctica de la presente invención. No obstante, se contempla que se pueden concebir varios mecanismos que pueden justificar la susceptibilidad a la EC en individuos que alojan la mutación 3020InsC u otra variante de Nod2. Nod2 es una proteína citosólica cuya expresión está restringida a los monocitos, sin expresión detectada en linfocitos (véase el Ejemplo 3, más adelante). Un déficit en notar las bacterias dentro de los monocitos/macrófagos podría dar como resultado una respuesta inflamatoria exagerada por el sistema inmunológico adaptativo. Una posibilidad alternativa es que la Nod2 de tipo silvestre puede participar en la inducción de los genes de citocinas,

tales como interleucina-10 que puede regular por disminución la respuesta inflamatoria (Moore et al., *Ann. Rev. Immunol.* 11:165 [1993]; Berg et al., *J. Clin. Invest.* 98:1010 [1996]).

5 En este escenario, una deficiencia en la función de Nod2 puede conducir a una sobreproducción relativa de citocinas proinflamatorias en el intestino. Por último, se ha demostrado que la variación en las LRR de los homólogos de Nod2 de plantas tiene como resultado el reconocimiento de nuevas especificidades para los componentes patógenos (Pamiske et al., *Cell* 91:821 [1997]; Ellis et al., *Plant Cell* 11:495 [1999]). Por tanto, también es posible que 3020InsC pueda actuar como una ganancia de función mutante para un patógeno desconocido. Los presentes estudios implican a la Nod2 en la susceptibilidad a la EC y sugieren una relación entre una respuesta innata a los componentes bacterianos y el desarrollo de la enfermedad. Los resultados pueden explicar la observación de que la disminución de la flora bacteriana intestinal puede conducir a una mejora clínica y a disminución de la inflamación intestinal (Fiocchi, *Gastroenterology* 115:182 [1998]).

10 Los experimentos realizados durante el desarrollo de la presente invención (véase, por ejemplo, los ejemplos 9 y 10) identificaron varios polimorfismos de Nod2 que se encuentran con una prevalencia mayor en individuos con enfermedad de Crohn. Además, se encontró que varios polimorfismos están asociados con un incremento del riesgo de desarrollar enfermedad de Crohn. De acuerdo con lo anterior, en algunas realizaciones, la presente invención proporciona procedimientos para determinar la susceptibilidad de un individuo a la enfermedad de Crohn. En algunas realizaciones, los procedimientos incluyen procedimientos bioinformáticos.

Definiciones

Para facilitar la comprensión de la invención, a continuación se definen una serie de términos.

20 Como se usa en el presente documento, el término "Nod2", cuando se usa en referencia a una proteína o ácido nucleico, se refiere a una proteína o ácido nucleico que codifica una proteína que, en su forma silvestre, activa NF-κB y contiene dos CARD (dominios de reclutamiento de caspasa). El término Nod2 abarca tanto las proteínas que son idénticas a Nod2 silvestre como las que derivan de la Nod2 silvestre (p. ej., variantes de Nod2 o genes quiméricos construidos con porciones de regiones de codificación de Nod2).

25 Como se usa en el presente documento, la expresión "activa NF-κB", cuando se usa en referencia a cualquier molécula que activa NF-κB, se refiere a una molécula (p. ej., una proteína) que induce la actividad del factor de transcripción de NF-κB a través de una vía de señalización celular. Los ensayos para determinar si una molécula activa NF-κB usan, por ejemplo, constructos de genes indicadores respondedores a NF-κB. Ensayos adecuados incluyen, entre otros, los descritos en los Ejemplos 4 y 5.

30 Como se usa en el presente documento, la expresión "actividad de Nod2" se refiere a cualquier actividad de la Nod2 silvestre. Con la expresión se pretende abarcar todas las actividades de Nod2 (p. ej., incluidas, entre otras, la activación de NF-κB, la unión a RICK y la potenciación de la apoptosis).

35 Como se usa en el presente documento, la expresión "individuo con mayor riesgo de desarrollar enfermedad de Crohn" se refiere a un individuo para el cual "el riesgo relativo al genotipo" o el "riesgo atribuible a la población" de desarrollar enfermedad de Crohn es mayor que el riesgo promedio en una población dada (p. ej., un grupo étnico).

Como se usa en el presente documento, la expresión "riesgo relativo al genotipo" y "riesgo atribuible a la población" hacen referencia a mediciones relativas del riesgo de desarrollar un estado de enfermedad (p. ej., enfermedad de Crohn). Ejemplos de cómo calcular el "riesgo relativo al genotipo" y "riesgo atribuible a la población" se proporcionan en el Ejemplo 10.

40 El término "apoptosis" significa muerte celular no necrótica que tiene lugar en las células de animales metazoos tras la activación de un programa de suicidio celular intrínseco. La apoptosis es un proceso normal en el desarrollo y la homeostasis de los animales metazoos. La apoptosis implica cambios morfológicos y bioquímicos característicos, incluidos la disminución de tamaño de la célula, zeiosis o formación de burbujas, de la membrana plasmática, y el colapso celular y la fragmentación de la cromatina nuclear, en lugares dentro del nucleosoma, debido a la activación de una nucleasa endógena.

45 El término "gen" se refiere a una secuencia de ácido nucleico (p. ej., ADN), que comprende secuencias de codificación necesarias para la producción de un polipéptido, ARN (p. ej., entre otros, ARNm, ARNt y ARNr) p precursor (p. ej., Nod2). El polipéptido, ARN o precursor puede estar codificado por una secuencia de codificación de longitud completa o por cualquier parte de la secuencia de codificación siempre se conserven la actividad deseada o las propiedades funcionales (p. ej., actividad enzimática, unión del ligando, transducción de la señal, etc.) de l longitud completa o del fragmento. El término también abarca la región de codificación de un gen estructural y las secuencias incluidas localizadas adyacentes a la región de codificación en los extremos tanto 5' como 3' para una distancia de aproximadamente 1 kb en cualquier extremo de modo que el gen corresponda a la longitud del ARNm de longitud completa. Las secuencias localizadas en 5' de la región de codificación y que están presentes en el ARNm se denominan secuencias 5' no traducidas. Las secuencias localizadas en 3' o cadena abajo de la región de codificación y que están presentes en el ARNm se denominan secuencias 3' no traducidas. El término "gen" abarca tanto ADNc como las formas genómicas de un gen. Una forma genómica o clon de un gen contiene la región de

codificación interrumpida con secuencias no codificantes denominadas “intrones” o “regiones intermedias” o “secuencias intermedias”. Los intrones son segmentos de un gen que se transcribe a ARN nuclear (ARNhn); los intrones pueden contener elementos reguladores, como los potenciadores. Los intrones se eliminan o “cortan” del transcrito nuclear o primario; por tanto, los intrones no están en el transcrito de ARN mensajero (ARNm). El ARNm funciona durante la traducción para especificar la secuencia u ordenar los aminoácidos en un polipéptido naciente.

En concreto, la expresión “gen Nod2” se refiere a la secuencia nucleotídica de Nod2 de longitud completa (p. ej., contenida en la SEC ID N° 1). No obstante, también se pretende que término abarque fragmentos de la secuencia de Nod2, así como otros dominios dentro de la secuencia de nucleótidos de Nod2 de longitud completa. Además, las expresiones “secuencia de nucleótidos de Nod2” o “secuencia de polinucleótidos de Nod2” abarca secuencias de ADN, ADNc y ARN (p. ej., ARNm).

Cuando en el presente documento se cita “secuencia de aminoácidos” para hacer referencia a una secuencia de aminoácidos de una molécula proteica natural, con “secuencia de aminoácidos” y términos similares, tales como “polipéptido” o “proteína” no se pretende limitar la secuencia de aminoácidos a la secuencia de aminoácidos nativa completa asociada con la molécula proteica citada.

Además de contener intrones, las formas genómicas de un gen también pueden incluir secuencias localizadas en ambos extremos, 5' y 3', de las secuencias presentes en el transcrito de ARN. Estas secuencias se denominan secuencias o regiones “flanqueantes” (estas secuencias flanqueantes se localizan en 5' o 3' de las secuencias no traducidas presentes en el transcrito de ARN,). La región flanqueante en 5' puede contener secuencias reguladoras, tales como promotores y potenciadores, que controlen o influyan sobre la transcripción del gen. La región flanqueante en 3' puede contener secuencias que dirijan la terminación de la transcripción, la escisión postranscripcional y la poliadenilación.

La expresión “de tipo silvestre” se refiere a un gen o producto génico que tiene las características de ese gen o producto génico cuando se aísla de una fuente de origen natural. Un gen de tipo silvestre es el que se observa más frecuentemente en una población y, por lo tanto, se denomina arbitrariamente la forma “normal” o “de tipo silvestre” del gen. Por el contrario, los términos “modificado”, “mutante”, “polimorfismo” y “variante” se refiere a un gen o producto génico que muestra modificaciones en secuencia y/o propiedades funcionales (es decir, características alteradas) cuando se compara con el gen o producto génico silvestre. Cabe destacar que se pueden aislar mutantes de origen natural; los mismos se identifican por el hecho de que tienen características alteradas cuando se compararon con el gen o producto génico de tipo silvestre.

Como se usa en el presente documento, las expresiones “molécula de ácido nucleico codificante”, “secuencia de ADN codificante” y “ADN codificante” se refieren a el orden o secuencia de desoxirribonucleótidos a lo largo de una hebra de ácido desoxirribonucleico. El orden de estos desoxirribonucleótidos determina el orden de los aminoácidos a lo largo de la cadena polipeptídica (proteína). Por tanto, la secuencia de ADN codifica la secuencia de aminoácidos.

Se dice que las moléculas de ADN tienen extremos “5'” y extremos “3'” porque los mononucleótidos reaccionan para fabricar oligonucleótidos o polinucleótidos de un modo tal que el 5'fosfato de un anillo mononucleótidos fosfato está unido al oxígeno en 3' de su vecino en una dirección mediante un enlace fosfodiéster. Por tanto, un extremo de un oligonucleótido o polinucleótido, al que se hace referencia como “extremo 5'” si su fosfato en 5' no está unido al oxígeno en 3' de un anillo pentosa de mononucleótido y como “extremo 3'” si su oxígeno en 3' no está unido a un fosfato en 5' de un anillo pentosa de mononucleótido posterior. Como se usa en el presente documento, también se puede decir que una secuencia de ácido nucleico, incluso si es interna a un oligonucleótido más grande, también tiene extremos 5' y 3'. En una molécula de ADN lineal o circular, se hace referencia a los elementos pequeños estando “cadena arriba” o en 5' de los elementos “cadena abajo” o en 3'. Esta terminología refleja el hecho de que la transcripción progresa en dirección 5' a 3' a lo largo de la hebra de ADN. El promotor y los elementos potenciadores que dirigen la transcripción de un gen ligado se localizan, en general, en 5' o cadena arriba de la región de codificación. No obstante, los elementos potenciadores pueden ejercer su efecto incluso cuando se localizan en 4' del elemento promotor y la región de codificación. Las señales de terminación de la transcripción y de poliadenilación se localizan en 3' o cadena debajo de la región de codificación.

Como se usa en el presente documento, las expresiones “oligonucleótido que tiene una secuencia de nucleótidos que codifica un gen” y “polinucleótido que tiene una secuencia de nucleótidos que codifica un gen” significan una secuencia de ácido nucleico que comprende la región de codificación de un gen o, en otras palabras, la secuencia de ácido nucleico que codifica un producto génico. La región de codificación puede estar presente en una forma de ADNc, ADN genómico o ARN. Cuando está presente en forma de ADN, el oligonucleótido o polinucleótido puede ser monocatenario (es decir, la hebra sentido) o bicatenario. Elementos de control adecuados, tales como potenciadores/promotores, uniones de corte y empalme, señales de poliadenilación etc., pueden colocarse cerca de la región de codificación del gen si es necesario para permitir el inicio adecuado de la transcripción y/o el correcto procesamiento del transcrito de ARN primario. Como alternativa, la región de codificación usada en los vectores de expresión de la presente invención puede contener potenciadores/promotores endógenos, uniones de corte y empalme, secuencias intermedias, señales de poliadenilación etc., o una combinación de elementos de control tanto endógenos como exógenos.

Como se usa en el presente documento, la expresión “elemento regulador” se refiere a un elemento genético que controla algún aspecto de la expresión de secuencias de ácido nucleico. Por ejemplo, un promotor es un elemento regulador que facilita la iniciación de la transcripción de una región de codificación unida operablemente. Otros elementos reguladores incluyen señales de corte y empalme, señales de poliadenilación, señales de terminación etc.

5 Como se usa en el presente documento, los términos “complementario” o “complementariedad” se usan en referencia a polinucleótidos (es decir, una secuencia de nucleótidos) de acuerdo con las reglas de apareamiento de bases. Por ejemplo, para la secuencia "A-G-T," es complementaria a la secuencia "T-C-A." La complementariedad puede ser “parcial”, en la que sólo algunas de las bases nucleotídicas están apareadas de acuerdo con las reglas de apareamiento de bases. O puede haber complementariedad “completa” o “total” entre los ácidos nucleicos. El grado
10 de complementariedad entre las hebras de ácido nucleico tiene efectos significativos sobre la eficiencia y la fuerza de la hibridación entre hebras de ácido nucleico. Esto es de particular importancia en las reacciones de amplificación, así como en los procedimientos de detección que dependen de la unión entre ácidos nucleicos.

El término “homología” se refiere a un grado de complementariedad. Puede haber homología parcial u homología completa (es decir, identidad). Una secuencia parcialmente complementaria es una que inhibe al menos
15 parcialmente la hibridación de una secuencia completamente complementaria con un ácido nucleico diana y se denomina usando la expresión funcional “sustancialmente homólogo”. La expresión “inhibición de la unión”, cuando se usa en referencia a la unión del ácido nucleico, se refiere a la inhibición de la unión debido a la competición de las secuencias homólogas por la unión a una secuencia diana. La inhibición de la hibridación de la secuencia completamente complementaria con la secuencia diana se puede analizar usando un ensayo de hibridación (de
20 transferencia Southern o Northern, hibridación en solución y similares) en condiciones de rigurosidad baja. Una secuencia o sonda sustancialmente homóloga competirá e inhibirá la unión (es decir, la hibridación) de una completamente homóloga a una diana en condiciones de rigurosidad baja. Cabe mencionar que las condiciones de rigurosidad baja son tales que se permita la unión inespecífica; las condiciones de rigurosidad baja requieren que la unión de dos secuencias entre sí sea una interacción específica (es decir, selectiva). La ausencia de unión
25 inespecífica puede analizarse mediante el uso de una segunda diana que carece de incluso un grado parcial de complementariedad (es decir, una identidad inferior a aproximadamente el 30%); en ausencia de unión inespecífica, la sonda no hibridará con la segunda diana no complementaria.

En la técnica se sabe que se pueden emplear numerosas condiciones equivalentes para comprender condiciones de rigurosidad baja; se consideran factores tales como la longitud y la naturaleza (ADN, ARN, composición en bases)
30 de la sonda y la naturaleza de la diana (ADN, ARN, composición en bases, presente en solución o inmovilizados etc.) y la concentración de las sales y otros componentes (p. ej., la presencia o ausencia de formamida, sulfato de dextrano, polietilenglicol) y la solución de hibridación se puede variar para generar condiciones de hibridación de rigurosidad baja diferentes, pero equivalentes, de las condiciones enumeradas anteriormente. Además, en la técnica se conocen condiciones que estimulan la hibridación en condiciones de rigurosidad alta (p. ej., incrementar la
35 temperaturas de la hibridación y/o etapas de lavado, el uso de formamida en la solución de hibridación etc.).

Cuando se usa en referencia a una secuencia de ácido nucleico bicatenario, tal como un ADNc o clon genómico, la expresión “sustancialmente homólogo” se refiere a cualquier sonda que pueda hibridar con una o ambas hebras de la secuencia de ácido nucleico bicatenario en condiciones de baja rigurosidad, como se ha descrito anteriormente.

Un gen puede producir múltiples especies de ARN que se generan mediante corte y empalme diferencial del
40 transcrito de ARN primario. Los ADNc que son variantes de corte y empalme del mismo gen contendrán regiones de identidad de secuencia u homología completa (que representan la presencia del mismo exón o porción del mismo exón sobre ambos ADNc) y regiones de no identidad completa (por ejemplo, que representan la presencia del exón “A” sobre el ADNc 1 en el que el ADNc 2 contiene el exón “B” en su lugar). Dado que los dos ADNc contienen regiones de identidad de secuencia, ambos hibridarán con una sonda que proceda de todo el gen o de porciones del
45 gen que contienen secuencias encontradas en ambos ADNc; las dos variantes de corte y empalme son, por tanto, sustancialmente homólogas a dicha sonda y entre sí.

Cuando se usa en referencia a una secuencia de ácido nucleico monocatenario, la expresión “sustancialmente homólogo” se refiere a cualquier sonda que pueda hibridar (es decir, es la complementaria de) con la secuencia de ácido nucleico bicatenario en condiciones de baja rigurosidad, como se ha descrito anteriormente.

50 Como se usa en el presente documento, la expresión “compite por la unión” se usa en referencia a un primer polipéptido con una actividad que se une al mismo sustrato que un segundo polipéptido con una actividad, en la que el segundo polipéptido es una variante del primer polipéptido o un polipéptido diferente o relacionado. La eficiencia (p. ej., cinética o termodinámica) de la unión por el primer polipéptido puede ser igual o mayor o menor que la eficiencia de la unión del sustrato por el segundo polipéptido. Por ejemplo, la constante de unión en el equilibrio (K_D)
55 para la unión al sustrato puede ser diferente para los dos polipéptidos. El término “ K_m ” como se usa en este documento se refiere a la constante de Michaelis-Menten para una enzima se define como la concentración del sustrato específico a la que una enzima dada produce la mitad de su velocidad máxima en una reacción catalizada por enzima.

Como se usa en el presente documento, el término “hibridación” se usa con referencia al emparejamiento de ácidos

nucleicos complementarios. La hibridación y la fuerza de la hibridación (es decir, la fuerza de la asociación entre los ácidos nucleicos) están influidas por factores tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicadas y la T_m del híbrido formado y la proporción G:C dentro de los ácidos nucleicos.

5 Como se usa en el presente documento, el término “ T_m ” se usa con referencia a la “temperatura de fusión”. La temperatura de fusión es la temperatura a la que la población de moléculas de ácido nucleico bicatenario se disocia a la mitad en hélices únicas. La ecuación para calcular la T_m de ácidos nucleicos es bien conocida en la técnica. Como se indica por referencias convencionales, una estimación simple del valor de T_m se puede calcular con la ecuación: $T_m = 81,5 + 0,41 (\% G + C)$, cuando un ácido nucleico está en solución acuosa a NaCl 1 M (véase, por ejemplo, Anderson y Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985). Otras referencias incluyen cálculos más sofisticados que tienen en cuenta características estructurales, así como características de secuencia para el cálculo de T_m .

10 Como se usa en el presente documento, el término “rigurosidad” se usa con referencia a las condiciones de temperatura, fuerza iónica y presencia de otros compuestos, tales como disolventes orgánicos, en las que se realizan las hibridaciones de ácidos nucleicos. Los expertos en la técnica reconocerán que condiciones de “rigurosidad” se pueden alterar variando los parámetros que se acaban de describir individualmente o en conjunto. En condiciones de “alta rigurosidad”, el emparejamiento de bases de ácidos nucleicos tendrá lugar solamente entre fragmentos de ácido nucleico que tienen una alta frecuencia de secuencias de bases complementarias (p. ej., hibridación en condiciones de “alta rigurosidad” se puede producir entre homólogos con una identidad de aproximadamente 85 – 100 %, preferentemente una identidad de aproximadamente 70 – 100 %). En condiciones de “rigurosidad media”, el emparejamiento de bases de ácidos nucleicos tendrá lugar entre ácidos nucleicos con una frecuencia intermedia de las secuencias de bases complementarias (p. ej., la hibridación en condiciones de “rigurosidad media” se puede producir entre homólogos con una identidad de aproximadamente 50 – 70 %). Por tanto, a menudo se requieren condiciones de rigurosidad “débil” o “baja” con ácidos nucleicos que derivan de organismos que son genéticamente diversos, ya que la frecuencia de las secuencias de complementariedad suele ser menor.

15 “Condiciones de rigurosidad alta”, cuando se usan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5X SSPE (43,8 g/l NaCl, 6,9 g/l $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ y 1,85 g/l EDTA, pH ajustado a 7,4 con NaOH), 0,5% de SDS, 5X reactivo de Denhardt y 100 $\mu\text{g/ml}$ de ADN desnaturalizado de esperma de salmón seguido de lavado en una solución que comprende 0,1X SSPE, 1,0% SDS a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

20 “Condiciones de rigurosidad media”, cuando se usan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5X SSPE (43,8 g/l NaCl, 6,9 g/l $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ y 1,85 g/l EDTA, pH ajustado a 7,4 con NaOH), 0,5% de SDS, 5X reactivo de Denhardt y 100 $\mu\text{g/ml}$ de ADN desnaturalizado de esperma de salmón seguido de lavado en una solución que comprende 1,0X SSPE, 1,0% SDS a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

25 “Condiciones de rigurosidad baja” comprenden condiciones equivalentes a la unión o hibridación a 42 °C en una solución que consiste en 5X SSPE (43,8 g/l NaCl, 6,9 g/l $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ y 1,85 g/l EDTA, pH ajustado a 7,4 con NaOH), 0,1% SDS, 5X reactivo de Denhardt (50X reactivo de Denhardt contiene por 500 ml: 5 g de Ficoll (Tipo400, Pharmacia), 5 g de BSA (Fracción V; Sigma) y 100 $\mu\text{g/ml}$ de ADN de esperma de salmón desnaturalizado, seguido de lavado en una solución que comprende 5X SSPE, 0,1 % SDS a 42 °C cuando se emplea una sonda de aproximadamente 500 nucleótidos de longitud.

30 Los siguientes términos se usan para describir las relaciones de secuencia entre dos o más polinucleótidos. “secuencia de referencia”, “identidad de secuencia”, “porcentaje de identidad de secuencia” e “identidad sustancial”. Una “secuencia de referencia” es una secuencia definida usada como base para una comparación de secuencias; una secuencia de referencia puede ser un subconjunto de una secuencia más grande, por ejemplo como segmento de una secuencia de ADNc de longitud completa dada en un listado de secuencias o puede comprender una secuencia génica completa. En general, una secuencia de referencia tiene una longitud de al menos 20 nucleótidos, con frecuencia una longitud de al menos 25 nucleótidos, y a menudo una longitud de al menos 50 nucleótidos. Dado que dos polinucleótidos pueden cada uno (1) comprender una secuencia (es decir, una porción de la secuencia de polinucleótidos completa) que es similar entre los dos polinucleótidos y (2) puede además comprender una secuencia que es divergente entre los dos polinucleótidos, las comparaciones de secuencia entre dos (o más) polinucleótidos normalmente se realizan comparando secuencias de los dos polinucleótidos sobre una “ventana de comparación” para identificar y comparar regiones locales de similitud de secuencia. Una “ventana de comparación”, como se usa en el presente documento, se refiere a un segmento conceptual de al menos 20 posiciones de nucleótidos contiguos en el que una secuencia de polinucleótidos se puede comparar con una secuencia de referencia de al menos 20 nucleótidos contiguos y en el que la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) del 20 por ciento o menos en comparación como la secuencia de referencia (que no comprende adiciones o deleciones) para una alineación óptima de las dos secuencias. La alineación óptima se secuencias para alinear una ventana de comparación se puede realizar mediante el logaritmo de homología local de Smith y Waterman [Smith y Waterman, Adv. Appl. Math.

2: 482 (1981)] mediante el algoritmo de alineación de la homología de Needleman y Wunsch [Needleman y Wunsch, J. Mol. Biol. 48:443 (1970)], mediante la búsqueda del procedimiento de similitud de Pearson y Lipman [Pearson y Lipman, Proc. Natl. Acad. Sci. (U.S.A.) 85:2444 (1988)], mediante implementaciones computerizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el paquete de software Wisconsin Genetics Software Package, versión 7,0, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante inspección y se selecciona la mejor alineación (es decir, la que tiene como resultado el mayor porcentaje de homología sobre la ventana de comparación) generada mediante los diversos procedimientos. La expresión "identidad de secuencia" significa que dos secuencias de polinucleótidos son idénticas (es decir, nucleótido por nucleótido) sobre la ventana de comparación. La expresión "porcentaje de identidad de secuencia" se calcula comparando dos secuencias alineadas óptimamente sobre la ventana de comparación, determinando el número de posiciones en las que la base de ácido nucleico idéntica (p. ej., A, T, C, G, U o I) se produce en ambas secuencias, para dar el número de posiciones coincidentes, dividiendo el número de las posiciones coincidentes por el número total de posiciones en la ventana de comparación (es decir, el tamaño de la ventana) y multiplicando el resultado por 100 para proporcionar el porcentaje de identidad de secuencia. Las expresiones "identidad sustancial" como se usa en el presente documento indica una característica de una secuencia de polinucleótidos, en la que el polinucleótido comprende una secuencia que tiene una identidad de secuencia de al menos un 99 por ciento en comparación con una secuencia de referencia sobre una ventana de comparación de al menos 20 posiciones de nucleótidos, con frecuencia sobre una ventana de al menos 25-50 nucleótidos, en la que el porcentaje de identidad de secuencia se calcula comparando la secuencia de referencia con la secuencia del polinucleótido que puede incluir deleciones o adiciones que totalizan un 20 por ciento o menos de la secuencia de referencia sobre la ventana de comparación. La secuencia de referencia puede ser un subconjunto de una secuencia más grande, por ejemplo como un segmento de las secuencias de longitud completa de las composiciones reivindicadas en la presente invención (p. ej., Nod2).

Como se aplica a polipéptidos, la expresión "identidad sustancial" significa que dos secuencias peptídicas, cuando están alineadas óptimamente, tal como mediante los programas GAP o BESTFIT usando pesos de huecos por defecto, comparten una identidad de secuencia de al menos 80%, preferentemente una identidad de secuencia de al menos 90%, más preferentemente una identidad de secuencia de al menos 95% o más (p. ej., una identidad de secuencia del 99 por ciento). Preferentemente, las posiciones de los residuos que no son idénticas difieren en sustituciones conservadoras de aminoácidos. Las sustituciones conservadoras de aminoácidos hacen referencia a la capacidad de intercambio de residuos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos que tienen cadenas laterales alifáticas es glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tienen cadenas laterales alifáticas-hidroxilo es serina y treonina; un grupo de aminoácidos que tienen cadenas laterales que contienen amida es asparagina y glutamina; un grupo de aminoácidos que tienen cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas es lisina, arginina e histidina; y un grupo de aminoácidos que tienen cadenas laterales que contienen azufre es cisteína y metionina. Grupos preferidos para la sustitución conservadora de aminoácidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparagina-glutamina.

El término "fragmento", tal como se usa en el presente documento, se refiere a un polipéptido que tiene una deleción en el amino terminal y/o el carboxi terminal en comparación con la proteína nativa, pero en el que la secuencia de aminoácidos restante es idéntica a las correspondientes posiciones en la secuencia de aminoácidos deducida de una secuencia de ADNc de longitud completa. Normalmente, los fragmentos tienen una longitud de al menos 4 aminoácidos, preferentemente una longitud de al menos 20 aminoácidos, normalmente una longitud de al menos 50 aminoácidos o mayor, y abarcan la porción del polipéptido requerida para la unión intermolecular de las composiciones (reivindicadas en la presente invención) con sus diversos ligandos y/o sustratos.

La expresión "locus polimórfico" es un locus presente en una población que muestra variación entre miembros de la población (p. ej., el alelo más común tiene una frecuencia inferior a 0,95). Por el contrario, un "locus monomórfico" es un locus genético con pocas o sin variaciones observadas entre miembros de la población (generalmente se asume que es un locus en el que el alelo más común supera una frecuencia de 0,95 en el grupo de genes de la población).

Como se usa en el presente documento, la expresión "información de variación genética" o "información de variante genética" hace referencia a la presencia o ausencia de una o más secuencias de ácido nucleico variantes (p. ej., polimorfismo o mutaciones) en un alelo dado de un gen concreto (p. ej., un gen de Nod2).

Como se usa en el presente documento, la expresión "ensayo de detección" hace referencia a un ensayo para detectar la presencia o ausencia de secuencias de ácido nucleico variantes (p. ej., polimorfismos o mutaciones) en un alelo dado de un gen concreto (p. ej., el gen de Nod2). Ejemplos de ensayos de detección adecuados incluyen, entre otros, los descritos más adelante en la Sección IIIB.

La expresión "de origen natural", como se usa en el presente documento aplicada a un objeto, se refiere al hecho de que un objeto se puede encontrar en la naturaleza. Por ejemplo, una secuencia polipeptídica o polinucleotídica que está presente en un organismo (incluidos virus) que se pueden aislar de una fuente de la naturaleza y que no se ha modificado intencionalmente por el hombre en el laboratorio es natural.

"Amplificación" es un caso especial de replicación de ácido nucleico que implica especificidad de molde. Se tiene que contrastar con replicación inespecífica del molde (es decir, replicación dependiente de molde pero no

dependiente de un molde específico). En el presente documento, especificidad de molde se distingue de fidelidad de la replicación (es decir, síntesis de la secuencia de polinucleótidos adecuada) y especificidad de nucleótidos (ribo o desoxirribo). La especificidad de molde con frecuencia se describe en términos de especificidad de "diana". Las secuencias diana son "dianas" en el sentido de que se buscan separarlas de otros ácidos nucleicos. Las técnicas de
5 amplificación se han diseñado principalmente para esta ordenación.

La especificidad de molde se consigue en la mayoría de las técnicas de amplificación mediante la elección de la enzima. Las enzimas de amplificación son encimas que, en las condiciones en las que se usan, procesarán únicamente secuencias específicas de ácido nucleico en una mezcla heterogénea de ácido nucleico. Por ejemplo, en el caso de la Φ replicasa, el ARN de MDV-1 es el molde específico para la replicasa (D.L. Kacian y col., Proc. Natl. Acad. Sci. USA 69:3038 [1972]). Otros ácidos nucleicos no se replicarán mediante esta enzima de
10 amplificación. De un modo similar, en el caso de la ARN polimerasa de T7, esta enzima de amplificación tiene una especificidad rigurosa por sus propios promotores (M. Chamberlin y col., Nature 228:227 [1970]). En el caso de la ADN ligasa de T4, la enzima no ligará los dos oligonucleótidos o polinucleótidos, en los que hay una falta de apareamiento entre el sustrato oligonucleótido o polinucleótido y el molde en el punto de unión (D.Y. Wu y R B. Wallace, Genomics 4:560 [1989]). Por último, se ha encontrado que las polimerasas *Taq* y *Pfu*, en virtud de su
15 capacidad para funcional a temperatura alta, muestran una especificidad alta por las secuencias unidas y, por tanto, definidas por los cebadores; las alta temperatura da lugar a condiciones termodinámicas que favorecen la hibridación del cebador con las secuencias diana y no la hibridación con secuencias no diana (H.A. Erlich (ed.), PCR Technology, Stockton Press [1989]).

20 Como se usa en el presente documento, el término "ácido nucleico amplificable" se usa con referencia a ácidos nucleicos que se pueden amplificar mediante cualquier procedimiento de amplificación. Se contempla que "ácido nucleico amplificable" normalmente comprender "molde de la muestra".

Como se usa en el presente documento, la expresión "molde de la muestra" se refiere al ácido nucleico de una muestra en el que se analiza la presencia de la "diana" (definido más adelante). Por el contrario, "molde básico" se
25 usa en referencia al ácido nucleico que no sea el molde de la muestra que puede o no estar presente en una muestra. El molde básico con frecuencia es inadvertido. Puede ser el resultado de un arrastre o puede deberse a la presencia de contaminantes del ácido nucleico que se desean eliminar de la muestra mediante purificación. Por ejemplo, los ácidos nucleicos de organismos distintos a los que se van a detectar pueden estar presentes como fondo en una muestra de ensayo.

30 Como se usa en el presente documento, el término "cebador" se refiere a un oligonucleótido, natural como en un digesto de restricción purificado o producido sintéticamente, que es capaz de actuar como punto de inicio de la síntesis cuando se introduce en las condiciones en las que la síntesis de un producto de extensión del cebador, que es complementario a una hebra de ácido nucleico, se induce, es decir, en presencia de nucleótidos y un agente inductor tal como una ADN polimerasa y a una temperatura y pH adecuados. Preferentemente, el cebador es
35 monocatenario para una máxima eficiencia en la amplificación, pero, como alternativa, puede ser bicatenario. Si es bicatenario, el cebador se trata primero para separar las hebras antes de usar para preparar productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo bastante largo como para cebar la síntesis de los productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluidos la temperatura, la fuente del cebador y el uso del
40 procedimiento.

Como se usa en el presente documento, el término "sonda" se refiere a un oligonucleótido (es decir, una secuencia de nucleótidos), ya sean naturales como en un digesto de restricción purificado, o producidos de forma sintética, de forma recombinante o mediante amplificación por OCR, que es capaz de hibridar con otro oligonucleótido de interés. Una sonda puede ser bicatenaria o monocatenaria. Las sondas son útiles en la detección, identificación y
45 aislamiento de secuencias génicas concretas. Se contempla que cualquier sonda usada en la presente invención esté marcada con cualquier "molécula indicadora", de modo que sea detectable en cualquier sistema de detección, incluidos, entre otros, enzimas (p. ej., ELISA, así como ensayos histoquímicos basados en enzimas), sistemas fluorescentes, radioactivos y luminiscentes. No se pretende que la presente invención esté limitada a ningún sistema de detección o marcador concreto.

50 Como se usa en el presente documento, el término "diana" se refiere a una secuencia o estructura de ácido nucleico que se va a detectar o caracterizar. Por tanto, se busca separar la "diana" de otras secuencias de ácido nucleico. Un "segmento" se define como una región de ácido nucleico dentro de la secuencia diana.

Como se usa en el presente documento, la expresión "reacción en cadena de la polimerasa" ("PCR") se refiere al procedimiento de K.B. Mullis, patentes de EE.UU. nº 4.683.195, 4.683.202, y 4.965.188, que se incorporan en el
55 presente documento por referencia, que describen un procedimiento para aumentar la concentración de un segmento de secuencia diana en una mezcla de ADN genómico sin clonación o purificación. Este procedimiento para amplificar la secuencia diana consiste en introducir un gran exceso de dos cebadores oligonucleotídicos en la mezcla de ADN que contiene la secuencia diana deseada, seguida de una secuencia precisa de ciclado térmico en presencia de una ADN polimerasa. Los dos cebadores son complementarios de sus respectivas hebras de la secuencia diana bicatenaria. Para efectuar la amplificación, la mezcla se desnaturaliza y los cebadores se hibridan
60

después con sus secuencias complementarias dentro de la molécula diana. Después de la hibridación, los cebadores se extienden con polimerasa para formar un nuevo par de hebras complementarias. Las etapas de desnaturalización, hibridación del cebador y extensión por polimerasa se pueden repetir muchas veces (es decir, desnaturalización, hibridación y extensión constituyen un "ciclo", puede haber numerosos "ciclos") para obtener una concentración alta de un segmento amplificado de la secuencia diana deseada. La longitud del segmento de la secuencia diana deseada se determina mediante las posiciones relativas de los cebadores entre sí y, por lo tanto, esta longitud es un parámetro controlable. En virtud del aspecto de repetición del procedimiento, el procedimiento se denomina "reacción en cadena de la polimerasa" (en lo sucesivo, "PCR"). Debido a que los segmentos deseados amplificados de la secuencia diana se convierten en las secuencias dominantes (en términos de concentración) en la mezcla, se dice que se "amplifican por PCR".

Con la PCR, es posible amplificar una única copia de una secuencia diana específica en el ADN genómico hasta un nivel detectable mediante varias metodologías diferentes (p. ej., hibridación con una sonda marcada; incorporación de los cebadores biotinilados seguido de detección del conjugado avidina-enzima; incorporación de desoxinucleótidos trifosfatos marcados con ³²P, tales como dCTP o dATP, en el segmento amplificado). Además del ADN genómico, se puede amplificar cualquier secuencia de oligonucleótido o polinucleótido con el conjunto adecuado de moléculas cebadoras. En concreto, los segmentos amplificados creados mediante el propio procedimiento de PCR son, ellos mismos, moldes eficientes para las posteriores amplificaciones por PCR.

Como se usa en el presente documento, las expresiones "producto de PCR", "fragmento de PCR" y "producto de amplificación" hacen referencia a la mezcla resultante de compuestos después de completar dos o más ciclos de las etapas de PCR de desnaturalización, hibridación y extensión. Estas expresiones abarcan el caso en el que se ha producido amplificación de uno o más segmentos de una o más secuencias diana.

Como se usa en el presente documento, la expresión "reactivos de amplificación" se refiere a los reactivos (desoxirribonucleótidos trifosfato, tampón, etc.) necesarios para la amplificación excepto los cebadores, molde de ácido nucleico y la enzima de amplificación. Normalmente, los reactivos de amplificación junto con otros componentes de reacción se colocan y contienen en un vaso de reacción (tubo de ensayo, micropocillo etc.).

Como se usa en el presente documento, las expresiones "endonucleasas de restricción" y "enzimas de restricción" se refieren a enzimas bacterianas, cada una de las cuales corta el ADN bicatenario en o cerca de una secuencia de nucleótidos específica.

Como se usa en el presente documento, la expresión "molécula de ADN recombinante" como se usa en el presente documento se refiere a una molécula de ADN que está comprendida por segmentos de ADN unidos entre sí mediante técnicas de biología molecular.

Como se usa en el presente documento, el término "antisentido" se usa con referencia a secuencias de ARN que son complementarias a una secuencia de ARN específica (p. ej., ARNm). Incluidas en esta definición son moléculas de ARN antisentido ("ARNas") implicadas en la regulación génica por las bacterias. El ARN antisentido se puede producir por cualquier procedimiento, incluida la síntesis mediante corte y empalme del o los genes de interés en una orientación inversa a un promotor viral que permite la síntesis de una hebra de codificación. Una vez introducida en un embrión, esta hebra transcrita se combina con el ARNm natural producido por el embrión para formar dúplex. Estos dúplex bloquean después la transcripción posterior del ARNm o su traducción. De este modo se pueden generar fenotipos mutantes. La expresión "hebra antisentido" se usa en referencia a una hebra de ácido nucleico que es complementaria a la hebra "sentido". La designación (-) (es decir, "negativa") en ocasiones se usa en referencia a la hebra antisentido, usándose a veces la designación (+) en referencia a la hebra sentido (es decir, "positiva").

El término "aislado", cuando se usa en relación con un ácido nucleico, como en "un oligonucleótido aislado" o "polinucleótido aislado" se refiere a una secuencia de ácido nucleico que se identifica y separa de al menos un ácido nucleico contaminante con el que normalmente se asocia en su forma natural. El ácido nucleico aislado está presente en una forma o contexto que es diferente del que se encuentra en la naturaleza. Por el contrario, los ácidos nucleicos no aislados son ácidos nucleicos, tales como ADN y ARN, presentes en el estado en el que existen en la naturaleza. Por ejemplo, una secuencia de ADN diana (p. ej., un gen) se encuentra en los cromosomas de células huésped en proximidad a los genes vecinos; las secuencias de ARN, tales como una secuencia de ARN, específica que codifica una proteína específica, se encuentran en la célula como una mezcla con otros numerosos ARN, que codifican una multitud de proteínas. No obstante, el ácido nucleico aislado que codifica Nod2 incluye, a modo de ejemplo, dicho ácido nucleico en células que habitualmente expresan Nod2 cuando el ácido nucleico está en una localización cromosómica diferente de la de las células naturales o, por otro lado, está flanqueado por una secuencia de ácido nucleico diferente de la encontrada en la naturaleza. El ácido nucleico, oligonucleótido o polinucleótido aislado puede estar presente en forma monocatenaria o bicatenaria. Cuando un ácido nucleico oligonucleótido o polinucleótido aislado se va a usar para expresar una proteína, el oligonucleótido o polinucleótido contendrá como mínimo la hebra sentido o de codificación (es decir, el oligonucleótido o polinucleótido puede ser monocatenario), pero puede contener las hebras sentido y antisentido (Es decir, el oligonucleótido o polinucleótido puede ser bicatenario).

Como se usa en el presente documento, una "porción de un cromosoma" se refiere a una pequeña sección del

5 cromosoma. Los citogenetistas dividen los cromosomas en sitios o secciones del siguiente modo: el brazo corto (con respecto al centrómero) de un cromosoma se denomina el brazo “p”; el brazo largo se denomina el brazo “q”. Después, cada brazo se divide en 2 regiones denominadas región 1 y región 2 (la región 2 es la más cercana al centrómero). Cada región se divide además en bandas. Las bandas se pueden dividir después en sub-bandas. Por ejemplo, la porción 11p15.5 del cromosoma 11 humano es la porción que se localiza en el cromosoma 11 (11) del brazo corto (p) en la primera región (1) en la quinta banda (5) en la sub-banda 5 (.5). Una porción de un cromosoma se puede “alterar”, por ejemplo la totalidad de la porción puede estar ausente debido a una delección o puede reorganizarse (p. ej., inversiones, translocaciones, expandirse o contraerse debido a cambios en las regiones de repetición). En el caso de una delección, un intento de hibridar (es decir, unirse específicamente) una sonda homóloga con una porción concreta de un cromosoma podría tener como resultado un resultado negativo (es decir, la sonda no podría unirse a la muestra que contiene material genético del que se sospecha que contiene la porción que falta del cromosoma). Por tanto, la hibridación de una sonda homóloga a una porción concreta de un cromosoma se puede usar para detectar alteraciones en una porción de un cromosoma.

15 La expresión “secuencias asociadas con un cromosoma” significa preparaciones de cromosomas (p. ej., extensiones de cromosomas en metafase), ácido nucleico extraído de una muestra que contiene ADN cromosómico (p. ej., preparaciones de ADN genómico); el ARN que se produce mediante transcripción de genes localizados en un cromosoma (p. ej., ARNhn y ARNm) y copias de ADNc del ARN transcrito del ADN localizado sobre un cromosoma. Las secuencias asociadas con un cromosoma se pueden detectar mediante numerosas técnicas, incluida el sondaje de transferencias Southern y Northern, e hibridación *in situ* a ARN, ADN o cromosomas en metafase con sondas que contienen secuencias homólogas con los ácidos nucleicos en las preparaciones indicadas anteriormente.

20 Como se usa en este documento, el término “porción” cuando se hace referencia a una secuencia de nucleótidos (como en “una porción de una secuencia de nucleótidos dada”) se refiere a fragmentos de dicha secuencia. El tamaño de los fragmentos puede variar desde cuatro nucleótidos a la totalidad de la secuencia de nucleótidos menos un nucleótido (10 nucleótidos, 20, 30, 40, 50, 100, 200, etc.).

25 Como se usa en el presente documento, la expresión “región de codificación”, cuando se usa en referencia a un gen estructural, se refiere a las secuencias de nucleótidos que codifican los aminoácidos encontrados en el polipéptido naciente como resultado de la traducción de una molécula de ARNm. La región de codificación está unida, en eucariotas, en el lado 5’ por el triplete de nucleótidos “ATG” que codifica el iniciador metionina y en el lado 3’ por uno de los tres tripletes que especifican codones de terminación (es decir, TAA, TAG, TGA).

30 Como se usa en este documento, el término “purificado” o “para purificar” se refiere a la retirada de contaminantes de una muestra. Por ejemplo, los anticuerpos de Nod2 se purifican mediante eliminación de proteínas no inmunoglobulínicas contaminantes; también se purifican mediante eliminación de inmunoglobulina que no se une a Nod2. La eliminación de proteínas no inmunoglobulínicas y/o la eliminación de inmunoglobulinas que no se unen a Nod2 tienen como resultado un incremento en el porcentaje de las inmunoglobulinas reactivas a Nod2 en la muestra. En otro ejemplo, los polipéptidos recombinantes de Nod2 se expresan en células huésped bacterianas y los polipéptidos se purifican mediante la eliminación de proteínas de célula huésped; de este modo se incrementa el porcentaje los de polipéptidos recombinantes de Nod2 en la muestra.

35 La expresión “molécula de ADN recombinante” como se usa en el presente documento se refiere a una molécula de ADN que está comprendida por segmentos de ADN unidos entre sí mediante técnicas de biología molecular.

40 La expresión “proteína recombinante” o “polipéptido recombinante” como se usa en este documento se refiere a una molécula de proteína que se expresa a partir de una molécula de ADN recombinante.

45 La expresión “proteína nativa”, como se usa en el presente documento, se usa para indicar que una proteína no contiene residuos de aminoácidos codificados por secuencias de vectores; es decir, la proteína nativa contiene únicamente los aminoácidos encontrados en la proteína como se encuentran en la naturaleza. Una proteína nativa se puede producir por medios recombinantes o se puede aislar de una fuente natural.

Como se usa en este documento, el término “porción” cuando se hace referencia a una proteína (como en “una porción de una proteína dada”) se refiere a fragmentos de esa proteína. El tamaño de los fragmentos puede variar desde cuatro restos aminoacídicos consecutivos a toda la secuencia de aminoácidos menos un aminoácido.

50 La expresión “transferencia Southern” se refiere al análisis de ADN en geles de agarosa o de acrilamida para fraccionar el ADN de acuerdo con el tamaño, seguido de transferencia del ADN del gel a un soporte sólido, tal como nitrocelulosa o una membrana de nylon. Después, el ADN inmovilizado se sonda con una sonda marcada para detectar especies de ADN complementarias a la sonda usada. El ADN se puede escindir con enzimas de restricción antes de la electroforesis. Tras la electroforesis, el ADN se puede despurinar y desnaturalizar parcialmente antes o durante la transferencia al soporte sólido. Las transferencias Southern son una herramienta estándar de los biólogos moleculares (J. Sambrook y col., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9,31-9,58 [1989]).

55 La expresión “transferencia Northern”, como se usa en el presente documento, se refiere al análisis de ARN mediante la electroforesis de ARN en geles de agarosa para fraccionar el ARN de acuerdo con el tamaño, seguido

de transferencia del ARN del gel a un soporte sólido, tal como nitrocelulosa o una membrana de nylon. Después, el ARN inmovilizado se sonda con una sonda marcada para detectar especies de ARN complementarias a la sonda usada. Las transferencias Northern son una herramienta estándar de los biólogos moleculares (J. Sambrook y col., *citado anteriormente*, pág. 7,39-7,52: [1989]).

5 La expresión “transferencia Western” se refiere al análisis de proteína(s) (o polipéptidos) inmovilizadas sobre un soporte tal como nitrocelulosa o una membrana. Las proteínas se pasan por los geles de acrilamida para separar las proteínas, seguido de transferencia de la proteína del gel a un soporte sólido, tal como nitrocelulosa o una membrana de nylon. Después, las proteínas inmovilizadas se exponen a anticuerpos que tiene reactividad contra un antígeno de interés. La unión de los anticuerpos se puede detectar por varios procedimientos, incluido el uso de anticuerpos radiomarcados.

10 La expresión “determinante antigénico”, como se usa en el presente documento, se refieren a la porción del antígeno que entra en contacto con un anticuerpo concreto (es decir, un epítipo). Cuando una proteína o fragmento de una proteína se usa para inmunizar un animal huésped, numerosas regiones de la proteína pueden inducir la producción de anticuerpos que se unen específicamente a una región dada o estructura tridimensional sobre la proteína; estas regiones o estructuras se denominan determinantes antigénicos. Un determinante antigénico puede competir con el antígeno intacto (es decir, el “inmunógeno” usado para producir la respuesta inmunitaria) para unirse a un anticuerpo.

15 El término “transgén”, como se usa en el presente documento, se refiere a un gen extraño, heterólogo o autólogo que se introduce en un organismo no humano introduciendo el gen en embriones recién fertilizados o en embriones en fase temprana. La expresión “gen extraño” se refiere a cualquier ácido nucleico (p. ej., una secuencia génica) que se introduce en el genoma de un animal mediante manipulaciones experimentales y puede incluir secuencias génicas encontradas en dicho animal siempre que el gen introducido no resida en la misma localización que el gen natural. Con la expresión “gen autólogo” se pretende abarcar variantes (p. ej., polimorfismos o mutantes) del gen natural. Por tanto, el término transgén abarca la sustitución del gen natural con una forma variante del gen.

20 Como se usa en el presente documento, el término “vector” se usa con referencia a moléculas de ácido nucleico que transfieren segmento(s) de ADN de una célula a otra. El término “vehículo” se usa, en ocasiones, de forma intercambiable con “vector”.

25 La expresión “vector de expresión”, como se usa en el presente documento, se refiere a una molécula de ADN recombinante que contiene una secuencia de codificación deseada y secuencias de ácido nucleico adecuadas necesarias para la expresión de la secuencia de codificación unida operablemente en un organismo huésped concreto. Las secuencias de ácido nucleico necesarias para la expresión en procariotas normalmente incluyen un promotor, un operador (opcional) y un sitio de unión al ribosoma, a menudo junto con otras secuencias. Se sabe que las células eucariotas utilizan promotores, potenciadores y señales de poliadenilación y de terminación.

30 Como se usa en el presente documento, la expresión “célula huésped” se refiere a cualquier célula eucariótica o procariótica (p. ej., células bacterianas tal como *E. coli*, células de levadura, células de mamífero, células de ave, células de anfibio, células vegetales, células de peces y células de insecto), localizadas *in vitro* o *in vivo*. Por ejemplo, las células huésped se pueden localizar en un animal transgénico.

35 Los términos “sobreexpresión” y “que sobreexpresa” y sus equivalentes gramaticales se usan en referencia a niveles de ARNm para indicar un nivel de expresión aproximadamente 3 veces mayor que lo que normalmente se observa en un tejido dado en un animal control o no transgénico. Los niveles de ARNm se miden usando cualquiera de una serie de técnicas conocidas por los expertos en la técnica incluyendo, entre otros, análisis de transferencia Northern (véase el Ejemplo 10, para un protocolo de realización de análisis de transferencia Northern). Los controles adecuados se incluyen en la transferencia Northern para controlar las diferencias en la cantidad de ARN cargado de cada tejido analizado (p. ej., la cantidad de ARNr 28S, un transcrito de ARN abundante presente en esencialmente la misma cantidad en todos los tejidos, presente en cada muestra se puede usar como medio de normalizar o estandarizar la señal específica de ARNm de RAD50 observada en las transferencias Northern). Se cuantifica la cantidad de ARNm presente en la banda correspondiente en tamaño al ARN del transgén de Nod2 cortado correctamente; otras especies minoritarias de ARN que hibridan con la sonda transgénica no se consideran en la cuantificación de la expresión del ARNm transgénico.

40 El término “transfección”, como se usa en el presente documento, se refiere a la introducción de ADN extraño en células eucariotas. La transfección se puede acompañar de varios medios conocidos en la técnica, incluidos coprecipitación de fosfato cálcico-ADN, transfección mediada por DEAE dextrano, transfección mediada por polibreno, electroporación, microinyección, fusión de liposomas, lipofección, fusión de protoplastos, infección con retrovirus y biolística.

45 La expresión “transfección estable” o “transfectado de forma estable” se refiere a la introducción e integración de ADN extraño en el genoma de la célula transfectada. La expresión “transfectante estable” se refiere a una célula que ha integrado de forma estable el ADN extraño en el ADN genómico.

La expresión “transfección transitoria” o “transfectado de forma transitoria” se refiere a la introducción de ADN

extraño en una célula en la que el ADN extraño no se integra en el genoma de la célula transfectada. El ADN extraño se queda en el núcleo de la célula transfectada durante varios días. Durante este tiempo, el ADN extraño está sujeto a los controles reguladores que gobiernan la expresión de genes endógenos en los cromosomas. La expresión "transfectante transitorio" se refiere a una célula que han captado ADN extraño pero que han integrado este ADN.

La expresión "coprecipitación con fosfato cálcico" se refiere a una técnica para la introducción de ácidos nucleicos en una célula. La captación de ácidos nucleicos por las células se potencia cuando el ácido nucleico se presenta como coprecipitado de ácido nucleico en fosfato cálcico. La técnica original de Graham y de van der Eb (Graham y van der Eb, *Virology*, 52:456 [1973]), ha sido modificada por varios grupos para optimizar las condiciones para tipos de células concretas. La técnica conoce bien estas numerosas modificaciones.

Una "composición que comprende una secuencia de polinucleótidos dada", como se usa en el presente documento, se refiere ampliamente a cualquier composición que contiene la secuencia de polinucleótidos dada. La composición puede comprender una solución acuosa. Las composiciones que comprenden secuencias polinucleotídicas que codifican Nod2 (p. ej., SEC ID N° 1) o fragmentos de las mismas se pueden usar como sondas de hibridación. En este caso, las secuencias polinucleotídicas que codifican Nod2 normalmente se usan en una solución acuosa que contienen sales (p. ej., NaCl), detergentes (p. ej., SDS) y otros componentes (p. ej., solución de Denhardt, leche seca, ADN de espermatozoos de salmón etc.).

La expresión "compuesto de ensayo" se refiere a cualquier entidad química, sustancia farmacéutica, fármaco y similares que se pueden usar para tratar o prevenir una enfermedad, patología, afección o trastorno de la función corporal o, de otro modo, para alterar el estado fisiológico o celular de una muestra. Los compuestos de ensayo comprenden compuestos terapéuticos conocidos y potenciales. Se puede determinar que un compuesto de ensayo es terapéutico mediante detección selectiva usando procedimientos de detección selectiva de la presente invención. Un "compuesto terapéutico conocido" se refiere a un compuesto terapéutico que se ha demostrado (p. ej., mediante ensayos con animales o experiencia previa con administración a seres humanos) que es eficaz en dicho tratamiento o prevención.

El término "muestra" como se usa en el presente documento se utiliza en su sentido más amplio. Una muestra que se sospecha que contiene un cromosoma humano o secuencias asociadas con un cromosoma humano puede comprender una célula, cromosomas aislados de una célula (p. ej., una extensión de cromosomas en metafase), ADN genómico (en solución o unido a un soporte sólido tal como, por ejemplo, análisis de transferencia Southern), ARN (en solución o unido a un soporte sólido tal como, por ejemplo, análisis de transferencia Northern), ADNc (en solución o unido a un soporte sólido) y similares. Una muestra que se sospecha que contiene una proteína puede comprender una célula, una porción de tejido, un extracto que contiene una o más proteínas y similares.

Como se usa en el presente documento, el término "respuesta", cuando se usa en referencia a un ensayo, se refiere a la generación de una señal detectable (p. ej., acumulación de proteína indicadora, incremento de la concentración de iones, acumulación de un producto químico detectable).

Como se usa en el presente documento, la expresión "proteína receptora de membrana" se refiere a proteínas que abarcan la membrana que se unen a un ligando (p. ej., una hormona o un neurotransmisor). Como se conoce en la técnica, la fosforilación de proteínas es un mecanismo regulador de uso habitual por las células para modificar de forma selectiva las proteínas portadoras de señales reguladoras desde fuera de la célula al núcleo. Las proteínas que ejecutan estas modificaciones bioquímicas son un grupo de enzimas conocidas como proteínas quinasas. Pueden definirse adicionalmente por el residuo del sustrato al que está dirigida la fosforilación. Un grupo de proteínas quinasas son las tirosina quinasas (TK) que fosforilan de forma selectiva una proteína diana en sus residuos de tirosina. Algunas tirosina quinasas son receptores unidos a la membrana (RTK) y, tras la activación por un ligando, pueden autofosforilarse así como modificar sustratos. El inicio de la fosforilación secuencial mediante estimulación de logando es un paradigma que subyace a la acción de dichos efectores tales como, por ejemplo, el factor de crecimiento epidérmico (EGF), insulina, el factor de crecimiento derivado de plaquetas (PDGF) y el factor de crecimiento de fibroblastos (FGF). Los receptores de estos ligandos son tirosina quinasas y proporcionan la interfaz entre la unión de un ligando (hormona, factor de crecimiento) a una célula diana y la transmisión de una señal en la célula mediante la activación de una o más vías bioquímicas. La unión del ligando a un receptor tirosina quinasa activa su actividad enzimática intrínseca. Las tirosina quinasas también pueden ser enzimas citoplasmáticas que no son receptores y actúan como componente posterior en la vía de transducción de señal.

Como se usa en el presente documento, la expresión "proteína de transducción de señal" se refiere a una proteína que se activa o, de otro modo, ejerce su efecto mediante la unión de ligando a una proteína receptora de membrana o citosólica o algún otro estímulo. Ejemplos de proteína de transducción de señal incluyen la adenilato ciclasa, la fosfolipasa C y las proteínas G. Muchas proteínas receptoras de membrana se acoplan a proteínas G (es decir, receptores acoplados a proteína G (GPCR); para una revisión, véase Neer, 1995, *Cell* 80:249-257 [1995]). Normalmente, los GPCR contienen siete dominios transmembrana. Los supuestos GPCR se pueden identificar en base a la homología de secuencia con GPCR conocidos.

Los GPCR participan en la transducción de señal a través de una membrana celular tras la unión de un ligando a una porción extracelular de un GPCR. La porción intracelular de un GPCR interactúa con una proteína G para

modular la transducción de señal desde fuera a dentro de una célula. Por tanto, se dice que un GPCR está “acoplado” a una proteína G. Las proteínas G están compuestas por tres subunidades polipeptídicas: una subunidad α , que se une e hidroliza el GTP, y una subunidad dimérica $\beta\gamma$. En estado basal inactivo, la proteína G existe como heterotrímero de las subunidades α y $\beta\gamma$. Cuando la proteína G está inactiva, el guanosina bifosfato (GDP) se asocia con la subunidad α de la proteína G. Cuando un ligando se une a un GPCR y lo activa, el GPCR se une a la proteína G heterotrimérica y disminuye la afinidad de la subunidad $G\alpha$ por el GDP. En su estado activo, la subunidad G intercambia GDP por guanina trifosfato (GTP) y la subunidad $G\alpha$ activa se disocia del receptor y de la subunidad dimérica $\beta\gamma$. La subunidad $G\alpha$ activa desasociada transduce las señales a efectores que están “después” en la vía de señalización de la proteína G dentro de la célula. En última instancia, la actividad GTPasa endógena de la proteína pasa la subunidad G activa a su estado inactivo, en el que se asocia con GDP y la subunidad $\beta\gamma$ dimérica.

Se han clonado numerosos miembros de la familia de proteína G heterotrimérica, incluidos más de 20 genes que codifican varias subunidades $G\alpha$. Las diversas subunidades G se han clasificado en cuatro familias en base a las secuencias de aminoácidos y la homología funcional. Estas cuatro familias se denominan $G\alpha_s$, $G\alpha_i$, $G\alpha_q$ y $G\alpha_{12}$. Funcionalmente, estas cuatro familias difieren con respecto a las vías de señalización intracelular que activan y el GPCR al que se acoplan.

Por ejemplo, ciertos GPCR normalmente se acoplan con las $G\alpha_s$ y, a través de las $G\alpha_s$, estos GPCR estimulan la actividad de la adenilato ciclasa. Otros GPCR normalmente se acoplan con $G\alpha_q$, y a través de $G\alpha_q$, estos GPCR pueden activar la fosfolipasa C (PLC), tal como la isoforma β de la fosfolipasa C (es decir, PLC β , Sternweis and Smrcka, Trends in Biochem. Sci. 17:502-506 [1992]).

Como se usa en el presente documento, la expresión “proteína de unión al ácido nucleico” se refiere a proteínas que se unen a ácido nucleico y, en concreto, a proteínas que aumentan (es decir, activadores o factores de transcripción) o disminuyen (es decir, inhibidores) la transcripción de un gen.

Como se usa el presente documento, la expresión “proteína de canal iónico” se refiere a proteínas que controlan la entrada o la salida de iones a través de membranas celulares. Ejemplos de proteínas de canal iónico incluyen, entre otros, la bomba ATPasa Na⁺-K⁺, la bomba de Ca²⁺ y el canal de fuga de K⁺.

Como se usa el presente documento, la expresión “proteína quinasa” se refiere a proteínas que catalizan la adición de un grupo fosfato desde un nucleósido trifosfato a una cadena lateral aminoacídica en una proteína. Las quinasas comprenden la superfamilia de enzimas más grande conocida y varían ampliamente en lo que respecta a sus proteínas diana. Las quinasas se pueden clasificar como proteínas tirosina quinasas (PTK), que fosforilan residuos de tirosina, y proteínas serina/treonina quinasas (STK), que fosforilan los residuos de serina y/o treonina. Algunas quinasas tienen especificidad doble por residuos de serina/treonina y tirosina. Casi todas las quinasas contienen un dominio catalítico conservado de 250-300 aminoácidos. Este dominio se puede dividir adicionalmente en 11 subdominios. Los subdominios I-IV en N-terminal se pliegan en una estructura de dos lóbulos que se une y orienta la molécula donante de ATP y el subdominio V abarca los dos lóbulos. Los subdominios VI-XI en C-terminal se unen al sustrato proteico y transfieren el fosfato gamma desde el ATP al grupo hidroxilo de un residuo de serina, treonina o tirosina. Cada uno de los 11 subdominios contiene residuos catalíticos específicos o motivos de aminoácidos característicos de dicho subdominio. Por ejemplo, el subdominio I contiene un motivo consenso de 8 aminoácidos de unión a ATP rico en glicina, el subdominio II contiene un residuo de lisina crucial necesario para una actividad catalítica máxima y los subdominios VI a IX comprenden el núcleo catalítico altamente conservado. Las STK y las PTK también contienen distintos motivos de secuencia en los subdominios VI y VIII, que pueden conferir especificidad por hidroxiaminoácido. Algunas STK y PTK poseen características estructurales de ambas familias. Además, las quinasas también se pueden clasificar mediante secuencias de aminoácidos adicionales, en general entre 5 y 100 residuos, que flanquean o están dentro del dominio quinasa.

Las PTK que no son transmembrana forman complejos de señalización con los dominios citosólicos de receptores de membrana plasmática. Los receptores que señalan a través de PTK que no son transmembrana incluyen citocinas, hormonas y receptores linfocíticos específicos de antígeno. Muchas PTK se identificaron primero como productos oncogénicos en células cancerosas en las que la activación de PTK ya no estaba sometida a los controles celulares normales. De hecho, aproximadamente un tercio de los oncogenes conocidos codifican PTK. Además, la transformación celular (oncogénesis) a menudo se acompaña de un incremento de la actividad de la fosforilación de la tirosina (véase, por ejemplo, Carbonneau, H. and Tonks, Annu. Rev. Cell Biol. 8:463-93 [1992]). Por tanto, la regulación de la actividad PTK puede ser una estrategia importante en el control de algunos tipos de cáncer.

Como se usa en el presente documento, la expresión “proteína fosfatasa” se refiere a proteínas que eliminan un grupo fosfato de una proteína. Las proteínas fosfatasas se dividen, en general, en dos grupos, proteínas de tipo receptor y de tipo no receptor. La mayoría de las proteínas tirosina fosfatasas de tipo receptor contienen dos dominios catalíticos conservados, cada uno de los cuales abarca un segmento de 240 residuos de aminoácidos. (Véase, por ejemplo, Saito y col., Cell Growth and Diff. 2:59-65 [1991]). Las proteínas tirosina fosfatasas de tipo receptor se pueden subclasificar además en base a la diversidad de la secuencia de aminoácidos de sus dominios extracelulares. (Véase, por ejemplo, Krueger y col., Proc. Natl. Acad. Sci. USA 89:7417-7421 [1992]).

Como se usa en el presente documento, “gen indicador” se refiere a un gen que codifica una proteína que se puede someter a ensayo. Ejemplos de genes indicadores incluyen, entre otros, luciferasa (véase, por ejemplo, de Wet y

co., Mol. Cell. Biol. 7:725 [1987] y las patentes de EE.UU. N° 6.074.859; 5.976.796; 5.674.713; y 5.618.682, todos ellos incorporados en el presente documento por referencia), la proteína fluorescente verde (p.ej., número de registro en GenBank U43284; una serie de variantes de GFP están disponibles comercialmente en CLONTECH Laboratories, Palo Alto, CA), cloranfenicol acetiltransferasa, β -galactosidasa, fosfatasa alcalina y peroxidasa de rábano.

Como se usa en el presente documento, el término "purificado" se refiere a moléculas, secuencias de ácido nucleico o de aminoácidos, que se eliminan de su ambiente natural, se aíslan o se separan. Una "secuencia de ácido nucleico aislada" es, por tanto, una secuencia de ácido nucleico purificada. Moléculas "sustancialmente purificadas" carecen de al menos un 60 %, preferentemente de al menos un 75 % y, más preferentemente de al menos un 90 %, de otros componentes con los que se asocian de forma natural.

Como se usa en el presente documento, las expresiones "memoria informática" y "dispositivo de memoria informática" hacen referencia a cualquier medio de almacenamiento legible por un procesador informático. Ejemplos de memoria informática incluyen, entre otros, RAM, ROM, circuitos de ordenador, discos de vídeo digital (DVD), discos compactos (CD), discos duros (HDD) y cintas magnéticas.

Como se usa en el presente documento, la expresión "medio legible por ordenador" hace referencia a cualquier dispositivo o sistema para almacenar y proporcionar información (p. ej., datos e instrucciones) a un procesador informático. Ejemplos de medios legibles por ordenador incluyen, entre otros, DVD, CD, discos duros, cintas magnéticas y servidores para transferir los medios sobre redes.

Como se usa en el presente documento, el término "introducir" como en "introducir dicha información sobre variación genética en dicho ordenador" hace referencia a la transferencia de información a un "medio legible por ordenador". La información se puede transferir mediante cualquier procedimiento adecuado, incluido, entre otros, manualmente (p. ej., tecleando en un ordenador) o automáticamente (p. ej., transferido a otro "medio legible por ordenador" mediante un "procesador").

Como se usa en el presente documento, los términos "procesador" y "unidad central de procesamiento" o "CPU" se usan de forma intercambiable y hacen referencia a un dispositivo que puede leer un programa de una memoria informática (p. ej., ROM u otra memoria informática) y realizar un conjunto de etapas de acuerdo con el programa.

Como se usa en el presente documento, la expresión "procedimiento implementado por ordenador" hace referencia a un procedimiento que usa una "CPU" y "medio legible por ordenador".

Descripción detallada de la invención

La presente invención está definida en las reivindicaciones y se refiere a moléculas de señalización intracelular, en concreto a variantes de la proteína y a los ácidos nucleicos que codifican las variantes de la proteína Nod2. Se divulgan formas tanto nativa como silvestre recombinante de Nod2, así como las formas mutante y variante, algunas de las cuales poseen características alteradas relativas a la Nod2 silvestre. La presente divulgación también se refiere a procedimientos de uso de Nod2, incluida la expresión alterada en organismos transgénicos y la expresión en procariontes y en sistemas de cultivo celular. La presente divulgación también abarca procedimientos para seleccionar fármacos que inhiben o potencian la acción de Nod2. La presente invención también se refiere a procedimientos de selección de la susceptibilidad a la enfermedad inflamatoria intestinal y la enfermedad de Crohn.

I. Polinucleótidos de Nod2

Como se ha descrito anteriormente se ha descubierto una familia nueva de proteínas que activan NF- κ B. Esta familia se identificó mediante detección selectiva en bases de datos públicas de secuencias de ácido nucleico que tienen homología con Nod1. De acuerdo con esto, la presente divulgación proporciona ácidos nucleicos que codifican genes de Nod2, homólogos, variantes (p. ej., mutaciones y polimorfismos), incluyendo, entre otros, los descritos en las SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88. La presente divulgación proporciona secuencias de polinucleótidos que son capaces de hibridar con las SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88 en condiciones de rigurosidad baja a alta, siempre que la secuencia de polinucleótidos capaz de hibridar codifique una proteína que conserva una actividad biológica de la proteína Nod2 de origen natural. La proteína que conserva una actividad biológica de Nod2 de origen natural puede tener una homología del 70 % con la Nod2 de tipo silvestre, una homología del 80 % con la Nod2 de tipo silvestre, una homología del 90 % con la Nod2 silvestre o una homología del 95 % con la Nod2 de tipo silvestre. Las condiciones de hibridación se basan en la temperatura de fusión (T_m) del complejo de unión al ácido nucleico y confieren una "rigurosidad" definida, como se ha explicado anteriormente (véase, por ejemplo, Wahl, et al., Meth. Enzymol., 152:399-407 [1987], que se incorpora en el presente documento por referencia).

Se proporcionan alelos adicionales de Nod2. En formas de realización preferidas, los alelos son el resultado de un polimorfismo o mutación (es decir, un cambio en la secuencia de ácido nucleico) y generalmente producen ARNm o polipéptidos alterados cuya estructura o función puede o no estar alterada. Cualquier gen dado puede tener ninguna, una o muchas formas alélicas. Los cambios mutacionales frecuentes que dan lugar a alelos generalmente se atribuyen a deleciones, adiciones o sustituciones de ácidos nucleicos. Cada uno de estos tipos de cambios se puede

producir solo o en combinación con los demás y a la velocidad de uno o mas veces en una secuencia dada. Los alelos de la presente invención incluyen los codificados por las SEC ID N° 33, 54, 56, 58, 84, y 88 (variantes).

Los ácidos nucleicos pueden codificar dos dominios CARD correspondientes (p. ej., secuencias de ácido nucleico que codifican los péptidos de SEC ID N° 5 y 6) o al menos un dominio seleccionado del grupo que consiste en un dominio NBD (p. ej., SEC ID N° 7) y un dominio de LRR (p. ej., las SEC ID N° 8-17) y el bucle P y los dominios de unión a Mg²⁺ (SEC ID N° 18-19).

Las secuencias nucleotídicas de la presente invención se pueden modificar mediante ingeniería con el fin de alterar una secuencia de codificación de Nod2 por diversas razones, incluyendo, entre otras, alteraciones que modifican la clonación, el procesamiento y/o la expresión del producto génico. Por ejemplo, se pueden introducir mutaciones usando técnicas bien conocidas en la materia (p. ej., mutagénesis dirigida a sitio para insertar nuevos sitios de restricción, para alterar los patrones de glucosilación, para cambiar la preferencia de codones etc.).

La secuencia de polinucleótidos de Nod2 se puede extender usando las secuencias de nucleótidos (p. ej., las SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88) en varios procedimientos conocidos en la materia para detectar secuencias cadena arriba, tales como promotores y elementos reguladores. Por ejemplo, se contempla que reacción en cadena de la polimerasa (PCR) en sitio de restricción encontrará utilidad en la presente invención. Este es un procedimiento directo que usa cebadores universales para recuperar la secuencia desconocida adyacente a un locus conocido (Gobinda y col., PCR Methods Applic., 2:318-22 [1993]). En primer lugar, el ADN genómico se amplifica en presencia de un cebador hasta una secuencia de unión y un cebador específico de la región conocida. Las secuencias amplificadas se someten a una segunda ronda de PCR con el mismo cebador de unión y otro cebador específico interno con respecto al primero. Los productos de cada ronda de PCR se transcriben con una ARN polimerasa adecuada y se secuencias usando la transcriptasa inversa.

Se puede usar PCR inversa para amplificar o extender las secuencias usando cebadores divergentes basados en una región conocida (Triglia y col., Nucleic Acids Res., 16:8186 [1988]). Los cebadores se pueden diseñar usando Oligo 4.0 (National Biosciences Inc., Plymouth, Minn.) u otro programa adecuado, de modo que tengan una longitud de 22-30 nucleótidos, un contenido de GC del 50% o más y que hibriden con la secuencia diana a temperaturas de aproximadamente 68-72°C. El procedimiento usa varias enzimas de restricción para generar un fragmento adecuado en la región conocida de un gen. A continuación, el fragmento se circulariza mediante ligadura intramolecular y se usa como molde para la PCR. En otros procedimientos más se usa la PCR circulante. La PCR circulante es un procedimiento para genes objetivo circulantes que permite la recuperación de secuencias desconocidas (Parker y col., Nucleic Acids Res., 19:3055-60 [1991]). El kit PROMOTERFINDER (Clontech) usa PCR, cebadores anidados y bibliotecas especiales para "entrar" en el ADN genómico. Este procedimiento evita la necesidad de buscar en bibliotecas y es útil para encontrar uniones intrones/exones.

Bibliotecas preferidas para la detección selectiva de ADNc de longitud completa incluyen bibliotecas de mamíferos que se han seleccionado según el tamaño de modo que incluyan ADNc de mayor tamaño. Asimismo, se prefieren las bibliotecas cebadas al azar porque contienen más secuencias que las que contienen en las regiones génicas 5' y cadena arriba. Una biblioteca cebada al azar puede ser especialmente útil en el caso en el que una biblioteca de oligo d(T) no da ADNc de longitud completa. Las bibliotecas genómicas de mamífero son útiles para obtener intrones y extender la secuencia en 5'.

En otras formas de realización de la presente invención se proporcionan variantes de las secuencias de Nod2 divulgadas. En formas de realización preferidas, las variantes son el resultado de polimorfismos o mutaciones (es decir, un cambio en la secuencia de ácido nucleico) y generalmente producen ARNm o polipéptidos alterados cuya estructura o función puede o no estar alterada. Cualquier gen dado puede tener ninguna, una o muchas formas variantes. Los cambios mutacionales habituales que dan lugar a variantes generalmente se atribuyen a deleciones, adiciones o sustituciones de ácidos nucleicos. Cada uno de estos tipos de cambios se puede producir solo o en combinación con los demás y a la velocidad de uno o mas veces en una secuencia dada.

Se ha contemplado que es posible modificar la estructura de un péptido que tiene una función (p. ej., función de Nod2) para propósitos tales como alterar (p. ej., aumentar o disminuir) la afinidad de unión de la Nod2 por RICK u otro regulador. Dichos péptidos modificados se consideran equivalentes funcionales de péptidos que tienen una actividad de Nod2 como se define en el presente documento. Se puede producir un péptido modificado en el que la secuencia de nucleótidos que codifica el polipéptido se ha alterado, tal como mediante sustitución, deleción o adición. En formas de realización particularmente preferidas, estas modificaciones no reducen significativamente la actividad sintética de la Nod2 modificada. En otras palabras, el constructor "X" se puede evaluar con el fin de determinar si es un miembro del género de Nod2 modificada o variante de la presente invención tal como se ha definido funcionalmente más que estructuralmente. La actividad de los polipéptidos Nod2 variantes se puede evaluar mediante los procedimientos descritos en el Ejemplo 4. De acuerdo con lo anterior, la presente divulgación proporciona ácidos nucleicos que codifican una Nod2 que activa NF-κB (p. ej., activa una respuesta inflamatoria). La actividad de una variante de Nod2 se puede evaluar transfectando las células HEK293T con una construcción de expresión que codifica la variante de Nod2 mutante. Las células pueden contener una construcción de luciferasa indicador que contiene regiones potenciadoras que responden a NF-κB. La variante de Nod2 puede ser capaz de unirse a una proteína (p. ej., RICK) pero no de activar NF-κB. Estas variantes se pueden someter a detección

selectiva mediante los procedimientos de inmunoprecipitación descritos en el Ejemplo 6.

Además, como se ha descrito anteriormente, las formas variantes de Nod2 también se contemplan como equivalentes de dichos péptidos y las moléculas de ADN que se establecen con más detalle en el presente documento. Por ejemplo, se contempla una sustitución aislada de leucina con isoleucina o valina, un aspartato con un glutamato, una treonina con una serina o una sustitución conservadora similar de un aminoácido con un aminoácido estructuralmente relacionado (es decir, mutaciones conservadoras), no tendrá un efecto importante sobre la actividad biológica de la molécula resultante. De acuerdo con esto, algunas realizaciones de la presente invención proporcionan variantes de Nod2 divulgadas en el presente documento que contienen sustituciones conservadoras. Las sustituciones conservadoras son las que tienen lugar dentro de una familia de aminoácidos que están relacionados por sus cadenas laterales. Los aminoácidos codificados genéricamente se pueden dividir en cuatro familias: (1) ácidos (aspartato, glutamato); (2) básicos (lisina, arginina, histidina); (3) no polares (alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano); y (4) polares sin carga (glicina, asparagina, glutamina, cisteína, serina, treonina, tirosina). En ocasiones, los aminoácidos fenilalanina, triptófano y tirosina se clasifican en conjunto como aminoácidos aromáticos. De un modo similar, el repertorio de aminoácidos se puede agrupar como (1) ácidos (Aspartato, glutamato), (2) básicos (lisina, arginina, histidina), (3) alifáticos (glicina, alanina, valina, leucina, isoleucina, serina, treonina), estando la serina y la treonina opcionalmente agrupadas por separado como alifáticos-hidroxilo; (4) aromáticos (fenilalanina, tirosina, triptófano); (5) amida (asparagina, glutamina); y (6) que contienen azufre (cisteína y metionina) (p. ej., Stryer ed., Biochemistry, pág. 17-21, 2ª ed, WH Freeman y Co., 1981). Si un cambio en la secuencia de aminoácidos de un péptido tiene como resultado un polipéptido funcional se puede determinar fácilmente evaluando la capacidad del péptido variante para funcionar de un modo similar a la proteína de tipo silvestre. Los péptidos que tienen más de una sustitución se pueden someter a ensayo fácilmente del mismo modo.

Más infrecuentemente, una variante incluye cambios “no conservadores” (p. ej., sustitución de una glicina con un triptófano). Variaciones minoritarias análogas también pueden incluir deleciones o inserciones de aminoácidos, o ambas. Usando programas informáticos (p. ej., LASERGENE software, DNASTAR Inc., Madison, Wis.) se pueden encontrar guías para determinar qué residuos de aminoácidos se pueden sustituir, insertar o delecionar sin anular la actividad biológica.

Como se describe con mayor detalle más adelante se pueden producir variantes mediante procedimientos tales como evolución dirigida u otras técnicas para producir bibliotecas combinatorias de variantes, descritas con mayor detalle más adelante. Las secuencias nucleotídicas de la presente invención se pueden modificar mediante ingeniería con el fin de alterar una secuencia de codificación de Nod2, incluyendo, entre otras, alteraciones que modifican la clonación, el procesamiento, la localización y/o la expresión del producto génico. Por ejemplo, se pueden introducir mutaciones usando técnicas que son bien conocidas en la técnica (p. ej., mutagénesis dirigida a sitio para insertar nuevos sitios de restricción, para alterar los patrones de glicosilación, para cambiar la preferencia de codones, etc.).

II. Polipéptidos de Nod2

La presente divulgación proporciona secuencias de polinucleótidos de Nod2 que codifican las secuencias de polipéptidos de Nod2. Los polipéptidos de Nod2 (p. ej., las SEC ID N° 2, 3, 34, 55, 57, 59, 61, 63, 65, 67, 69, 85, 87, and 89) se describen en las Figuras 13, 14, 15, 22, 24, 28, 30, 32, 34, 36, 38, 40, 42 y 44. También se divulgan fragmentos, proteínas de fusión o equivalentes funcionales de estas proteínas Nod2. Las secuencias de ácido nucleico correspondientes a variantes, homólogos y mutantes de Nod2 para generar moléculas de ADN recombinantes que dirigen la expresión de las variantes homólogos y mutantes de Nod2 en las células huésped adecuadas. El polipéptido puede ser un producto purificado natural o un producto de procedimientos sintéticos químicos o se puede producir mediante técnicas recombinantes usando un huésped procarionta o eucariota (p. ej., usando células bacterianas, de levadura, de plantas superiores, de insectos y de mamíferos en cultivo). Dependiendo del huésped empleado en un procedimiento de producción recombinante, el polipéptido puede estar glicosilado o no glicosilado. Los polipéptidos pueden también incluir un residuo de aminoácido metionina inicial.

Debido a la degeneración inherente del código genético, se pueden usar secuencias de ADN distintas de las secuencias de polinucleótidos de la SEC ID N° 1 que codifican sustancialmente la misma secuencia de aminoácidos o una funcionalmente equivalente para clonar y expresar Nod2. En general, dichas secuencias de polinucleótidos hibridan con la SEC ID N° 1 en condiciones de rigurosidad alta a media como se ha descrito en lo que antecede. Como los expertos en la técnica entenderán, puede ser ventajoso producir secuencias de nucleótidos que codifican Nod2 que poseen codones no naturales. Por tanto, los codones preferidos por un huésped procarionta o eucariota concreto (Murray y col., Nucl. Acids Res., 17 [1989]) se seleccionan, por ejemplo, para incrementar el índice de expresión de Nod2 o para producir transcritos de ARN recombinante que posee propiedades deseables tales como una semivida mayor que la de los transcritos producidos a partir de la secuencia natural.

1. Vectores para la producción de Nod2

Los polinucleótidos de la presente invención se pueden usar para producir polipéptidos mediante técnicas recombinantes. Por tanto, por ejemplo, el polinucleótido se puede incluir en uno cualquiera de diversos vectores de expresión para expresar un polipéptido. En algunas formas de realización de la presente invención, los vectores

incluyen, entre otros, secuencias de ADN cromosómico, no cromosómico y sintético, derivados de SV40; plásmidos bacterianos, ADN de fago, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN de fago, ADN viral, tal como de vaccinia, adenovirus, virus de la viruela aviar y seudorrabia. Se contempla que se puede usar cualquier vector siempre que sea replicable y viable en la célula huésped.

5 En concreto, algunas formas de realización de la presente invención proporcionan constructos recombinantes que comprenden una o más de las secuencias como se ha descrito ampliamente en lo que antecede (p. ej., las SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88). En algunas formas de realización de la presente invención, los constructos comprenden un vector, tal como un plásmido o vector viral, en el que se ha insertado una secuencia de la invención en orientación directa o inversa. En otras formas de realización, la secuencia estructura heteróloga (p. ej., SEC ID N° 1) se ensambla en la fase adecuada con secuencias de inicio y de terminación de la traducción. En formas de realización preferidas de la presente invención, la secuencia de ADN adecuada se inserta en el vector usando cualquiera de diversos procedimientos. En general, la secuencia de ADN se inserta en un sitio(s) de endonucleasa de restricción adecuado(s) mediante procedimientos conocidos en la técnica.

15 Los expertos en la técnica conocen un número elevado de vectores adecuados y comercialmente disponibles. Dichos vectores incluyen, entre otros, los vectores siguientes: 1) Bacterianos: pQE70, pQE60, pQE-9 (Qiagen), pBS, pD10, phagescript, six174, pbluescript SK, pBSKS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); 2) Eukaryotic-pWLNEO, pSV2CAT, pOG44, PXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia); y 3) Baculovirus - pPbac y pMbac (Stratagene). Se puede usar otro plásmido o vector siempre que sea replicable y viable en la célula huésped. En algunas formas de realización preferidas de la presente invención, los vectores de expresión en mamíferos comprenden un origen de replicación, un promotor y potenciador adecuados, y, también, cualquier sitio de unión al ribosoma necesario, sitios de poliadenilación, sitios donantes y aceptores de corte y empalme, secuencias de terminación de la transcripción y secuencias no transcritas flanqueantes en 5'. En otras formas de realización, las secuencias de ADN derivadas de los sitios de corte y empalme y de poliadenilación de SV40 se pueden usar para proporcionar los elementos genéticos no transcritos requeridos.

20 En ciertas formas de realización de la presente invención, la secuencia de ADN en el vector de expresión está unida operablemente a una o más secuencias de control de la expresión (promotores) adecuadas para dirigir la síntesis directa de ARNm. Los promotores útiles en la presente invención incluyen, entre otros, el promotor de LTR o SV40, el *E. coli lac* o *trp*, el fago *lambda* P_L y P_R, los promotores de T3 y T7 y los promotores temprano inmediato de citomegalovirus (CMV), de timidina quinasa del virus del herpes simple (VHS) y de la metalotioneína-I de ratón, y otros promotores conocidos porque controlan la expresión génica en células procariotas y eucariotas o sus virus. En otras formas de realización de la presente invención, los vectores de expresión recombinante incluyen orígenes de replicación y marcadores seleccionables que permiten la transformación de la célula huésped (p. ej., la dihidrofolato reductasa o resistencia a la neomicina para cultivo celular eucariótico, o resistencia a tetraciclina o ampicilina en *E. coli*).

35 En algunas formas de realización de la presente invención, la transcripción del ADN que codifica los polipéptidos de la presente invención por eucariotas superiores aumenta insertando una secuencia potenciadora en el vector. Los potenciadores son elementos de ADN que actúan en cis, normalmente de una longitud de aproximadamente 10 a 300 pb, que actúan sobre el promotor para aumentar la transcripción. Potenciadores útiles en la presente invención incluyen, entre otros, el potenciador de SV40 en el lado tardío del origen de replicación, pb 100 a 270 pb, un potenciador del promotor temprano del citomegalovirus, el potenciador del polioma en el lado tardío del origen de replicación, y los potenciadores de adenovirus.

40 En otras formas de realización, el vector de expresión también contiene un sitio de unión al ribosoma para la iniciación de la traducción y un terminador de la transcripción. En otras formas de realización más de la presente invención, el vector puede incluir también secuencias adecuadas para amplificar la expresión.

2. Células huésped para la producción de Nod2

45 En una forma de realización adicional, la presente invención proporciona células huésped que contienen las construcciones descritas en lo que antecede. En algunas formas de realización de la presente invención, la célula huésped es una célula eucariota superior (p. ej., una célula de mamífero o de insecto). En otras formas de realización de la presente invención, la célula huésped es una célula eucariota inferior (p. ej., una célula de levadura). En otras formas de realización más de la presente invención, la célula huésped es una célula procariota (p. ej., una célula bacteriana). Ejemplos específicos de células huésped incluyen, entre otras, *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*, y varias especies de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*, así como *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, células de *Drosophila* S2, células de *Spodoptera Sf9*, células de ovario de hámster chino (CHO), líneas celulares COS-7 de fibroblastos de riñón de mono, (Gluzman, Cell 23:175 [1981]), líneas celulares C127, 3T3, 293, 293T, HeLa y BHK.

50 Las construcciones en las células huésped se pueden usar de un modo convencional para producir el producto génico codificado por la secuencia recombinante. En algunas formas de realización, la introducción de la construcción en la célula huésped puede efectuarse mediante transfección mediada por fosfato cálcico, transfección mediada por DEAE-dextrano o electroporación (Véase, por ejemplo, Davis y col., Basic Methods in Molecular

Biology, [1986]). Como alternativa, en algunas formas de realización de la presente invención, los polipéptidos de la invención se pueden producir de forma sintética mediante sintetizadores peptídicos convencionales.

Las proteínas se pueden expresar en células de mamífero, levaduras, bacterias u otras células bajo el control de los promotores adecuados. También se pueden emplear sistemas de traducción acelulares para producir dichas proteínas usando ARN derivados de las construcciones de ADN de la presente invención. En, Sambrook y col., Molecular Cloning: A Laboratory Manual, 2ª Ed., Cold Spring Harbor, N.Y., (1989) se describen vectores de clonación y expresión adecuados para usar con huéspedes procariotas y eucariotas. A Laboratory Manual, Segunda Edición, Cold Spring Harbor, N. Y., [1989].

En algunas formas de realización de la presente invención, tras la transformación de una cepa huésped adecuada y el crecimiento de la cepa huésped hasta una densidad celular adecuada, el promotor seleccionados de induce mediante medios adecuados (p. ej., cambio de temperatura o inducción química) y las células se cultivan durante un periodo adicional. En otras formas de realización de la presente invención, normalmente las células se recogen mediante centrifugación, se rompen por medios físicos o químicos y el extracto bruto resultante se conserva para su posterior purificación. En otras formas de realización más de la presente invención, las células microbianas usadas en la expresión de proteínas se pueden romper mediante cualquier procedimiento conveniente, incluidos ciclos de congelación-descongelación, sonicación, fragmentación mecánica o uso de agentes de lisis celular.

3. Purificación de Nod2

La presente invención también proporciona procedimientos para recuperar y purificar polipéptidos de Nod2 de cultivos de células recombinantes incluidos, entre otros, precipitación con sulfato amónico o etanol, extracción ácida, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía de hidroxilapatita y cromatografía de lectina. En otras formas de realización de la presente invención, las etapas de replegamiento de proteínas se pueden usar según sea necesario, para completar la configuración de la proteína madura. En otras formas de realización más de la presente invención, la cromatografía de líquidos de alto rendimiento (HPLC) se puede usar en las etapas de purificación final.

La presente invención proporciona además polinucleótidos que tiene la secuencia de codificación (p. ej., las SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88) condensada en el marco con una secuencia marcadora que permite la purificación del polipéptido de la presente invención. Un ejemplo no limitante de una secuencia marcadora es una cola de hexahistidina que puede ser suministrada por un vector, preferentemente un vector pQE-9, que proporciona la purificación del polipéptido condensado con el marcador en el caso de un huésped bacteriano o, por ejemplo, la secuencia marcadora puede ser una cola de hemaglutinina (HA) cuando se usa un huésped mamífero (p. ej., células COS-7). La cola de HA corresponde a un epítipo derivado de la proteína hemaglutinina de la gripe (Wilson y col., Cell, 37:767 [1984]).

4. Mutantes de truncamiento de Nod2

Además, la presente invención proporciona fragmentos de Nod2 (es decir, mutantes de truncamiento, por ejemplo la SEC ID N° 3). En algunas formas de realización de la presente invención, cuando se desea la expresión de una porción de la proteína Nod2 puede ser necesario añadir un codón de iniciación (ATG) al fragmento oligonucleotídico que contiene la secuencia deseada que se va a expresar. En la técnica se conoce bien que una metionina en la posición N-terminal se puede escindir enzimáticamente mediante el uso de la enzima metionina aminopeptidasa (MAP). La MAP se ha clonado en *E. coli* (Ben-Bassat et al., J. Bacteriol., 169:751 [1987]) y *Salmonella typhimurium* y su actividad *in vitro* se ha demostrado en proteínas recombinantes (Miller et al., Proc. Natl. Acad. Sci. USA 84:2718-1722 [1990]). Por tanto, la eliminación de una metionina en N-terminal, si se desea, se puede conseguir *in vivo* expresando dichos polipéptidos recombinantes en un huésped que produce MAP (p. ej., *E. coli* o CM89 o *S. cerevisiae*), o *in vitro* mediante el uso de MAP purificada.

5. Proteínas de fusión que contienen Nod2

La presente invención también proporciona proteínas de fusión que incorporan toda o parte de Nod2. De acuerdo con esto, en algunas formas de realización de la presente invención, las secuencias de codificación para el polipéptido se pueden incorporar como parte de un gen de fusión que incluye una secuencia de nucleótidos que codifica un polipéptido diferente. Se contempla que este tipo de sistema de expresión encontrará utilidad en las condiciones en las que sea deseable producir un fragmento inmunogénico de una proteína Nod2. En algunas formas de realización de la presente invención, la proteína de la cápsida VP6 de rotavirus se usa como proteína vehículo inmunológica para porciones del polipéptido Nod2, bien en forma monomérica o bien en forma de una partícula viral. En otras formas de realización de la presente invención, las secuencias de ácido nucleico correspondientes a la porción de Nod2 contra la cual se deben producir anticuerpos se pueden incorporar en una construcción génica de fusión que incluye secuencias de codificación para una proteína estructural tardía del virus vacunal para producir un conjunto de virus recombinantes que expresan las proteínas de fusión que comprenden una porción de Nod2 como parte del virión. Se ha demostrado, con el uso de proteínas de fusión inmunogénicas que usan las proteínas de fusión del antígeno de superficie de la hepatitis B, que los viriones de la hepatitis B recombinantes se pueden usar también en esta función. De un modo similar, en otras formas de realización de la presente invención, las construcciones quiméricas que codifican proteínas de fusión que contienen una porción de Nod2 y la proteína de la

cápside del poliovirus se crean para potenciar la inmunogenicidad del conjunto de antígenos polipeptídicos (véase, por ejemplo, la publicación EP N° 025949; y Evans y col., *Nature* 339:385 [1989]; Huang y col., *J. Virol.*, 62:3855 [1988]; y Schlienger y col., *J. Virol.*, 66:2 [1992]).

5 En otras formas de realización más de la presente invención se puede usar el sistema de múltiples antígenos peptídicos para inmunización basada en péptidos. En este sistema se obtiene una porción deseada de Nod2 directamente de la síntesis órgano-química del péptido sobre un núcleo de lisina ramificado oligomérico (véase, por ejemplo, Posnett y col., *J. Biol. Chem.*, 263:1719 [1988]; y Nardelli y col., *J. Immunol.*, 148:914 [1992]). En otras formas de realización de la presente invención también se pueden expresar determinantes antigénicos de las proteínas Nod2, que pueden estar presentados por células bacterianas.

10 Además de usar proteínas de fusión para potenciar la inmunogenicidad, está ampliamente apreciado que las proteínas de fusión pueden también facilitar la expresión de proteínas, tales como la proteína Nod2 de la presente invención. De acuerdo con esto, en algunas realizaciones de la presente invención, Nod2 se puede generar como glutatión-S-transferasa (es decir, proteína de fusión GST). Se contempla que dichas proteínas de fusión GST permitirán una fácil purificación de Nod2, tal como mediante el uso de matrices derivadas de glutatión (véase, por ejemplo, Ausabel y col., (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, NY [1991]). En otra forma de realización de la presente invención, un gen de fusión que codifica una secuencia líder de purificación, tal como una secuencia del sitio de escisión de la poli-(His)/enteroquinasa en el extremo N de la porción deseada de Nod2 puede permitir la purificación de la proteína de fusión Nod2 expresada mediante cromatografía de afinidad usando una resina metálica de Ni²⁺. En otra forma de realización más de la presente invención, la secuencia líder de purificación se puede eliminar después mediante tratamiento con enteroquinasa (véase, por ejemplo, Hochuli y col., *J. Chromatogr.*, 411:177 [1987]; y Janknecht y col., *Proc. Natl. Acad. Sci. USA* 88:8972).

25 Las técnicas para fabricar genes de fusión son bien conocidas. Esencialmente, la unión de varios fragmentos de ADN que codifica diferentes secuencias polipeptídicas se realiza de acuerdo con técnicas convencionales usando extremos romos o escalonados para la ligadura, digestión con enzimas de restricción para proporcionar extremos adecuados, llenado con extremos cohesivos según sea adecuado, tratamiento con fosfatasa alcalina para evitar uniones no deseadas y ligadura enzimática. En otra forma de realización de la presente invención, el gen de fusión se puede sintetizar mediante técnicas convencionales, incluidos sintetizadores de ADN automáticos. Como alternativa, en otras formas de realización de la presente invención se puede llevar a cabo la amplificación de fragmentos génicos usando cebadores de ancla que dan lugar a salientes complementarios entre dos fragmentos génicos consecutivos que después se pueden hibridar para generar una secuencia génica quimérica (véase, por ejemplo, *Current Protocols in Molecular Biology*, citado anteriormente).

6. Variantes de Nod2

35 Otras formas de realización más de la presente invención proporcionan formas mutantes o variantes de Nod2 (es decir, muteínas). Es posible modificar la estructura de un péptido que tiene una actividad de Nod2 para fines tales como potenciar la eficacia terapéutica o profiláctica, o la estabilidad (p. ej., duración *ex vivo* y/o resistencia a la degradación proteolítica *in vivo*). Dichos péptidos modificados se consideran equivalentes funcionales de péptidos que tienen una actividad de las proteínas Nod2 objeto como se define en el presente documento. Se puede producir un péptido modificado en el que la secuencia de aminoácidos se ha alterado, tal como mediante sustitución, delección o adición de aminoácidos.

40 Además, como se ha descrito anteriormente, las formas variantes (p. ej., secuencias mutantes o polimórficas) de las proteínas Nod2 sujeto también se contemplan como equivalentes de dichos péptidos y las moléculas de ADN que se establecen con más detalle en el presente documento. Por ejemplo, como se ha descrito anteriormente, la presente invención abarca proteínas mutantes y variantes que contienen sustituciones de aminoácidos conservadoras o no conservadoras.

45 La presente invención contempla adicionalmente un procedimiento de generar conjuntos de mutantes combinatorias de las presentes proteínas Nod2, así como mutantes de truncamiento, y es especialmente útil para identificar potenciales secuencias variantes (es decir, mutantes o secuencias polimórficas) que son funcionales en la unión a RICK u otros reguladores en la vía de señalización de NF-κB y la señalización de una respuesta inflamatoria. El objetivo de la detección selectiva de dichas bibliotecas combinatorias es generar, por ejemplo, nuevas variantes de Nod2 que pueden actuar como agonistas o antagonistas, o, como alternativa, poseen nuevas actividades en conjunto.

55 Por tanto, en algunas formas de realización de la presente invención, las variantes de Nod2 se someten a ingeniería mediante el presente procedimiento para proporcionar una activación alterada (aumentada o disminuida) de NF-κB (es decir, generar una respuesta inflamatoria). En otras formas de realización de la presente invención se generan variantes derivadas mediante combinatoria que tienen una potencia selectiva respecto a una Nod2 natural. Dichas proteínas, cuando se expresan en construcciones de ADN recombinantes, se pueden usar en protocolos de terapia génica.

Otras formas de realización más de la presente invención proporcionan variantes de Nod2 que tienen semividas intracelulares espectacularmente diferentes a las correspondientes a la proteína silvestre. Por ejemplo, la proteína

alterada se puede dar más estable o menos estable a la degradación proteolítica o a otros procesos celulares que tengan como resultado la destrucción de Nod2 o inactivarla. Dichas variantes, y los genes que las codifican, se pueden usar para alterar la localización de la expresión de Nod2 mediante modulación de la semivida de la proteína. Por ejemplo, una semivida corta puede dar lugar a efectos biológicos de Nod2 más transitorios y, cuando son parte de un sistema de expresión inducible, pueden permitir un control más estrecho de los niveles de Nod2 dentro de la célula. Como se ha indicado anteriormente, dichas proteínas, y particularmente sus construcciones de ácido nucleico recombinante, se pueden usar en protocolos de terapia génica.

En otras formas de realización más de la presente invención, las variantes de Nod2 se generan mediante el abordaje combinatorio para actuar como antagonistas, en cuanto a que son capaces de interferir en la capacidad de la correspondiente proteína silvestre para regular la función celular.

En algunas formas de realización del abordaje de mutagénesis combinatoria de la presente invención, las secuencias de aminoácidos para una población de homólogos, variantes de Nod2 u otras proteínas relacionadas se alinean, preferentemente para estimular la mayor homología posible. Dicha población de variantes puede incluir, por ejemplo, homólogos de Nod2 de una o más especies o variantes de Nod2 de la misma especie pero que difieren debido a la mutación o polimorfismos. Los aminoácidos que aparecen en cada posición de las secuencias alineadas se seleccionan para crear un conjunto degenerado de secuencias combinatorias.

En una forma de realización preferida de la presente invención, la biblioteca de Nod2 combinatoria se produce mediante una biblioteca degenerada de genes que codifica una biblioteca de polipéptidos cada uno de los cuales incluye al menos una porción de posibles secuencias de la proteína Nod2. Por ejemplo, una mezcla de oligonucleótidos sintéticos se pueden ligar enzimáticamente en secuencias génicas de modo que el conjunto degenerado de posibles secuencias de Nod2 se puedan expresar como polipéptidos individuales o, como alternativa, como un conjunto de proteínas de fusión más grandes (p. ej., para expresión en fagos) que contienen el conjunto de secuencias de Nod2 en las mismas.

Hay muchos modos por los que se puede generar la biblioteca de posibles homólogos y variantes de Nod2 a partir de una secuencia de oligonucleótidos degenerada. En algunas formas de realización, la síntesis química de una secuencia génica degenerada se lleva a cabo en un sintetizador de ADN automático y los genes sintéticos se ligan en un gen adecuado para expresión. El fin de un conjunto degenerado de genes es proporcionar, en una mezcla, todas las secuencias que codifican el conjunto deseado de potenciales secuencias de Nod2. La síntesis de oligonucleótidos degenerados se conoce bien en la técnica (véase, por ejemplo, Narang, *Tetrahedron Lett.*, 39:39-9 [1983]; Itakura y col., *Recombinant DNA*, en Walton (ed.), *Proceedings of the 3rd Cleveland Symposium on Macromolecules*, Elsevier, Amsterdam, pp 273-289 [1981]; Itakura y col., *Annu. Rev. Biochem.*, 53:323 [1984]; Itakura y col., *Science* 198:1056 [1984]; Ike y col., *Nucl. Acid Res.*, 11:477 [1983]). Dichas técnicas se han empleado en la evolución dirigida de otras proteínas (véase, p. ej., Scott et al., *Science* 249:386 [1980]; Roberts et al., *Proc. Natl. Acad. Sci. USA* 89:2429 [1992]; Devlin et al., *Science* 249: 404 [1990]; Cwirla et al., *Proc. Natl. Acad. Sci. USA* 87: 6378 [1990]; así como en las patentes de EE.UU. N° 5.223.409; 5.198.346 y 5.096.815, cada una de las cuales incorporada en el presente documento por referencia).

Se contempla que los ácidos nucleicos de Nod2 (p. ej., SEC ID N° 1 y fragmentos y variantes de los mismos) se pueden usar como ácidos nucleicos de partida para la evolución dirigida. Esas técnicas se pueden usar para desarrollar variantes de Nod2 que tienen propiedades deseables, tales como una afinidad de unión incrementada o disminuida por RICK.

En algunas formas de realización, la evolución artificial se realiza mediante mutagénesis aleatoria (p. ej., usando PCR propensa a error para introducir mutaciones aleatorias en una secuencia de codificación dada). Este procedimiento requiere un ajuste fino de la frecuencia de mutación. Como norma general, las mutaciones beneficiosas son raras, mientras que las mutaciones perjudiciales son infrecuentes. Esto es porque la combinación de una mutación perjudicial y una mutación beneficiosa a menudo tiene como resultado una enzima inactiva. El número ideal de sustituciones de bases para genes objetivo normalmente está entre 1,5 y 5 (Moore y Arnold, *Nat. Biotech.*, 14, 458-67 [1996]; Leung y col., *Technique*, 1:11-15 [1989]; Eckert y Kunkel, *PCR Methods Appl.*, 1:17-24 [1991]; Caldwell y Joyce, *PCR Methods Appl.*, 2:28-33 [1992]; y Zhao y Arnold, *Nuc. Acids. Res.*, 25:1307 [1997]). Tras la mutagénesis, los clones resultantes se seleccionan según la actividad deseable (p. ej., detección selectiva según la actividad de Nod2). A menudo son necesarias rondas sucesivas de mutagénesis y selección para desarrollar enzimas con propiedades deseables. Cabe destacar que solo las mutaciones útiles se pasan a la siguiente ronda de mutagénesis.

En otras realizaciones de la presente invención, los polinucleótidos de la presente invención se usan en procedimientos de transposición genética o PCR sexual (p. ej., Smith, *Nature*, 370:324-25 [1994]; patentes de EE.UU. n° 5.837.458, 5.830.721, 5.811.238, 5.733.731, todas las cuales se incorporan en el presente documento por referencia). La transposición genética implica fragmentación aleatoria de varios ADN mutantes, seguido de su reensamblaje mediante PCR en moléculas de longitud completa. Ejemplos de varios procedimientos de transposición genética incluyen, entre otros, ensamblaje tras tratamiento con ADNasa, el procedimiento de extensión escalonada (STEP) y cebado aleatorio en recombinación *in vitro*. En el procedimiento mediado por la ADNasa, los segmentos de ADN aislados de un conjunto de mutantes positivos se escinden en fragmentos aleatorios con

ADNasa I y se someten a múltiples rondas de PCR sin cebador añadido. Las longitudes de fragmentos aleatorios se acercan a las del segmento sin escindir a medida que progresan los ciclos de PCR, lo que tiene como resultado la mezcla de mutaciones presentes en clones diferentes y la acumulación en algunas de las secuencias resultantes. Múltiples ciclos de selección y transposición han conducido a la potenciación funcional de una serie de enzimas (Stemmer, *Nature*, 370:398-91 [1994]; Stemmer, *Proc. Natl. Acad. Sci. USA*, 91:10747 [1994]; Cramer et al., *Nat. Biotech.*, 14:315 [1996]; Zhang et al., *Proc. Natl. Acad. Sci. USA*, 94:4504 [1997]; and Cramer et al., *Nat. Biotech.*, 15:436 [1997]). Las variantes producidas mediante evolución dirigida se puede someter a detección selectiva según la actividad de Nod2 mediante los procedimientos descritos en los Ejemplos 4-8.

En la materia se conoce una amplia gama de técnicas para la detección selectiva de productos génicos de bibliotecas combinatorias fabricados mediante mutaciones puntuales y para la detección selectiva de bibliotecas de ADNc de productos génicos que tienen una propiedad determinada. En general, dichas técnicas serán adaptables para una detección selectiva rápida de bibliotecas génicas generadas mediante la mutagénesis combinatoria o recombinación de homólogos o variantes de Nod2. Las técnicas más usadas para la detección selectiva de grandes bibliotecas génicas habitualmente comprende la clonación de la biblioteca génica en vectores de expresión replicables, transformando las células adecuadas con la biblioteca de vectores resultante y expresando los genes combinatorios en condiciones en las que la detección de una actividad deseada facilita un aislamiento relativamente fácil del vector que codifica el gen cuyo producto se detectó.

7. Síntesis química de Nod2

En una forma de realización alternativa de la invención, la secuencia de codificación de Nod2 se sintetiza, completa o en parte, usando procedimientos químicos bien conocidos en la técnica (véase, por ejemplo, Caruthers y col., *Nucl. Acids Res. Symp. Ser.*, 7: 215 [1980]; Crea and Horn, *Nucl. Acids Res.*, 9: 2331 [1980] ; Matteucci and Caruthers, *Tetrahedron Lett.*, 21: 719 [1980]; y Chow and Kempe, *Nucl. Acids Res.*, 9: 2807 [1981]). En otras formas de realización de la presente invención, la propia proteína se produce usando procedimientos químicos para sintetizar una secuencia de aminoácidos completa de Nod2 o una porción de la misma. Por ejemplo, los péptidos se pueden sintetizar mediante técnicas de fase sólida, se escinden de la resina y se purifican mediante cromatografía de líquidos preparativa de alto rendimiento (véase, por ejemplo, Creighton, *Proteins Structures And Molecular Principles*, W H Freeman y Co, New York N.Y. [1983]). En otras formas de realización de la presente invención, la composición de los péptidos sintéticos se confirma mediante análisis o secuenciación de aminoácidos (véase, por ejemplo, Creighton, citado anteriormente).

La síntesis peptídica directa se puede realizar usando varias técnicas de fase sólida (Roberge y col., *Science* 269: 202-204 [1995]) y la síntesis automática se puede conseguir, por ejemplo, usando el Sintetizador Peptídico ABI431A (Perkin Elmer) de acuerdo con las instrucciones proporcionadas por el fabricante. Adicionalmente, durante la síntesis directa, la secuencia de aminoácidos de Nod2, o cualquier parte de la misma, se puede alterar y/o combinar mediante procedimientos químicos con otras secuencias para producir una variante del polipéptido.

III. Detección de alelos de Nod2

A. Alelos de Nod2

La presente invención incluye alelos de Nod2 que incrementan la susceptibilidad de un paciente a la enfermedad de Crohn (p. ej., incluidas, entre otras, las SEC ID N° 33. El análisis de los alelos de Nod2 humana natural reveló que los pacientes con mayor susceptibilidad a la enfermedad de Crohn tienen un alelo mutante de Nod2 que, por ejemplo, contiene un residuo citosina adicional (p. ej., SEC ID N° 33;3020InsC o Nod Δ 33). El residuo citosina adicional produce una mutación de desplazamiento del marco que tiene como resultado la generación de un codón de terminación que produce la delección de una porción del dominio LRR. La expresión del alelo de Nod2 de la enfermedad de Crohn de SEC ID N° 33 en ausencia de LPS indujo la activación de NF- κ B (Fig. 19a). También se analizó la capacidad de las proteínas Nod2 para potenciar la activación de NF- κ B tras la incubación con LPS de varias bacterias. El LPS de varias bacterias indujo activación de NF- κ B en células que expresan Nod2 de tipo silvestre, mientras que no se observó una inducción significativa de la actividad del gen indicador en células transfectadas con plásmido control (Fig. 19b). De forma significativa, la capacidad del mutante de Nod2 para conferir capacidad de respuesta al LPS disminuyó considerablemente cuando se comparó con la Nod2 de tipo silvestre (Fig. 19b).

La presente invención no está limitada a ningún mecanismo de acción concreto. De hecho, no es necesario conocer el mecanismo de acción para la práctica de la presente invención. No obstante, se contempla que el sistema de señalización del que Nod2 es un componente reconoce patógenos bacterianos y virales e inicia una respuesta para matar al patógeno o célula infectada (es decir, señalización celular que activa un factor de transcripción que, a su vez, activa una respuesta inflamatoria. Se contempla que, en la enfermedad de Crohn, la activación de la vía de señalización se produce en ausencia de estimulación por patógenos por la presencia de la forma truncada de Nod2. Esto conduce a la inflamación asociada con la enfermedad de Crohn.

La presente divulgación no se limita a la mutación descrita en las SEC ID N° 3 y 33, pero considera que cualquier mutación que tiene como resultado el fenotipo indeseado (p. ej., un grado elevado de activación de NF- κ B en ausencia de otros estímulos de señalización o sensibilidad aumentada a la enfermedad de Crohn. Los ensayos para

determinar si un polipéptido dado tiene dichas actividades se proporcionan en los ejemplos 4 y 5.

5 Por ejemplo, la presente divulgación proporciona alelos que contienen uno o más cambios de un solo nucleótido de Nod2 (p. ej., secuencias mutantes o polimórficas) (p. ej., incluyendo, entre otros, las secuencias de ácidos nucleicos descritos en las SEC ID N° 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88. El ejemplo 10 y las Figuras 26 y 27 describen polimorfismos ilustrativos y su prevalencia solo y en combinación. Los ejemplos 9 y 19 describen la asociación de tres polimorfismos de Nod2 con enfermedad de Crohn. El ejemplo 10 describe la asociación de una o más de las mutaciones descritas anteriormente con un mayor riesgo de desarrollar enfermedad de Crohn. La Tabla 1 describe secuencias de Nod2 para los ácidos nucleicos y polipéptidos de tipo silvestre y mutantes.

Tabla 1		
Secuencia	SEC ID Número	Número de Figura
ADNc de Nod2 de tipo silvestre	1	12
Inserción de citosina en ADNc de Nod2	33	11
Aminoácido de Nod2a de tipo silvestre	2	13
Aminoácido de Nod2b de tipo silvestre	3	14
Aminoácido de Nod2a Δ 33	34	15
Exón 11 de Nod2 de tipo silvestre	35	16
Exón 11 de Nod2 Δ 33	36	16
ADNc de Nod2 de tipo silvestre	53	20
Ácido nucleico Nod2 c802t + 3020Insc	54	21
Aminoácido Nod2a P268S + cambio de marco 1007	55	22
Ácido nucleico Nod2 c2104t (SNP 20)	56	23
Aminoácido Nod2a R702W (SNP 20)	57	24
Ácido nucleico Nod2 g2722c (SNP 17)	58	25
Nod2a G908R (SNP 17)	59	28
Ácido nucleico Nod2 c802t (SNP 4)	60	29
Aminoácido Nod2a P268S (SNP 4)	61	30
Ácido nucleico Nod2 g2377a (SNP 7)	62	31
Aminoácido Nod2a V793M (SNP 7)	63	32
Ácido nucleico Nod2 a2555g (SNP 18)	64	33
Aminoácido Nod2a N852S (SNP 18)	65	34
Ácido nucleico Nod2 g2863a (SNP 23)	66	35
Aminoácido Nod2a V955I (SNP 23)	67	36
Ácido nucleico Nod2 a2587g (SNP 25)	68	37
Aminoácido Nod2a M863V (SNP 25)	69	38
Ácido nucleico Nod2 c802t + g2722c (SNP 4 + SNP 17)	84	39
Aminoácido Nod2a P268S + G908R (SNP 4 + SNP 17)	85	40
Ácido nucleico Nod2 c802t + a2555g (SNP 4 + SNP 18)	86	41
Aminoácido Nod2a P268S + N852S (SNP 4 + SNP 18)	87	42
Ácido nucleico Nod2 c802t + c2104t (SNP 4 + SNP 20)	88	43

Tabla 1		
Secuencia	SEC ID Número	Número de Figura
Aminoácido Nod2a P268S + R702W (SNP 4 + SNP 20)	89	44

B. Detección de alelos de Nod2

De acuerdo con esto, la presente invención proporciona procedimientos para determinar si un paciente tiene una mayor susceptibilidad a la enfermedad intestinal inflamatoria o a la enfermedad de Crohn determinando si el individuo tiene o no un alelo de Nod2 variante. En otras formas de realización, la presente invención proporciona procedimientos para proporcionar un pronóstico de un riesgo incrementado de sufrir la enfermedad de Crohn a un individuo en base a la presencia o ausencia de uno o más alelos variantes de Nod2. En formas de realización preferidas, la variación produce un truncamiento del dominio LRR. En otras formas de realización preferidas, la variación tiene como resultado un incremento de la activación de NF- κ B y la consiguiente respuesta inflamatoria. En formas de realización particularmente preferidas, la variación es un polimorfismo de un solo nucleótido causado por una inserción de un residuo citosina o una sustitución de un único nucleótido.

Se dispone de numerosos procedimientos para el análisis de las secuencias de ácido nucleico variante (p. ej., mutantes o polimórficas). Los ensayos para la detección de variantes (p. ej., polimorfismos o mutaciones) entran dentro de varias categorías, incluidos, entre otros, ensayos de secuenciación directa, ensayos de polimorfismo de fragmentos, ensayos de hibridación y análisis de datos por ordenador. Se dispone de protocolos y kits o servicios comercialmente disponibles para realizar múltiples variaciones de estos ensayos. En algunas formas de realización, los ensayos se realizan en combinación o en híbridos (p. ej., diferentes reactivos o tecnologías de varios ensayos se combinan para dar un ensayo). Los siguientes ensayos son útiles en la presente invención.

1. Ensayos de secuenciación directa

En algunas formas de realización de la presente invención se detectan secuencias variantes usando una técnica de secuenciación directa. En estos ensayos, las muestras de ADN se aíslan primero de un sujeto usando cualquier procedimiento adecuado. En algunas formas de realización, la región de interés se clona en un vector adecuado y se amplifica mediante crecimiento en una célula huésped (p. ej., una bacteria). En otras formas de realización, el ADN en la región de interés se amplifica usando PCR.

Tras la amplificación, el ADN en la región de interés (p. ej., la región que contiene el SNP o la mutación de interés) se secuencia usando cualquier procedimiento adecuado, incluidos, entre otros, secuenciación manual usando nucleótidos marcadores radiactivos o secuenciación automática. Los resultados de la secuenciación se muestran usando cualquier procedimiento adecuado. Se analiza la secuencia y se determina la presencia o ausencia de un SNP o mutación dada.

2. Ensayo PCR

En algunas formas de realización de la presente invención se detectan secuencias variantes usando un ensayo basado en PCR. En algunas formas de realización, el ensayo PCR comprende el uso de cebadores oligonucleotídicos que hibridan únicamente con el alelo variante o silvestre de Nod2 (p. ej., con la región del polimorfismo o la mutación). Ambos conjuntos de cebadores se usan para amplificar una muestra de ADN. Si solo los cebadores mutantes dan como resultado un producto de PCR, el paciente tiene el alelo mutante de Nod2. Si solo los cebadores silvestres dan como resultado un producto de PCR, el paciente tiene el alelo silvestre de Nod2.

3. Ensayos de polimorfismos de longitud de fragmento

En algunas formas de realización de la presente invención se detectan las secuencias variantes usando un ensayo de polimorfismo de longitud de fragmento. En un ensayo de polimorfismo de longitud de fragmento se genera un patrón de bandas de ADN único basado en la escisión del ADN en una serie de posiciones usando una enzima (p. ej., una enzima de restricción o una enzima CLEAVASE I [Third Wave Technologies, Madison, WI]). Los fragmentos de una muestra que contiene un SNP o una mutación tendrán un patrón de bandas diferente al de los silvestres.

a. Ensayo RFLP

En algunas formas de realización de la presente invención se detectan las secuencias variantes usando un ensayo de polimorfismo de longitud de fragmento de restricción (RFLP). Primero, se aísla la región de interés usando PCR. Después, los productos de la PCR se escinden con enzimas de restricción que se sabe que dan un fragmento de longitud único para un polimorfismo dado. Los productos de la PCR digeridos con la enzima de restricción se separan mediante electroforesis y se visualizan mediante tinción con bromuro de etidio. La longitud de los fragmentos se compara con marcadores de peso molecular y fragmentos generados de controles silvestres y mutantes.

b. Ensayo CFLP

En otras formas de realización, las secuencias variantes se detectan usando un ensayo de polimorfismo de longitud de fragmento CLEAVASE (CFLP; Third Wave Technologies, Madison, WI; véase, por ejemplo, las patentes de EE.UU. nº 5.843.654, 5.843.669, 5.719.208 y 5.888.780, cada una de las cuales se incorpora en el presente documento por referencia). Este ensayo se basa en la observación de que cuando las hebras sencillas de ADN se pliegan sobre sí mismas, asumen estructuras de orden superior que son altamente individuales de la secuencia precisa de la molécula de ADN, Estas estructuras secundarias implican regiones de ADN parcialmente en dúplex de modo que las regiones monocatenarias están en yuxtaposición con las horquillas de ADN bicatenario, La enzima CLEAVASE I es una nucleasa termoestable específica de estructura que reconoce y escinde las uniones entre las regiones monocatenarias y bicatenarias.

Primero, se aísla la región de interés, por ejemplo, usando PCR. Después, las hebras de ADN se separan con calor. A continuación, las reacciones se enfrían para permitir la formación de la estructura secundaria intracatenaria. Después, los productos de PCR se tratan con la enzima CLEAVASE I para generar una serie de fragmentos que son únicos de un SNP o mutación dados. Los productos de la PCR tratados con la enzima CLEAVASE se separan y se detectan (p. ej., mediante electroforesis en gel) y se visualizan (p. ej., mediante tinción con bromuro de etidio). La longitud de los fragmentos se compara con marcadores de peso molecular y fragmentos generados de controles silvestres y mutantes.

4. Ensayos de hibridación

En formas de realización preferidas de la presente invención se detectan secuencias variantes usando un ensayo de hibridación. En un ensayo de hibridación, la presencia o ausencia de un SNP o mutación dados se determina en base a la capacidad del ADN de la muestra para hibridar con una molécula de ADN complementario (p. ej., una sonda oligonucleotídica). Se dispone de diversos ensayos de hibridación que usan varias tecnologías de hibridación y detección. Más adelante se proporciona una descripción de una selección de ensayos.

a. Detección directa de hibridación

En algunas formas de realización, la hibridación de una sonda con la secuencia de interés (p. ej., un SNP o una mutación) se detecta directamente visualizando una sonda unida (p. ej., un ensayo Northern o Southern; véase, por ejemplo, Ausabel y col., (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, NY [1991]). En estos ensayos, el ADN genómico (Southern) o el ARN (Northern) se aísla de un sujeto. Después, el ADN o el ARN se fragmenta con una serie de enzimas de restricción que escinden con poca frecuencia el genoma y no están cerca de ninguno de los marcadores que se están analizando. Después, el ADN o el ARN se separa (p. ej., en un gel de agarosa) y se transfiere a una membrana. Una o más sondas marcadas (p. ej., mediante incorporación de un radionúclido) específicas del SNP o de la mutación que se está detectando se deja entrar en contacto con la membrana en una condiciones de rigurosidad baja, media o alta. La sonda no unida se elimina y se detecta la presencia de unión mediante visualización de la sonda marcada.

b. Detección de hibridación usando ensayos en “circuitos de ADN”

En algunas formas de realización de la presente invención se detectan secuencias variantes usando un ensayo de hibridación en circuito de ADN. En este ensayo, una serie de sondas oligonucleotídicas se fija a un soporte sólido. Las sondas oligonucleotídicas están diseñadas para ser únicas de un SNP o una mutación dada. La muestra de ADN de interés se pone en contacto con el “circuito” de ADN se detecta la hibridación.

En algunas formas de realización, el ensayo con circuito de ADN es un GeneChip (Affymetrix, Santa Clara, CA; véase, por ejemplo, las patentes de EE.UU. Nº 6.045.996, 5.925.525 y 5.858.659. La tecnología GeneChip usa matrices miniaturizadas de alta densidad de sondas oligonucleotídicas fijadas a un “circuito”. Las matrices de las sonadas las fabrica Affymetrix mediante un procedimiento de síntesis química dirigida por luz que combina síntesis química en fase sólida con técnicas de fabricación fotolitográfica usadas en la industria de los semiconductores. Usando una serie de máscaras fotolitográficas para definir los sitios de exposición del circuito, seguido de etapas síntesis química específica, el procedimiento construye matrices de alta densidad de oligonucleótidos con cada sonda en una posición predefinida en la matriz. Múltiples matrices de sondas se sintetizan de forma simultánea en una oblea de cristal grande. Después se cortan las obleas y las matrices de sondas individuales se empaquetan en cartuchos plásticos moldeados por inyección que los protegen del ambiente y sirven como cámaras de hibridación.

El ácido nucleico a analizar se aísla, amplifica mediante PCR y se marca con un grupo indicador fluorescente. El ADN marcado se incuba después con la matriz usando una estación fluida. Después, la matriz se inserta en el escáner, en el que se detectan los patrones de hibridación. Los datos de hibridación se recogen como luz emitida desde los grupos indicadores fluorescentes ya incorporados en la diana, que está unida a la matriz de la sonda. Las sondas que coinciden perfectamente con la diana producen, en general, señales más fuertes que los que tienen apareamientos erróneos. Dado que la secuencia y la posición de cada sonda sobre la matriz se conocen, mediante complementariedad se puede determinar la identidad del ácido nucleico diana aplicado a la matriz de la sonda.

En otras formas de realización se usa un microcircuito de ADN que contiene sondas capturadas electrónicamente

(Nanogen, San Diego, CA) (véase, por ejemplo, las patentes de EE.UU. N° 6.017.696, 6.068.818 y 6.051.380). Mediante el uso de microelectrónica, la tecnología Nanogen permite el movimiento activo y la concentración de moléculas cargadas hacia y desde los sitios de ensayo designados en su microcircuito de semiconductores. Las sondas de captura de ADN únicas de un SNP o mutación dados se colocan electrónicamente en, o “dirigen” a, sitios
5 específicos sobre el microcircuito. Dado que el ADN tiene una fuerte carga negativa, puede moverse electrónicamente a un área de carga positiva.

Primero, un sitio de ensayo o una hilera de sitios de ensayo sobre el microcircuito se activa electrónicamente con una carga positiva. Después, una solución que contiene las sondas de ADN se introduce en el microcircuito. Las sondas cargadas negativamente se mueven rápidamente hacia los sitios cargados positivamente, en los que se
10 concentran y se unen químicamente a un sitio sobre un microcircuito. Después, el microcircuito se lava y se añade otra solución de sondas de ADN distintas hasta que la matriz de sondas de ADN unidas específicamente está completa.

Después, una muestra de ensayo se analiza para determinar la presencia de moléculas de ADN diana determinando cuál de las sondas de captura de ADN hibridan on el ADN complementario en la muestra de ensayo (p. ej., un gen de interés amplificado por PCR). También se usa una carga electrónica para mover y concentrar las moléculas diana a uno o más sitios de ensayo sobre el microcircuito. La concentración electrónica del ADN de la muestra en cada sitio de ensayo estimula la rápida hibridación de un ADN de la muestra con sondas de captura complementarias (la hibridación puede producirse en minutos). Para eliminar todo ADN no unido o unido inespecíficamente de cada sitio, la polaridad o la carga del sitio se invierten a negativa, de modo que se fuerza todo el ADN no unido o unido
15 inespecíficamente de vuelta a la solución lejos de las sondas de captura. Para detectar la unión se usa un escáner de fluorescencia basado en láser.

En otras formas de realización adicionales se usa una tecnología de matriz basada en la segregación de fluidos sobre una superficie plana (circuito) por diferencias en la tensión superficial (ProtoGene, Palo Alto, CA) (véase, por ejemplo, las patentes de EE.UU. N° 6.001.311, 5.985.551 y 5.474.796). La tecnología de Protogene se basa en el hecho de que los fluidos se pueden segregar sobre una superficie plana por diferencias en la tensión superficial que se han impartido mediante recubrimientos químicos. Una vez segregadas de este modo, las sondas oligonucleotídicas se sintetizan directamente sobre el circuito mediante impresión con chorro de tinta de reactivos. La matriz con sus sitios de reacción definida por tensión superficial se monta sobre una etapa de traducción X/Y bao un conjunto de cuatro boquillas piezoeléctricas, una por cada una de las cuatro bases convencionales del ADN. La etapa de traducción se mueve a lo largo de cada una de las filas de la matriz y el reactivo adecuado se libera en cada sitio de reacción. Por ejemplo, la amidita A se libera únicamente en los sitios en los que la amidita A se va a acoplar durante dicha etapa de síntesis, y así sucesivamente. Los reactivos y lavados habituales se liberan mediante inundación de la totalidad de la superficie y, después, eliminándolos mediante centrifugación.
25

Las sondas de ADN únicas para el SNP o la mutación de interés se fijan en el circuito usando tecnología de Protogene. Después, el circuito se pone en contacto con los genes de interés amplificados por PCR. Tras la hibridación, el ADN no unido se elimina y la hibridación se detecta usando cualquier procedimiento adecuado (p. ej., mediante destemplado con fluorescencia de un grupo fluorescente incorporado).
30

En otras formas de realización más se usa una “matriz de perlas” para la detección de polimorfismos (Illumina, San Diego, CA; véase, por ejemplo, las publicaciones PCT WO 99/67641 y WO 00/3958). Illumina usa tecnología de MATRIZ DE PERLAS que combina haces de fibra óptica y perlas que se autoensamblan en una matriz. Cada haz de fibra óptica contiene de miles a millones de fibras individuales dependiendo del diámetro del haz. Las perlas se recubren con un oligonucleótido específico de la detección de un SNP o mutación dados. Se combinan lotes de perlas para formar un conjunto específico de la matriz. Para realizar un ensayo, la MATRIZ DE PERLAS se pone en contacto con una muestra sujeto preparada (p. ej., ADN). La hibridación se detecta usando cualquier procedimiento adecuado.
35

c. Detección enzimática de hibridación

En algunas formas de realización de la presente invención la hibridación se detecta mediante escisión enzimática de estructuras específicas (ensayo INVADER, Third Wave Technologies; véase, por ejemplo, las patentes de EE.UU. N° 5.846.717, 6.090.543, 6.001.567, 5.985.557 y 5.994.069). El ensayo INVADER detecta secuencias específicas de
40 ADN y ARN usando enzimas específicas de estructura para escindir un complejo formado mediante la hibridación de sondas oligonucleotídicas solapantes. La temperatura elevada y un exceso de una de las sondas permiten escindir múltiples sondas para cada secuencia diana presente sin ciclos de temperatura. Después, estas sondas escindidas dirigen la escisión de una segunda sonda marcada. La sonda oligonucleotídica secundaria puede estar marcada en el extremo 5' con fluoresceína que se inactiva mediante un pigmento interno. Tras la escisión, el producto marcado con fluoresceína destemplado se puede detectar usando un lector de placas de fluorescencia estándar.
45

El ensayo INVADER detecta mutaciones específicas y SNP en ADN genómico no amplificado. La muestra de ADN aislada se pone en contacto con la primera sonda específica para un SNP/mutación o una secuencia silvestre y se deja hibridar. Después, una sonda secundaria específica de la primera sonda y que contiene el marcador fluoresceína se hibrida y se añade la enzima. La unión se detecta usando un lector de placas fluorescente y comparando la señal de la muestra de ensayo con controles negativos y positivos conocidos.
50

En algunas formas de realización, la hibridación de una sonda unida se detecta usando un ensayo TaqMan (PE Biosystems, Foster City, CA; véase, por ejemplo, las patentes de EE.UU. N° 5.962.233 y 5.538.848). El ensayo se realiza durante una reacción de PCR. El ensayo TaqMan explota la actividad de exonucleasa 5'-3' de la ADN polimerasa AMPLITAQ GOLD. Una sonda específica de un alelo o mutación dados se incluye en la reacción de PCR. La sonda consiste en un oligonucleótido con un pigmento indicador en 5' (p. ej., un pigmento fluorescente) y un pigmento inactivador en 3'. Durante la PCR, si la sonda se une a su diana, la actividad nucleotídica 5'-3' de la polimerasa AMPLITAQ GOLD escinde la sonda entre el pigmento indicador y el inactivador. La separación del pigmento indicador del pigmento inactivador tiene como resultado un incremento de la fluorescencia. La señal se acumula con cada ciclo de PCR y se puede monitorizar con un fluorímetro.

En otras formas de realización más, los polimorfismos se detectan usando un ensayo de extensión de cebadores SNP-IT (Orchid Biosciences, Princeton, NJ; véanse las patentes de EE.UU. N° 5.952.174 y 5.919.626). En este ensayo, los SNP se identifican usando un cebador de ADN sintetizado especialmente y una ADN polimerasa para extender de forma selectiva la cadena de ADN en una base en la posible localización del SNP. El ADN en la región de interés se amplifica y se desnatura. Después, las reacciones de la polimerasa se realizan usando sistemas miniaturizados denominados microfluidos. La detección se lleva a cabo añadiendo un marcador al nucleótido que se sospecha que está en la localización del SNP o la mutación. La incorporación del marcador en el ADN se puede detectar mediante cualquier procedimiento adecuado (p. ej., si el nucleótido contiene un marcador de biotina, la detección se realiza mediante un anticuerpo marcado con fluorescencia específico de la biotina).

5. Ensayo de espectroscopia de masas

En algunas formas de realización se usa un sistema MassARRAY (Sequenom, San Diego, CA) para detectar secuencias variantes (véanse, por ejemplo, las patentes de EE.UU. N° 6.043.031, 5.777.324 y 5.605.798). El ADN se aísla de muestras de sangre usando procedimientos estándar. Después, las regiones de ADN específicas que contienen la mutación o el SNP de interés, de una longitud de aproximadamente 200 pares de bases, se amplifican mediante PCR. Los fragmentos amplificados se unen después por una hebra a una superficie sólida y las hebras no inmovilizadas se eliminan mediante desnaturación y lavado estándar. La hebra única inmovilizada restante sirve después como molde para reacciones enzimáticas automatizadas que producen productos para diagnóstico específicos de genotipo.

Después, cantidades muy pequeñas de productos enzimáticos, normalmente de cinco a diez nanolitros, se transfieren a una matriz SpectroCHIP para su posterior análisis automatizado con el espectrómetros de masas SpectroREADER. Cada mancha se carga previamente con cristales absorbentes de luz que forman una matriz con el producto diagnóstico dispensado. El sistema MassARRAY usa espectrometría de masas MALDI-TOF (ionización-desorción láser asistida por matriz—tiempo de vuelo). En un procedimiento conocido como desorción, se aplica a la matriz un pulso de un haz láser. La energía del haz láser se transfiere a la matriz y se vaporiza, lo que tiene como resultado la expulsión de una cantidad pequeña del producto diagnóstico en un tubo de vuelo. Como el producto diagnóstico se carga cuando posteriormente se aplica un pulso de campo eléctrico al tubo, se lanzan en sentido descendente por el tubo de vuelo hacia un detector. El tiempo entre la aplicación del pulso de campo eléctrico y la colisión del producto diagnóstico con el detector se denomina tiempo de vuelo. Esta es una medida muy precisa del peso molecular del producto, ya que la masa de una molécula se correlaciona directamente con el tiempo de vuelo con moléculas similares que vuelan más rápido que las moléculas más grandes. Todo el ensayo se completa en menos de una milésima de segundo, lo que permite analizar las muestras en un total de 3-5 segundos, incluida la recolección de datos repetitivos. Después, el software SpectroTYPER calcula, registra, compara e informa los genotipos a la velocidad de tres segundos por muestra.

6. Análisis de la variante mediante unión diferencial de anticuerpos

En otras formas de realización de la presente invención, los anticuerpos (véase más adelante la producción de anticuerpos) se usan para determinar si un individuo contiene un alelo que codifica un gen variante de Nod2. En formas de realización preferidas, se usan anticuerpos que discriminan entre proteínas variantes (es decir, proteínas truncadas) y de tipo silvestre (SEC ID N° 2 y 3). En algunas formas de realización particularmente preferidas, los anticuerpos están dirigidos al extremo C de Nod2.

7. Kits para analizar el riesgo de enfermedad de Crohn

La presente invención también proporciona kits para determinar si un individuo contiene un alelo silvestre o variante (p. ej., polimórfico o mutante) de Nod2. En algunas formas de realización, los kits son útiles para determinar si un sujeto está en riesgo de desarrollar la enfermedad e Crohn. Los kits para diagnóstico se producen de diversos modos. En algunas formas de realización, los kits contienen al menos un reactivo para detectar específicamente un alelo o proteína mutante de Nod2. En formas de realización preferidas, los kits contienen reactivos para detectar un SNP producido mediante una inserción de un residuo citosina o una sustitución de un único nucleótido del gen de tipo silvestre. En formas de realización preferidas, el reactivo es un ácido nucleico que hibrida con ácidos nucleicos que contienen el SNP y que no se une a ácidos nucleicos que no contienen el SNP. En otras formas de realización preferidas, los reactivos son cebadores para amplificar la región del ADN que contiene el SNP. En otras formas de realización más, los reactivos son anticuerpos que se unen, preferentemente, a las proteínas Nod2 silvestres o truncadas. En algunas formas de realización, el kit contiene instrucciones para determinar si un sujeto está en riesgo

de desarrollar la enfermedad e Crohn. En formas de realización preferidas, las instrucciones especifican que el riesgo de desarrollar enfermedad de Crohn se determina detectando la presencia o ausencia de un alelo mutante de Nod2 en el sujeto, en el que los sujetos que tienen un alelo que contiene una mutación por inserción de citosina o sustitución de un solo nucleótido tienen un mayor riesgo de desarrollar la enfermedad de Crohn. En algunas formas de realización, los kits incluyen reactivos auxiliares tales como agentes tampón, reactivos de estabilización de ácido nucleico, reactivos de estabilización de proteínas y sistemas de producción de señal (p. ej., sistemas de generación de fluorescencia como sistemas Fret). El kit de ensayo puede estar envasado de cualquier manera adecuada, normalmente con los elementos en un único contenedor o varios contenedores según sea necesario, junto con una hoja de instrucciones para llevar a cabo el ensayo. En algunas formas de realización, los kits también incluyen, preferentemente, una muestra de control positivo.

8. Bioinformática

En algunas formas de realización, la presente invención proporciona procedimientos de determinar el riesgo de un individuo de desarrollar la enfermedad de Crohn en base a la presencia de uno o más alelos variantes de Nod2. En algunas formas de realización, el análisis de los datos de las variantes se procesa con un ordenador usando la información almacenada en un ordenador (p. ej., en una base de datos). Por ejemplo, en algunas formas de realización, la presente invención proporciona un sistema de investigación de bioinformática que comprende una pluralidad de ordenadores que dirigen un idioma de programación orientado a objetos de multiplataforma (véase, por ejemplo, la patente de EE.UU. 6.125.383, incorporada en el presente documento por referencia). En algunas realizaciones, uno de los ordenadores almacena los datos genéticos (p. ej., el riesgo de poner en contacto la enfermedad de Crohn asociada con un polimorfismo dado, así como las secuencias). En algunas formas de realización, uno de los ordenadores almacena programas de aplicación (p. ej., para analizar los datos de desequilibrio de transmisión o determinar los riesgos relativos al genotipo y los riesgos atribuibles a la población (véanse los ejemplos 9 y 10). Después, los resultados se entregan al usuario (p. ej., mediante uno de los ordenadores o mediante Internet).

IV. Generación de anticuerpos frente a Nod2

Estos anticuerpos se pueden generar para permitir la detección de la proteína Nod2. Los anticuerpos se pueden preparar usando varios inmunógenos. En una forma de realización, el inmunógeno es un péptido Nod2 humano para generar anticuerpos que reconocen la Nod2 humana. Dichos anticuerpos incluyen, entre otros, policlonales, monoclonales, quiméricos, de cadena sencilla, fragmentos Fab y una biblioteca de expresión de Fab.

Se pueden usar varios procedimientos conocidos en la técnica para la producción de anticuerpos policlonales dirigidos contra Nod2. Para la producción de un anticuerpo se pueden inmunizar varios animales huéspedes mediante inyección con el péptido correspondiente en el epitopo de Nod2, incluidos, entre otros, conejos, ratones, ratas, ovejas, cabras etc. En una realización preferida, el péptido se conjuga con un vehículo inmunogénico (por ejemplo, toxoide diftérico, seroalbúmina bovina (BSA) o hemocianina de lapa californiana (KLH)). Se pueden usar varios adyuvantes para aumentar la respuesta inmunológica dependiendo de la especie huésped, incluidos, entre otros, adyuvante de Freund (completo e incompleto), geles minerales (por ejemplo, hidróxido de aluminio), sustancias de superficie activa (por ejemplo, lisolecitina, polioles plurónicos, polianiones, péptidos, emulsiones en aceite, hemocianina de lapa californiana, dinitrofenol) y adyuvantes humanos potencialmente útiles, tales como BCG (bacilo de *Calmette-Guerin*) y *Corynebacterium parvum*.

Para la preparación de anticuerpos monoclonales dirigidos a Nod2, se contempla que cualquier técnica que proporcione la producción de moléculas de anticuerpos mediante cultivos continuos de líneas celulares encontrarán utilidad con la presente invención (véase, por ejemplo, Harlow y Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Estas incluyen, entre otras, la técnica del hibridoma inicialmente desarrollada por Köhler y Milstein (Köhler y Milstein 1975, *Nature* 256:495-497 [1975]), así como la técnica del trioma, la técnica del hibridoma de células B humanas (véase, por ejemplo, Kosbor, y col., *Immunol. Tod.* 4:72 [1983]) y la técnica del hibridoma con EBV para producir anticuerpos monoclonales humanos (Cole y col., 1985, en *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pág. 77-96 [1985]).

En una realización adicional de la invención, los anticuerpos monoclonales se producen en animales sin gérmenes usando la tecnología como la descrita en el documento PCT/US90/02545. Adicionalmente, se contempla que los anticuerpos humanos se generen mediante hibridomas humanos (Cote y col., 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80: 2026-2030[1983]) o transformando linfocitos B humanos con el virus de EBV *in vitro* (Cole y col., 1985, en *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, pág. 77-96 [1985]).

Además, se contempla que las técnicas descritas para la producción de anticuerpos de cadena sencilla (patente de EE.UU. 4.946.778) encontrarán utilidad en la producción de anticuerpos de cadena sencilla específicos de Nod2. Una forma de realización adicional de la invención usa las técnicas descritas para la construcción de bibliotecas de expresión de Fab (Huse, y col., *Science*, 246: 1275-1281 [1989]) para permitir la identificación rápida y fácil de fragmentos Fb monoclonales con la especificidad deseada por Nod2.

Se contempla que cualquier técnica adecuada para producir fragmentos de anticuerpos encontrará utilidad en la generación de fragmentos de anticuerpos que contienen el idiotipo (región de unión al antígeno) de la molécula de

anticuerpo. Por ejemplo, entre estos fragmentos se incluyen: El fragmento F(ab')₂ que se puede producir mediante digestión con pepsina de la molécula de anticuerpo; los fragmentos Fab' que se pueden generar mediante la reducción de los puentes disulfuro de los fragmentos F(ab')₂, y los fragmentos Fab que se pueden generar tratando la molécula de anticuerpo con papaína y un agente reductor.

5 En la producción de anticuerpos, se contempla que la detección selectiva del anticuerpo deseado se efectuará mediante técnicas conocidas en la materia (por ejemplo, radioinmunoensayo, ELISA (ensayo de inmunosorción ligada a enzimas), inmunoensayos de tipo "sándwich", ensayos inmunoradiométricos, reacciones de precipitación con difusión en gel, ensayos de inmunodifusión, inmunoensayos *in situ* (usando oro coloidal, marcadores enzimáticos o radioisótopos, por ejemplo), transferencias de tipo western, reacciones de precipitación, ensayos de aglutinación (p. ej., ensayos de aglutinación en gel, ensayos de hemaglutinación), ensayos de fijación del complemento, ensayos de inmunofluorescencia, ensayos de proteína A y ensayos de inmunoelectroforesis etc.

10 En una forma de realización, la unión del anticuerpo se detecta detectando un marcador en el anticuerpo primario. En otra forma de realización, el anticuerpo primario se detecta detectando la unión de un anticuerpo secundario o reactivo frente al anticuerpo primario. En una forma de realización adicional, el anticuerpo secundario está marcado:
15 En la técnica se conocen muchos medios para detectar la unión en un inmunoensayo y entran dentro del alcance de la presente invención. Como es bien conocido en la técnica, el péptido inmunogénico debería proporcionarse sin la molécula vehicular usada en cualquier protocolo de inmunización. Por ejemplo, si el péptido estaba conjugado con KLH, se puede conjugar con BSA o usarse directamente en un ensayo de detección selectiva).

20 Los anteriores anticuerpos de la divulgación se pueden usar en procedimientos conocidos en la técnica en relación con la localización y la estructura de Nod2 (p. ej., para transferencia de tipo western), medición de los niveles de los mismos en muestras biológicas adecuadas etc. Los anticuerpos se pueden usar para detectar Nod2 en una muestra biológica de un individuo. La muestra biológica puede ser un fluido biológico tal como, ente otros, sangre, suero, plasma, fluido intersticial, orina, líquido cefalorraquídeo y similares, que contienen las células.

25 Después se puede analizar las muestras biológicas directamente para determinar la presencia de Nod2 humana usando una estrategia (p. ej., ELISA o radioinmunoensayo) y formato (p. ej., micropocillos, tira reactiva (p. ej., como se describe en la publicación de patente internacional WO 93/03367), etc. Como alternativa, las proteínas en la muestra se pueden separar por tamaño (p. ej., mediante electroforesis en gel de poliacrilamida (PAGE) en presencia no de dodecilsulfato sódico (SDS) y la presencia de Nod2 detectada mediante inmunotransferencia (transferencia Western). Las técnicas de inmunotransferencia son, en general, más eficaces con anticuerpos generados contra un péptido correspondiente a un epítipo de una proteína y, por tanto, son particularmente adecuadas para la presente invención.

30 Otro procedimiento usa anticuerpos como agentes para alterar la transducción de la señal. Se pueden usar anticuerpos específicos que se unen a los dominios de unión de Nod2 u otras proteínas implicadas en la señalización intracelular para inhibir la interacción entre las diversas proteínas y su interacción con otros ligandos.
35 Los anticuerpos que se unen al complejo también se pueden usar terapéuticamente para inhibir las interacciones del complejo proteico en las vías de transducción de señal que conducen a los diversos efectos fisiológicos y celulares de NF- κ B. Dichos anticuerpos también se pueden usar en diagnóstico para medir la expresión anómala de Nod2 o la formación aberrante de complejos proteicos, que puede ser indicativo de un estado de enfermedad.

V. Terapia génica usando Nod2

40 La presente invención es útil para procedimientos y composiciones adecuados para terapia génica para alterar la expresión, producción o función de Nod2. Como se ha descrito en lo que antecede, la presente divulgación proporciona genes humanos de Nod2 y proporciona procedimientos de obtención de genes de Nod2 de otras especies. Por tanto, los procedimientos que se describen más adelante son, en general, aplicables en muchas especies. La terapia génica se puede realizar proporcionando a un sujeto un alelo de tipo silvestre de Nod2 (es decir, un alelo que no contiene una mutación por inserción de citosina u otro cambio de ácido nucleico (p. ej., polimorfismos o mutaciones). Los sujetos que necesitan dicha terapia se identifican mediante los procedimientos descritos en lo que antecede. Como se ha descrito anteriormente, la Nod2 se expresa principalmente en los monocitos. De acuerdo con esto, un procedimiento preferido de terapia génica es eliminar los monocitos de los sujetos (p. ej., mediante radiación) y sustituir los monocitos defectivos con monocitos que expresan la Nod2 silvestre mediante un trasplante de médula ósea. Los monocitos defectuosos de los sujetos se pueden recolectar antes del tratamiento de radiación, transfectar con un vector (que se describe más adelante) que codifica monocitos silvestres, amplificar mediante cultivos *in vitro* y reintroducir en el sujeto.

55 Los vectores virales que normalmente se usa para los procedimientos dirigidos y de terapia *in vivo* o *ex vivo* son vectores basados en ADN y vectores retrovirales. Los procedimientos para construir y usar vectores virales se conocen en la técnica (véase, por ejemplo, Miller y Rosman, BioTech., 7:980-990 [1992]). Preferentemente, los vectores virales son defectuosos en la replicación, es decir son incapaces de replicarse de forma autónoma en la célula diana. En general, el genoma de los vectores virales de replicación defectuosa que se usan dentro del alcance de la presente invención carece de al menos una región que es necesaria para la replicación de los virus en la célula infectada. Estas regiones se pueden eliminar (en su totalidad o en parte) o se pueden hacer no funcionales por cualquier técnica conocida por un experto en la técnica. Estas técnicas incluyen la eliminación total, sustitución (por

otras secuencias, en particular por el ácido nucleico insertado), delección parcial o adición de una o más bases en una región esencial (para replicación). Dichas técnicas pueden realizarse *in vitro* (es decir, en el ADN aislado) o *in situ*, usando las técnicas de manipulación genética o mediante tratamiento con agentes mutagénicos.

5 Preferentemente, el virus de replicación defectuosa conserva las secuencias de su genoma que son necesarias para encapsidar las partículas virales. Los vectores virales de ADN incluyen virus de ADN atenuados o defectuosos, incluidos, entre otros, el virus del herpes simple (VHS), el papilomavirus, el virus de Epstein Barr (EBV), el adenovirus, el virus adenoasociado (AAV) y similares. Se prefieren virus defectuosos que carecen completamente o casi completamente de genes virales, ya que los virus defectuosos no son infecciosos después de su introducción en una célula. El uso de vectores virales defectivos permite la administración a las células en un área localizada
10 específica, sin tener que preocuparse de que el vector pueda infectar a otras células. Por tanto, un tejido específico puede ser una diana específica. Ejemplos de vectores concretos incluyen, entre otros, un vector del virus herpes 1 defectuoso (HSV1) Kaplitt et al., *Mol. Cell. Neurosci.*, 2: 320-330 (1991)], vector del virus del herpes defectuoso que carece del ge de la glicoproteína L (véase, por ejemplo, la publicación de patente RD 371005 A), u otros vectores del virus del herpes defectuoso (véase, por ejemplo, los documentos WO 94/21807 y WO 92/05263); un vector
15 adenovirus atenuado, tal como el vector descrito por Stratford-Perricaudet y col. (*J. Clin. Invest.*, 90: 626-630 [1992]; véase también La Salle et al., *Science* 259: 988-990 [1993]); y un vector de virus adenoasociado defectuoso (Samulski y col., *J. Virol.* 61: 3096-3101 [1987]; Samulskiet al., *J. Virol.*, 63: 3822-3828 [1989]; y Lebkowski et al., *Mol. Cell. Biol.*, 8: 3988-3996 [1988]).

20 Preferentemente, para la administración *in vivo*, se empela un tratamiento inmunosupresor adecuado junto con le vector viral, por ejemplo vector adenovirus, para evitar la inmunodesactivación del vector viral y las células transfectadas. Por ejemplo, las citocinas inmunosupresoras, como la interleucina-12 (IL12), el interferón gamma (IFN- γ) o el anticuerpo anti-CD4, se pueden administrar para bloquear las respuestas humorales o celulares a los vectores virales. Además, es ventajoso emplear un vector viral que se ha sometido a ingeniería para expresar un número mínimo de antígenos.

25 El vector puede ser un vector adenovirus. Los adenovirus son virus de ADN eucarióticas que se pueden modificar para liberar de forma eficiente un ácido nucleico de la invención a diversos tipos celulares. Existen varios serotipos de adenovirus. De estos serotipos se da preferencia, dentro del alcance de la presente invención, a los adenovirus humanos de tipo 2 o de tipo 5 (Ad 2 o Ad 5), o adenovirus de origen animal (véase, por ejemplo, el documento WO 94/26914). Estos adenovirus de origen animal que se pueden usar incluyen adenovirus de origen canino, bovino,
30 murino (p. ej., Mav1, Beard et al., *Virol.*, 75-81 [1990]), ovino, porcino, aviar y simio (p. ej., SAV). Preferentemente, el adenovirus de origen animal es un adenovirus canino, más preferentemente un adenovirus CAV2 (p. je., la cepa Manhattan o A26/61 (ATCC VR-800)).

35 Preferentemente, los vectores adenovirus de replicación defectuosa comprenden los ITR, una secuencia de encapsidación y el ácido nucleico de interés. Todavía más preferentemente, al menos la región E1 del vector adenovirus es no funcional. La delección en la región E1 se extiende, preferentemente, desde los nucleótidos 455 a 3329 en la secuencia del adenovirus Ad5 (fragmento *PvuII-BglII*) o de 382 a 3446 (fragmento *HinflI-Sau3A*). Otras regiones también se pueden modificar, en concreto la región E3 (p. ej., el documento WO95/02697), la región E2 (p. ej., el documento WO94/28938), la región E4 (p. ej., los documentos WO94/28152, WO94/12649 y WO95/02697), o en cualquiera de los genes tardíos L1-L5.

40 El vector adenoviral puede tener una delección en la región E1 (Ad 1,0). Ejemplos de adenovirus con delección de E1 se divulgan en el documento EP 185.573. El vector adenoviral puede tener una delección en las regiones E1 y E4 (Ad 3,0). Ejemplos de adenovirus con delección en E1/E4 se divulgan en los documentos WO95/02697 y WO96/22378. El vector adenoviral puede tener una delección en la región E1 en la que se insertan la región E4 y la secuencia de ácido nucleico.

45 Los adenovirus recombinantes de replicación defectuosa de acuerdo con la invención pueden prepararse mediante cualquier técnica conocida por el experto en la técnica (véase por ejemplo, Levrero y col., *Gene* 101: 195 [1991]; EP 185 573; y Graham, *EMBO J.*, 3: 2917 [1984]). En particular, se pueden preparar mediante recombinación homóloga entre un adenovirus y un plásmido que porta, entre otros, la secuencia de ADN de interés. La recombinación homóloga se consigue tras la cotransfección del adenovirus y el plásmido en una línea celular adecuada. La línea celular que se usa deberá ser, preferentemente, (i) transformable por los elementos que se van a usar y (ii) contener las secuencias que son capaces para complementar la parte del genoma del adenovirus de replicación defectuosa, preferentemente en forma integrada con el fin de evitar los riesgos de recombinación. Ejemplos de líneas celulares que se pueden usar son la línea de células renales embrionarias humanas 293 (Graham y col., *J. Gen. Virol.*, 36: 59 [1977]), que contiene la porción izquierda del genoma de un adenovirus Ad5 (12%) integrada en su genoma, y líneas celulares que son capaces de complementar las funciones de E1 y E4, como se ha descrito en las solicitudes WO 94/26914 y WO 95/02697. Los adenovirus recombinantes se recuperan y purifican usando técnicas de biología molecular estándar que son bien conocidas por los expertos en la técnica.

60 Los virus adenoasociados (AAV) son virus de ADN de tamaño relativamente pequeño que se pueden integrar, de un modo estable y específico de sitio, en el genoma de las células que infectan. Son capaces de infectar un amplio espectro de células sin inducir ningún efecto sobre el crecimiento celular, su morfología o diferenciación, y no

parecen estar implicados en enfermedades humanas. El genoma de AAV se ha clonado, secuenciado y caracterizado. Abarca aproximadamente 4.700 bases y contiene una región de repetición terminal invertida (ITR) de aproximadamente 145 bases en cada extremo, que sirve como origen de replicación para el virus. El resto del genoma se divide en dos regiones esenciales que portan las funciones de encapsidación: la parte izquierda del genoma, que contiene el gen *rep* implicado en la replicación viral y la expresión de los genes virales; y la parte derecha del genoma, que contiene el gen *cap* que codifica las proteínas de la cápside del virus.

Se ha descrito el uso de vectores derivados de los AAV para transferir genes *in vitro* e *in vivo* (véase, por ejemplo, los documentos WO 91/18088; WO 93/09239; la patente de EE.UU. N° 4.797.368; la patente de EE.UU. N° 5.139.941; y EP 488 528). Estas publicaciones describen varias construcciones derivadas de AAV en las que los genes *rep* y/o *cap* se delecionan y sustituyen con un gen de interés, y el uso de estas construcciones para transferir el gen de interés *in vitro* (en células cultivadas) o *in vivo* (directamente en un organismo). Los AAV recombinantes de replicación defectuosa de acuerdo con la invención se pueden preparar mediante cotransfección de un plásmido que contiene la secuencia de ácido nucleico de interés flanqueada por dos regiones de repetición terminal invertidas (ITR) de AAV y un plásmido portador de los genes de encapsidación del AAV (genes *rep* y *cap*) en una línea celular que se infecta con un virus colaborador humano (por ejemplo un adenovirus). Los AAV recombinantes que se producen se purifican después mediante técnicas estándar.

El gen se puede introducir en un vector retroviral (p. ej., como se describe en las patentes de EE.UU. N° 5.399.346, 4.650.764, 4.980.289 y 5.124.263; Mann et al., Cell 33:153 [1983]; Markowitz et al., J. Virol., 62: 1120 [1988]; PCT/US95/14575 ; EP 453242; EP178220 ; Bernstein et al. Genet. Eng., 7: 235 [1985]; McCormick, BioTechnol., 3: 689 [1985]; documento WO 95/07358; y Kuo et al., Blood 82: 845 [1993]). Los retrovirus son virus de integración que infectan las células en división. El genoma del retrovirus incluye dos LTR, una secuencia de encapsidación y tres regiones de codificación (*gag*, *pol* y *env*). En los vectores retrovirales recombinantes, los genes *gag*, *pol* y *env* normalmente se delecionan, completamente o en parte, y se sustituyen con una secuencia de ácido nucleico heterólogo de interés. Estos vectores se pueden construir a partir de diferentes tipos de retrovirus, tales como, VIH, MoMuLV ("virus Moloney de la leucemia murina" MSV ("virus del sarcoma Moloney murino"), HaSV ("virus del sarcoma de Harvey virus"); SNV ("virus de la necrosis del bazo"); RSV ("virus del sarcoma de Rous") y virus Friend. Vectores retrovirales defectuosos también se divulgan en el documento WO95/02697.

En general, con el fin de construir retrovirus recombinantes que contienen una secuencia de ácido nucleico, se construye un plásmido que contiene las LTR, la secuencia de encapsidación y la secuencia de codificación. Esta construcción se usa para transfectar una línea celular concentrada, en la que la línea celular puede suministrar en trans las funciones retrovirales que son deficientes en el plásmido. En general, las líneas celulares concentradas son, por tanto, capaces de expresar los genes *gag*, *pol* y *env*. Dichas líneas celulares concentradas se han descrito en la técnica anterior, en particular la línea celular PA317 (patente de EE.UU. N° 4.861.719, la línea celular PsiCRIP (véase el documento WO90/02806) y la línea celular GP+envAm-12 (véase el documento WO89/07150). Además, los vectores retrovirales recombinantes pueden contener modificaciones dentro de las LTR para suprimir la actividad transcripcional, así como secuencias de encapsidación extensas que pueden incluir una parte del gen *gag* (Bender y col., J. Virol., 61: 1639 [1987]). Los vectores retrovirales recombinantes se purifican mediante técnicas estándar conocidas por los expertos en la técnica.

Como alternativa, el vector se puede introducir *in vivo* mediante lipofección. Durante la última década se ha producido un uso creciente de liposomas para encapsulación y transfección de ácidos nucleicos *in vitro*. Los lípidos catiónicos sintéticos diseñados para limitar las dificultades y peligros encontrados con la transfección mediada por liposomas se pueden usar para preparar liposomas para la transfección *in vivo* de un gen que codifica un marcador (Felgner, y col., Proc. Natl. Acad. Sci. USA 84: 7413-7417 [1987]; véase también Mackey, et al., Proc. Natl. Acad. Sci. USA 85: 8027-8031 [1988]; Ulmer et al., Science 259: 1745-1748 [1993]). El uso de lípidos catiónicos puede estimular la encapsulación de ácidos nucleicos con carga negativa y también promueve la fusión con membranas celulares cargadas negativamente [Felgner y Ringold, 1989, Science 337: 387-388 [1989]). Compuestos y composiciones lipídicas particularmente útiles para transferir ácidos nucleicos se describen en los documentos WO95/18863 y WO96/17823, y en la patente de EE.UU. N° 5.459.127.

Otras moléculas también son útiles para facilitar la transfección de un ácido nucleico *in vivo*, tal como un oligopéptido catiónico (p. ej., el documento WO95/21931), péptidos derivados de proteínas de unión a ADN (p. ej., el documento WO96/25508), o un polímero catiónico (p. ej., el documento WO95/21931).

También es posible introducir el vector *in vivo* como un plásmido de ADN desnudo. Procedimientos para formular y administrar ADN desnudo al tejido muscular de mamífero se divulgan en las patentes de EE.UU. N° 5.580.859 y 5.589.466.

Los vectores de ADN para terapia génica se pueden introducir en las células huésped deseadas mediante procedimientos conocidos en la técnica, incluidos, entre otros, transfección, electroporación, microinyección, transducción, fusión celular, DEAE dextrano, precipitación en fosfato cálcico, el uso de una pistola génica o el uso de un transportador del vector de ADN (véase, por ejemplo, Wu y col., J. Biol. Chem., 267: 963 [1992]; Wu y Wu, 1988, J. Biol. Chem. 263: 14621 [1988]; y Williams et al., Proc. Natl. Acad. Sci. USA 88: 2726 [1991]). La liberación de ADN mediada por receptores también se pueden usar [Curiel y col., Hum. Gene Ther., 3: 147 [1992]; y Wu and Wu, J.

Biol. Chem., 262: 4429 [1987]).

VI. Animales transgénicos que expresan genes exógenos de Nod2 y homólogos, mutantes y variantes de los mismos

5 La presente invención es útil en la generación de animales transgénicos que comprenden un gen exógeno de Nod2 u homólogos, mutantes o variantes de los mismos. El animal transgénico exhibe un fenotipo alterado en comparación con los animales silvestres. El fenotipo alterado es la sobreexpresión de ARNm para un gen de Nod2 en comparación con los niveles silvestres de expresión de Nod2 o el fenotipo alterado es la expresión disminuida de ARNm para un gen endógeno de Nod2 en comparación con los niveles silvestres de expresión de Nod2 endógeno. Los procedimientos para analizar la presencia o ausencia de dichos fenotipos incluyen transferencia Northern, ensayos de protección de ARNm y RT-PCR. Los ratones transgénicos pueden tener una mutación defectiva del gen Nod2. La expresión de un gen variante de Nod2 (p. ej., la SEC ID N° 33), (el mutante de inserción c), variantes de sustitución de un solo nucleótido (p.ej., SEC ID. N° 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88) o mutantes que contienen delecciones de una o más repeticiones LRR). Los animales transgénicos pueden mostrar un fenotipo de enfermedad de Crohn

15 Los animales transgénicos encuentran utilidad en detecciones selectivas de la dieta, fármacos y patógenos (p. ej., bacterias entéricas). Los animales transgénicos no humanos (p. ej., animales que muestran un fenotipo de la enfermedad de Crohn) son alimentos de ensayo o dietas control y se evalúa la respuesta de los animales a las dietas. Los compuestos de ensayo (p. ej., un fármaco que se sospecha que es útil para la enfermedad de Crohn) y los compuestos control (p. ej., un placebo) se pueden administrar a los animales transgénicos y los animales control y se evalúan los efectos. Los animales transgénicos y control se pueden infectar con una bacteria entérica y se evalúa el efecto sobre los síntomas de enfermedad de Crohn y los animales transgénicos y control se infectan con bacterias entéricas que se ha hallado que producen o incrementan la gravedad de los síntomas de enfermedad, seguido de la administración de compuestos de ensayo y compuestos control. Después, se evalúan los efectos de los compuestos de ensayo y control sobre los síntomas de enfermedad.

25 Los animales transgénicos se pueden generar mediante diversos procedimientos. Se pueden usar células embrionarias en varias etapas del desarrollo para introducir transgenes para la producción de animales transgénicos. Se usan diferentes procedimientos en función de la etapa de desarrollo de la célula embrionaria. El cigoto es la mejor diana para la microinyección. En el ratón, el pronúcleo masculino alcanza el tamaño de aproximadamente 20 micrómetros de diámetro que permite la inyección reproducible de 1-2 picolitros (pl) de solución de ADN. El uso de cigotos como diana para la transferencia génica tiene una gran ventaja en cuando a que, en la mayoría de los casos, el ADN inyectado se incorporará en el genoma del huésped antes de la primera escisión (Brinster y col., Proc. Natl. Acad. Sci. USA 82: 4438-4442 [1985]). Como consecuencia, todas las células del animal transgénico no humano portarán el transgén incorporado. En general, esto también se reflejará en la transmisión eficiente del transgén de la descendencia del inicial, ya que el 50 % de las células germinales alojará el transgén. La patente de EE.UU. N° 4.873.191 describe un procedimiento para la microinyección de cigotos.

35 Se puede usar la infección retroviral para introducir transgenes en un animal no humano. El vector retroviral se puede utilizar para transfectar oocitos inyectando el vector retroviral en el espacio perivitelino del oocito (patente de EE.UU. N° 6.080.912). El embrión no humano en desarrollo se puede cultivar *in vitro* hasta la etapa de blastocisto. Durante este tiempo, los blastómeros pueden ser dianas para la infección retroviral (Janenich, Proc. Natl. Acad. Sci. USA 73: 1260 [1976]). Una infección eficiente de los blastómeros se obtiene mediante tratamiento enzimático para eliminar la zona pelúcida (Hogan y col., en *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [1986]). El sistema de vector viral usado para introducir el transgén normalmente es un retrovirus de replicación defectuosa portador del transgén (Jahner y col., Proc. Natl. Acad. Sci. USA 82: 6927 [1985]). La transfección se obtiene fácil y eficientemente mediante cultivo de los blastómeros en una monocapa de las células productoras de virus (Van der Putten, citado anteriormente; Stewart, y col., EMBO J., 6: 383 [1987]). Como alternativa, la infección se puede realizar en una etapa posterior. El virus o las células productoras de virus se pueden inyectar en el blastocele (Jahner y col., Nature 298: 623 [1982]). La mayoría de los fundadores será mosaico para el transgén, ya que la incorporación solo se produce en una subpoblación de células que forman el animal transgénico. Adicionalmente, el fundador puede contener varias inserciones retrovirales del transgén en diferentes posiciones en el genoma, que generalmente se segregarán en la descendencia. Además, también es posible introducir transgenes en la línea germinal, pero con baja eficiencia, mediante infección retroviral intrauterina del embrión a mitad de la gestación (Jahner y col., citado anteriormente [1982]). Medios adicionales de usar retrovirus o vectores retrovirales para crear animales transgénicos conocidos en la técnica implican la microinyección de partículas retrovirales o células tratadas con mitomicina C que producen retrovirus en el espacio perivitelino de huevos fertilizados o embriones tempranos (solicitud internacional de PCR WO 90/08832 [1990], y Haskell y Bowen, Mol. Reprod. Dev., 40: 386 [1995])

60 El transgén se puede introducir en células madre embrionarias y las células madre transfectadas se usan para formar un embrión. Las células ES se obtienen cultivando embriones preimplantación *in vitro* en las condiciones adecuadas (Evans et al., Nature 292: 154[1981]; Bradley et al, Nature 309: 255 [1984]; Gossler et al., Proc. Acad. Sci. USA 83: 9065 [1986]; y Robertson et al., Nature 322: 445 [1986]). Los transgenes se pueden introducir de forma eficiente en las células ES mediante transfección de ADN a través de procedimientos conocidos en la técnica,

incluidos coprecipitación en fosfato cálcico, fusión de protoplastos y esferoplastos, lipofección y transfección mediada por DEAE-dextrano. Los transgenes también se pueden introducir en células ES no humanas mediante transducción mediada por retrovirus o mediante microinyección. Dichas células ES transfectadas pueden colonizar después un embrión tras su introducción en el blastocele de un embrión en estadio de blastocisto y contribuye a la línea germinal del animal quimérico resultante (para revisión, véase Jaenisch, Science 240: 1468 [1988]). Antes de la introducción de células ES transfectadas en el blastocele, las células ES transfectadas pueden someterse a varios protocolos de la selección para enriquecer las células ES que tienen integrado el transgén suponiendo que el transgén proporciona un medio para dicha selección. Como alternativa, la reacción en cadena de la polimerasa se puede usar para la detección selectiva de células ES que tienen integrado el transgén. Esta técnica obvia la necesidad de crecimiento de las células ES transfectadas en condiciones selectivas adecuadas antes de la transferencia al blastocele.

Se puede usar recombinación homóloga para inactivar la función de los genes o crear mutantes de delección (p. ej., mutantes en los que se han delecionado las LRR de Nod2). Los procedimientos para la recombinación homóloga se describen en la solicitud de patente de EE.UU. Nº 5.614.396.

15 VII. Plantas transgénicas que expresan genes exógenos de Nod2 y homólogos, mutantes y variantes de los mismos.

Como se ha descrito anteriormente, los homólogos de Nod2 en plantas comparten homología con una clase de productos génicos R resistentes a enfermedades vegetales. La presente invención es útil en procedimientos para crear plantas transgénicas que tienen respuestas alteradas y/o resistencia a patógenos. Las plantas transgénicas expresan un gen exógeno de Nod2 o un homólogo, mutante o variante del mismo (p. ej., las SEC ID Nº 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88). La planta transgénica puede exhibir un fenotipo alterado en comparación con las plantas silvestres. El fenotipo alterado puede ser la sobreexpresión de ARNm para un gen de Nod2 en comparación con los niveles silvestres de expresión de Nod2. El fenotipo alterado puede ser la expresión disminuida de ARNm para un gen endógeno de Nod2 en comparación con los niveles silvestres de expresión de Nod2 endógeno. Los procedimientos para analizar la presencia o ausencia de dichos fenotipos incluyen transferencia Northern, ensayos de protección de ARNm y RT-PCR. El incremento de la expresión génica de Nod2 en la planta transgénica puede conferir una mayor resistencia a los patógenos. El fenotipo observado simula la respuesta inflamatoria inducida por Nod2 en animales. Las plantas transgénicas que expresan este fenotipo se pueden someter a detección selectiva exponiendo las plantas con un patógeno y seleccionando las plantas que muestran resistencia en comparación con las plantas no transgénicas control.

Los vectores son útiles para la transfección de huéspedes vegetales para crear plantas transgénicas. En general, estos vectores comprenden un ácido nucleico de Nod2 (p. ej., las SEC ID Nº 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88) unido operablemente a un promotor y otras secuencias reguladoras (p. ej., potenciadores, señales de poliadenilación etc.) requeridas para la expresión en una planta. El ácido nucleico de Nod2 se puede orientar para producir transcritos sentido o antisentido en función del uso deseado. El promotor es un promotor constitutivo (p. ej., superpromotor o promotor SD). El promotor puede ser un promotor específico de semilla (p. ej., promotor de faseolina [véase, por ejemplo, la patente de EE.UU. Nº 5.589.616], el promotor de napina [véase, por ejemplo, la patente de EE.UU. Nº 5.608.152] o el promotor de la proteína transportadora de acil-CoA [véase, por ejemplo, el documento 5.767.363]).

El vector puede estar adaptado para usar en un proceso de transfección mediado por *Agrobacterium* (véase, por ejemplo, las patentes de EE.UU. 5.981.839, 6.051.757, 5.981.840, 5.824.877, y 4.940.838). La construcción de plásmidos Ti y Ri recombinantes sigue, en general, los procedimientos habitualmente usados con los vectores bacterianos más frecuentes, tales como pBR322. Se puede hacer un uso adicional de elementos genéticos auxiliares que en ocasiones se encuentran con los plásmidos nativos y en ocasiones se construyen a partir de secuencias extrañas. Estos pueden incluir, entre otros, genes estructurales para resistencia a antibióticos como genes de selección.

Estos son dos sistemas de sistemas de vectores plasmídicos Ti y Ri recombinantes actualmente en uso.. El primer sistema se denomina el sistema "cointegrado". En este sistema, el vector lanzadera que contiene el gen de interés se inserta mediante recombinación genética en un plásmido Ti no oncogénico que contiene los elementos tanto de acción en cis como de acción en trans requeridos para la transformación de plantas, como, por ejemplo, en el vector lanzadera pMLJ1 y el plásmido Ti no oncogénico pGV3850. El segundo sistema se denomina el sistema "binario" en el que se usan dos plásmidos; el gen de interés se inserta en un vector lanzadera que contiene los elementos de acción en cis requeridos para la transformación de plantas. Las otras funciones necesarias se proporcionan en trans mediante el plásmido Ti no oncogénico, como pone de manifiesto el vector lanzadera pBIN19 y el plásmido Ti no oncogénico PAL4404. Algunos de estos vectores están disponibles comercialmente.

Puede ser deseable dirigir la secuencia de ácido nucleico de interés a un locus concreto en el genoma de la planta. La integración dirigida a sitio de la secuencia de ácido nucleico de interés en el genoma de la célula vegetal se puede conseguir mediante, por ejemplo, recombinación homóloga usando secuencias derivadas de *Agrobacterium*. En general, las células vegetales se incuban con una cepa de *Agrobacterium* que contiene un vector dirigible en el que las secuencias que son homólogas a una secuencia de ADN dentro del locus diana están flanqueadas por

secuencias de ADN de transferencia (T-ADN) de *Agrobacterium*, como se ha descrito anteriormente (patente de EE.UU. nº 5.501.967). Un experto en la técnica sabe que la recombinación homóloga se puede conseguir usando vectores dirigibles que contienen secuencias que son homólogas a cualquier parte del gen vegetal diana, pertenezcan o no a los elementos reguladores del gen, o las regiones de codificación del gen. La recombinación homóloga se puede conseguir en cualquier región de un gen vegetal siempre que se conozca la secuencia de ácido nucleico de las regiones flanqueantes del sitio diana.

Los ácidos nucleicos de la presente invención también se pueden usar para construir vectores derivados de virus de ARN (+) vegetal (p. ej., virus del mosaico del bromo, virus del mosaico del tabaco, virus del mosaico de la alfalfa, virus del mosaico del pepino, virus del mosaico del tomate y combinaciones e híbridos de los mismos). En general, el polinucleótido de Nod2 insertado se puede expresar de estos vectores como una proteína de fusión (p. ej., proteína de fusión de la proteína de recubrimiento) o de su propio promotor subgenómico u otro promotor. Los procedimientos para la construcción y uso de dichos virus se describen en las patentes de EE.UU. Nº 5.846.795, 5.500.360, 5.173.410, 5.965.794, 5.977.438 y 5.866.785.

Como alternativa, los vectores se pueden construir para expresión en huéspedes distintos a plantas (p. ej., células procarióticas tales como *E. coli*, células de levaduras, *C. elegans* y cultivos celulares de células de mamífero). Los vectores pueden incluir, entre otros, secuencias de ADN cromosómico, no cromosómico y sintético (p. ej., derivados de SV40; plásmidos bacterianos, ADN de fago, baculovirus, plásmidos de levadura, vectores derivados de combinaciones de plásmidos y ADN de fago, y ADN viral, tal como de virus vacunal, adenovirus, virus de la viruela aviar y seudorrabia). Los expertos en la técnica conocen un número elevado de vectores adecuados que son replicables y viables en la célula huésped y que están comercialmente disponibles. Se puede usar otro plásmido o vector siempre que sea replicable y viable en la célula huésped.

Vectores de expresión en bacterias pueden comprender un origen de replicación, un promotor y, opcionalmente, un potenciado adecuados, y, también, cualquier sitio de unión al ribosoma necesario, sitios de poliadenilación, secuencias de terminación de la transcripción y secuencias no transcritas flanqueantes en 5'. Los promotores útiles en la presente invención incluyen, entre otros, LTR retrovirales, promotor de SV40, promotor de CMV, promotor de RSV, promotores *lac* o *trp* de *E. coli*, promotores P_L y P_R del fago lambda, promotores T3, SP6 y T7. Los vectores de expresión recombinante pueden incluir orígenes de replicación y marcadores seleccionables que permiten la transformación de la célula huésped (p. ej., resistencia a tetraciclina o ampicilina en *E. coli* o gen de neomicina fosfotransferasa para la selección en células eucarióticas).

Los vectores descritos anteriormente se pueden usar para expresar la Nod2 en plantas transgénicas. Se conocen varios procedimientos para producir plantas transgénicas.

En algunas formas de realización, la transfección mediada por *Agrobacterium* se usa para crear plantas transgénicas. Dado que la mayoría de las plantas dicotiledóneas son huéspedes naturales para *Agrobacterium*, casi todas las plantas dicotiledóneas pueden transformarse con *Agrobacterium in vitro*. Aunque las plantas monocotiledóneas, y en particular los cereales y las hierbas, no son huéspedes naturales de *Agrobacterium*, también se han llevado cabo trabajos para transformarlas usando *Agrobacterium* (Hooymas-Van Slogteren y col., Nature 311: 763-764 [1984]). Los géneros de plantas que se pueden transformar mediante *Agrobacterium* incluyen *Arabidopsis*, *Chrysanthemum*, *Dianthus*, *Gerbera*, *Euphorbia*, *Pelargonium*, *Ipomoea*, *Passiflora*, *Cyclamen*, *Malus*, *Prunus*, *Rosa*, *Rubus*, *Populus*, *Santalum*, *Allium*, *Lilium*, *Narcissus*, *Ananas*, *Arachis*, *Phaseolus* y *Pisum*.

Para la transformación con *Agrobacterium*, las células de *Agrobacterium* desarmadas se transforman con plásmidos T_i recombinantes de *Agrobacterium tumefaciens* o plásmidos Ri de *Agrobacterium rhizogenes* (como los descritos en la patente de EE.UU. Nº 4.940.838). La secuencia de ácidos nucleicos de interés se integra de forma estable en el genoma de plantas mediante infección con la cepa transformada de *Agrobacterium*. Por ejemplo, las secuencias de ácido nucleico heterólogo se han introducido en tejidos vegetales usando el sistema de transferencia de ADN natural de las bacterias *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes* (para una revisión, véase, Klee y col., Ann. Rev. Plant Phys. 38: 467-486 [1987]).

Estos son procedimientos habituales para transformar células vegetales con *Agrobacterium*. El primer procedimiento es el co-cultivo de *Agrobacterium* con protoplastos aislados cultivados. Este procedimiento requiere un sistema de cultivo establecido que permite cultivar protoplastos y la regeneración de plantas de protoplastos cultivados. El segundo procedimiento es la transformación de células o tejidos con *Agrobacterium*. Este procedimiento requiere (a) que las células o tejidos vegetales se puede transformar con *Agrobacterium* y (b) que las células o tejidos transformados se pueden inducir para regenerar en plantas completas. El tercer procedimiento es la transformación de semillas, ápices o meristemos con *Agrobacterium*. Este procedimiento requiere micropropagación.

Un experto en la técnica sabe que la eficiencia de la transformación por *Agrobacterium* puede potenciarse usando una serie de procedimientos conocidos en la técnica. Por ejemplo, se ha demostrado que la inclusión de una molécula de respuesta natural tal como acerosiringona (AS) al cultivo de *Agrobacterium* potencia la eficiencia de la transformación con *Agrobacterium tumefaciens* [Shahla y col., Plant Molec. Biol. 8: 291 [1987]]. Como alternativa, la eficiencia de transformación se puede potenciar dañando el tejido diana que se va a transformar. El daño del tejido vegetal se puede conseguir mediante, por ejemplo, perforación, maceración, bombardeo con microproyectiles etc. [Véase, por ejemplo, Bidney y col., Plant Molec. Biol. 18: 301 [1992]].

Las células vegetales se pueden transfectar con vectores mediante bombardeo de partículas (es decir, con una pistola génica). Los procedimientos de transferencia génica mediada por partículas se conocen en la técnica, están disponibles comercialmente e incluyen, entre otros, el instrumento de liberación génica dirigida por gas en McCabe, patente de EE.UU. n° 5.584.807. Este procedimiento implica el recubrimiento de la secuencia de ácido nucleico de interés sobre partículas de metal pesado y acelerando las partículas recubiertas bajo la presión del gas comprimido para liberar en el tejido diana.

Otros procedimientos de bombardeo de partículas también están disponibles para la introducción de secuencias de ácido nucleico heterólogas en células vegetales. En general, estos procedimientos implican depositar la secuencia de ácido nucleico de interés en la superficie de partículas pequeñas y densas de un material tal como oro, platino o tungsteno. Las partículas recubiertas se recubren después sobre una superficie rígida, tal como una placa metálica o sobre una lámina transportadora hecha de un material frágil, tal como Mylar. La lámina recubierta se acelera después hacia el tejido biológico diana. El uso de la lámina plana genera una extensión uniforme de partículas aceleradas que maximiza el número de células que reciben partículas en condiciones uniformes, que tiene como resultado la introducción de la muestra de ácido nucleico en el tejido diana.

Las plantas, células y tejidos vegetales transformados con una secuencia de ácido nucleico heterólogo de interés se detectan fácilmente usando procedimientos conocidos en la técnica, incluidos, entre otros, mapas de restricción del ADN genómico, análisis PCR, hibridación ADN-ADN, hibridación ADN-ARN, análisis de la secuencia del ADN y similares.

Adicionalmente, la selección de células vegetales transformadas se puede conseguir usando un gen marcador de selección. Se prefiere, aunque no es necesario, usar un gen marcador de selección para seleccionar células vegetales transformadas. Un gen marcador de selección puede conferir selección positiva o negativa.

Un gen marcador de selección positivo se puede usar en construcciones para la integración aleatoria y la integración dirigida a sitio. Los genes marcadores de selección positiva incluyen genes de resistencia a antibióticos y genes de resistencia a herbicidas y similares. En una forma de realización, el marcador de selección positiva es el gen NPTII que confiere resistencia a la geneticina (G418) o a la kanamicina. En otra forma de realización, el marcador de selección positiva es el gen HPT que confiere resistencia a la higromicina. La elección del gen marcador de selección positiva no es crucial para la invención siempre que codifique un producto polipeptídico funcional. Los genes de selección positiva conocidos en la técnica incluyen, entre otros, el gen de ALS (resistencia al clorosulfurón) y el gen de DHFR (resistencia al metotrexato).

También se pueden incluir en las construcciones un gen marcador de selección negativa. El uso de uno o más genes marcadores de selección negativa en combinación con un gen marcador de selección positiva se prefiere en las construcciones usadas para recombinación homóloga. Los genes marcadores de selección negativa se colocan, en general, fuera de las regiones implicadas en el acontecimiento de recombinación homóloga. El gen marcador de selección negativa sirve para proporcionar una desventaja (preferentemente letalidad) para las células que tienen integrado estos genes en su genoma de modo que se pueden expresar. Las células en las que los vectores dirigibles para recombinación homóloga se integran aleatoriamente en el genoma serán dañadas o morirán debido a la presencia del gen marcador de selección negativa. Cuando un gen marcador de selección positiva se incluye en la construcción, solo sobrevivirán las células que tienen el gen marcador de selección positiva integrado en su genoma.

La elección del gen marcador de selección negativa no es crucial para la invención siempre que codifique un polipéptido funcional en la célula vegetal transformada. El gen de selección negativa puede elegirse, por ejemplo, del gen *aux-2* del plásmido Ti de *Agrobacterium*, el gen *tk* de SV40, el citocromo P450 de *Streptomyces griseolus*, el gen *Adh* de *Maize* o *Arabidopsis*, etc. Se puede usar cualquier gen que codifica una enzima capaz de convertir una sustancia que, por otro lado, es dañina para las células vegetales en una sustancia que es dañina para las células vegetales. Se contempla que los polinucleótidos de Nod2 de la presente invención se pueden usar para incrementar o disminuir el nivel de ARNm de Nod2 y/o la proteína en células transfectadas en comparación con los niveles en las células silvestres. De acuerdo con esto, en algunas formas de realización, la expresión en plantas mediante los procedimientos descritos anteriormente conduce a la sobreexpresión de Nod2 en plantas transgénicas, tejidos vegetales o células vegetales.

En otras formas de realización de la presente invención, los polinucleótidos Nod2 se usan para disminuir el nivel de la proteína o el ARNm de Nod2 en plantas transgénicas, tejidos vegetales o células vegetales en comparación con las plantas silvestres, tejidos vegetales o células vegetales. Un procedimiento de reducir la expresión de Nod2 usa la expresión de transcritos antisentido. El ARN antisentido se ha usado para inhibir los genes diana de las plantas de un modo específico de tejido (p. ej., van der Krol y col., *Biotechniques* 6:958 [1988]). La inhibición antisentido se ha demostrado usando toda la secuencia de ADNc además de una secuencia de ADNc parcial (p. ej., Sheehy y col., *Proc. Natl. Acad. Sci. USA* 85:8805 [1988]; Cannon y col., *Plant iMol. Biol.* 15:39 [1990]). También existen pruebas de que el fragmento de la secuencia no codificante en 3' y los fragmentos de secuencia de codificación en 5', que contienen los 41 pares de bases de un ADNc de 1,87 kb, pueden desempeñar importantes papeles en la inhibición antisentido (Ch'ng y col., *Proc. Natl. Acad. Sci. USA* 86:10006 [1989]).

De acuerdo con esto, en algunas formas de realización, los ácidos nucleicos de Nod2 de la presente invención (p. ej., la SEC ID N° 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88, y los fragmentos y variantes de los mismos) se

orientan en un vector y se expresan para producir transcritos antisentido. Para conseguir esto, se transcribirá un segmento de ácido nucleico del que el gen deseado está clonado y unido operablemente a un promotor tal como la hebra antisentido del ARN. El casete de expresión se transforma después en plantas y se produce la hebra antisentido de ARN. El segmento de ácido nucleico que se va a introducir generalmente será sustancialmente idéntico a al menos una porción del gen o genes endógenos que se van a reprimir. No obstante, la secuencia no tiene que ser perfectamente idéntica para inhibir la expresión. Los vectores de la presente invención se pueden diseñar de un modo tal que el efecto inhibitor se aplique a otras proteínas dentro de una familia de genes que exhiben homología o una homología sustancial con el gen diana.

Además, para la supresión antisentido, la secuencia introducida tampoco tiene que tener la longitud completa con respecto al producto de transcripción primario o al ARNm completamente procesado. En general, la homología mayor se puede usar para compensar el uso de una secuencia más corta. Además, la secuencia introducida no tiene que tener el mismo patrón de intrones o exones y la homología de los segmentos no codificantes puede ser igualmente eficaz. Normalmente se usará una secuencia de entre aproximadamente 30 o 40 nucleótidos y nucleótidos de aproximadamente la longitud completa, aunque se prefiere una secuencia de al menos aproximadamente 100 nucleótidos, es más preferida una secuencia de al menos aproximadamente 200 nucleótidos y es especialmente preferida una secuencia de al menos aproximadamente 500 nucleótidos.

Las moléculas de ARN catalíticas o ribozimas también se pueden usar para inhibir la expresión del gen o genes diana. Es posible diseñar ribozimas que se aparean específicamente con casi cualquier ARN diana y escinden la estructura fosfodiéster en una localización específica, de un modo que inactivan funcionalmente el ARN diana. Al llevar a cabo esta escisión, la ribozima no se altera y, por tanto, es capaz de reciclarse y escindir otras moléculas, lo que la convierten en una verdadera enzima. La inclusión de secuencias de ribozima dentro de ARN antisentido confiere actividad de escisión de ARN, de modo que aumenta la actividad de las construcciones.

Se ha identificado una serie de clases de ribozimas. Una clase de ribozimas deriva de una serie de ARN circulares pequeños que pueden autoescindir y replicarse en plantas. Los ARN se replican solos (ARN viroides) o con un virus colaborador (ARN satélites). Ejemplos incluyen ARN del viroide de la mancha solar del aguacate y los ARN satélites del virus de la mancha anular del tabaco, el virus del veteado transitorio de la alfalfa, el virus del moteado de terciopelo del tabaco, el virus del moteado de *Solanum nodiflorum* y el virus del moteado del trébol subterráneo. El diseño y el uso de ribozimas específicas de ARN diana se describen en Haseloff, y col., *Nature* 334: 585 [1988].

Otro procedimiento de reducir la expresión de *Nod2* usa el fenómeno de cosupresión o silenciamiento génica (véase, p. ej., la patente de EE.UU. nº 6.063.947). El fenómeno de cosupresión también se ha usado para inhibir los genes diana de la planta de un modo específico de tejido. La cosupresión de un gen endógeno usando una secuencia de ADNc de longitud completa, así como una secuencia de ADNc parcial (730 pb de un ADNc de 1770 pb) se conocen (p. ej., Napoli y col., *Plant Cell* 2: 279 [1990]; van der Krol et al., *Plant Cell* 2: 291 [1990]; Smith et al., *Mol. Gen. Genetics* 224: 477[1990]). De acuerdo con esto, en algunas formas de realización, los ácidos nucleicos de *Nod2* (p. ej., las SEC ID Nº 1, 33, 54, 56, 58, 60, 62, 64, 66, 68, 84, 86 y 88) y fragmentos y variantes de los mismos se expresan en otras especies de plantas para efectuar la cosupresión de un gen homólogo.

En general, cuando se desea inhibir la expresión se produce alguna transcripción de la secuencia introducida. El efecto se puede producir cuando la secuencia introducida no contiene una secuencia de codificación *per se*, sino solo secuencias intrónicas o no traducidas homólogas a secuencias presentes en el transcrito primario de la secuencia endógena. En general, la secuencia introducida será sustancialmente idéntica a la secuencia endógena destinada para reprimirse. Esta mínima identidad normalmente será mayor que aproximadamente 65 %, pero una identidad mayor podría ejercer una represión más eficaz de la expresión de las secuencias endógenas. Se prefiere una identidad sustancialmente mayor de más de aproximadamente el 80 %, aunque sería lo más preferido aproximadamente un 95 % de identidad absoluta. Como con la regulación antisentido, el efecto deberá aplicarse a cualquier otra proteína dentro de una familia similar de genes que exhiben homología o una homología sustancial.

Para la cosupresión, la secuencia introducida en el casete de expresión, que necesita menos que una identidad absoluta, tampoco tiene que tener la longitud completa con respecto al producto de transcripción primario o al ARNm completamente procesado. Esto puede preferirse para evitar la producción concurrente de algunas plantas que sobreexpresan. Una identidad mayor en una secuencia más corta que la longitud completa compensa por una secuencia más larga menos idéntica.

Además, la secuencia introducida no tiene que tener el mismo patrón de intrones o exones y la identidad de los segmentos no codificantes será igualmente eficaz. Normalmente se usa una secuencia de los intervalos de tamaño que se han indicado anteriormente para la regulación antisentido.

VIII. Detección selectiva de fármacos usando *Nod2*

La presente invención es útil en procedimientos y composiciones para usar *Nod2* como diana para la detección selectiva de fármacos que pueden alterar, por ejemplo, la señalización de RICK y, por tanto, los efectos fisiológicos de NF-κB (p. ej., la respuesta inflamatoria). Por ejemplo, los fármacos que inducen o inhiben las respuestas inflamatorias mediadas por NF-κB se pueden identificar mediante detección selectiva para compuestos dirigidos a *Nod2* o que regulan la expresión génica de *Nod2*.

La presente invención no está limitada a ningún mecanismo de acción concreto. De hecho, no es necesario conocer el mecanismo de acción para la práctica de la presente invención. No obstante, se contempla que Nod2 se une a RICK y esta unión tiene como resultado la activación de NF- κ B. De acuerdo con esto, se contempla que los ensayos de unión son útiles para la detección selectiva de compuestos que bloquean la unión de Nod2 a RICK. En concreto, se contempla que dichas detecciones selectivas pueden identificar compuestos que son útiles para inhibir la actividad de NF- κ B y, por tanto, para tratar la enfermedad de Crohn. La unión no necesita usar RICK y Nod2 de longitud completa. De hecho, las porciones de RICK y Nod2 se pueden usar en los ensayos de unión. Por ejemplo, en algunas formas de realización se usa un fragmento de Nod2 que contiene los dos dominios CARD en el ensayo de unión.

La presente invención se puede usar en procedimientos de detección selectiva para compuestos que incrementan o disminuyen la unión de Nod2 a patógenos, componentes patogénicos o proteínas de unión a patógeno y, en consecuencia, afectan a la señalización cadena abajo y a la activación de NF- κ B. Se puede usar Nod2 de tipo silvestre o un fragmento de la misma o se puede usar Nod2 que contiene una o más variaciones (p. ej., mutaciones o polimorfismos).

En un procedimiento de detección selectiva, el sistema de dos híbridos se usa para la detección selectiva de compuestos (p. ej., fármacos) capaces de alterar (p. ej., inhibir) las funciones Nod2 (p. ej., la transducción de señal mediada por NF- κ B) *in vitro* o *in vivo*. Un sitio de unión a GAL4, unido a un gen indicador tal como *lacZ*, se pone en contacto en presencia y ausencia de un compuesto candidato con un dominio de unión a GAL4 unido a un fragmento de Nod2 y un dominio de transactivación de GAL4 II unido a un fragmento de NF- κ B. La expresión del gen indicador se monitoriza y una disminución en la expresión es una indicación de que el compuesto candidato inhibe la interacción de Nod2 con NF- κ B. Como alternativa, el efecto de los compuestos candidatos sobre la interacción de Nod2 con otras proteínas (p. ej., proteínas que se sabe que interaccionan directa o indirectamente con NF- κ B) se puede analizar de un modo similar.

En otro procedimiento de detección selectiva, los compuestos candidatos se evalúan por su capacidad para alterar la señalización de Nod2 poniendo en contacto Nod2, NF- κ B, las proteínas asociadas con NF- κ B o fragmentos de los mismos con el compuesto candidato y determinando la unión del compuesto candidato al péptido. La proteína o fragmentos proteicos se inmovilizan usando procedimientos conocidos en la técnica, tal como unión de una proteína de fusión GST-Nod2 a una perla polimérica que contiene glutatión. Un gen quimérico que codifica una proteína de fusión GST se construye condensando el ADN que codifica el polipéptido o fragmento polipeptídico de interés con el ADN que codifica el extremo carboxilo de GST (véase, por ejemplo, Smith y col., Gene 67:31 [1988]). La construcción de fusión se transforma después en un sistema de expresión adecuado (p. ej., *E. coli* XA90) en el que la expresión de la proteína de fusión GST se puede inducir con isopropil- β -D-tiogalactopiranosido (IPTG). La inducción con IPTG debería dar la proteína de fusión como constituyente mayoritario de proteínas celulares solubles. Las proteínas de fusión se pueden purificar mediante procedimientos conocidos por los expertos en la técnica, incluida la purificación mediante cromatografía de afinidad con glutatión. La unión del compuesto candidato a las proteínas o fragmentos proteicos se correlaciona con la capacidad del compuesto para alterar la vía de transducción de la señal y, de este modo, regular los efectos fisiológicos de Nod2 (p. ej., apoptosis).

En otro procedimiento de detección selectiva, uno de los componentes del sistema de señalización de Nod2/NF- κ B, tal como Nod2 o un fragmento de Nod2, se inmoviliza. Los polipéptidos se pueden inmovilizar usando procedimientos conocidos en la técnica, tales como adsorción en una placa de microtitulación de plástico o unión específica de una proteína de fusión GST a una perla polimérica que contiene glutatión. Por ejemplo, GST-Nod2 se une a perlas de glutatión-seferosa. Después, el péptido inmovilizado se pone en contacto con otro péptido con el que es capaz de unirse en presencia o ausencia de un compuesto candidato. Después, el péptido no unido se elimina y el complejo se solubiliza y analiza para determinar la cantidad de péptido marcado unido. Una disminución de la unión es una indicación de que el compuesto candidato inhibe la interacción de Nod2 con el otro péptido. Una variación de este procedimiento permite la detección selectiva de compuestos que son capaces de romper un complejo proteína/proteína formada previamente. Por ejemplo, se puede inmovilizar un complejo que comprende Nod2 o un fragmento de Nod2 unido a otro péptido, tal como se ha descrito anteriormente y se pone en contacto con un compuesto candidato. La disolución del complejo por el compuesto candidato se correlaciona con la capacidad del compuesto para alterar o inhibir la interacción entre Nod2 y el otro péptido.

Otra técnica que se puede usar para la detección selectiva del fármaco proporciona detección selectiva de alto rendimiento para compuestos que poseen una afinidad de unión adecuada por los péptidos de Nod2 y se describe con detalle en el documento WO84/03564. En pocas palabras, se sintetizan grandes números de pequeños compuestos peptídicos de ensayo diferentes sobre un sustrato sólido, tal como agujas de plástico o alguna otra superficie. Los compuestos de prueba peptídicos se hacen reaccionar con péptidos de Nod2 y se lavan. Los péptidos de Nod2 unidos se detectan por medio de procedimientos bien conocidos en la técnica.

Otra técnica usa anticuerpos contra Nod2 generados como se ha tratado anteriormente. Dichos anticuerpos capaces de unirse específicamente a los péptidos de Nod2 compiten con un compuesto de ensayo por la unión a Nod2. De este modo, se pueden usar anticuerpos para detectar la presencia de algún péptido que compara uno o más determinantes antigénicos del péptido de Nod2.

Los compuestos se someten a detección selectiva por su capacidad para inhibir la unión de los componentes patógenos (p. ej., incluidos, entre otras, proteínas de superficie celular bacteriana, proteínas de hongos, proteínas de parásitos y proteínas de virus) a Nod2. Cualquier ensayo de detección selectiva se puede usar, incluidos, entre otros, los descritos en el presente documento.

5 La presente invención contempla el uso de líneas celulares transfectadas con Nod2 y variantes de la misma para la detección selectiva de compuestos según su actividad y, en particular, para la detección selectiva de alto rendimiento de compuestos de bibliotecas combinatorias (p. ej., bibliotecas que contienen más de 10^4 compuestos). Las líneas celulares se pueden usar en diversos procedimientos de detección selectiva. Las células se pueden usar en ensayos de segundo mensajero que monitorizan la transducción de señal tras la activación de receptores de superficie celular o las células se pueden usar en ensayos de gen indicador que monitorizan las respuestas celulares a nivel de transcripción/traducción o las células se pueden usar en ensayos de proliferación celular para monitorizar la respuesta de crecimiento global/no crecimiento de las células a estímulos externos.

10 En ensayos de segundo mensajero, las células huésped de transfectan preferentemente como se ha descrito anteriormente con vectores que codifican Nod2 o variantes o mutantes de la misma. Después, las células huésped se tratan con un compuesto o pluralidad de compuestos (p. ej., de una biblioteca combinatoria) y se analiza la presencia o ausencia de una respuesta. Se contempla que al menos algunos de los compuestos de la biblioteca combinatoria pueden servir como agonistas, antagonistas, activadores o inhibidores de la proteína o proteínas codificadas por los vectores. También se contempla que al menos algunos de los compuestos de la biblioteca combinatoria pueden servir como agonistas, antagonistas, activadores o inhibidores de la proteína que actúa antes o después de la proteína codificada por el vector en una vía de transducción de señal.

15 Los ensayos de segundo mensajero pueden medir las señales fluorescentes de moléculas indicadoras que responden a cambios intracelulares (p. ej., concentración de Ca^{2+} , potencial de membrana, pH, IP_3 , $AMPc^+$, liberación de ácido araquidónico) debido a la estimulación de los receptores de membrana y canales iónicos (p. ej., canales iónicos dependientes de ligando; véase Denyer et al., *Drug Discov. Today* 3: 323 [1998]; y Gonzales et al., *Drug. Discov. Today* 4: 431-39 [1999]). Ejemplos de moléculas indicadoras incluyen, entre otras, sistemas FRET (transferencia de energía de resonancia de fluorescencia) (p. ej., Cuo-lípidos y oxonoles, EDAN/DABCYL), indicadores sensibles a calcio (p. ej., Fluo-3, FURA 2, INDO 1, and FLUO3/AM, BAPTA AM), indicadores sensibles a cloruro (p. ej., SPQ, SPA), indicadores sensibles a potasio (p. ej., PBFI), indicadores sensibles a sodio (p. ej., SBFI) e indicadores sensibles a pH (p. ej., BCECF).

20 En general, las células huésped se cargan con el indicador antes de la exposición al compuesto. Las respuestas de las células huésped al tratamiento con los compuestos se pueden detectar mediante procedimientos conocidos en la técnica, incluidos, entre otros, microscopia de fluorescencia, microscopia confocal (p. ej., sistemas FCS), citometría de flujo, dispositivos microfluídicos, sistemas FLIPR (véase, por ejemplo, Schroeder and Neagle, *J. Biomol. Screening* 1: 75 [1996]), y sistemas de lectura de placas. En algunas realizaciones preferidas, la respuesta (p. ej., incremento de la intensidad de la fluorescencia) producida por el compuesto de actividad desconocida se compara con la respuesta generada por un agonista conocido y se expresa como un porcentaje de la respuesta máxima del agonista conocido. La respuesta máxima producida por un agonista conocido se define como una respuesta de 100 %. Asimismo, la respuesta máxima registrada tras la adición de un agonista a una muestra que contiene un antagonista conocido o de ensayo es detectablemente menor que la respuesta del 100 %.

30 Las células también son útiles en ensayos génicos indicadores. Los ensayos génicos indicadores implican el uso de células huésped transfectadas con vectores que codifican un ácido nucleico que comprende elementos de control de la transcripción de un gen diana (es decir, un gen que controla la expresión biológica y la función de una enfermedad diana) sometidas a corte y empalme a una secuencia de codificación para un gen indicador. Por tanto, la activación del gen diana tiene como resultado la activación del producto génico indicador. Como se ha descrito anteriormente se contempla que Nod2 se une a RICK y esta unión tiene como resultado la activación de NF- κ B. Por tanto, en algunas formas de realización, la construcción del gen indicador comprende la región reguladora en 5' (p. ej., promotores y/o potenciadores) de una proteína cuya expresión está controlada por NF- κ B en asociación operable con un gen indicador (véase en el Ejemplo 4 e Inohara y col., *J. Biol. Chem.* 275: 27823 [2000] una descripción de la construcción indicadora de luciferasa pBVlx-Luc). Ejemplos de genes indicadores que encuentran utilidad en la presente invención incluyen, entre otros, cloranfenicol transferasa, fosfatasa alcalina, luciferasas de luciérnaga y bacterianas, beta-galactosidasa, beta-lactamasa y proteína fluorescente verde. La producción de estas proteínas, con la excepción de la proteína fluorescente verde, se detecta mediante el uso de productos quimioluminiscentes, colorimétricos, bioluminiscentes de sustratos específicos (p. ej., X-gal y luciferina). Las comparaciones entre compuestos de actividades conocidas y desconocidas se pueden realizar como se ha descrito en lo que antecede.

55 **IX. Composiciones farmacéuticas que contienen ácido nucleico de Nod2, péptidos y análogos**

La presente invención se puede usar para composiciones farmacéuticas que pueden comprender todas o porciones de secuencias polinucleotídicas de Nod2, polipéptidos de Nod2, inhibidores o antagonistas de la bioactividad de Nod2, incluidos anticuerpos solos o en combinación con al menos otro agente, tal como un compuesto estabilizante, que puede administrarse en cualquier transportador farmacéutico estéril biocompatible, incluidos, entre otros, solución salina, solución salina tamponada, dextrosa y agua.

Los polipéptidos de la presente invención se pueden usar en el tratamiento de enfermedades o la alteración de estados fisiológicos caracterizados por apoptosis de células u otros efectos mediados por NF- κ B y son útiles para inhibir la interacción de Nod2 con NF- κ B y proteínas asociadas a NF- κ B administrando péptidos o fragmentos peptídicos de Nod2. Los péptidos se pueden administrar al paciente por vía intravenosa en un vehículo farmacéuticamente aceptable, tal como solución fisiológica salina. Se pueden usar procedimientos estándar para la liberación intracelular de péptidos (p. ej., liberación mediante liposomas). Dichos procedimientos son bien conocidos para los expertos habituales en la técnica. Las formulaciones de la presente invención son útiles para administración parenteral, tal como intravenosa, subcutánea, intramuscular e intraperitoneal. La administración terapéutica de un polipéptido intracelularmente también se puede conseguir usando terapia génica como se ha descrito en lo que antecede.

Como es bien conocido en las técnicas médicas, las dosis para un paciente cualquiera pueden depender de muchos factores, incluidos el tamaño del paciente, el área de superficie corporal, la edad, el compuesto concreto que se va a administrar, el sexo, la hora y la vía de administración, el estado de salud general y la interacción con otros fármacos que se estén administrando de forma concurrente.

De acuerdo con lo anterior, las secuencias de nucleótidos de Nod2 y de aminoácidos de Nod2 se pueden administrar a un paciente en monoterapia o en combinación con otras secuencias de nucleótidos, fármacos u hormonas, en composiciones farmacéuticas en las que se mezcla con excipiente(s) o transportadores farmacéuticamente aceptable. El vehículo farmacéuticamente aceptable puede ser farmacéuticamente inerte. Las secuencias de polinucleótidos de Nod2 o las secuencias de de aminoácidos de Nod2 se pueden administrar en monoterapia a un sujeto individual o que padece una enfermedad.

Dependiendo de la afección que se esté tratando, estas composiciones farmacéuticas se pueden formular y administrar por vía sistémica o local. Las técnicas de formulación y administración se pueden encontrar en la última edición de Remington's Pharmaceutical Sciences (Ed. Maack Publishing Co, Easton, Pa.). Las vías adecuadas pueden incluir, por ejemplo, administración oral o transmucosa, así como liberación parenteral, incluida la administración intramuscular, subcutánea, intramedular, intratecal, intraventricular, intravenosa, intraperitoneal, o intranasal.

Para inyección, las composiciones farmacéuticas se pueden formular en soluciones acuosas, preferentemente en tampones fisiológicamente compatibles, tales como solución de Hank, solución de Ringer o solución salina tamponada fisiológicamente. Para la administración tisular o celular, en la formulación se usan penetrantes adecuados para atravesar la barrera concreta. Tales penetrantes se conocen generalmente en la técnica.

Las composiciones farmacéuticas se pueden formular usando soportes farmacéuticamente aceptables bien conocidos en la técnica con posologías adecuadas para la administración oral. Dichos vehículos permiten formular las composiciones farmacéuticas en forma de comprimidos, píldoras, cápsulas, líquidos, geles, jarabes, suspensiones espesas y similares, para ingestión oral o nasal por el paciente que se va a tratar.

Composiciones farmacéuticas incluyen composiciones en las que los ingredientes activos están contenidos en una cantidad eficaz para alcanzar el propósito previsto.. Por ejemplo, una cantidad eficaz de Nod2 puede ser la cantidad que suprime la apoptosis. La determinación de las cantidades eficaces entra bien en de la capacidad de los expertos en la técnica, especialmente a la luz de la divulgación proporcionada en el presente documento.

Además de los ingredientes activos, estas composiciones farmacéuticas pueden contener vehículos farmacéuticamente aceptables que comprenden excipientes y sustancias auxiliares que facilitan el procesamiento de los compuestos activos en preparaciones que pueden usarse farmacéuticamente. Las preparaciones formuladas para administración oral, puede estar en forma de comprimidos, grageas, cápsulas, o soluciones.

Las composiciones farmacéuticas pueden fabricarse de un modo conocido en la técnica, por ejemplo por medio de procedimientos convencionales mezclado, disolución, granulación, formación de grageas, trituración, emulsificación, encapsulación, atrapamiento o liofilización.

Las formulaciones farmacéuticas para administración parenteral incluyen soluciones acuosas de los compuestos activos en forma hidrosoluble. Adicionalmente se pueden preparar suspensiones de los compuestos activos como suspensiones oleosas para inyección adecuadas. Los disolventes o vehículos lipófilos adecuados incluyen aceites grasos tales como aceite de sésamo, o ésteres de ácidos grasos sintéticos, tales como oleato de etilo o triglicéridos, o liposomas. Las suspensiones inyectables acuosas pueden contener sustancias que aumentan la viscosidad de la suspensión, tales como carboximetilcelulosa sódica, sorbitol, o dextrano. Opcionalmente, la suspensión también puede contener estabilizantes o agentes adecuados para aumentar la solubilidad de los compuestos y permitir la preparación de soluciones muy concentradas.

Las preparaciones farmacéuticas para uso oral se pueden obtener mediante combinación de compuestos activos con excipientes sólidos, opcionalmente moliendo una mezcla resultante y procesando la mezcla de gránulos, tras añadir las sustancias auxiliares adecuadas, si se desea, para obtener núcleos de comprimidos o núcleos de grageas. Excipientes adecuados son cargas de hidratos de carbono o de proteínas, tales como azúcares, incluidos lactosa, sacarosa, manitol o sorbitol; almidón de maíz, trigo, arroz, patata; celulosa tal como metilcelulosa,

hidroxipropilmetilcelulosa o carboximetilcelulosa sódica; y gomas, incluidas arábica y de tragacanto; y proteínas tales como gelatina y colágeno. Si se desea, se pueden añadir agentes disgregantes o solubilizantes, tales como la polivinilpirrolidona reticulada, agar, ácido alginico o una sal de los mismos, tal como alginato de sodio.

5 Los núcleos de grageas se proporcionan con recubrimientos adecuados, tales como soluciones de azúcar concentrado, que también pueden contener goma arábica, talco, polivinilpirrolidona, gel de carbopol, polietilenglicol y/o dióxido de titanio, soluciones de laca y disolventes orgánicos adecuados o mezclas de disolventes. Se pueden añadir colorantes o pigmentos a los recubrimientos de comprimidos o grageas para la identificación de productos o para caracterizar la cantidad de compuesto activo (es decir dosificación).

10 Las preparaciones farmacéuticas que se pueden usar por vía oral incluyen cápsulas duras fabricadas de gelatina, así como cápsulas blandas selladas fabricadas de gelatina y un revestimiento tal como glicerol o sorbitol. Las cápsulas duras pueden contener ingredientes activos mezclados con una carga o aglutinantes, tales como lactosa o almidones, lubricantes, tales como talco o estearato de magnesio, y, opcionalmente, estabilizantes. En cápsulas blandas, los compuestos activos pueden disolverse o suspenderse en líquidos adecuados, tales como aceites grasos, parafina líquida o polietilenglicol líquido con o sin estabilizantes.

15 Las composiciones que comprenden un compuesto de la invención formulado en un vehículo farmacéutico aceptable se pueden preparar, introducir en un contenedor adecuado y etiquetarse para el tratamiento de una afección para la que están indicadas. Para secuencias de polinucleótidos o de aminoácidos de Nod2, las afecciones indicadas en la etiqueta pueden incluir el tratamiento de una afección relacionada con la apoptosis.

20 La composición farmacéutica puede proporcionarse en forma de una sal y se puede formar con ácidos, incluidos, entre otros, ácido clorhídrico, sulfúrico, acético, láctico, tartárico, málico, succínico etc. Las sales tienden a ser más solubles en disolventes acuosos u otros protónicos que son las correspondientes formas de base libre. En otros casos, la preparación preferida puede ser un polvo liofilizado en histidina 1 mM-50 mM, sacarosa 0,1%-2%, manitol 2%-7% a un intervalo de pH de 4,5 a 5,5 que se combina con tampón antes de usar.

25 Para cualquier compuesto usado en el procedimiento de la invención, la dosis terapéuticamente eficaz se puede estimar inicialmente a partir de ensayos con cultivos celulares. Después, preferentemente, la dosis se puede formular en modelos animales (particularmente modelos murinos) para conseguir un intervalo de concentración circulante deseable que ajuste los niveles de Nod2.

30 Una dosis terapéuticamente eficaz se refiere a la cantidad de Nod2 que alivia los síntomas del estado de enfermedad. La eficacia terapéutica y la toxicidad de dichos compuestos se pueden determinar mediante procedimientos farmacéuticos estándar en cultivos celulares o animales experimentales, por ejemplo, para determinar la DL₅₀ (la dosis letal para el 50% de la población) y la DE₅₀ (la dosis terapéuticamente eficaz en el 50% de la población). La relación de dosis entre efectos tóxicos y terapéuticos es el índice terapéutico y puede expresarse como la relación DE₅₀/DL₅₀. Se prefieren los compuestos que exhiben índices terapéuticos grandes. Los datos obtenidos de estos ensayos de cultivos celulares y estudios animales adicionales se usan en la formulación de
35 in intervalo de dosificación para uso humano. Las dosis de tales compuestos están, preferentemente, dentro de un intervalo de concentraciones en circulación que incluyen la DE₅₀ con poca o nula toxicidad. La dosificación varía dentro de este intervalo dependiendo de la forma de dosificación empleada, la sensibilidad del paciente y la vía de administración.

40 El médico individual escoge la dosificación exacta según el paciente que va a tratar. La dosificación y la administración se ajustan para proporcionar niveles suficientes del resto activo o para mantener el efecto deseado. Entre los factores adicionales que se pueden tener en cuenta se incluyen la gravedad de la enfermedad, la edad, el peso y el sexo del paciente, la dieta, la hora y la frecuencia de administración, la(s) combinaciones de fármacos, las sensibilidades de la reacción y la tolerancia/respuesta al tratamiento. Las composiciones farmacéuticas de acción prolongada podrían administrar cada 3 a 4 días, cada semana o una vez cada dos semanas dependiendo de la
45 semivida y el índice de aclaramiento de la formulación concreta.

Las cantidades de dosificación normales varían de 0,1 a 100.000 microgramos, hasta una dosis total de aproximadamente 1 g, en función de la vía de administración. En la literatura se proporcionan guías en cuanto a dosificaciones y procedimientos de liberación (véase, por ejemplo, las patentes de EE.UU. N° 4.657.760, 5.206.344, o 5.225.212).

50 Los expertos en la técnica usarán diferentes formulaciones para Nod2 y para los inhibidores de Nod2. La administración en la médula ósea puede requerir liberación de un modo diferente de inyecciones intravenosas.

Parte experimental

Los ejemplos siguientes se proporcionan con el fin de demostrar e ilustrar adicionalmente determinadas realizaciones y aspectos preferidos de la presente invención y no deben interpretarse como limitantes de su ámbito.

55 En las divulgación experimental siguiente se aplican las siguientes abreviaturas: eq (equivalentes); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); g (gramos); mg

(miligramos); µg (microgramos); ng (nanogramos); l o L (litros); ml (mililitros); µl (microlitros); cm (centímetros); mm (milímetros); µm (micrómetros); nm (nanómetros); °C (grados centígrados); U (unidades), mU (miliunidades); min. (minutos); s (segundos); % (por ciento); kb (kilobase); pb (par de bases); PCR (reacción en cadena de la polimerasa); BSA (seroalbúmina bovina); Fisher (Fisher Scientific, Pittsburgh, PA); Sigma (Sigma Chemical Co., St. Louis, MO.); Promega (Promega Corp., Madison, WI); Perkin Elmer (Perkin-Elmer/Applied Biosystems, Foster City, CA); Boehringer Mannheim (Boehringer Mannheim, Corp., Indianapolis, IN); Clontech (Clontech, Palo Alto, CA); Qiagen (Qiagen, Santa Clarita, CA); Stratagene (Stratagene Inc., La Jolla, CA); National Biosciences (National Biosciences Inc, Plymouth Minn.) y NEB (New England Biolabs, Beverly, MA), CARD (dominio de reclutamiento de caspasa); EST (marca de secuencia expresada); HA (hemaglutinina); IκB (inhibidor de NF-κB); IKK (IκB quinasa); LRR (repetición ricas en leucina); NBD (dominio de unión al nucleótido); NF-κB (factor nuclear κB); TNFα (factor α de necrosis tumoral); wt (silvestre); Ab (anticuerpo); IL-1 (interleucina 1); IL-1R (receptor de la IL-1); LPS (lipopolisacárido); LTA (ácido lipoteicoico); PGN (peptidoglicano); SBLP (lipoproteína bacteriana sintética); y TLR (receptor de tipo toll).

Metodología

15 **Reactivos.** El LPS de varias fuentes en este estudio se obtuvo de Sigma (St. Louis, MO). El PGN de *Staphylococcus aureus* se obtuvo de Fluka-Chemie (Buchs, Alemania). El manano de *Candida albicans* 20A fue un regalo de P. Lehmann (Medical College of Ohio). PaM3CysSerLyS4, un análogo de la lipoproteína bacteriana sintética fue un regalo de A. Zychlinsky (New York University School of Medicine).

20 **Aislamiento del ADNc de Nod2.** Las secuencias de nucleótidos que codifican péptidos con homología con Nod1 (números de acceso en GeneBank AC007728 y AQ534686) se encontraron en la base de datos de genoma pública usando el programa TBLASTN. La región de codificación de la Nod2 humana se obtuvo mediante amplificación con PCR-transcriptasa inversa (RT) y RACE en 5' usando los fragmentos de ADNc de cebadores oligonucleotídicos específicos de Nod2 y ARN, de tejido de mama primario como molde. La 5' RACE se realizó usando un kit comercial (Roche Molecular Biochemicals, Indianapolis, IN). Para la PCR se usaron tres conjuntos de cebadores: 5'-ATGTGCTCGCAGGAGGCTTTTCAGGCA-3' (SEC ID N°:37) and 5'-CGCCTCACCCACCACCAGCACAGTGT-3' (SEC ID N°:38); 5'-CATGGCTGGACCCCGCAGAAGAGCCCA-3' (SEC ID N°:39) y 5'-CATGCCCGGGTTCATCTGGCTCATCCGG-3' (SEC ID N°:40); 5'-GCCATGCCCGGGTTCATCTGGCTCATC-3' (SEC ID N°:41) y 5'-TGAGTCGAGACATGGGAAAGCTGCTTC-3' (SEC ID N°:42). Para 5' RACE, se usó el cebador inicial 5'AGCAGCTCGACCAGCTGGCTCCTCTGT-3' (SEC ID N° 43) y el producto se amplificó mediante PCR con el cebador anclado y el segundo cebador específico de Nod2. 5'-GACAGGCCCAAGTACCCTTATTCCAGA-3' (SEC ID N° 44). Los fragmentos de ADNc resultantes se digirieron con enzimas de restricción y se ligaron para generar un ADNc único que contenía todo el marco de lectura abierto de Nod2. La secuencia de ADNc se verificó mediante secuenciación de nucleótidos.

35 **Análisis mediante transferencia Northern y RT-PCR de la expresión de Nod2.** Un fragmento de 3,7 kb que contiene toda la región de codificación de Nod2 se radiomarcó mediante cebado aleatorio usando un kit comercial (Roche Molecular Biochemicals) y se aplicó para el análisis de transferencias de ARN poli(A)' humana de varios tejidos (Clontech Laboratories, Palo Alto, CA) de acuerdo con las instrucciones del fabricante Los leucocitos de sangre periférica se obtuvieron de sangre venosa heparinizada de voluntarios sanos mediante centrifugación en gradiente de densidades con Ficoll-Paque (Amersham Pharmacia Biotech, Uppsala, Suecia). Los granulocitos se separaron de los glóbulos rojos mediante incubación breve con tampón de lisis hipotónico. La población de células mononucleares se fraccionó en linfocitos y monocitos mediante adherencia a placas de plástico. Para el análisis RT-PCR, 2 µg del ARN total de cada preparación celular se usaron para generar el primer ADNc de una hebra usando un kit comercialmente disponible (Gibco BRL; Gaithersburg, MD). Los fragmentos de ADNc de Nod2 correspondientes a la región de codificación de Nod2 se amplificaron mediante PCR usando dos conjuntos de cebadores específicos; P1: 5'-ATGTGCTCGCAGGAGGCTTTTCAGGCA-3' (SEC ID N° 45); P2: 5'-CGCCTCACCCACCACCAGCACAGTGT-3' (SEC ID N° 46); P3 5'-ATGTGCTCGCAGGAGGCTTTTCAGGCA-3' (SEC ID N° 47) y P4: 5'-CGCCTCACCCACCACCAGCACAGTGT-3' (SEC ID N° 48). Como control, un fragmento de ADNc de la gliceraldehído-3-fosfato deshidrogenasa humana se amplificó usando los cebadores 5'-GAGTCAACGGATTTGGTCGTAT-3' (SEC ID N° 49) y 5'-AGTCTTCTGGGTGGCAGTGAT-3' (SEC ID N° 50).

50 **Construcción de plásmidos de expresión.** El ADNc de Nod2 se clonó en pcDNA3-HA y pcDNA3-Fpk3-Myc (Inohara y col., [2000], citado anteriormente). Los mutantes de Nod2 dirigidos a sitio y por delección (129-1040, A125-214, 1-125, 1-301, 1-744, 265-1040, 126-301, 265744, 744-1040, K305R, 1-744K305R) se construyeron mediante un procedimiento de PCR y se clonaron en pcDNA3-HA y pcDNA3-Fpk3-Myc (Inohara y col., [2000], citado anteriormente) La autenticidad de todas las construcciones se confirmó mediante secuenciación. Anteriormente se han descrito pcDNA3-Flag-RICK, pcDNA3-Flag-RICK(1-374), pcDNA3-Flag-RICK(374-540), pcDNA3-Myc-RICK (406-540), pcDNA3-Myc-RIP(558-671), pRK7-Flag-IKKa, pRK7-FlagIKKa-K44A, RSVMad-3MSS(IK-Ba-S32A/S36A), pRK7-Flag-IKKj3, pRK7-Flag-IKKp-K44A, y pcDNA3-Flag-IKKy (134-419) (Inohara y col., citado anteriormente, 10). Los plásmidos de expresión pcDNA3-Nod1-Flag, pcDNA3-Nod1 (1-648)-Flag, pcDNA3-Flag-IKKi, pcDNA3CIPER-Flag, pCMV-ILIR. pCMV-TLR4-Flag, pcDNA3-Flag-RIP, pcDNA3-MyD88 DN (aminoácidos 1-109), pcDNA3-CD14, pCMV-MD2-FLAG y pcDNA3-p-gal también se han descrito previamente (Inohara y col., [1999], citado anteriormente; Inohara y col., [1999], citado anteriormente; Inohara y col., [2000], citado anteriormente; Shimada y col., Int. Immunol., 11:1357-1362 [1999]; Huang y col., PNAS, 94:12829-12832 [1997]; Medzhitov y col., Mol. Cell,

2:253-258 [1998]; Hsu y col., *Immunity*, 4:387-396 [1996]). Para construir el plásmido de expresión que produce interleucina 1 madura marcada con HA en el extremo C (IL1 β), pcDNA3-mIL1p-HA, la región madura de la IL1 β de ratón se amplificó mediante PCR y se insertó en pcDNA3-HA-pro que contiene la secuencia señal de protripsina y la marca de HA.

5 **Transfección, expresión, inmunoprecipitación e inmunodetección de proteínas marcadas.** Las células HEK293T se co-transfectaron con pcDNA3-Nod2-HA y varios plásmidos de expresión como se ha descrito (Inohara y col., [1999], citado anteriormente). Para analizar la interacción entre la RICK silvestre y las proteínas mutantes de Nod2, las células HEK293T se cotransfectaron con pcDNA3-Flag-RICK y con plásmidos de expresión de Nod2 de tipo silvestre o mutante. Las proteínas co-inmunoprecipitadas con el anticuerpo anti-HA se detectaron con el anticuerpo anti-Flag. Para analizar la interacción entre Nod2 silvestre y los mutantes de RICK, las células HEK293T se cotransfectaron con pcDNA3-HA-Nod2 and pcDNA3-Flag-RICK, pcDNA3-Flag-RICK(1-374) o pcDNA3-Flag-RICK (374-540) (Inohara y col., [1999] citado anteriormente). Las proteínas co-inmunoprecipitadas con el anticuerpo anti-HA se detectaron con el anticuerpo anti-Flag. Las proteínas en el lisado total se detectaron con el anticuerpo monoclonal anti-Flag y anti-HA, respectivamente.

15 **Ensayos de activación de NF- κ B.** Los ensayos de activación de NF- κ B se realizaron como ha descrito (Inohara y col., [1999] citado anteriormente, Inohara y col., [2000], citado anteriormente). Brevemente, fibroblastos de Rat1 y su línea celular 5R derivada (Yamaoka y col., *Cell* 93: 1231-1240 [1998]) así como las células HEK293T se cotransfectaron con 12 mg de la construcción indicadora pBVix-Luc, más cantidades indicadas de cada plásmido de expresión y 120 ng de pEF-BOS- β -gal por triplicado. 24 horas después de la transfección se prepararon los extractos celulares y se midió su actividad relativa de luciferasa como ha descrito (Inohara y col., [1999] citado anteriormente, Inohara y col., [2000], citado anteriormente). Los resultados se normalizaron para la eficiencia de la transfección, con valores obtenidos con pEF-BOS- β -gal.

25 **Ensayo in vitro desunió a LPS** 1 x 10⁸ células HEK293T se co-transfectaron con los plásmidos de expresión indicados en las leyendas de las figuras, como se ha descrito (Inohara y col., [2000], citado anteriormente). 24 horas después de la transfección se prepararon fracciones S100 de células transfectadas como se describe usando tampón A (Poltorak y col., [1998], citado anteriormente). Para la Fig. 10A, el lisado S100 que contenía 5 mg de proteína se incubó con 300 ng de [³H] LPS (1 X 10⁵ Bq, 347 Bq/ng, List Biological Laboratories, Campbell, CA) de *Escherichia coli* K12 KCD25, 6 μ g de anticuerpo anti-FLAG M2 (Sigma Chemical), 10 μ l de proteína A-Sepharose y 10 μ l de Proteína G-Sepharose a 4°C durante 2 horas. Las proteínas unidas a la matriz se lavaron 5 veces con 1 ml de tampón A. La radiactividad unida se midió usando un contador de centelleo líquido Counter Beckman LS5000LD. Para la Figura 3B las proteínas se inmunopurificaron primero a partir de 20 mg del lisado S100 como se ha descrito anteriormente y se incubaron con [³H] LPS en presencia de 10 mg de seroalbúmina bovina, fracción V (Sigma Chemical) a 4 °C durante 2 horas. Tras 5 lavados con 1 ml de tampón A se midió la radiactividad unida. Para monitorizar la expresión de la proteína en 50 μ g del lisado S100 se detectó mediante inmunotransferencia con el Ac anti-FLAG.

Ejemplo 1

Este ejemplo describe la identificación de Nod2. Para identificar nuevas moléculas de tipo Nod1/Apaf-1 se realizaron búsquedas en bases de datos genómicas públicas de genes que codifican proteínas con homología con Nod1 (Inohara y col., citado anteriormente). Se identificó una secuencia genómica en el cromosoma 16 humano (número de acceso en GeneBank AC007728) que codifica un péptido con una homología significativa con el NBD de Nod1. El análisis con GeneFinder de la región genómica predijo un gen que codificaba una nueva proteína con una homología significativa con Nod1. Para determinar los extremos de la región de codificación se realizó un 5' RACE usando un oligonucleótido complementario a las secuencias que codifican el extremo N de la proteína predicha y secuenció varios ADNc de EST que contienen secuencias parciales del gen (números de acceso en GeneBank AA775466, AA910520, A1090427). Para amplificar el ADNc que contiene la totalidad del marco de lectura abierto, se realizó RT-PCR con tres conjuntos de cebadores correspondientes a las secuencias solapantes de la región de codificación del gen. El marco de lectura abierto predicho codifica una proteína de 1040 aminoácidos. Una búsqueda BLAST de bases de datos de proteínas indicó que la proteína codificada por el nuevo marco de lectura abierto era casi homóloga a Nod1 (34 % de identidad de aminoácidos). Esta proteína se denominó Nod2 dado su elevado nivel de homología con Nod1 y, por tanto, representa un nuevo miembro de la superfamilia de Apaf-1/Nod1 (Fig. 1). El análisis de la secuencia de nucleótidos reveló dos posibles sitios de iniciación de la traducción en el marco separados por 81 nucleótidos. Análisis adicionales revelaron que ambos sitios de iniciación de la traducción se pueden usar en las células, aunque preferentemente se usa el marco de lectura abierto más largo (véase más adelante). Por simplicidad, el marco de lectura abierto más largo se denomina Nod2a y el producto codificado por el marco de lectura abierto más corto se denomina Nod2b. Una búsqueda BLAST y análisis de dominios revelaron que Nod2 está compuesta por dos CARD en NH2 terminal (residuos 28-220) condensados con un dominio NBD localizado centralmente (residuos 273-577) que contienen los motivos de unión al nucleótido consenso, seguidos de diez LRR en tándem (residuos 744-1020) (Figs 1 y 2). Cada una de las 10 LRR de Nod2 contenía secuencias predichas de α hélice y lámina β consistentes con la estructura en forma de herradura prototípica de LRR (Kobe y Deisenhofer, *Curr. Opin. Struct. Biol.* 5: 409-416 [1995]) (Fig. 2C). Nod2 es la primera molécula que se sabe que codifica dos CARD.

Ejemplo 2

Este ejemplo describe la localización en el cromosoma y la organización genómica del gen humano de Nod2. Se identificaron dos clones BAC humanos, RP11-327F22 y RP11-40IP9, que contenían la secuencia genómica de Nod2 humana (números de acceso en GenBank AC007728 and AC007608, respectivamente). Estos clones BAC se localizaron en el cromosoma 16 en q12. La comparación del ADNc de Nod2 y las secuencias genómicas reveló que el gen de Nod2 contiene doce exones de codificación.

Ejemplo 3

Este ejemplo demuestra que la expresión de Nod2 es más abundante en monocitos. El análisis de transferencia Northern mostró que Nod2 se expresaba como dos transcritos de 7,0 y 5,5 kb en leucocitos de sangre periférica con poco o ninguna expresión detectable en varios tejidos humanos (Fig. 3A). Este patrón de expresión altamente restringido contrasta con el de Nod1 and Apaf-1, que se expresan en casi todos los tejidos adultos, aunque a diferentes niveles (Inohara y col., citado anteriormente). Para determinar las células que expresan Nod2, los leucocitos de sangre periférica se fraccionaron en poblaciones de granulocitos, linfocitos y monocitos, y se analizaron mediante análisis RT-PCR con dos conjuntos diferentes de cebadores oligonucleotídicos complementarios a las secuencias de codificación de Nod2. El análisis mostró que Nod2 se expresaba principalmente en monocitos (Fig. 3B). Dado que la secuencia de Nod2 contenía dos posibles sitios de iniciación de la traducción en el marco separados por 81 nucleótidos (Fig. 3C), su uso se determinó mediante transfección de una construcción de Nod2 que contiene ambos sitios de iniciación de la traducción en las células HEK293T. Dado que la diferencia en tamaño entre ambos productos de Nod2 predichos es de solo 27 aminoácidos, los inventores han expresado una Nod2 truncada en el COOH terminal que carece de los residuos 302-1040 para facilitar la identificación de los sitios de iniciación de la traducción. Como control, los plásmidos de Nod2 se sometieron a ingeniería para expresar cada sitio de iniciación de la traducción por separado dentro de un motivo canónico de iniciación de la traducción de Kozak. El análisis reveló que se usaron ambos sitios de iniciación de la traducción en el marco de lectura abierto de Nod2, aunque el codón de iniciación de la traducción más en NH2 era más eficiente evaluado mediante inmunotransferencia de extractos celulares con un anticuerpo que reconoce una cola de HA en el COOH terminal (Fig. 3D).

Ejemplo 4

Este ejemplo describe la activación de NF- κ B por Nod2. Dada la homología entre Nod1 y Nod2, las pruebas se realizaron para determinar si la expresión de Nod2 activa NF- κ B mediante la transfección de los plásmidos de Nod2 en células HEK293. La transfección del ADNc de Nod2 silvestre indujo una potente activación de NF- κ B, medida con una construcción indicadora de luciferasa (véase más adelante). Además, los inventores probaron el ADNc de Nod2b y obtuvieron resultados similares a los observados con Nod2. Se generó un panel de mutantes de Nod2 para determinar las regiones de Nod2 requeridas para la activación de NF- κ B (Fig. 4A). El análisis de inmunotransferencia reveló que estas construcciones mutantes se expresaban cuando se transfectaban transitoriamente en células HEK293T (Fig. 4B). La expresión de tan poco como 3 ng de la Nod2 silvestre indujo una activación por 18 de NF- κ B (Fig. 4C). La expresión de una forma mutante de Nod2 que carece de LRR tuvo como resultado una activación potenciada de NF- κ B, mientras que los mutantes que expresan las LRR o el NBD solo eran inactivos (Fig. 4C). La actividad potenciada del mutante de Nod2 que carece de LRR no se pudo explicar mediante la expresión incrementada del mutante (Fig. 4A). Consistente con estos resultados, se demostró anteriormente que la delección de las LRR de Nod1 y las repeticiones WD-40 de Apaf-1 tiene como resultado una mayor activación de NF- κ B y un incremento de la capacidad para activar la procaspasa-9, respectivamente (Inohara y col., citado anteriormente, Srinivasula y col., citado anteriormente, Hu y col., citado anteriormente). La delección de los CARD de Nod2, solos o en combinación, tuvo como resultado una pérdida total de la actividad de NF- κ B (Fig. 4C). No obstante, la expresión de ambos CARD solos, pero no de cada CARD por separado, fue suficiente para activar NF- κ B (Fig. 4C). Por tanto, ambos CARD de Nod2 son necesarios y suficientes para activar NF- κ B, lo que sugiere que las CARD actúa como dominio efector en la señalización de Nod2. El residuo de lisina conservado en el bucle P de Nod1 y Apaf-1 es importante para las actividades de estas proteínas (Inohara y co., [1999] citado anteriormente, Inohara y col., [2000], citado anteriormente, Hu y col., EMBO J. 18: 3586-3595 [1999]). De un modo similar, la sustitución de la correspondiente lisina por arginina en Nod2 tuvo como resultado una disminución de la actividad de NF- κ B que se rescató al menos en parte mediante la delección de las LRR (Fig. 4C)

También se investigó la capacidad de la Nod2 para inducir apoptosis. La sobreexpresión de Nod2 no indujo apoptosis por sí misma pero potenció la apoptosis inducida por la expresión de caspasa-9. Estos resultados son similares a los indicados para Nod1 y Apaf-1 (Bertin y col., citado anteriormente, Inohara y col., [1999] citado anteriormente).

Ejemplo 5

Este ejemplo demuestra que la activación de NF- κ B inducida por Nod2 requiere IKK y es inhibida por formas negativas dominantes de las IKK y RICK. Una vía principal de la activación de NF- κ B está mediada por las IkB quinasa (IKKs) resultante en la fosforilación de IkB y la liberación de NF- κ B citoplasmático (Karin, J. Biol. Chem. 274: 27339-27342 [1999]). Para determinar si Nod2 activa una vía dependiente de IKK, la Nod2 se co-expresó con formas mutantes de IKK α , Ikk β e IkB que se ha demostrado que actúan como inhibidores dominantes de sus

correspondientes homólogos endógenos y/p el complejo IKK (Karin, *citado anteriormente*). Además, se usó un mutante truncado de IKK γ /NEMO (residuos 134-419) que es defectuoso en la unión de IKK α y IKK β y actúa como inhibidor de la activación de NF- κ B inducida por RIP y RICK (Inohara y *col.*, [2000], *citado anteriormente*). La actividad de NF- κ B inducida por Nod2, así como la inducida por la estimulación del TNF α se vieron considerablemente inhibidas por las mutantes IKK α , IKK γ , IKK β e I κ B α (Fig. 5A). Dado que se ha demostrado que RICK sirve como diana cadena abajo de Nod1 (Bertin y *col.*, *citado anteriormente*, Inohara y *col.*, [1999] *citado anteriormente*, Inohara y *col.*, [2000], *citado anteriormente*), se usó una forma truncada de RICK que contiene sus CARD (residuos 406-540) que actúa como inhibidor dominante de la actividad de Nod1 (Bertin y *col.*, *citado anteriormente*) para analizar su la activación de NF- κ B inducida por Nod2 es inhibida de un modo similar por este mutante de RICK. La activación de NF- κ B inducida por Nod2 fue inhibida por el mutante de RICK pero no por una forma mutante de RIP que expresa su dominio efector de muerte (Fig. 5A). La inhibición por el CARD de RICK fue específica en cuanto a que no interfería con la capacidad del TNF α para inducir NF- κ B, una actividad inhibida por el mutante RIP (Fig. 5A). Para verificar que Nod2 actúa cadena arriba del complejo IKK para activar NF- κ B, los inventores analizaron la capacidad de Nod2 para activar NF- κ B en fibroblastos Rat1 parentales y células 5R, una línea celular derivada de Rat1 que es defectiva en IKK γ , una subunidad esencial de los IKK (Yamaoka y *col.*, *citado anteriormente*). La Nod2, así como la Nod1 y TNF α , indujo actividad NF- κ B en células Rat1 parentales pero no en células 5R deficientes en IKK γ (Fig. 5B). Como control, la expresión de IKK β , que funciona cadena debajo de IKK γ , indujo actividad NF- κ B en líneas celulares Rat1 y 5R (Fig. 5B). Estos resultados indican que Nod2 actúa a través de IKK γ /IKK α /IKK β para activar NF- κ B.

20 Ejemplo 6

Este ejemplo demuestra que Nod2 se asocia con RICK a través de una interacción CARD-CARD homófila. El motivo CARD funciona como dominio efector que media en la interacción homófila específica con las moléculas que contienen CARD cadena abajo (Hofmann y *col.*, Trends Biochem. Sci. 22: 155-156 [1997]). Dado que la activación de NF- κ B inducida por Nod2 se inhibía por acción de un mutante truncado de RICK, se analizó la capacidad de RICK para actuar como mediador directo cadena debajo de la señalización de Nod2. Para analizar una asociación física entre Nod2 y RICK, las células HEK293T se co-transfectaron con plásmidos que expresan formas mutante o silvestre marcadas con Ha de Nod2 y RICK marcadas con flag y los extractos celulares se inmunoprecipitaron con un anticuerpo anti-HA. La inmunotransferencia con el anticuerpo anti-Flag reveló que RICK se asociaba con Nod2 (Fig. 6A). La asociación estaba mediada por ambos CARD de Nod2, ya que solo las proteínas de Nod2 que contenían ambos CARD eran capaces de interactuar con RICK (Fig. 6A, B). La asociación de Nod2 con RICK fue específica en cuanto a que Nod2 no se asociaba con varias proteínas que contenían CARD, incluidas Apaf-1, caspasa-1, caspasa-4, c-IAP-1, c-IAP2, procaspasa-9, Bc1-10, RAIDD y Ced-4 ni con varias moléculas que activan NF- κ B incluidas TRAF-1, TRAF-2, TRAF-5, TRAF-6, RIP, NIK, TRADD, IKK α , IKK β o IKK γ . Para determinar la región de RICK que se asocia con Nod2, las formas mutantes de RICK que expresan el CRD (residuos 374-540) o que carecen del CARD (residuos 1-374) se coexpresaron con Nod2 y los extractos celulares se inmunoprecipitaron con anticuerpos anti-Flag. El análisis mostró que solo el CARD de RICK se co-inmunoprecipitó con Nod2 (Fig. 6C). Por tanto, Nod2 y RICK se asocian a través de una interacción CARD-CARD homófila.

35 Ejemplo 7

Este ejemplo demuestra que la oligomerización forzada de Nod2 induce la activación de NF- κ B. Los estudios previos mostraron que el NBD de Nod1 y Apaf-1 participa la oligomerización de estas moléculas, una actividad que es crucial para activación de NF- κ B y de la caspasa-9, respectivamente (Srinivasula y *col.*, *citado anteriormente*; Hu y *col.*, [1998] *citado anteriormente*, Inohara y *col.*, [2000], *citado anteriormente*). En el caso de Nod2, su oligomerización parece estimular la proximidad de la activación de RICK y NF- κ B. Para analizar un papel similar de Nod2 se construyeron plásmidos para expresar proteínas quiméricas en las que Nod2 silvestre o mutantes de Nod2 se condensaron con tres dominios de dimerización repetidos en tándem de Fpk (Fpk3), que se pueden oligomerizar por el ligando permeable a la célula AP1510 (MacCorkle y *col.*, Proc. Nat. Acad. Sci. U. S. A. 95: 3655 [1998]). El análisis de inmunotransferencia mostró que las construcciones Fpk3-Nod2 quiméricas se expresaban cuando se transfectaban en células HEK293T (Fig. 7A). Dado que la Nod2 silvestre sola induce activación de NF- κ B, los inventores expresaron cantidades subóptimas de las construcciones Fpk3-Nod2 quiméricas en células HEK293T. En estas condiciones experimentales, la expresión de Nod2-Fpk3 indujo activación de NF- κ B de un modo dependiente de ligando (Fig. 7B). Consistente con los resultados mostrados en la Fig. 4C, la oligomerización forzada de ambos CARD, pero no de cada CARD por separado, indujo activación de NF- κ B (Fig. 7B). De un modo similar la activación de NF- κ B inducida por el mutante del bucle P de Nod2 que carece de las LRR (K305RALRR), que tiene menor capacidad para inducir activación de NF- κ B, se potenció mediante la oligomerización forzada (Fig. 7C). Una construcción Nod2-Fpk3 que carece de las LRR indujo activación de NF- κ B en ausencia y presencia de AP1510 (Fig. 7B). El último resultado podía explicarse por las observaciones de los inventores de que la Nod 2 carece de las LRR tiene una mayor actividad para autoasociarse e inducir NF κ B (Fig. 4C).

50 Ejemplo 8

Este ejemplo describe el papel de Nod1 en la respuesta celular a los componentes microbianos. Las células HEK293T de riñón embrionario humano se cotransfectaron transitoriamente con un plásmido de expresión de Nod1 o un plásmido control y la construcción indicadora de NF- κ B en presencia de productos bacterianos o fúngicos. No

se observó una inducción significativa de la actividad del gen indicador cuando las células transfectadas con el plásmido control se expusieron a LPS, PGN, LTA, lipopéptido bacteriano sintético (SBLP) o manano (Fig. 8A). Estos resultados están de acuerdo con las observaciones previas en las células HEK293 (Yang y col., *Nature*, 395:284-288 [1998]; Aliprantis y col., *citado anteriormente*; Chow y col., *citado anteriormente*; Schwandner y col., *J. Biol. Chem.*, 274:17406-17409 [2000]). Dado que la sobreexpresión de Nod1 induce activación de NF- κ B (Zou y col., *Cell*, 90:405-413 [1997]; Bertin y col., *J. Biol. Chem.*, 274:12955-12858 [1999]), la células HEK293T se transfectaron con 0,3 ng de Nod1 y se midió la activación de NF- κ B tras la incubación con varios componentes patógenos. El LPS, pero no los otros productos microbianos analizados, indujo una activación significativa de NF- κ B (por aproximadamente 12) en células que expresan cantidades mínimas de Nod1 (Fig. 8A). Para demostrar que la activación de NF- κ B por el LPS es específica de las células que expresan Nod1, las células HEK293T se transfectaron con plásmidos de expresión que producen el receptor de interleucina 1 (IL1R) y su ligando interleucina 1 β (IL1 β) o RIP, un mediador de la vía de señalización del TNF α (Huang y col., [1997] *citado anteriormente*; Hsu y col., [1996], *citado anteriormente*). Como estaba previsto, la estimulación de IL1R y la expresión de RIP indujeron la activación de NF- κ B en ausencia de LPS (Fig. 8B). Significativamente, el LPS no potenció la activación de NF- κ B inducida por la estimulación de IL1R o RIP (Fig. 8B).

Las proteínas resistentes a enfermedades de plantas tienen LRR en C-terminal que son cruciales para las respuestas específicas de patógenos (Dixon y col., [2000], *citado anteriormente*). Las alteraciones en sus LRR tiene como resultado una falta de respuesta a patógenos concretos (Dixon y col., [2000], *citado anteriormente*), lo que sugiere que las LRR de Nod2 podrían también ser necesarias para la respuesta a LPS. Para comprobar esta hipótesis, las células HEK293T se transfectaron con plásmidos que expresan Nod1 silvestre o mutante truncada que carece de las LRR (Nod1 Δ LRR) y tratados con LPS. La expresión de Nod1 Δ LRR indujo mayor activación de NF- κ B que la Nod1 silvestre en ausencia de LPS, como se ha indicado previamente (Inohara y col., [1999], *citado anteriormente*). Significativamente, el LPS no potenció la activación de NF- κ B inducida por Nod1 Δ LRR (Fig. 9). Por tanto las LRR son esenciales para que Nod1 responda al LPS.

En varios estudios se han proporcionado pruebas concluyentes de que TLR4 es un receptor de superficie celular para LPS (Aderam y Ulevitch, *citado anteriormente*, Poltorak y col., *Science*, 282:2085 [1998]; Chow y col., *citado anteriormente*; Takeuchi y col., *Immunity*, 4:443 [1999]). Por tanto, es posible que la expresión de Nod1 confiera capacidad de respuesta a LPS a través de TLR4. Para comprobar esta posibilidad, las células HEK293T se co-transfectaron con un plásmido de expresión de TLR4 y la actividad de NF- κ B se midió en presencia y ausencia de LPS. La expresión de TLR4 solo no indujo activación de NF- κ B en presencia de LPS, lo que es consistente con recientes informes de que moléculas de superficie celular adicionales, tales como MD2 y CD14, se requieren para las respuestas de LPS mediadas por TLR4 (Chow y col., *citado anteriormente*; Takeuchi y col., *citado anteriormente*). De acuerdo con lo último, la co-transfección de los plásmidos de expresión TLR4, CD14 y MD2 indujo una activación multiplicada por 8 de NF- κ B (véase la leyenda de la figura de Fig. 8C). Para verificar adicionalmente que Nod1 confiere capacidad de respuesta a LPS con independencia de TLR, un mutante negativo dominante de MyD88, una molécula de señalización frecuente de IL-1 y receptores relacionados con Toll incluido el TLR4, se co-expresión con Nod1 o TLR4, CD14 y MD2 como control, y las células transfectadas se estimularon con LPS. La coexpresión del mutante MyD88 suprimió la activación de NF- κ B inducida por la estimulación de TLR4 e IL1R, pero no afectó a la activación de NF- κ B mediada por LPS inducida por Nod1 (Fig. 8C). Adicionalmente, la expresión de un mutante negativo dominante de TRAF6, una molécula de señalización de las vías de señalización de TLR no bloqueó la activación de NF- κ B inducida por Nod2 pero inhibió la activación de NF- κ B mediada por TLR4 (Inohara y col., [1999], *citado anteriormente*). Estos resultados indican que la activación de NF- κ B en células que expresan Nod1 inducida por LPS no está mediada por la vía de señalización de TLR4. Consistente con esta idea son las observaciones recientes que muestran que la vía de señalización de Nod1 que conduce a la activación de NF- κ B es distinta a la de los TLR. Nod1 activa NF- κ B mediante su asociación con RICK, una proteína quinasa que interacciona directamente con IKK γ /NEMO, la subunidad reguladora del complejo I κ B quinasa (Inohara y col., [2000], *citado anteriormente*).

El LPS de diferentes bacterias gramnegativas tiene diversas estructuras (Rietschel y col., *Curr Top. Microbiol. Immunol.*, 216:39-81 [1997]). Para determinar si Nod1 confiere capacidad de respuesta a LPS de varias fuentes bacterianas. Las células que expresan Nod1 se estimularon con LPS de seis bacterias patogénicas o TNF α como control positivo. Todas las preparaciones de LPS indujeron activación de NF- κ B en células que expresan Nod1, per diferentes fuentes de LPS diferían en su capacidad para potenciar la activación de NF- κ B mediada por Nod1 (Fig. 9). Como se ha encontrado con el LPS de *Escherichia coli* 055:B5 (Fig. 8B), ninguna de las preparaciones de LPS indujo una activación significativa de NF- κ B en células que expresan un mutante de Nod1 que carece de las LRR (Fig. 9).

Las plantas tiene numerosos genes de resistencia R a enfermedad y las células de mamífero, además de las de insectos, tienen múltiples miembros de la familia de TLR que responden a diferentes patógenos (Dixon y col., [2000], *citado anteriormente*). Notablemente, Nod2, otra proteína de tipo Nod1, que es homóloga a Nod1 (identidad de aminoácidos del 34 5) está compuesta por CARD, NBD y LRR terminales. La presencia de múltiples familias de Nod sugiere que Nod1 y Nod2 pueden tener diferentes especificidades por componentes patógenos. Para comprobar esto, las células HEK293T se co-transfectaron con plásmidos que expresan Nod2 silvestre o mutante que carece de las LRR. Como se observó con Nod1, todas las preparaciones de LPS, incluidas las de bacterias invasivas tales como *Salmonella* y *Shigella*, estimularon la activación de NF- κ B en células que expresan Nod2 silvestre pero no

Nod2 mutante (Fig. 9). Notablemente, el LPS de *Serratia macreeseans* y *Salmonella typhimurium* fue más eficaz en la inducción de la activación de NF- κ B en células que expresan Nod2 que Nod1 (Fig. 9). Además, la preparación de PGN de *Staphylococcus aureus* estimuló la activación de NF- κ B en células que expresan Nod2 pero no en Nod1 (Fig. 9). La base molecular de la respuesta diferencial de Nod1 y Nod2 al LPS y al PGN no está clara. Otros análisis bioquímicos y determinación de la estructura del resto LPS reconocido por Nod1 y Nod2 se requieren para entender la respuesta diferencial de las proteínas Nod1 a componentes bacterianos.

Apaf-1, una molécula de tipo Nod1 que desempeña un papel crucial en la apoptosis, media en la capacidad de respuesta al citocromo c de la mitocondria (Dixon y col., citado anteriormente; Li y col., Cell, 91:479 [1997]). Apaf-1 se une directamente al citocromo c (Li y col., citado anteriormente). Para determinar si Nod1 se une al LPS, se prepararon lisados de células S100 de células HEK293T que expresan Nod1 y la capacidad de Nod1 para unirse al LPS radiomarcado se analizó mediante un ensayo de inmunoprecipitación modificado. El LPS se co-inmunoprecipitó con Nod1 marcado con Flag, pero no con otras proteínas control marcadas con Flag (Fig. 10A). Por tanto, Nod1 se asocia con una actividad de unión a LPS presente en la fracción citosólica de las células HEK293T. No obstante, es posible que Nod1 no se una directamente al LPS y que la asociación requiera otros factores citosólicos. Por ejemplo, se requieren dATP o ATP para la respuesta de Apaf-1 al citocromo c (Li y col., citado anteriormente). Para comenzar a analizar esto, primero se inmunoprecipitó Nod1 o IKK β , como proteína control, con un anticuerpo anti-Flag y se analizó la capacidad de las proteínas inmunoprecipitadas para unirse a LPS en tampón sin nucleótidos. La Nod1 inmunopurificada exhibió actividad de unión a LPS, pero IKK β control no (Fig. 10). Estos resultados sugieren que Nod1 se une directamente al LPS. No obstante, no se puede excluir la posibilidad de que Nod1 interaccione con LPS a través de uno o más factores citosólicos intrínsecos que está estrechamente unidos a Nod1 y co-inmunoprecipita con Nod1 en ausencia de LPS. En plantas, el producto génico de RPS2 de resistencia a la enfermedad por *Arabidopsis thaliana* que está estructuralmente relacionado con Nod1 y Nod2 puede formar un complejo proteico *in vivo* con el producto del gen *avrRpt2* de la bacteria fitopatógena *Pseudomonas syringae*, pero el complejo proteico también contenía al menos una proteína vegetal adicional de aproximadamente 75 kDa (Leister y Katagiri, Plant J., 22:345 [2000]).

Ejemplo 9

Caracterización de los alelos de la enfermedad de Crohn

Este ejemplo describe la identificación de alelos mutantes de Nod2 en pacientes con enfermedad intestinal inflamatoria (EII). La Nod2 se ha mapeado hasta el cromosoma 16q12 y está estrechamente relacionada con los marcadores D16S3396, D16S416, and D16S419 (Fig. 17a), un sitio que solapa de forma precisa con la región del pico IBD1 del enlace (Hugot et al., Nature 379:821 [1996]). Por tanto, se investigó la posibilidad de que Nod2 podría funcionar como un gen de susceptibilidad a la EC. La organización genómica de doce exones del gen Nod2 se determinó alineando la secuencia de ADN (Nº de acceso en Genbank: AF178930) con un clon de BAC genómico, RP111-327F22 (Nº de acceso en Genbank: AC007728) (Fig. 17a).

A. Determinación de las familias de EII

Los pacientes se seleccionaron en EE.UU. Las familias de EII se determinaron según estudios de unión y de asociación (niños afectados con ambos padres) en la Universidad de Chicago, el Johns Hopkins Hospital y la Universidad de Pittsburgh. En todos los casos se obtuvo el consentimiento informado para un estudio genético molecular y los comités de revisión institucional individuales lo aprobaron.

Usando cebadores complementarios a las secuencias intrónicas se amplificaron todos los exones de codificación e intrones flanqueantes en las muestras de ADN de pacientes con EC y controles. En el análisis inicial, un individuo afectado de doce familias con EC puras con incremento en las puntuaciones de unión en D16S3396 se seleccionó para la secuenciación, junto con cuatro casos-controles. En tres individuos afectados no relacionados se observó una inserción de citosina en el exón 11 en el nt 3020 (3020InsC); dos sujetos eran heterocigotos y un sujeto fue homocigoto para la mutación (Fig. 17b). La confirmación de los resultados de secuenciación se realizó mediante subclonación del amplicón de la PCR de individuos heterocigotos, transformando y secuenciando clones individuales. Esta inserción tuvo como resultado un cambio de marco en el segundo nucleótido del codón 1007 y una sustitución de Leu1007Pro en la décima LRR, seguida de un codón de terminación prematuro (Fig. 17c). La proteína Nod2 truncada predicha contenía 1007 aminoácidos en lugar de los 1040 aminoácidos de la proteína Nod2 de tipo silvestre, eliminado 33 aminoácidos que incluían parte de la LRR más en C-terminal (Fig. 17d).

B. PCR específica de alelo

Se usó un ensayo de PCR específica de alelo para tipificar la 3020InsC en pacientes determinadas en tres centros de EE.UU. y controles y casos. La Figura 18 muestra una familia que ilustra la detección de alelos 3020InsC homocigotos y heterocigotos mediante el ensayo de PCR.

Los cebadores que enmarcan una región de 533 pares de bases que rodean al alelo 3020InsC se usaron para amplificar mediante PCR el ADN genómico aislado de controles y pacientes (sentido: 5'-CTGAGCCTTTGTTGATGAGC-3'; SEC ID Nº 51 y antisentido: 5'-TCTTCAACCCACATCCCCATT-3'; SEC ID Nº 52). Además, cada reacción de PCR contenía dos cebadores adicionales diseñados para detectar el alelo de tipo

silvestre (sentido: 5'-CAGAAGCCCTCCTGCAGGCCCT-3'; SEC ID N° 92) y otro cebador diseñado para detectar el alelo 3020InsC (antisentido: 5'-CGCGTGTTCATTCTTTTCATGGGGC-3'; SEC ID N° 93). El 3020InsC se confirmó mediante secuenciación de ADN. Para la detección del alelo C802T se usaron cuatro cebadores: dos sentido: 5'-AGTGCACAGCTTGTGAATGG-3' (SEC ID N° 94) y 5'-CGCGGGCAGATGTGGGCATGGCTAGAC-3' (SEC ID N° 95); y dos antisentido: 5'-GCAGCTGAATGGGAAGACA-3' (SEC ID N° 96) y 5'-GCCGTGGCTGGGCTCTTCTGCGAGGA-3' (SEC ID N° 97). Se realizó PCR multiplex con los cuatro cebadores en un tubo. Los productos de la PCR se aislaron en geles de agarosa al 2% y se visualizaron con bromuro de etidio.

C. Análisis de Datos

Las frecuencias de los alelos en los pacientes de EC se calcularon a partir de una submuestra de individuos no relacionados y a partir de la muestra entera de pacientes con EC. Cuando se usaron todas las muestras, las frecuencias se estimaron usando un enfoque de probabilidades. La probabilidad de los genotipos se calculó para cada familia (en el cálculo solo se usaron los genotipos de EC y la información de unión se ignoró). La solución de la función de la puntuación se estimó numéricamente. Los valores p para el ensayo TDT se calcularon usando una prueba exacta binomial. Se usó Sib-TDT para 500.000 duplicados para calcular las probabilidades empíricas para la estadística de chi-cuadrado cuando se contaron todas las familias independientes (Teng and Risch, Genome Res. 9:234 [1999]). Este cálculo se realizó permutando los alelos parentales fijando el estado de EII de los hermanos de una familia.. El equilibrio de Hardy-Weinberg se analizó usando una prueba de cociente de probabilidades. La probabilidad de los genotipos de EC no relacionados se calculó según las restricciones de Hardy-Weinberg y las frecuencias generales del genotipo. La significación del cociente de probabilidades-log se evaluó usando una distribución chi-cuadrado con un grado de libertad. La proporción de penetrancia homocigota 3020InsC y la penetrancia heterocigota se estimó suponiendo que 3020InsC está en equilibrio de Hardy-Weinberg en la población general. La frecuencia de los genotipos en los individuos afectados se estimó en 416 pacientes de EC no relacionada.

Para el análisis de casos y controles se realizó el genotipado de un individuo de EC de 416 familias independientes y la frecuencia del alelo entre todos los grupos fue 8,2 % (Tabla 2). Las frecuencias del alelo son comparables en los dos grupos étnicos principales, judíos (8,4%) y caucásicos no judíos (8,1%). Entre los casos y controles (Tabla 2), la frecuencia del alelo en cuatro cohortes caucásicas separadas de 4,0% fue significativamente menor que en los pacientes de EC (p = 0,0018, mediante las grandes aproximaciones a una prueba binomial de dos muestras). Un valor similar para la frecuencia del alelo (8,3%) se estimó de los 797 individuos de todas las familias. la frecuencia del alelo del 3020InsC entre 182 pacientes de CU no relacionados fue del 3,0% y fue significativamente inferior a la frecuencia entre los pacientes de EC (p = 0,0010). Dado que las frecuencias del alelo en los diferentes grupos control son comparables, no cabe esperar que la significativamente mayor frecuencia de 3020InsC en pacientes afectados por EC se deba a la subestructura de la población.

Tabla 2					
Frecuencia del alelo en pacientes de enfermedad de Crohn y controles no relacionados					
Enfermedad de Crohn			Controles		
Fuente	Tamaño de la muestra	3020InsC*	Fuente	Tamaño de la muestra	3020InsC*
Univ. de Chicago	212	7,3	Chicago	65	3,8
Johns Hopkins	88	6,8	Baltimore	46	3,2
Univ. de Pittsburgh	116	10,8	San Francisco	81	3,1
			Alemania	94	5,3
Total	416	8,2	Total	287	4,0

* Porcentaje de la frecuencia del alelo.

Para análisis adicional de que las diferencias observadas entre los individuos con EC y control no se deben a las diferencias en la población se realizó un análisis mediante la prueba del desequilibrio de la transmisión (TDT). Usando solo un paciente de EC por familia independiente se observó transmisión preferencial de 3020InsC, con 39 transmisiones y 17 no transmisiones de padres heterocigotos a niños afectados (p = 0,0046) (Tabla 2). Usando todos los tríos padres—hijos disponibles se observaron 68 transmisiones y 33 no transmisiones (Tabla 3). Mediante sib-TDT (Teng and Risch, Genome Res. 9:234 [1999]), que calcula las probabilidades empíricas para la estadística

chi-cuadrado y refleja con precisión la asociación independiente de la unión en las familias, el valor p empírico fue de 0,00071. La prueba de asociación basada en la familia (Lake et al., Am J Hum Genet 67:1515 [2000]) (que analiza la asociación bajo la hipótesis nula de la unión pero sin asociación) estima el valor p a 0,0014.

5 En la secuenciación de 36 individuos de EC no relacionados se identificó un polimorfismo de aminoácido adicional, C802T, que tiene como resultado una sustitución de prolina a serina en el codón 268. La frecuencia del alelo del alelo C802T es 32,9% en pacientes de EC en comparación con el 25,8% en casos y controles ($p = 0,008$). Esta diferencia está dirigida, al menos parcialmente, por el haplotipo (802T-3020InsC) Mientras que el cociente de probabilidades del desequilibrio de transmisión [O.R. = probabilidad (transmisión)/1-probabilidad (transmisión)] en
10 pacientes de EC para el haplotipo (C802T-3020InsC) es 2,29 ($p = 0,0033$), no se ve significativamente aumentado (O.R. =1,23, $p = 0,20$) para el haplotipo (C802T-3020 silvestre).

Las frecuencias del genotipo obtenidas usando individuos de EC no relacionados fueron 11 3020InsC homocigotos, 46 3020InsC heterocigotos y 359 homocigotos de tipo silvestre. Entre los casos y controles, había 23 individuos heterocigotos, siendo el resto homocigotos de tipo silvestre. Se produjo una desviación significativa del equilibrio de Hardy-Weinberg entre los pacientes de EC ($p = 1,46 \times 10^{-5}$, en base a la prueba de cociente de verosimilitudes). La penetrancia relativa de enfermedad de 3020InsC homocigoto en comparación con 3020InsC heterocigotos es de aproximadamente 11,5, lo que sugiere que la mutación 3020InsC sola puede funcionar de un modo recesivo.
15

Tabla 3					
La TDT demuestra transmisión preferencial de 3020InsC a pacientes de EC					
Un paciente de EC por familia				Todos los pacientes de EC	
Fuente	Transmitido	No transmitido	Valor p	Transmitido	No transmitido
Univ de Chicago	21	10		32	16
Johns Hopkins	4	4		10	8
Univ de Pittsburgh	14	3		26	9
Total	39	17	0,0046	68	33

D. Plásmidos de expresión e inmunotransferencia

20 Se han descrito los plásmidos de expresión pcDNA3-Nod2, pcDNA3-TLR4 y pcDNA3-MD-2. El plásmido de expresión que produce la mutante Nod2 3020InsC se generó mediante un procedimiento de PCR y se clonó en pcDNA3 (Invitrogen). Mediante secuenciación de ADN Se confirmó la autenticidad de la construcción. La expresión de proteínas Nod2 sin marcar en células transfectadas se determinó mediante inmunotransferencia usando anticuerpo anti-Nod2 de conejo purificado por afinidad, como se ha descrito anteriormente (véase la sección de Metodología). Para producir el anticuerpo, la proteína Nod2 recombinante (residuos 28-301) se sobreexpresó en la
25 cepa BL21 de E. coli (DE3) usando el vector pET-30a (Novagen). La proteína Nod2 recombinante que contenía una cola de histidina en C-terminal se purificó usando una columna de níquel (Novagen) y se inyectó en conejos.

E. Ensayos de activación de NF- κ B.

30 Los ensayos de activación de NF- κ B se realizaron como ha descrito anteriormente (Ejemplo 4 y sección Metodología). Brevemente, las células HEK293T se contransfectaron con 12 mg de la construcción indicadora pBVix-Luc, más cantidades indicadas de cada plásmido de expresión y 120 ng de pEF-BOS- β -gal por triplicado en presencia o ausencia de LPS. El LPS de varias fuentes en este estudio se obtuvo de Sigma (St. Louis, MO) o de varios investigadores. 24 horas después de la transfección se prepararon los extractos celulares y se midió su actividad relativa de luciferasa como ha descrito anteriormente. Los resultados se normalizaron para la eficiencia de la transfección, con valores obtenidos con pEF-BOS- β -gal.

35 Se ha demostrado que Nod2 activa NF- κ B y que confiere capacidad de respuesta a los lipopolisacáridos bacterianos (véase el Ejemplo 8 anterior). Para analizar la capacidad de una Nod2 de tipo silvestre y mutante para activar NF- κ B, las células (HEK) 293T de riñón embrionario humano se cotransfectaron de forma transitoria con los plásmidos de 3020InsC Nod2 y una construcción indicadora de NF- κ B En ausencia de LPS, la expresión de Nod2 tanto de tipo silvestre como mutante indujo activación de NF- κ B (Fig. 19a). Es importante el hecho de que niveles equivalentes de
40 expresión de proteína Nod2 de tipo silvestre y mutante (evaluado mediante inmunotransferencia de los lisados totales) tuvo como resultado niveles similares de activación de NF- κ B (Fig. 19a). Como Nod2, las proteínas R vegetales citosólicas tienen LRR en C-terminal que son cruciales para el reconocimiento de componentes patógenos y la inducción de respuestas específicas de patógenos (Pamiske et al., Cell 91: 821 [1997]; Ellis et al., Plant Cell. 11:495 [1999]; Dixon et al., [2000], citado anteriormente). Dado que 3020InsC rompe una LRR de Nod2, se analizó
45 la capacidad de las proteínas Nod2 de tipo silvestre y mutante para inducir la actividad de NF- κ B en respuesta al

LPS. Dado que la sobreexpresión de Nod2 induce una potente activación de NF-κB (Fig. 19a), las células se transfectaron con cantidades bajas de plásmidos de Nod2 de tipo silvestre y mutante para inducir niveles similares de expresión de proteínas y actividad basal de NF-κB (Fig. 19a). También se analizó la capacidad de las proteínas Nod2 para potenciar la activación de NF-κB tras la incubación con LPS de varias bacterias. El LPS de varias bacterias indujo activación de NF-κB en células que expresan Nod2 de tipo silvestre, mientras que no se observó una inducción significativa de la actividad del gen indicador en células transfectadas con plásmido control (Fig. 19b). La capacidad del mutante de Nod2 para conferir capacidad de respuesta al LPS disminuyó considerablemente cuando se comparó con la Nod2 de tipo silvestre (Fig. 19b). Ni Nod2 de tipo silvestre ni la mutante potenció la actividad de NF-κB inducida por LPS de *Salmonella minnesota*, *Neisseria meningitidis* y *Hemophilus influenzae*, lo que sugiere que existe una regulación diferencial de la función de Nod2 por LPS de bacterias diferentes. En experimentos control mostrados en la Fig. 19b, todas las preparaciones de LPS analizadas indujeron activación de NF-κB en las células transfectadas con el receptor 4 de tipo Toll (TLR4), un receptor de superficie celular para LPS 16. Por tanto, la Nod2 parece tener una preferencia diferencial por LPS de determinadas bacterias y la actividad de NF-κB inducida por LPS está marcadamente disminuida en Nod2 con 3020InsC.

15 **Ejemplo 10**

Alelos adicionales de la enfermedad de Crohn

Este ejemplo describe la identificación de mutaciones adicionales en el gen Nod2, así como la asociación de mutaciones adicionales con la enfermedad de Crohn.

A. Identificación de alelos adicionales de Crohn

20 La Figura 26 muestra polimorfismos adicionales que se identificaron en el gen Nod2 (SEC ID N° 1 y 33). La mutación 3020InsC se identificó como se ha descrito en el Ejemplo 9, citado anteriormente. Se identificaron mutaciones adicionales mediante secuenciación directa. La Figure 26 describe la mutación 3020InsC/Nod2Δ33, así como 7 polimorfismos adicionales que se identificaron. La Figure 27 describe las frecuencia de los alelos para cada haplotipo identificado. Las posiciones de los polimorfismos están indicadas usando la numeración del transcrito de Nod2a y la proteína.

B. Asociación significativa de Nod2033, G908R y R702W con enfermedad de Crohn

Se usó la prueba de desequilibrio de transmisión (descrita en el Ejemplo 9 anterior) para demostrar la transmisión de tres alelos que se sospecha que está asociada con enfermedad de Crohn. Los polimorfismos investigados se identificaron como se ha descrito en lo que antecede. Las Tabla 4 siguiente demuestra los resultados de la prueba de desequilibrio de transmisión (TDT) para las tres variantes asociadas usando un paciente de EC por 416 familias de EC independientes o contando todos los tríos. Contando todas las familias nucleares independientes usando la prueba de sib-TDT demuestra valores p comparables. Nod2Δ33 (p = 0,0007), G908R (p = 0,0005) y R702W (p= 0,0005).

35 Se obtuvo soporte adicional para la asociación de la enfermedad mediante análisis de casos y controles, en el que se observaron frecuencia de alelos significativamente más altas en pacientes de EC en comparación con los casos y controles. La Tabla 5 describe los resultados del análisis de casos y controles para pacientes con enfermedad de Crohn no relacionados.

La Tabla 6 describe los riesgos relativos al genotipo para alelos de riesgo heterocigotos y homocigotos. Los riesgos relativos al genotipo (RRG) se definen como la proporción de la penetrancia marginal entre los genotipos de riesgo homocigotos y heterocigotos y los homocigotos de tipo silvestre. Usando la norma de BAYES, los RRG se pueden expresar como una función de las frecuencias del alelo en el caso de los grupos control. Para el grupo control, se supone que los alelos están en equilibrio de Hardy-Weinberg. Cabe observar que las estimaciones para G908R están algo sesgadas debido a su frecuencia del alelo muy baja entre no-ashkenazi.

45 El riesgo atribuible a la población entre no-ashkenazi para Nod2Δ33, G908R, and R702W fueron 6,18%, 4,35% y 12,76% respectivamente. El riesgo atribuible a la población se calculó como (K-Kw)/K, en la que K es la prevalencia de la enfermedad de Crohn en la población general y Kw es la prevalencia de la enfermedad de Crohn en la subpoblación que consiste en individuos homocigotos para el alelo de tipo silvestre en la variante especificada.

Este ejemplo demuestra que existen múltiples polimorfismos en el gen de Nod2 y que al menos 3 de los polimorfismos se asocian con una mayor prevalencia y riesgo de desarrollar enfermedad de Crohn.

Tabla 4					
La TDT demuestra transmisión preferencial de los polimorfismos de Nod2					
	Una EC por familia			Todos los tríos con EC	
Polimorfismo	Transmitido	No transmitido	Valor p	Transmitido	No transmitido

Tabla 4					
La TDT demuestra transmisión preferencial de los polimorfismos de Nod2					
	Una EC por familia			Todos los tríos con EC	
Polimorfismo	Transmitido	No transmitido	Valor p	Transmitido	No transmitido
Nod2Δ33	39	17	0,0046	68	33
G908R	34	12	0,0016	53	20
R702W	43	16	0,00060	78	34

Tabla 5			
Frecuencia del alelo (%) en pacientes de enfermedad de Crohn y controles no relacionados			
Polimorfismo	Pacientes de EC (n=416)	Controles (n = 287)	Valor p
Nod2Δ33	8,2	1,0	0,0018
G908R	5,9	1,7	0,00010
R702W	8,5	4,0	0,0010

TABLA 6		
RIESGOS RELATIVOS DEL GENOTIPO PARA NO-ASHKENAZI		
Polimorfismo	Heterocigoto	Homocigoto
Nod2Δ33	1,5	17,6
G908R	2,0	34,5
R702W	2,6	14,3

Listado de secuencias

- 5 <110> Nunez, Gabriel
Inohara, Naohiro
Ogur, Yasunori
Cho, Judy
Nicolae, Dan L
- 10 Bonen, Denise

<120> Ácidos nucleicos y proteínas NOD2

<130> UM-06646
- 15 <160> 99

<170> PatentIn versión 3,1
- 20 <210> 1
<211> 4485
<212> ADN
<213> Homo sapiens
- 25 <400> 1

ES 2 486 340 T3

gtagacagat ccaggetcac cagtectgtg ccactgggct tttggcgttc tgcacaaggc 60
 ctaccgcag atgccatgcc tgctcccca gectaagggg ctttgatggg ggaagagggt 120
 ggttcagcct ctcacgatga ggaggaaaga gcaagtgtcc tectcggaca ttctccgggt 180
 tgtgaaatgt gctcgcagga ggcttttcag gcacagagga gccagctggt cgagctgctg 240
 gtctcagggt ccctggaagg ctccgagagt gtctggact ggctgctgtc ctgggaggtc 300
 ctctcctggg aggactacga ggcttccac ctctgggccc agcctctctc ccacttggcc 360
 aggcgccttc tggacaccgt ctggaataag ggtacttggg cctgtcagaa gctcatcgcg 420
 gctgcccagg aagcccaggc cgacagccag tcccccaagc tgcattgctg ctgggacccc 480
 cactcgctcc acccagcccg agacctgcag agtcaccggc cagccattgt caggaggctc 540
 cacagccatg tggagaacat gctggacctg gcatgggagc ggggtttcgt cagccagtat 600
 gaatgtgatg aatcagggt gccgatctc acaccgtccc agaggccaag aaggctgctt 660
 gatcttgcca cggtgaaagc gaatggattg gctgccttcc ttctacaaca tgttcaggaa 720
 ttaccagtcc cattggcctt gcctttgga gctgccacat gcaagaagta tatggccaag 780
 ctgaggacca cggtgtctgc tcagtctcgc ttctcagta cctatgatgg agcagagacg 840
 ctctgcctgg aggacatata cacagagaat gtctggagg tctgggcaga tgtgggcatg 900
 gctggacccc cgcagaagag cccagcccac ctgggcctgg aggagctctt cagcaccctt 960
 ggccaacctca atgacgatgc ggacactgtg ctgggtgggg gtgaggcggg cagtggcaag 1020
 agcacgctcc tgcagcggct gcaactgtg tgggctgcag ggcaagactt ccaggaattt 1080
 ctctttgtct tccattcag ctgccggcag ctgcagtga tggccaaacc actctctgtg 1140
 cggactctac tctttgagca ctgctgttg cctgatgtg gtcaagaaga catcttccag 1200
 ttactccttg accacctga ccgtgtctg ttaaccttg atggcttga cgagttcaag 1260
 ttcaggttca cggatcgtga acgccaactg tccccgaccg accccacctc tgtccagacc 1320
 ctgctcttca accttctgca gggcaacctg ctgaagaatg cccgcaaggt ggtgaccagc 1380
 cgtccggccg ctgtgtcggc gttcctcagg aagtacatcc gcaccgagtt caacctcaag 1440
 ggcttctctg aacagggcat cgagctgtac ctgaggaagc gccatcatga gcccggggtg 1500
 gcggaccgcc tcatccgctt gctccaagag acctcagccc tgcacggtt gtgccacctg 1560
 cctgtcttct catggatggt gtccaaatgc caccaggaac tgttgctgca ggaggggggg 1620
 tccccaaaga ccactacaga tatgtacctg ctgattctgc agcattttct gctgcatgcc 1680
 accccccag actcagcttc ccaaggtctg ggaccagtc ttcttcgggg ccgcctcccc 1740
 accctctgc acctgggcag actggctctg tggggcctgg gcatgtgctg ctacgtgttc 1800
 tcagcccagc agctccaggc agcacaggtc agcctgatg acatttctt tggettctct 1860
 gtgcgtgcc aagggtgtgt gccaggagc acggcgcccc tggaaattct tcacatcact 1920
 ttccagtgtc tctttgccgc gttctacctg gcactcagtg ctgatgtgcc accagctttg 1980
 ctcagacacc tcttcaattg tggcaggcca ggcaactcac caatggccag gctcctgccc 2040
 acgatgtgca tccaggcttc ggagggaaag gacagcagcg tggcagcttt gctgcagaag 2100

ES 2 486 340 T3

gccgagccgc acaaccttca gatcacagca gccttcctgg cagggctggt gtccccggag 2160
 cactggggcc tgctggctga gtgccagaca tctgagaagg ccctgctccg gcgccaggcc 2220
 tgtgcccgcct ggtgtctggc ccgcagcctc cgcaagcaact tccactccat cccgccagct 2280
 gcaccgggtg aggccaaagag cgtgcctgcc atgccgggt tcactctggt catccggagc 2340
 ctgtacgaga tgcaggagga gcggctggt cggaaggctg cacgtggcct gaatgttggg 2400
 cacctcaagt tgacattttg cagtgtgggc cccactgagt gtgctgccct ggcctttgtg 2460
 ctgcagcacc tccggcgcc cgtggccctg cagctggact acaactctgt gggtgacatt 2520
 ggcgtggagc agctgctgcc ttgccttggg gtctgcaagg ctctgtattt gcgcgataac 2580
 aatatctcag accgaggcat ctgcaagctc attgaatgtg ctcttcaactg cgagcaattg 2640
 cagaagttag ctctattcaa caacaaattg actgacggct gtgcacactc catggctaag 2700
 ctcttgcctg gcaggcagaa cttcttggca ttgaggctgg ggaataacta catcactgcc 2760
 gcgggagccc aagtgtctgc cgaggggctc cgaggcaaca cctccttgcg gttcctggga 2820
 ttctggggca acagagtggg tgacgagggg gccagggccc tggtgaagc cttgggtgat 2880
 caccagagct tgaggtggct cagcctggg ggaacaaca ttggcagtg gggtgccaa 2940
 gccttggcac tgatgctggc aaagaacgct atgctagaag aactctgcct ggaggagaac 3000
 catctccagg atgaaggtgt atgttctctc gcagaaggac tgaagaaaa ttcaagttg 3060
 aaaatcctga agttgtccaa taactgcctc acctacctag gggcagaagc cctcctgcag 3120
 gcccttgaaa ggaatgacac catcctggaa gtctggctcc gaggaacac tttctctcta 3180
 gaggaggttg acaagctcgg ctgcaggac accagactct tgctttgaag tctccgggag 3240
 gatgttcgct tcagtttgtt tgtgagcagg ctgtgagttt gggccccaga ggctgggtga 3300
 catgtgttgg cagcctcttc aaaatgagcc ctgtcctgcc taaggctgaa cttgttttct 3360
 gggaacacca taggtcacct ttattctggc agaggaggga gcatcagtg cctccaggat 3420
 agacttttcc caagcctact tttgccattg acttcttccc aagattcaat cccaggatgt 3480
 acaaggacag cccctcctcc atagtatggg actggcctct gctgacctc ccaggcttcc 3540
 gtgtgggtca gtggggccca tggatgtgct tgttaactga gtgcctttt gtggagaggc 3600
 ccggcctctc acaaaagacc ccttaccact gctctgatga agaggagtac acagaacaca 3660
 taattcagga agcagcttcc cccatgtctc gactcatcca tccaggccat tccccgtctc 3720
 tggttcctcc cctcctcctg gactcctgca cacgctcctt cctctgaggc tgaattcag 3780

ES 2 486 340 T3

```

aatattagtg acctcagctt tgatatttca cttacagcac ccccaacctt ggcacccagg 3840
gtgggaaggg ctacacctta gcctgcacctt ctttccggtg ttttaagacat ttttgggaagg 3900
ggacacgtga cagccgtttg ttcccccaaga cattctaggt ttgcaagaaa aatatgacca 3960
cactccagct gggatcacat gtggactttt atttccagtg aaatcagtta ctcttcagtt 4020
aagccttttg aaacagctcg actttaaaaa gctccaaatg cagctttaaa aaattaatct 4080
gggccagaat ttcaaaccggc ctacttaggc ttctgggtga tgctgtgaa ctgaactctg 4140
acaacagact tctgaaatag acccacaaga ggcagttcca tttcatttgt gccagaatgc 4200
tttaggatgt acagttatgg attgaaagtt tacaggaaaa' aaaattaggc cgttccttca 4260
aagcaaagt cttcctggat tattcaaaat gatgtatgtt gaagcctttg taaattgtca 4320
gatgctgtgc aaatgttatt attttaaaca ttatgatgtg tgaaaactgg ttaatattta 4380
taggtcactt tgttttactg tcttaagttt atactcttat agacaacatg gccgtgaact 4440
ttatgctgta aataatcaga gggaataaaa ctggtgagtc aaaac 4485

```

<210> 2
 <211> 1040
 <212> PRT
 <213> Homo sapiens

5

<400> 2

```

Met Gly Glu Glu Gly Gly Ser Ala Ser His Asp Glu Glu Glu Arg Ala
1           5           10           15

Ser Val Leu Leu Gly His Ser Pro Gly Cys Glu Met Cys Ser Gln Glu
           20           25           30

Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser Gly
           35           40           45

Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp Glu
           50           55           60

Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln Pro
65           70           75           80

```

10

ES 2 486 340 T3

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

ES 2 486 340 T3

305		310		315		320
Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln Leu		325		330		335
Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu His		340		345		350
Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu Leu		355		360		365
Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu Phe		370		375		380
Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp Pro			390		395	400
Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu Leu		405		410		415
Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser Ala		420		425		430
Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe Ser		435		440		445
Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro Gly		450		455		460
Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu His		465		470		475
Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys His		485		490		495
Gln Glu Leu Leu Leu Gln Glu Gly Gly Ser Pro Lys Thr Thr Thr Asp		500		505		510
Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro Pro		515		520		525
Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg Leu		530		535		540

ES 2 486 340 T3

Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly Met
 545 550 555 560

Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val Ser
 565 570 575

Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val Val
 580 585 590

Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln Cys
 595 600 605

Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro Ala
 610 615 620

Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro Met
 625 630 635 640

Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys Asp
 645 650 655

Ser Ser Val Ala Ala Leu Leu Gln Lys Ala Glu Pro His Asn Leu Gln
 660 665 670

Ile Thr Ala Ala Phe Leu Ala Gly Leu Leu Ser Arg Glu His Trp Gly
 675 680 685

Leu Leu Ala Glu Cys Gln Thr Ser Glu Lys Ala Leu Leu Arg Arg Gln
 690 695 700

Ala Cys Ala Arg Trp Cys Leu Ala Arg Ser Leu Arg Lys His Phe His
 705 710 715 720

Ser Ile Pro Pro Ala Ala Pro Gly Glu Ala Lys Ser Val His Ala Met
 725 730 735

Pro Gly Phe Ile Trp Leu Ile Arg Ser Leu Tyr Glu Met Gln Glu Glu
 740 745 750

Arg Leu Ala Arg Lys Ala Ala Arg Gly Leu Asn Val Gly His Leu Lys
 755 760 765

ES 2 486 340 T3

Leu Thr Phe Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala Phe
 770 775 780

Val Leu Gln His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr Asn
 785 790 795 800

Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly Val
 805 810 815

Cys Lys Ala Leu Tyr Leu Arg Asp Asn Asn Ile Ser Asp Arg Gly Ile
 820 825 830

Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys Leu
 835 840 845

Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met Ala
 850 855 860

Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly Asn
 865 870 875 880

Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu Arg
 885 890 895

Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val Gly
 900 905 910

Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln Ser
 915 920 925

Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly Ala
 930 935 940

Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu Leu
 945 950 955 960

Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu Ala
 965 970 975

Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser Asn
 980 985 990

Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu Glu
 995 1000 1005

Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe
 1010 1015 1020

Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu
 1025 1030 1035

Leu Leu
 1040

ES 2 486 340 T3

<211> 1013
 <212> PRT
 <213> Homo sapiens

5 <400> 3

Met Cys Ser Gln Glu Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu
 1 5 10 15

Leu Leu Val Ser Gly Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp
 20 25 30

Leu Leu Ser Trp Glu Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His
 35 40 45

Leu Leu Gly Gln Pro Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr
 50 55 60

Val Trp Asn Lys Gly Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala
 65 70 75 80

Gln Glu Ala Gln Ala Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp
 85 90 95

Asp Pro His Ser Leu His Pro Ala Arg Asp Leu Gln Ser His Arg Pro
 100 105 110

Ala Ile Val Arg Arg Leu His Ser His Val Glu Asn Met Leu Asp Leu

ES 2 486 340 T3

	115		120		125														
Ala	Trp	Glu	Arg	Gly	Phe	Val	Ser	Gln	Tyr	Glu	Cys	Asp	Glu	Ile	Arg				
	130					135					140								
Leu	Pro	Ile	Phe	Thr	Pro	Ser	Gln	Arg	Ala	Arg	Arg	Leu	Leu	Asp	Leu				
	145				150					155					160				
Ala	Thr	Val	Lys	Ala	Asn	Gly	Leu	Ala	Ala	Phe	Leu	Leu	Gln	His	Val				
				165					170					175					
Gln	Glu	Leu	Pro	Val	Pro	Leu	Ala	Leu	Pro	Leu	Glu	Ala	Ala	Thr	Cys				
			180					185						190					
Lys	Lys	Tyr	Met	Ala	Lys	Leu	Arg	Thr	Thr	Val	Ser	Ala	Gln	Ser	Arg				
		195					200					205							
Phe	Leu	Ser	Thr	Tyr	Asp	Gly	Ala	Glu	Thr	Leu	Cys	Leu	Glu	Asp	Ile				
	210					215					220								
Tyr	Thr	Glu	Asn	Val	Leu	Glu	Val	Trp	Ala	Asp	Val	Gly	Met	Ala	Gly				
	225				230					235					240				
Pro	Pro	Gln	Lys	Ser	Pro	Ala	Thr	Leu	Gly	Leu	Glu	Glu	Leu	Phe	Ser				
				245					250					255					
Thr	Pro	Gly	His	Leu	Asn	Asp	Asp	Ala	Asp	Thr	Val	Leu	Val	Val	Gly				
			260					265					270						
Glu	Ala	Gly	Ser	Gly	Lys	Ser	Thr	Leu	Leu	Gln	Arg	Leu	His	Leu	Leu				
		275					280					285							
Trp	Ala	Ala	Gly	Gln	Asp	Phe	Gln	Glu	Phe	Leu	Phe	Val	Phe	Pro	Phe				
	290					295					300								
Ser	Cys	Arg	Gln	Leu	Gln	Cys	Met	Ala	Lys	Pro	Leu	Ser	Val	Arg	Thr				
	305				310					315					320				
Leu	Leu	Phe	Glu	His	Cys	Cys	Trp	Pro	Asp	Val	Gly	Gln	Glu	Asp	Ile				
				325					330					335					
Phe	Gln	Leu	Leu	Leu	Asp	His	Pro	Asp	Arg	Val	Leu	Leu	Thr	Phe	Asp				
			340					345					350						

ES 2 486 340 T3

Gly Phe Asp Glu Phe Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys
 355 360 365

Ser Pro Thr Asp Pro Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu
 370 375 380

Gln Gly Asn Leu Leu Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro
 385 390 395 400

Ala Ala Val Ser Ala Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn
 405 410 415

Leu Lys Gly Phe Ser Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg
 420 425 430

His His Glu Pro Gly Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu
 435 440 445

Thr Ser Ala Leu His Gly Leu Cys His Leu Pro Val Phe Ser Trp Met
 450 455 460

Val Ser Lys Cys His Gln Glu Leu Leu Leu Gln Glu Gly Gly Ser Pro
 465 470 475 480

Lys Thr Thr Thr Asp Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu
 485 490 495

His Ala Thr Pro Pro Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu
 500 505 510

Leu Arg Gly Arg Leu Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu
 515 520 525

Trp Gly Leu Gly Met Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln
 530 535 540

Ala Ala Gln Val Ser Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg
 545 550 555 560

Ala Lys Gly Val Val Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His
 565 570 575

ES 2 486 340 T3

Ile Thr Phe Gln Cys Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala
580 585 590

Asp Val Pro Pro Ala Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro
595 600 605

Gly Asn Ser Pro Met Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala
610 615 620

Ser Glu Gly Lys Asp Ser Ser Val Ala Ala Leu Leu Gln Lys Ala Glu
625 630 635 640

Pro His Asn Leu Gln Ile Thr Ala Ala Phe Leu Ala Gly Leu Leu Ser
645 650 655

Arg Glu His Trp Gly Leu Leu Ala Glu Cys Gln Thr Ser Glu Lys Ala
660 665 670

Leu Leu Arg Arg Gln Ala Cys Ala Arg Trp Cys Leu Ala Arg Ser Leu
675 680 685

Arg Lys His Phe His Ser Ile Pro Pro Ala Ala Pro Gly Glu Ala Lys
690 695 700

Ser Val His Ala Met Pro Gly Phe Ile Trp Leu Ile Arg Ser Leu Tyr
705 710 715 720

Glu Met Gln Glu Glu Arg Leu Ala Arg Lys Ala Ala Arg Gly Leu Asn
725 730 735

Val Gly His Leu Lys Leu Thr Phe Cys Ser Val Gly Pro Thr Glu Cys
740 745 750

Ala Ala Leu Ala Phe Val Leu Gln His Leu Arg Arg Pro Val Ala Leu
755 760 765

Gln Leu Asp Tyr Asn Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu
770 775 780

Pro Cys Leu Gly Val Cys Lys Ala Leu Tyr Leu Arg Asp Asn Asn Ile
785 790 795 800

ES 2 486 340 T3

Ser Asp Arg Gly Ile Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu
805 810 815

Gln Leu Gln Lys Leu Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys
820 825 830

Ala His Ser Met Ala Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala
835 840 845

Leu Arg Leu Gly Asn Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu
850 855 860

Ala Glu Gly Leu Arg Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp
865 870 875 880

Gly Asn Arg Val Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu
885 890 895

Gly Asp His Gln Ser Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile
900 905 910

Gly Ser Val Gly Ala Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val
915 920 925

Met Leu Glu Glu Leu Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly
930 935 940

Val Cys Ser Leu Ala Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile
945 950 955 960

Leu Lys Leu Ser Asn Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu
965 970 975

Leu Gln Ala Leu Glu Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg
980 985 990

Gly Asn Thr Phe Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp
995 1000 1005

Thr Arg Leu Leu Leu
1010

<210> 4
<211> 1040
<212> PRT
<213> Homo sapiens

5

<400> 4

ES 2 486 340 T3

Met Gly Glu Glu Gly Gly Ser Ala Ser His Asp Glu Glu Glu Arg Ala
 1 5 10 15

Ser Val Leu Leu Gly His Ser Pro Gly Cys Glu Met Cys Ser Gln Glu
 20 25 30

Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser Gly
 35 40 45

Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp Glu
 50 55 60

Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln Pro
 65 70 75 80

Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr Val Trp Asn Lys Gly
 85 90 95

Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala Gln Glu Ala Gln Ala
 100 105 110

Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp Asp Pro His Ser Leu
 115 120 125

His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val Arg Arg
 130 135 140

Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu Arg Gly
 145 150 155 160

Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe Thr
 165 170 175

Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val Lys Ala

ES 2 486 340 T3

			180					185					190		
Asn	Gly	Leu	Ala	Ala	Phe	Leu	Leu	Gln	His	Val	Gln	Glu	Leu	Pro	Val
		195					200					205			
Pro	Leu	Ala	Leu	Pro	Leu	Glu	Ala	Ala	Thr	Cys	Lys	Lys	Tyr	Met	Ala
	210					215					220				
Lys	Leu	Arg	Thr	Thr	Val	Ser	Ala	Gln	Ser	Arg	Phe	Leu	Ser	Thr	Tyr
225					230					235					240
Asp	Gly	Ala	Glu	Thr	Leu	Cys	Leu	Glu	Asp	Ile	Tyr	Thr	Glu	Asn	Val
				245					250					255	
Leu	Glu	Val	Trp	Ala	Asp	Val	Gly	Met	Ala	Gly	Pro	Pro	Gln	Lys	Ser
			260					265					270		
Pro	Ala	Thr	Leu	Gly	Leu	Glu	Glu	Leu	Phe	Ser	Thr	Pro	Gly	His	Leu
		275					280						285		
Asn	Asp	Asp	Ala	Asp	Thr	Val	Leu	Val	Val	Gly	Glu	Ala	Gly	Ser	Gly
	290					295					300				
Lys	Ser	Thr	Leu	Leu	Gln	Arg	Leu	His	Leu	Leu	Trp	Ala	Ala	Gly	Gln
305					310					315					320
Asp	Phe	Gln	Glu	Phe	Leu	Phe	Val	Phe	Pro	Phe	Ser	Cys	Arg	Gln	Leu
				325					330					335	
Gln	Cys	Met	Ala	Lys	Pro	Leu	Ser	Val	Arg	Thr	Leu	Leu	Phe	Glu	His
			340					345					350		
Cys	Cys	Trp	Pro	Asp	Val	Gly	Gln	Glu	Asp	Ile	Phe	Gln	Leu	Leu	Leu
		355					360					365			
Asp	His	Pro	Asp	Arg	Val	Leu	Leu	Thr	Phe	Asp	Gly	Phe	Asp	Glu	Phe
	370					375					380				
Lys	Phe	Arg	Phe	Thr	Asp	Arg	Glu	Arg	His	Cys	Ser	Pro	Thr	Asp	Pro
385					390					395					400
Thr	Ser	Val	Gln	Thr	Leu	Leu	Phe	Asn	Leu	Leu	Gln	Gly	Asn	Leu	Leu
				405					410					415	

ES 2 486 340 T3

Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser Ala
 420 425 430

Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe Ser
 435 440 445

Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro Gly
 450 455 460

Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu His
 465 470 475 480

Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys His
 485 490 495

Gln Glu Leu Leu Leu Gln Glu Gly Gly Ser Pro Lys Thr Thr Thr Asp
 500 505 510

Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro Pro
 515 520 525

Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg Leu
 530 535 540

Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly Met
 545 550 555 560

Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val Ser
 565 570 575

Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val Val
 580 585 590

Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln Cys
 595 600 605

Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro Ala
 610 615 620

Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro Met
 625 630 635 640

ES 2 486 340 T3

Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys Asp
645 650 655

Ser Ser Val Ala Ala Leu Leu Gln Lys Ala Glu Pro His Asn Leu Gln
660 665 670

Ile Thr Ala Ala Phe Leu Ala Gly Leu Leu Ser Arg Glu His Trp Gly
675 680 685

Leu Leu Ala Glu Cys Gln Thr Ser Glu Lys Ala Leu Leu Arg His Gln
690 695 700

Ala Cys Ala Arg Trp Cys Leu Ala Arg Ser Leu Arg Lys His Phe His
705 710 715 720

Ser Ile Pro Pro Ala Ala Pro Gly Glu Ala Lys Ser Val His Ala Met
725 730 735

Pro Gly Phe Ile Trp Leu Ile Arg Ser Leu Tyr Glu Met Gln Glu Glu
740 745 750

Arg Leu Ala Arg Lys Ala Ala Arg Gly Leu Asn Val Gly His Leu Lys
755 760 765

Leu Thr Phe Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala Phe
770 775 780

Val Leu Gln His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr Asn
785 790 795 800

Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly Val
805 810 815

Cys Lys Ala Leu Tyr Leu Arg Asp Asn Asn Ile Ser Asp Arg Gly Ile
820 825 830

Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys Leu
835 840 845

Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met Ala
850 855 860

ES 2 486 340 T3

Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly Asn
865 870 875 880

Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu Arg
885 890 895

Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val Gly
900 905 910

Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln Ser
915 920 925

Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly Ala
930 935 940

Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu Leu
945 950 955 960

Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu Ala
965 970 975

Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser Asn
980 985 990

Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu Glu
995 1000 1005

Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe
1010 1015 1020

Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu
1025 1030 1035

Leu Leu
1040

<210> 5
<211> 97
<212> PRT
<213> Homo sapiens

<400> 5

5

ES 2 486 340 T3

Met Cys Ser Gln Glu Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu
1 5 10 15

Leu Leu Val Ser Gly Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp
20 25 30

Leu Leu Ser Trp Glu Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His
35 40 45

Leu Leu Gly Gln Pro Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr
50 55 60

Val Trp Asn Lys Gly Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala
65 70 75 80

Gln Glu Ala Gln Ala Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp
85 90 95

Asp

5 <210> 6
<211> 94
<212> PRT
<213> Homo sapiens

<400> 6

Ser Leu His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val
1 5 10 15

Arg Arg Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu
20 25 30

Arg Gly Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile
35 40 45

10 Phe Thr Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val
50 55 60

Lys Ala Asn Gly Leu Ala Ala Phe Leu Leu Gln His Val Gln Glu Leu
65 70 75 80

Pro Val Pro Leu Ala Leu Pro Leu Glu Ala Ala Thr Cys Lys
85 90

15 <210> 7
<211> 305
<212> PRT
<213> Homo sapiens

<400> 7

ES 2 486 340 T3

Pro Ala Thr Leu Gly Leu Glu Glu Leu Phe Ser Thr Pro Gly His Leu
 1 5 10 15
 Asn Asp Asp Ala Asp Thr Val Leu Val Val Gly Glu Ala Gly Ser Gly
 20 25 30
 Lys Ser Thr Leu Leu Gln Arg Leu His Leu Leu Trp Ala Ala Gly Gln
 35 40 45
 Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln Leu
 50 55 60
 Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu His
 65 70 75 80
 Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu Leu
 85 90 95
 Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu Phe
 100 105 110
 Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp Pro
 115 120 125
 Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu Leu
 130 135 140

ES 2 486 340 T3

Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser Ala
145 150 155 160

Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe Ser
165 170 175

Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro Gly
180 185 190

Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu His
195 200 205

Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys His
210 215 220

Gln Glu Leu Leu Leu Gln Glu Gly Gly Ser Pro Lys Thr Thr Thr Asp
225 230 235 240

Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro Pro
245 250 255

Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg Leu
260 265 270

Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly Met
275 280 285

Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val Ser
290 295 300

Pro
305

<210> 8
<211> 28
5 <212> PRT
<213> Homo sapiens

<400> 8

Arg Ser Leu Tyr Glu Met Gln Glu Glu Arg Leu Ala Arg Lys Ala Ala
1 5 10 15

Arg Gly Leu Asn Val Gly His Leu Lys Leu Thr Phe
20 25

10

15 <210> 9
<211> 28
<212> PRT
<213> Homo sapiens

<400> 9

ES 2 486 340 T3

Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala Phe Val Leu Gln
 1 5 10 15

His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr
 20 25

5 <210> 10
 <211> 25
 <212> PRT
 <213> Homo sapiens

<400> 10

Asn Ser Val Gly Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly
 1 5 10 15

Val Cys Lys Ala Leu Tyr Leu Arg Asp
 20 25

10 <210> 11
 <211> 28
 <212> PRT
 15 <213> Homo sapiens

<400> 11

Asn Asn Ile Ser Asp Arg Gly Ile Cys Lys Leu Ile Glu Cys Ala Leu
 1 5 10 15

His Cys Glu Gln Leu Gln Lys Leu Ala Leu Phe Asn
 20 25

20 <210> 12
 <211> 28
 <212> PRT
 25 <213> Homo sapiens

<400> 12

Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met Ala Lys Leu Leu Ala
 1 5 10 15

Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly Asn
 20 25

30 <210> 13
 <211> 28
 <212> PRT
 <213> Homo sapiens

35 <400> 13

Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu Arg
 1 5 10 15

Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly
 20 25

40 <210> 14
 <211> 28
 <212> PRT
 <213> Homo sapiens

ES 2 486 340 T3

<400> 14

Asn Arg Val Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly
1 5 10 15

Asp His Gln Ser Leu Arg Trp Leu Ser Leu Val Gly
20 25

5

<210> 15
<211> 28
<212> PRT
<213> Homo sapiens

10

<400> 15

Asn Asn Ile Gly Ser Val Gly Ala Gln Ala Leu Ala Leu Met Leu Ala
1 5 10 15

Lys Asn Val Met Leu Glu Glu Leu Cys Leu Glu Glu
20 25

15

<210> 16
<211> 28
<212> PRT
<213> Homo sapiens

20

<400> 16

Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu Ala Glu Gly Leu Lys
1 5 10 15

Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser Asn
20 25

25

<210> 17
<211> 28
<212> PRT
<213> Homo sapiens

30

<400> 17

Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu Glu
1 5 10 15

Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly
20 25

35

<210> 18
<211> 7
<212> PRT
<213> Homo sapiens

<400> 18

40

Glu Ala Gly Ser Gly Lys Ser
1 5

45

<210> 19
<211> 5
<212> PRT
<213> Homo sapiens

<400> 19

ES 2 486 340 T3

Leu Leu Thr Phe Asp
1 5

5 <210> 20
<211> 92
<212> PRT
<213> Homo sapiens

10 <400> 20

Glu Ser His Pro His Ile Gln Leu Leu Lys Ser Asn Arg Glu Leu Leu
1 5 10 15

Val Thr His Ile Arg Asn Thr Gln Cys Leu Val Asp Asn Leu Leu Lys
20 25 30

Asn Asp Tyr Phe Ser Ala Glu Asp Ala Glu Ile Val Cys Ala Cys Pro
35 40 45

Thr Gln Pro Asp Lys Val Arg Lys Ile Leu Asp Leu Val Gln Ser Lys
50 55 60

Gly Glu Glu Val Ser Glu Phe Phe Leu Tyr Leu Leu Gln Gln Leu Ala
65 70 75 80

Asp Ala Tyr Val Asp Leu Arg Pro Trp Leu Leu Glu
85 90

15 <210> 21
<211> 92
<212> PRT
<213> Homo sapiens

<400> 21

Gly Ile Ala Gln Gln Trp Ile Gln Ser Lys Arg Glu Asp Ile Val Asn
1 5 10 15

Gln Met Thr Glu Ala Cys Leu Asn Gln Ser Leu Asp Ala Leu Leu Ser
20 25 30

Arg Asp Leu Ile Met Lys Glu Asp Tyr Glu Leu Val Ser Thr Lys Pro
35 40 45

20 Thr Arg Thr Ser Lys Val Arg Gln Leu Leu Asp Thr Thr Asp Ile Gln
50 55 60

Gly Glu Glu Phe Ala Lys Val Ile Val Gln Lys Leu Lys Asp Asn Lys
65 70 75 80

Gln Met Gly Leu Gln Pro Tyr Pro Glu Ile Leu Val
85 90

25 <210> 22
<211> 93
<212> PRT
<213> Homo sapiens

ES 2 486 340 T3

<400> 22

Glu Arg Pro Ser Glu Thr Ile Asp Arg Glu Arg Lys Arg Leu Val Glu
 1 5 10 15
 Thr Leu Gln Ala Asp Ser Gly Leu Leu Leu Asp Ala Leu Val Ala Arg
 20 25 30
 Gly Val Leu Thr Gly Pro Glu Tyr Glu Ala Leu Asp Ala Leu Pro Asp
 35 40 45
 Ala Glu Arg Arg Val Arg Arg Leu Leu Leu Leu Val Gln Ser Lys Gly
 50 55 60
 Glu Ala Ala Cys Gln Glu Leu Leu Arg Cys Ala Gln Gln Thr Val Ser
 65 70 75 80
 Met Pro Asp Pro Ala Trp Asp Trp Gln His Val Gly Pro
 85 90

5 <210> 23
 <211> 94
 <212> PRT
 <213> Homo sapiens

10 <400> 23

Met Glu Ala Arg Asp Lys Gln Val Leu Arg Ser Leu Arg Leu Glu Leu
 1 5 10 15
 Gly Ala Glu Val Leu Val Glu Gly Leu Val Leu Gln Tyr Leu Tyr Gln
 20 25 30
 Glu Gly Ile Leu Thr Glu Asn His Ile Gln Glu Ile Asn Ala Gln Thr
 35 40 45
 Thr Gly Leu Arg Lys Thr Met Leu Leu Leu Asp Ile Leu Pro Ser Arg
 50 55 60
 Gly Pro Lys Ala Phe Asp Thr Phe Leu Asp Ser Leu Gln Glu Phe Pro
 65 70 75 80
 Trp Val Arg Glu Lys Leu Lys Lys Ala Arg Glu Glu Ala Met
 85 90

15 <210> 24
 <211> 91
 <212> PRT
 <213> Homo sapiens

20 <400> 24

ES 2 486 340 T3

Met His Pro His His Gln Glu Thr Leu Lys Lys Asn Arg Val Val Leu
 1 5 10 15

Ala Lys Gln Leu Leu Leu Ser Glu Leu Leu Glu His Leu Leu Glu Lys
 20 25 30

Asp Ile Ile Thr Leu Glu Met Arg Glu Leu Ile Gln Ala Lys Val Gly
 35 40 45

Ser Phe Ser Gln Asn Val Glu Leu Leu Asn Leu Leu Pro Lys Arg Gly
 50 55 60

Pro Gln Ala Phe Asp Ala Phe Cys Glu Ala Leu Arg Glu Thr Lys Gln
 65 70 75 80
 Gly His Leu Glu Asp Met Leu Leu Thr Leu
 85 90

5 <210> 25
 <211> 91
 <212> PRT
 <213> Homo sapiens

10 <400> 25

Met Arg Gln Asp Arg Arg Ser Leu Leu Glu Arg Asn Ile Met Met Phe
 1 5 10 15

Ser Ser His Leu Lys Val Asp Glu Ile Leu Glu Val Leu Ile Ala Lys
 20 25 30

Gln Val Leu Asn Ser Asp Asn Gly Asp Met Ile Asn Ser Cys Gly Thr
 35 40 45

Val Arg Glu Lys Arg Arg Glu Ile Val Lys Ala Val Gln Arg Arg Gly
 50 55 60

Asp Val Ala Phe Asp Ala Phe Tyr Asp Ala Leu Arg Ser Thr Gly His
 65 70 75 80

Glu Gly Leu Ala Glu Val Leu Glu Pro Leu Ala
 85 90

15 <210> 26
 <211> 90
 <212> PRT
 <213> Homo sapiens

<400> 26

20 Leu Cys Glu Ile Glu Cys Arg Ala Leu Ser Thr Ala His Thr Arg Leu
 1 5 10 15

ES 2 486 340 T3

Ile His Asp Phe Glu Pro Arg Asp Ala Leu Thr Tyr Leu Glu Gly Lys
 20 25 30
 Asn Ile Phe Thr Glu Asp His Ser Glu Leu Ile Ser Lys Met Ser Thr
 35 40 45
 Arg Leu Glu Arg Ile Ala Asn Phe Leu Arg Ile Tyr Arg Arg Gln Ala
 50 55 60
 Ser Glu Leu Gly Pro Leu Ile Asp Phe Phe Asn Tyr Asn Asn Gln Ser
 65 70 75 80
 His Leu Ala Asp Phe Leu Glu Asp Tyr Ile
 85 90

5 <210> 27
 <211> 93
 <212> PRT
 <213> Homo sapiens

<400> 27

Met Asp Glu Ala Asp Arg Arg Leu Leu Arg Arg Cys Arg Leu Arg Leu
 1 5 10 15
 Val Glu Glu Leu Gln Val Asp Gln Leu Trp Asp Val Leu Leu Ser Arg
 20 25 30
 Glu Leu Phe Arg Pro His Met Ile Glu Asp Ile Gln Arg Ala Gly Ser
 35 40 45
 Gly Ser Arg Arg Asp Gln Ala Arg Gln Leu Ile Ile Asp Leu Glu Thr
 50 55 60
 Arg Gly Ser Gln Ala Leu Pro Leu Phe Ile Ser Cys Leu Glu Asp Thr
 65 70 75 80
 Gly Gln Asp Met Leu Ala Ser Phe Leu Arg Thr Asn Arg
 85 90

10 <210> 28
 <211> 93
 <212> PRT
 15 <213> Homo sapiens

<400> 28

ES 2 486 340 T3

Met Asp Ala Lys Ala Arg Asn Cys Leu Leu Gln His Arg Glu Ala Leu
 1 5 10 15
 Glu Lys Asp Ile Lys Thr Ser Tyr Ile Met Asp His Met Ile Ser Asp
 20 25 30
 Gly Phe Leu Thr Ile Ser Glu Glu Glu Lys Val Arg Asn Glu Pro Thr
 35 40 45
 Gln Gln Gln Arg Ala Ala Met Leu Ile Lys Met Ile Leu Lys Lys Asp
 50 55 60
 Asn Asp Ser Tyr Val Ser Phe Tyr Asn Ala Leu Leu His Glu Gly Tyr
 65 70 75 80
 Lys Asp Leu Ala Ala Leu Leu His Asp Gly Ile Pro Val
 85 90

<210> 29
 <211> 92
 <212> PRT
 <213> Homo sapiens

5

<400> 29

Met Ala Ser Asp Asp Leu Ser Leu Ile Arg Lys Asn Arg Met Ala Leu
 1 5 10 15

Phe Gln Gln Leu Thr Cys Val Leu Pro Ile Leu Asp Asn Leu Leu Lys
 20 25 30

10

Ala Asn Val Ile Asn Lys Gln Glu His Asp Ile Ile Lys Gln Lys Thr
 35 40 45

Gln Ile Pro Leu Gln Ala Arg Glu Leu Ile Asp Thr Ile Trp Val Lys
 50 55 60

Gly Asn Ala Ala Ala Asn Ile Phe Lys Asn Cys Leu Lys Glu Ile Asp
 65 70 75 80

Ser Thr Leu Tyr Lys Asn Leu Phe Val Asp Lys Asn
 85 90

<210> 30
 <211> 320
 <212> PRT
 <213> Homo sapiens

15

<400> 30

ES 2 486 340 T3

Asn Glu Ser Leu Gly Ser Leu Asn Ser Leu Ala Cys Leu Leu Asp His
 1 5 10 15
 Thr Thr Gly Ile Leu Asn Glu Gln Gly Glu Thr Ile Phe Ile Leu Gly
 20 25 30
 Asp Ala Gly Val Gly Lys Ser Met Leu Leu Gln Arg Leu Gln Ser Leu
 35 40 45
 Trp Ala Thr Gly Arg Leu Asp Ala Gly Val Lys Phe Phe Phe His Phe
 50 55 60
 Arg Cys Arg Met Phe Ser Cys Phe Lys Glu Ser Asp Arg Leu Cys Leu
 65 70 75 80
 Gln Asp Leu Leu Phe Lys His Tyr Cys Tyr Pro Glu Arg Asp Pro Glu
 85 90 95
 Glu Val Phe Ala Phe Leu Leu Arg Phe Pro His Val Ala Leu Phe Thr
 100 105 110
 Phe Asp Gly Leu Asp Glu Leu His Ser Asp Leu Asp Leu Ser Arg Val
 115 120 125

ES 2 486 340 T3

Pro Asp Ser Ser Cys Pro Trp Glu Pro Ala His Pro Leu Val Leu Leu
 130 135 140

Ala Asn Leu Leu Ser Gly Lys Leu Leu Lys Gly Ala Ser Lys Leu Leu
 145 150 155 160

Thr Ala Arg Thr Gly Ile Glu Val Pro Arg Gln Phe Leu Arg Lys Lys
 165 170 175

Val Leu Leu Arg Gly Phe Ser Pro Ser His Leu Arg Ala Tyr Ala Arg
 180 185 190

Arg Met Phe Pro Glu Arg Ala Leu Gln Asp Arg Leu Leu Ser Gln Leu
 195 200 205

Glu Ala Asn Pro Asn Leu Cys Ser Leu Cys Ser Val Pro Leu Phe Cys
 210 215 220

Trp Ile Ile Phe Arg Cys Phe Gln His Phe Arg Ala Ala Phe Glu Gly
 225 230 235 240

Ser Pro Gln Leu Pro Asp Cys Thr Met Thr Leu Thr Asp Val Phe Val
 245 250 255

Leu Val Thr Glu Val His Leu Asn Arg Met Gln Pro Ser Ser Leu Val
 260 265 270

Gln Arg Asn Thr Arg Ser Pro Val Glu Thr Leu His Ala Gly Arg Asp
 275 280 285

Thr Leu Cys Ser Leu Gly Gln Val Ala His Arg Gly Met Glu Lys Ser
 290 295 300

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

<210> 31
 <211> 308
 <212> PRT
 <213> Homo sapiens
 <400> 31

5

ES 2 486 340 T3

Pro Val Val Phe Val Thr Arg Lys Lys Leu Val Asn Ala Ile Gln Gln
 1 5 10 15

Lys Leu Ser Lys Leu Lys Gly Glu Pro Gly Trp Val Thr Ile His Gly
 20 25 30

Met Ala Gly Cys Gly Lys Ser Val Leu Ala Ala Glu Ala Val Arg Asp
 35 40 45

His Ser Leu Leu Glu Gly Cys Phe Pro Gly Gly Val His Trp Val Ser
 50 55 60

Val Gly Lys Gln Asp Lys Ser Gly Leu Leu Met Lys Leu Gln Asn Leu
 65 70 75 80

Cys Thr Arg Leu Asp Gln Asp Glu Ser Phe Ser Gln Arg Leu Pro Leu
 85 90 95

Asn Ile Glu Glu Ala Lys Asp Arg Leu Arg Ile Leu Met Leu Arg Lys
 100 105 110

His Pro Arg Ser Leu Leu Ile Leu Asp Asp Val Trp Asp Ser Trp Val
 115 120 125

Leu Lys Ala Phe Asp Ser Gln Cys Gln Ile Leu Leu Thr Thr Arg Asp
 130 135 140

Lys Ser Val Thr Asp Ser Val Met Gly Pro Lys Tyr Val Val Pro Val
 145 150 155 160

Glu Ser Ser Leu Gly Lys Glu Lys Gly Leu Glu Ile Leu Ser Leu Phe
 165 170 175

Val Asn Met Lys Lys Ala Asp Leu Pro Glu Gln Ala His Ser Ile Ile
 180 185 190

Lys Glu Cys Lys Gly Ser Pro Leu Val Val Ser Leu Ile Gly Ala Leu
 195 200 205

ES 2 486 340 T3

Leu Arg Asp Phe Pro Asn Arg Trp Glu Tyr Tyr Leu Lys Gln Leu Gln
 210 215 220

Asn Lys Gln Phe Lys Arg Ile Arg Lys Ser Ser Ser Tyr Asp Tyr Glu
 225 230 235 240

Ala Leu Asp Glu Ala Met Ser Ile Ser Val Glu Met Leu Arg Glu Asp
 245 250 255

Ile Lys Asp Tyr Tyr Thr Asp Leu Ser Ile Leu Gln Lys Asp Val Lys
 260 265 270

Val Pro Thr Lys Val Leu Cys Ile Leu Trp Asp Met Glu Thr Glu Glu
 275 280 285

Val Glu Asp Ile Leu Gln Glu Phe Val Asn Lys Ser Leu Leu Phe Cys
 290 295 300

Asp Arg Asn Gly
 305

<210> 32
 <211> 315
 <212> PRT
 <213> Homo sapiens

5

<400> 32

Met Thr Cys Tyr Ile Arg Glu Tyr His Val Asp Arg Val Ile Lys Lys
 1 5 10 15

Leu Asp Glu Met Cys Asp Leu Asp Ser Phe Phe Leu Phe Leu His Gly
 20 25 30

Arg Ala Gly Ser Gly Lys Ser Val Ile Ala Ser Gln Ala Leu Ser Lys
 35 40 45

Ser Asp Gln Leu Ile Gly Ile Asn Tyr Asp Ser Ile Val Trp Leu Lys
 50 55 60

10

ES 2 486 340 T3

Asp Ser Gly Thr Ala Pro Lys Ser Thr Phe Asp Leu Phe Thr Asp Ile
65 70 75 80

Leu Leu Met Leu Lys Ser Glu Asp Asp Leu Leu Asn Phe Pro Ser Val
85 90 95

Glu His Val Thr Ser Val Val Leu Lys Arg Met Ile Cys Asn Ala Leu
100 105 110

Ile Asp Arg Pro Asn Thr Leu Phe Val Phe Asp Asp Val Val Gln Glu
115 120 125

Glu Thr Ile Arg Trp Ala Gln Glu Leu Arg Leu Arg Cys Leu Val Thr
130 135 140

Thr Arg Asp Val Glu Ile Ser Asn Ala Ala Ser Gln Thr Cys Glu Phe
145 150 155 160

Ile Glu Val Thr Ser Leu Glu Ile Asp Glu Cys Tyr Asp Phe Leu Glu
165 170 175

Ala Tyr Gly Met Pro Met Pro Val Gly Glu Lys Glu Glu Asp Val Leu
180 185 190

Asn Lys Thr Ile Glu Leu Ser Ser Gly Asn Pro Ala Thr Leu Met Met
195 200 205

Phe Phe Lys Ser Cys Glu Pro Lys Thr Phe Glu Lys Met Ala Gln Leu
210 215 220

Asn Asn Lys Leu Glu Ser Arg Gly Leu Val Gly Val Glu Cys Ile Thr
225 230 235 240

Pro Tyr Ser Tyr Lys Ser Leu Ala Met Ala Leu Gln Arg Cys Val Glu
245 250 255

Val Leu Ser Asp Glu Asp Arg Ser Ala Leu Ala Phe Ala Val Val Met
260 265 270

Pro Pro Gly Val Asp Ile Pro Val Lys Leu Trp Ser Cys Val Ile Pro
275 280 285

Val Asp Ile Cys Ser Asn Glu Glu Glu Gln Leu Asp Asp Glu Val Ala
290 295 300

Asp Arg Leu Lys Arg Leu Ser Lys Arg Gly Ala
305 310 315

<210> 33
5 <211> 4486
<212> ADN
<213> Homo sapiens

<400> 33

10

ES 2 486 340 T3

gtagacagat ccaggctcac cagtcctgtg ccactgggct tttggcgttc tgcacaaggc	60
ctaccogcag atgccatgcc tgctcccca gcctaattggg ctttgatggg ggaagagggt	120
ggttcagcct ctcacgatga ggaggaaaga gcaagtgtcc tcctcggaca ttctccgggt	180
tgtgaaatgt gctcgcagga ggcttttcag gcacagagga gccagctggt cgagctgctg	240
gtctcagggt ccctggaagg cttcgagagt gtctggact ggctgctgtc ctgggaggtc	300
ctctcctggg aggactacga gggcttcac ctctgggcc agcctctctc ccacttgcc	360
aggcgccttc tggacaccgt ctggaataag ggtacttggg cctgtcagaa gctcatcgcg	420
gctgcccag aagcccaggc cgacagccag tcccccaagc tgcatggctg ctgggacccc	480
cactcgtctc acccagcccg agacctgcag agtcaccggc cagccattgt caggaggtc	540
cacagccatg tggagaacat gctggacctg gcatgggagc ggggttctgt cagccagtat	600
gaatgtgatg aaatcagggt gccgatcttc acaccgtccc agagggcaag aaggctgctt	660
gatcttgcca cggtgaaagc gaatggattg gctgccttcc ttctacaaca tgttcaggaa	720
ttaccagtcc cattggccct gcctttggaa gctgccacat gcaagaagta tatggccaag	780
ctgaggacca cggtgtctgc tcagtctcgc ttctcagta cctatgatgg agcagagacg	840
ctctgcctgg aggacatata cacagagaat gtctggagg tctgggcaga tgtgggcatg	900
gctggacccc cgcagaagag cccagccacc ctgggcctgg aggagctctt cagcaccct	960
ggccacctca atgacgatgc ggacctgtg ctggtggtgg gtgaggcggg cagtggcaag	1020
agcacgctcc tgcagcggct gcaattgctg tgggctgcag ggcaagactt ccaggaattt	1080
ctctttgtct tccattcag ctgccggcag ctgcagtgca tggccaaacc actctctgtg	1140
cggactctac tctttgagca ctgctggttg cctgatgttg gtcagaaga catcttcag	1200

ES 2 486 340 T3

ttactccttg accaccctga ccgtgtcctg ttaacctttg atggctttga cgagttcaag 1260
 ttcaggttca cggatcgtga acgccactgc tccccgaccg accccacctc tgtccagacc 1320
 ctgctcttca accttctgca gggcaacctg ctgaagaatg cccgcaaggt ggtgaccagc 1380
 cgtccggccg ctgtgtcggc gttcctcagg aagtacatcc gcaccgagtt caacctcaag 1440
 ggcttctctg aacagggcat cgagctgtac ctgaggaagc gccatcatga gcccggggtg 1500
 gcggaccgcc tcattccgct gtcceaagag acctcagccc tgcaagggtt gtgccacctg 1560
 cctgtcttct catggatggt gtccaaatgc caccaggaac tgttgctgca ggaggggggg 1620
 tccccaaaga ccactacaga tatgtacctg ctgattctgc agcattttct gctgcatgcc 1680
 accccccag actcagcttc ccaaggctc ggaccagtc ttcttcgggg ccgcctcccc 1740
 acctcctgc acctgggag actggctctg tggggcctgg gcatgtgctg ctacgtgttc 1800
 tcagcccagc agctccaggc agcacaggtc agccctgatg acatttctct tggcttctg 1860
 gtgcgtgcca aagggtgctg gccagggagt acggcgcccc tggaaattct tcacatcact 1920
 ttccagtget tctttgccgc gttctacctg gcactcagtg ctgatgtgcc accagctttg 1980
 ctcagacacc tcttcaattg tggcaggcca ggcaactcac caatggccag gctcctgccc 2040
 acgatgtgca tccaggcctc ggagggaaag gacagcagcg tggcagcttt gctgcagaag 2100
 gccgagccgc acaaccttca gatcacagca gccttctctg cagggtctgt gtcccgggag 2160
 cactggggcc tgctggctga gtgccagaca tctgagaagg ccctgctccg gcgccaggcc 2220
 tgtgcccgtt ggtgtctggc ccgcagcctc cgcaagcact tccactccat cccgccagct 2280
 gcaccgggtg aggccaagag cgtgcatgcc atgccgggt tcactctggct catccggagc 2340
 ctgtacgaga tgcaggagga gcggctggct cgggaaggctg cacgtggcct gaatgttggg 2400
 cacctcaagt tgacattttg cagtgtgggc cccactgagt gtgctgccct ggcctttgtg 2460
 ctgcagcacc tccggcggcc cgtggccctg cagctggact acaactctgt gggtgacatt 2520
 ggcgtggagc agctgctgcc ttgccttggg gtctgcaagg ctctgtattt gcgcgataac 2580
 aatatctcag accgaggcat ctgcaagctc attgaatgtg ctcttcaactg cgagcaattg 2640
 cagaagttag ctctattcaa caacaaattg actgacggct gtgcacactc catggctaag 2700
 ctcttgcatt gcaggcagaa cttcttggca ttgaggctgg ggaataacta catcaactgc 2760
 gcgggagccc aagtgtctgc cgaggggctc cgaggcaaca cctccttgca gttcctggga 2820
 ttctggggca acagagtggg tgacgagggg gccagggccc tggctgaagc cttgggtgat 2880

ES 2 486 340 T3

caccagagct tgaggtggct cagcctggg ggaacaaca ttggcagtg ggggccc 2940
gccttggcac tgatgctggc aaagaacgtc atgctagaag aactctgcct ggaggagaac 3000
catctccagg atgaaggtgt atgttctctc gcagaaggac tgaagaaaa ttcaagtttg 3060
aaaatcctga agttgtccaa taactgcac accctac tag gggcagaagc cctcctgcag 3120
gccccctgaa aggaatgaca ccatcctgga agtctggctc cgagggaaaca ctttctctct 3180
agaggagggt gacaagctcg gctgcaggga caccagactc ttgctttgaa gtctccggga 3240
ggatgttcgt ctacagtttg ttgtgagcag gctgtgagtt tgggccccag aggctgggtg 3300
acatgtgttg gcagcctctt caaaatgagc cctgtcctgc ctaaggctga acttgttttc 3360
tgggaacacc ataggtcacc tttattctgg cagaggagg agcatcagtg ccctccagga 3420
tagacttttc ccaagcctac ttttgccatt gacttcttc caagattcaa tcccaggatg 3480
tacaaggaca gccctcctc catagtatgg gactggcctc tgctgatcct ccaggtctc 3540
cgtgtgggtc agtggggccc atggatgtgc ttgttaactg agtgcctttt ggtggagagg 3600
cccggcctct cacaaaagac cccttaccac tgctctgatg aagaggagta cacagaacac 3660
ataattcagg aagcagcttt ccccatgtct cgactcatcc atccaggcca tccccgtct 3720
ctggttcctc ccctcctect ggactcctgc acacgctect tctctgagg ctgaaattca 3780
gaatattagt gacctcagct ttgatatttc acttacagca cccccacc tggcaccag 3840
ggtgggaagg gctacacctt agcctgccct cctttccggt gttaagaca tttttggaag 3900
gggacacgtg acagccgttt gttccccaa acattctagg tttgcaagaa aaatatgacc 3960
aactccagc tgggatcaca tgtggacttt tatttccagt gaaatcagtt actcttcagt 4020
taagccttg gaaacagctc gactttaaaa agctccaaat gcagctttaa aaaattaatc 4080
tgggccagaa tttcaaacgg cctcactagg cttctggtg atgcctgtga actgaactct 4140
gacaacagac ttctgaaata gaccacaag aggcagttcc atttcatttg tgccagaatg 4200
ctttaggatg tacagttatg gattgaaagt ttacaggaaa aaaaattagg ccgttccttc 4260
aaagcaaatg tcttctgga ttattcaaaa tgatgtatgt tgaagccttt gtaaatgtc 4320
agatgctgtg caaatgttat tattttaaac attatgatgt gtgaaaactg gttaatattt 4380
ataggctact ttgttttact gtcttaagtt tatactctta tagacaacat ggccgtgaac 4440
tttatgctgt aaataatcag aggggaataa actggtgagt caaac 4486

<210> 34
<211> 1007
<212> PRT
<213> Homo sapiens

5

<400> 34

ES 2 486 340 T3

Met Gly Glu Glu Gly Gly Ser Ala Ser His Asp Glu Glu Glu Arg Ala
 1 5 10 15

Ser Val Leu Leu Gly His Ser Pro Gly Cys Glu Met Cys Ser Gln Glu
 20 25 30

Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser Gly
 35 40 45

Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp Glu
 50 55 60

Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln Pro
 65 70 75 80

Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr Val Trp Asn Lys Gly
 85 90 95

Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala Gln Glu Ala Gln Ala
 100 105 110

Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp Asp Pro His Ser Leu
 115 120 125

His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val Arg Arg
 130 135 140

Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu Arg Gly
 145 150 155 160

Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe Thr
 165 170 175

Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val Lys Ala
 180 185 190

ES 2 486 340 T3

Asn Gly Leu Ala Ala Phe Leu Leu Gln His Val Gln Glu Leu Pro Val
 195 200 205

Pro Leu Ala Leu Pro Leu Glu Ala Ala Thr Cys Lys Lys Tyr Met Ala
 210 215 220

Lys Leu Arg Thr Thr Val Ser Ala Gln Ser Arg Phe Leu Ser Thr Tyr
 225 230 235 240

Asp Gly Ala Glu Thr Leu Cys Leu Glu Asp Ile Tyr Thr Glu Asn Val
 245 250 255

Leu Glu Val Trp Ala Asp Val Gly Met Ala Gly Pro Pro Gln Lys Ser
 260 265 270

Pro Ala Thr Leu Gly Leu Glu Glu Leu Phe Ser Thr Pro Gly His Leu
 275 280 285

Asn Asp Asp Ala Asp Thr Val Leu Val Val Gly Glu Ala Gly Ser Gly
 290 295 300

Lys Ser Thr Leu Leu Gln Arg Leu His Leu Leu Trp Ala Ala Gly Gln
 305 310 315 320

Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln Leu
 325 330 335

Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu His
 340 345 350

Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu Leu
 355 360 365

Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu Phe
 370 375 380

Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp Pro
 385 390 395 400

Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu Leu
 405 410 415

ES 2 486 340 T3

Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser Ala
 420 425 430

Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe Ser
 435 440 445

Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro Gly
 450 455 460

Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu His
 465 470 475 480

Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys His
 485 490 495

Gln Glu Leu Leu Leu Gln Glu Gly Gly Ser Pro Lys Thr Thr Thr Asp
 500 505 510

Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro Pro
 515 520 525

Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg Leu
 530 535 540

Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly Met
 545 550 555 560

Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val Ser
 565 570 575

Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val Val
 580 585 590

Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln Cys
 595 600 605

Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro Ala
 610 615 620

Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro Met
 625 630 635 640

Ala Arg Leu Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys Asp

ES 2 486 340 T3

				645						650					655
Ser	Ser	Val	Ala	Ala	Leu	Leu	Gln	Lys	Ala	Glu	Pro	His	Asn	Leu	Gln
			660					665					670		
Ile	Thr	Ala	Ala	Phe	Leu	Ala	Gly	Leu	Leu	Ser	Arg	Glu	His	Trp	Gly
		675					680					685			
Leu	Leu	Ala	Glu	Cys	Gln	Thr	Ser	Glu	Lys	Ala	Leu	Leu	Arg	Arg	Gln
	690					695					700				
Ala	Cys	Ala	Arg	Trp	Cys	Leu	Ala	Arg	Ser	Leu	Arg	Lys	His	Phe	His
705					710					715					720
Ser	Ile	Pro	Pro	Ala	Ala	Pro	Gly	Glu	Ala	Lys	Ser	Val	His	Ala	Met
				725					730					735	
Pro	Gly	Phe	Ile	Trp	Leu	Ile	Arg	Ser	Leu	Tyr	Glu	Met	Gln	Glu	Glu
			740					745					750		
Arg	Leu	Ala	Arg	Lys	Ala	Ala	Arg	Gly	Leu	Asn	Val	Gly	His	Leu	Lys
		755					760					765			
Leu	Thr	Phe	Cys	Ser	Val	Gly	Pro	Thr	Glu	Cys	Ala	Ala	Leu	Ala	Phe
	770					775					780				
Val	Leu	Gln	His	Leu	Arg	Arg	Pro	Val	Ala	Leu	Gln	Leu	Asp	Tyr	Asn
785					790					795					800
Ser	Val	Gly	Asp	Ile	Gly	Val	Glu	Gln	Leu	Leu	Pro	Cys	Leu	Gly	Val
				805					810					815	
Cys	Lys	Ala	Leu	Tyr	Leu	Arg	Asp	Asn	Asn	Ile	Ser	Asp	Arg	Gly	Ile
			820					825					830		
Cys	Lys	Leu	Ile	Glu	Cys	Ala	Leu	His	Cys	Glu	Gln	Leu	Gln	Lys	Leu
		835					840						845		
Ala	Leu	Phe	Asn	Asn	Lys	Leu	Thr	Asp	Gly	Cys	Ala	His	Ser	Met	Ala
	850					855					860				
Lys	Leu	Leu	Ala	Cys	Arg	Gln	Asn	Phe	Leu	Ala	Leu	Arg	Leu	Gly	Asn
865					870					875					880

ES 2 486 340 T3

Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu Arg
 885 890 895

Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val Gly
 900 905 910

Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln Ser
 915 920 925

Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly Ala
 930 935 940

Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu Leu
 945 950 955 960

Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu Ala
 965 970 975

Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser Asn
 980 985 990

Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Pro
 995 1000 1005

<210> 35
 <211> 360
 <212> ADN
 <213> Homo sapiens

5

<400> 35

cagacatgag caggatgtgt ctaagggaca ggtgggcttc agtagactgg ctaactcctg 60

cagtctcttt aactggacag tttcaagagg aaaaccaaga atccttgaag ctccaccattg 120

tatcttcttt tccaggttgt ccaataactg catcacctac ctaggggcag aagccctcct 180

gcaggccctt gaaaggaatg acaccatcct ggaagtctgg taaggccctt gggcaggcct 240

gttttagctc tccgaacctc agtttttcta tctgtaaaat ggggtgacgg gagagaggaa 300

10 tggcagaatt ttgaggatcc cttctgattc tgacattcag tgagaatgat tctgcatgtg 360

REIVINDICACIONES

1. Un ácido nucleico aislado que comprende una secuencia que codifica un polipéptido seleccionado del grupo que consiste en las SEC ID N°: 34, 55, 57, 59, 85 y 89.
- 5 2. La secuencia de ácido nucleico de la reivindicación 1, en la que dicha secuencia está unida funcionalmente a un promotor heterólogo.
3. La secuencia de ácido nucleico de la reivindicación 2, en la que dicha secuencia está contenida dentro de un vector.
4. Una célula huésped que comprende el vector de la reivindicación 3.
- 10 5. La célula huésped de la reivindicación 4, en la que dicha célula huésped está localizada en un organismo seleccionado del grupo que consiste en una planta y un animal no humano.
6. El ácido nucleico aislado de la reivindicación 1, en el que dicha secuencia está seleccionada del grupo que consiste en las SEC ID N° 33, 54, 56, 58, 84 y 88.
7. Un polipéptido aislado que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en las SEC ID N° 34, 55, 57, 59, 85 y 89.
- 15 8. Un procedimiento implementado por ordenador de determinar el riesgo de un paciente de desarrollar enfermedad de Crohn, que comprende:
- a) proporcionar:
- un ordenador que comprende software para la predicción del riesgo de un paciente de desarrollar enfermedad de Crohn; y
- 20 b) detectar en una muestra de ácido nucleico que comprende un gen de Nod2 de un paciente en presencia de una o más variaciones en dicho gen de Nod2 del paciente, para generar información sobre variación genética, en la que dichas variaciones se seleccionan del grupo que consiste en Nod2 3020Insc, Nod2 c802t + 3020Insc, Nod2 c2104t, Nod2 g2722c, Nod2 c802t + g2722c, and Nod2 c802t + c2104t;
- c) introducir dicha información de variación genética en dicho ordenador; y
- 25 d) calcular el riesgo de dicho paciente con dicho software, en el que dicho cálculo es indicativo de que dicho sujeto puede estar expuesto al riesgo de desarrollar enfermedad de Crohn.
9. El procedimiento de la reivindicación 8, que además comprende la etapa de e) representar el riesgo de dicho paciente.
10. El procedimiento de la reivindicación 8, en el que dicho riesgo comprende un riesgo relativo del genotipo.
- 30 11. El procedimiento de la reivindicación 8, en el que dicho riesgo comprende un riesgo atribuible a la población,
12. El procedimiento de la reivindicación 8, en el que dicha detección en la etapa (b) comprende comparar la secuencia del ácido nucleico variante con la secuencia del ácido nucleico de Nod2 de tipo silvestre.
13. Un procedimiento de identificación de si un paciente tiene una susceptibilidad incrementada a la enfermedad de Crohn que está comprendida por la etapa de analizar las variantes de:

35 **Nod2 c802t + 3020Insc, Nod2 c802t + g2722c, y Nod2 c802t + c2104t.**

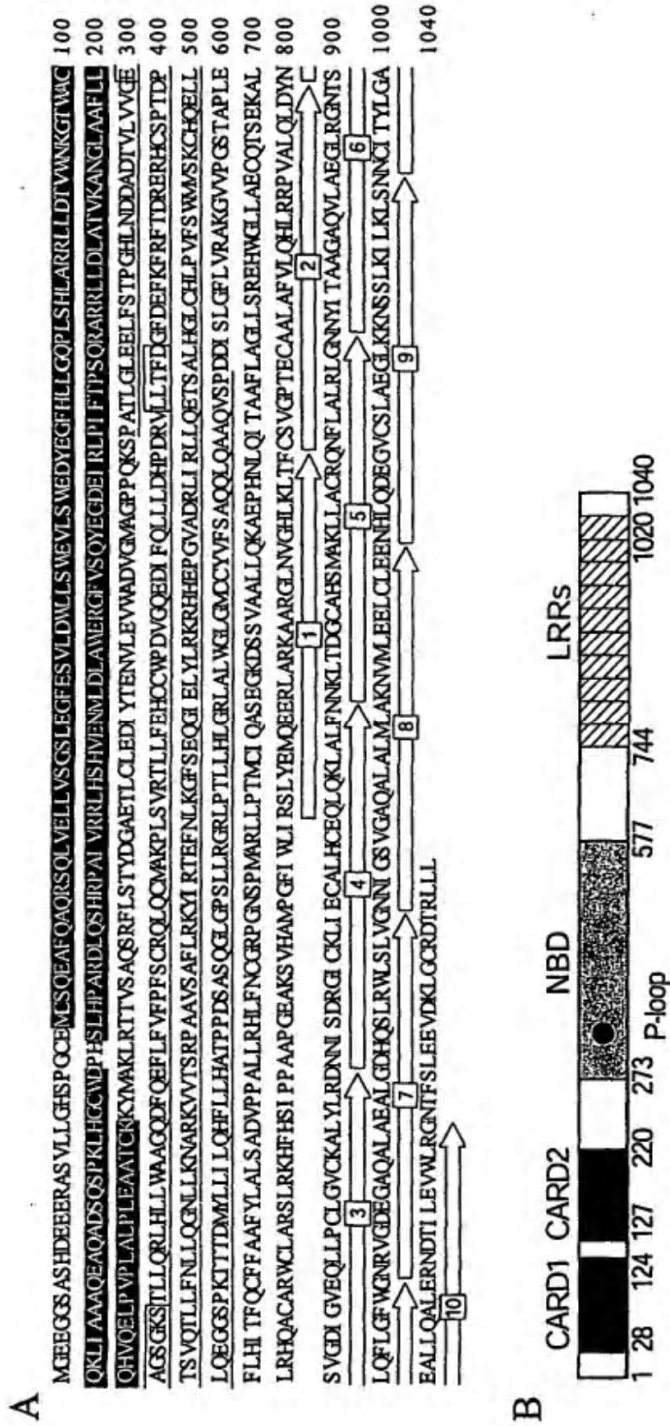


Figura 1

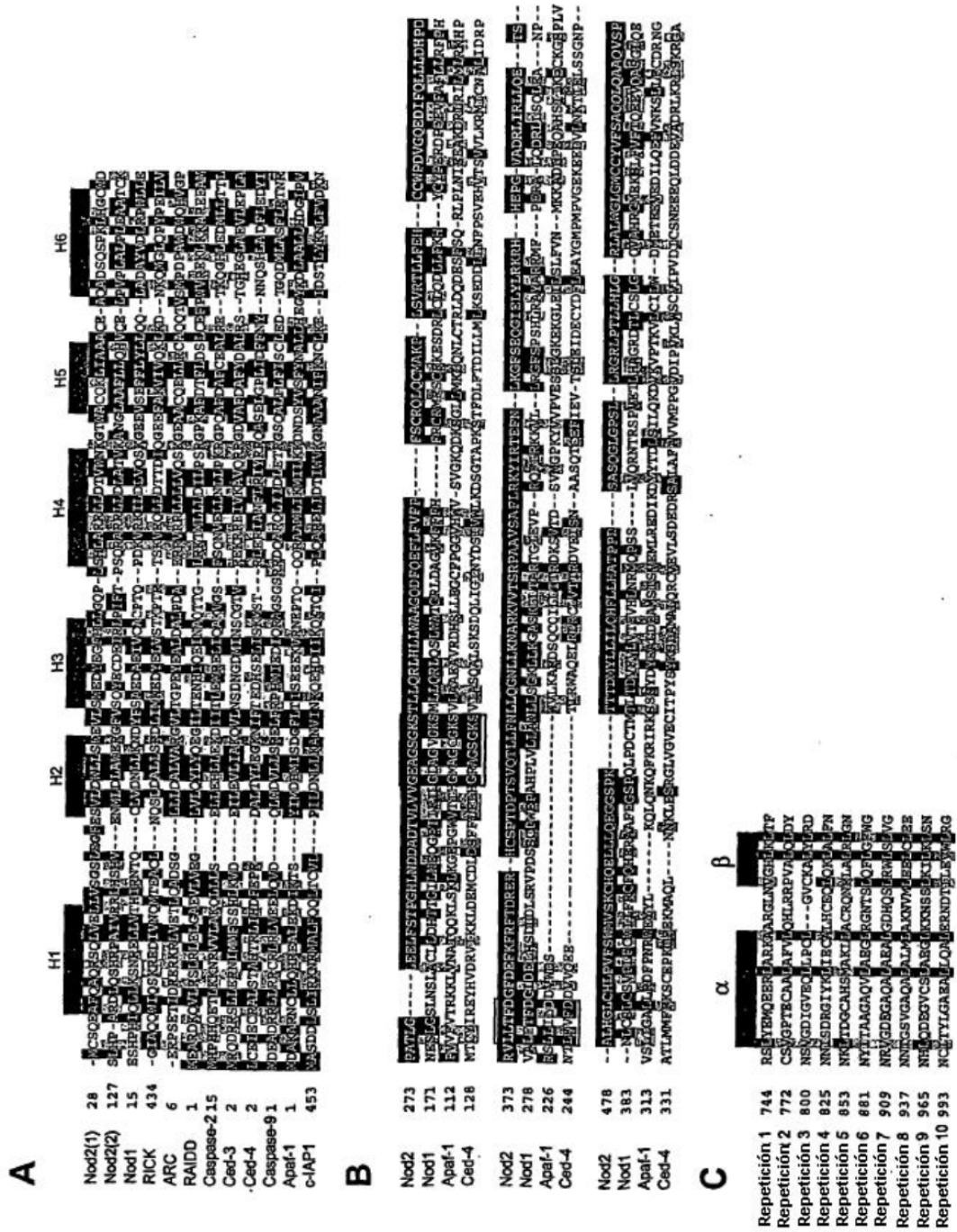


Figura 2

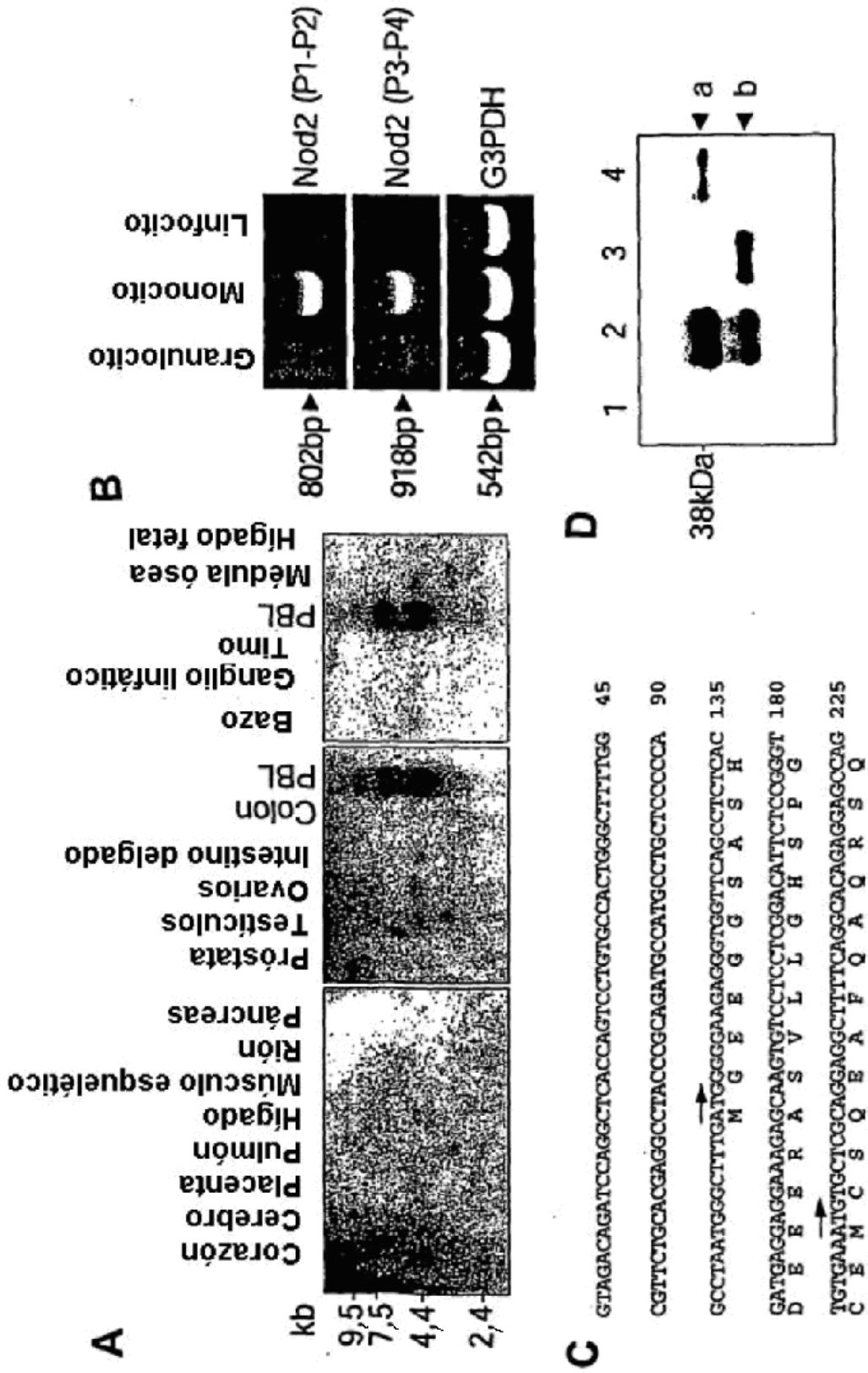


Figura 3

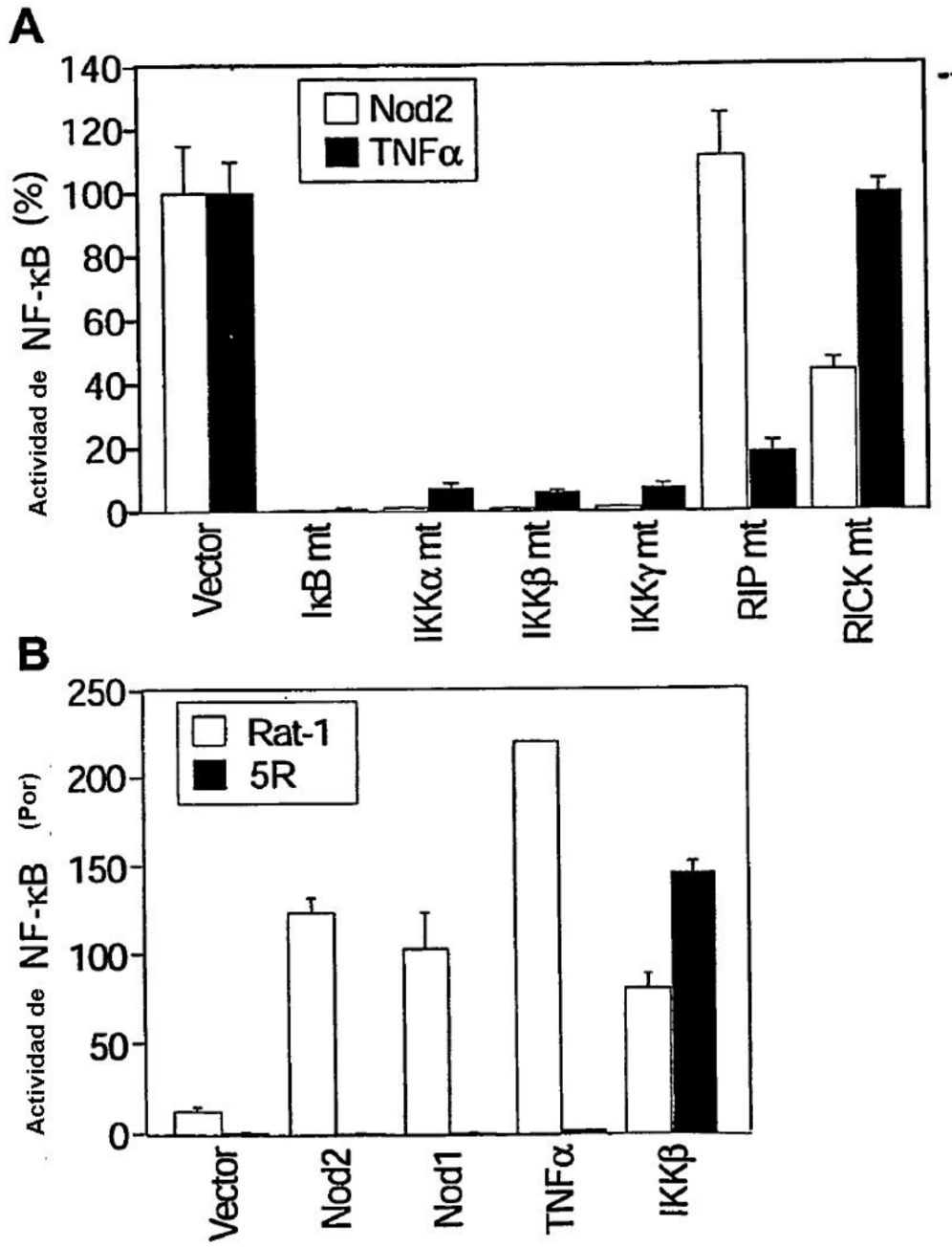


Figura 5

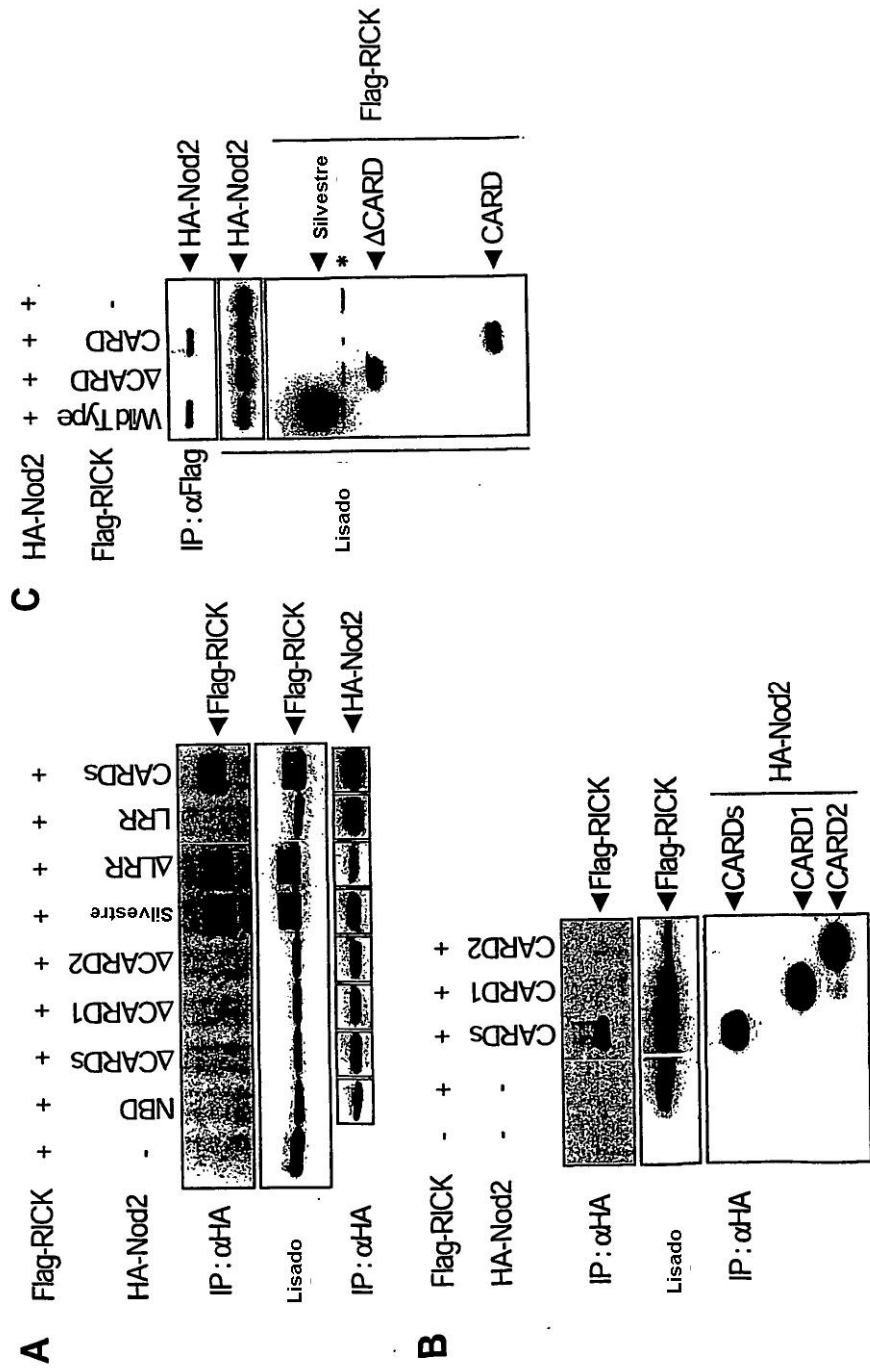


Figura 6

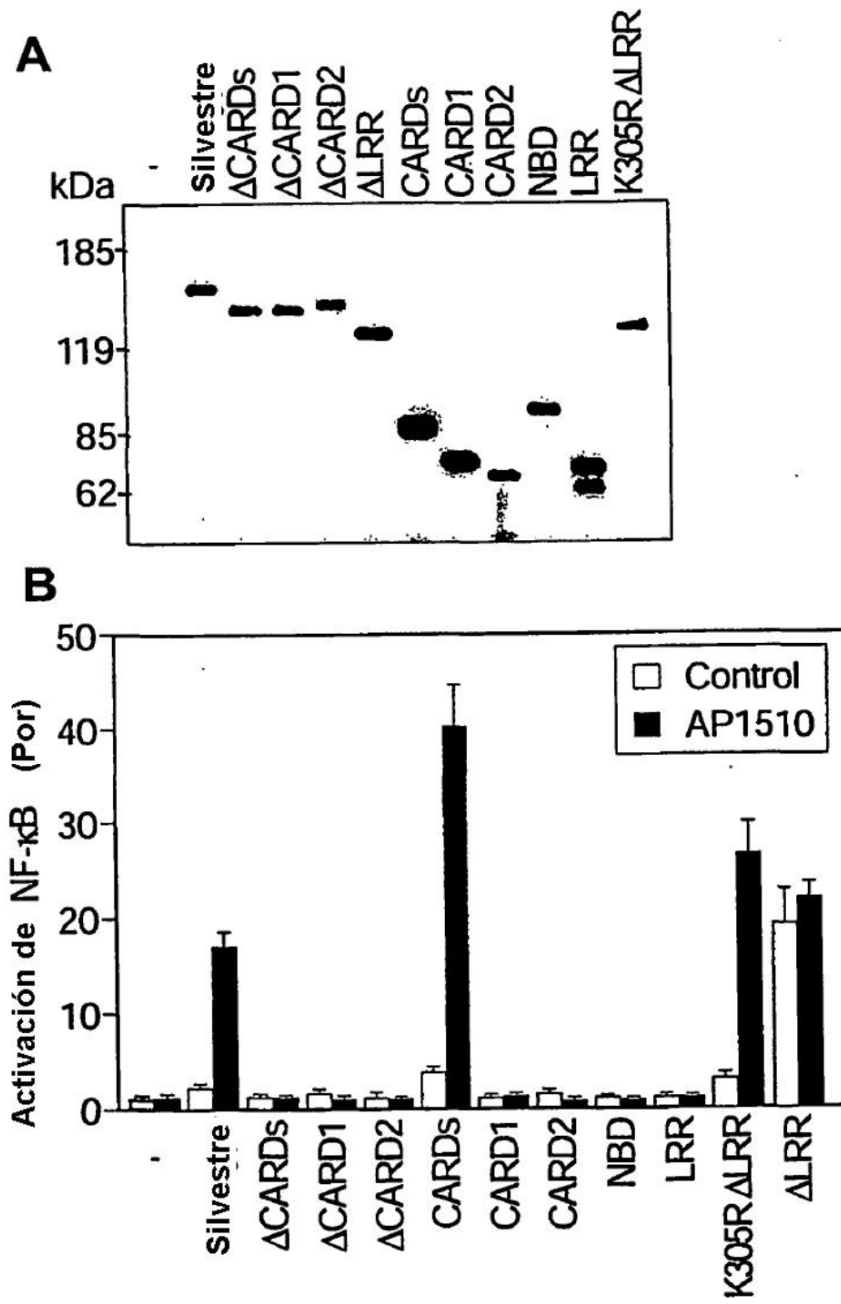


Figura 7

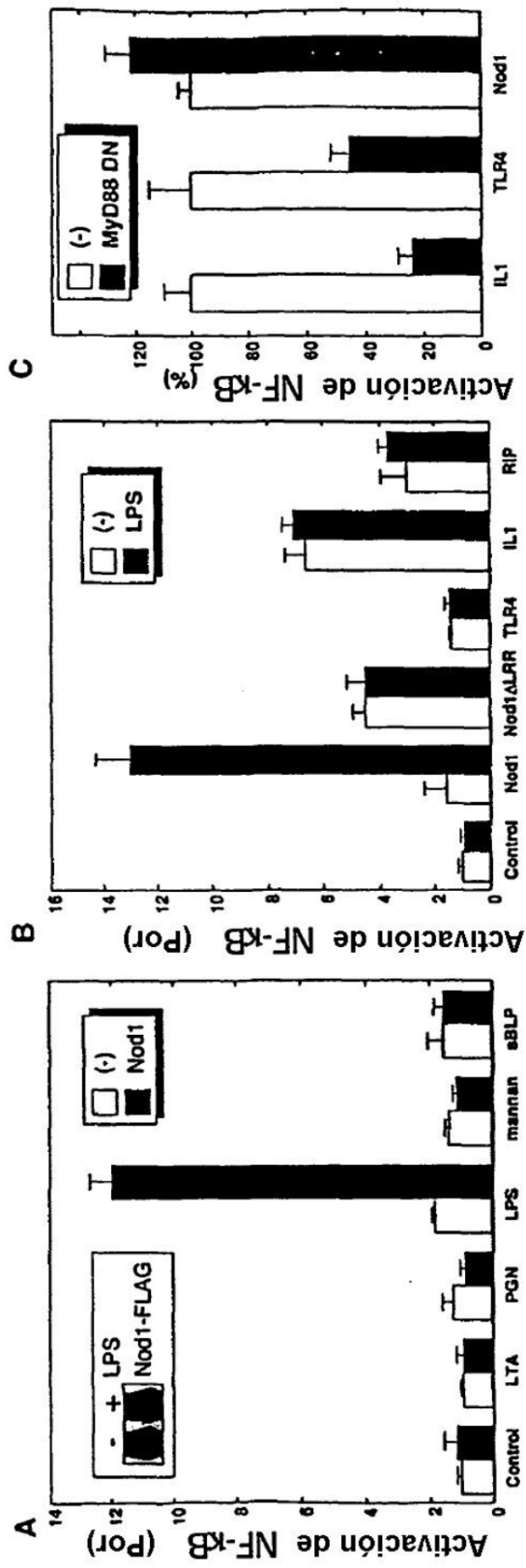


Figura 8

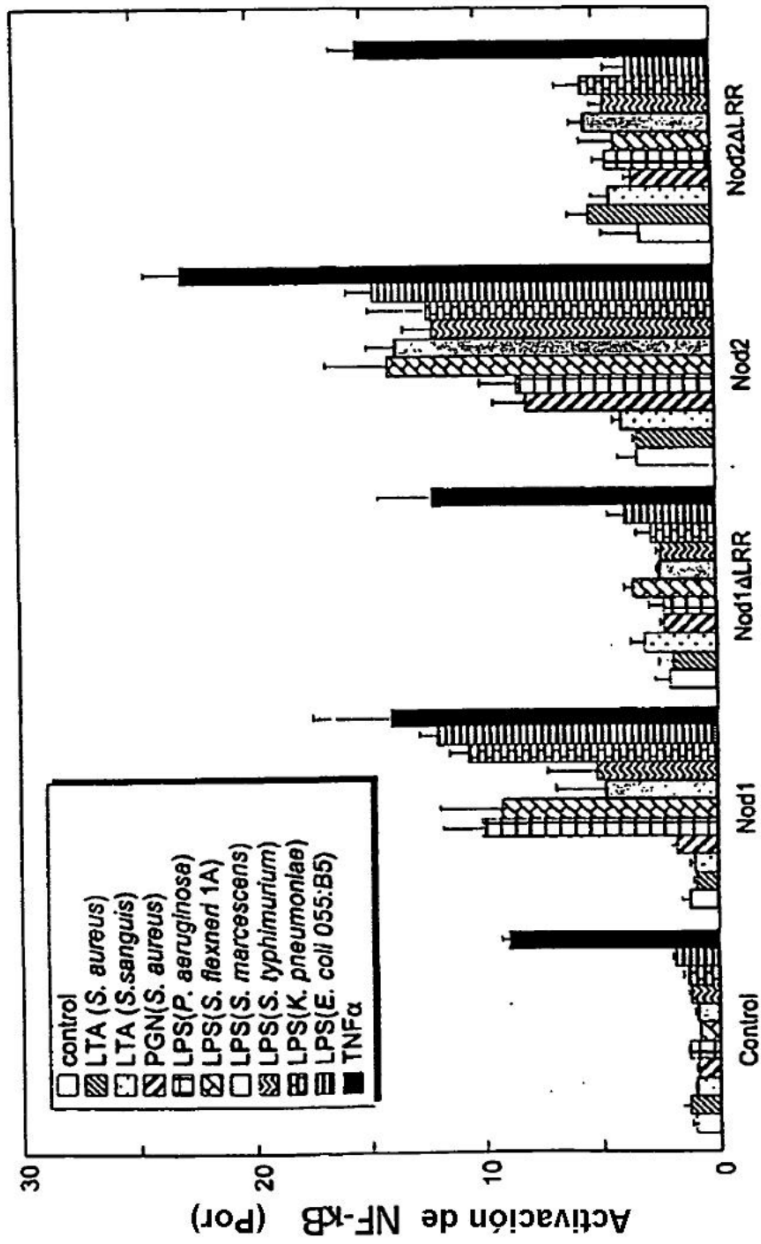


Figura 9

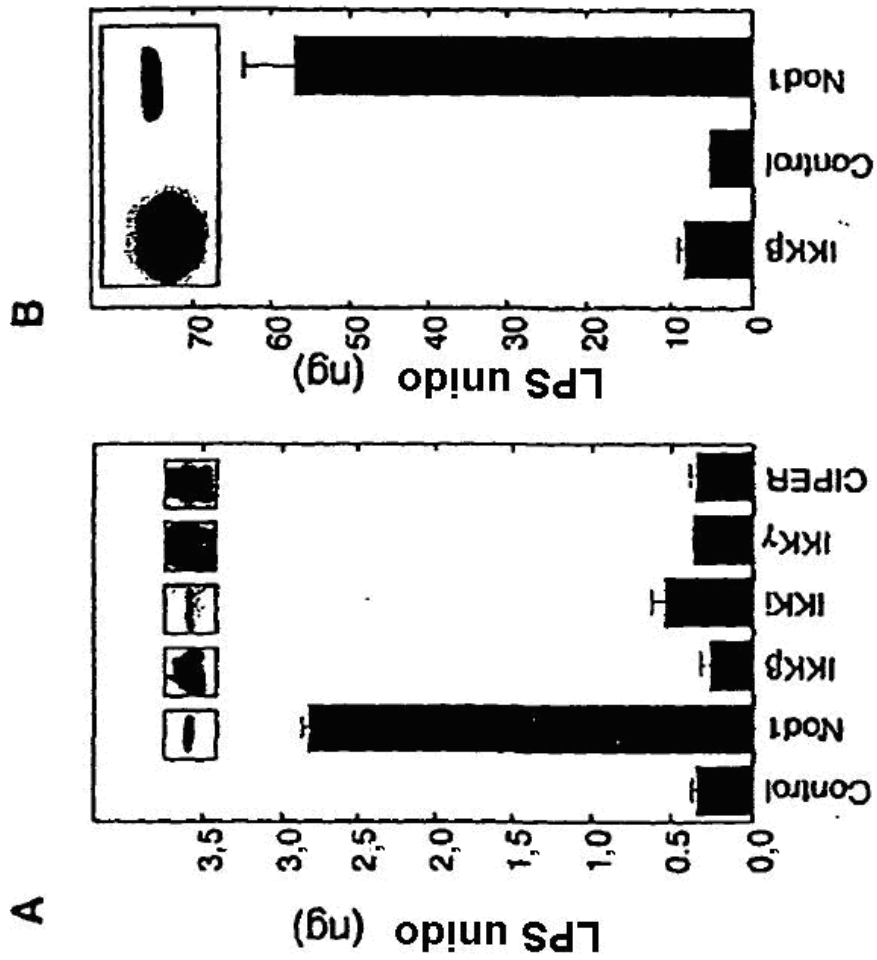


Figura 10

Figura 11

SEC ID N° 33

Secuencia del ADNc de Nod2

gtagacagat ccaggctcac cagtcctgtg ccaactgggt tttggcggtc tgcacaaggc	60
ctaccocgag atgccatgcc tgctcccca gcctaattggg ctttgatggg ggaagaggg	120
ggttcagcct ctacagatga ggaggaaaga gcaagtgtcc tcctcggaca ttctcgggt	180
tgtgaaatgt gctcgcagga ggcttttcag gcacagagga gccagctggg cgagctgctg	240
gtctcagggc ccctggaagg ctctcagaggt gtctcggact ggctgctgtc ctgggaggtc	300
ctctcctggg aggactacga gggcttccac ctctcgggcc agcctctctc ccacttggcc	360
aggcgccttc tggacaccgt ctggaataag ggtacttggg cctgtcagaa gctcatcgcg	420
gctgccaag aagcccaggc cgacagccag tcccccaagc tgcattggctg ctgggacccc	480
cactcgtctc acccagcccg agacctgcag agtcaccggc cagccattgt caggaggtc	540
cacagccatg tggagaacat gctggacctg gcatgggagc ggggtttcgt cagccagtat	600
gaatgtgatg aaatcaggtt gccgatctc acacogtccc agagggcaag aaggctgctt	660
gatcttgcca cggtgaaagc gaatggattg gctgccttcc ttctacaaca tgttcaggaa	720
ttaccagtc cattggccct gcctttggaa gctgccacat gcaagaagta tatggccaag	780
ctgaggacca cgggtgtctg tcagctctgc ttctcagta cctatgatgg agcagagacg	840
ctctgcctgg aggacatata cacagagaat gtctcggagg tctgggcaga tgtgggcatg	900
gctggacccc cgcagaagag cccagccacc ctgggcttgg aggagctctt cagcaccct	960
ggccacctca atgacgatgc ggacactgtg ctgggtgggg gtgaggcggg cagtggcaag	1020
agcacgctcc tgcagcggct gcacttgtg tgggctgcag ggcaagactt ccaggaattt	1080
ctctttgtct tcccattcag ctgccggcag ctgcagtgca tggccaaacc actctctgtg	1140
eggactctac tctttgagca ctgctgttg cctgatgttg gtcaagaaga catcttcag	1200
ttactccttg accaccctga ccgtgtcctg ttaaccttg atggcttga cgagttcaag	1260
ttcaggttca cggatcgtga acgccactgc tcccogaccg accccacctc tgtccagacc	1320

ES 2 486 340 T3

ctgctcttca accttctgca gggcaacctg ctgaagaatg ccogcaaggt ggtgaccagc 1380
 cgtccggcog ctgtgtcggc gttcctcagg aagtacatcc gcaccgagtt caacctcaag 1440
 ggcttctctg aacagggcat cgagctgtac ctgaggaagc gccatcatga gccoggggtg 1500
 ggggaccgoc tcatccgctt gctccaagag acctcagccc tgcacggttt gtgccacctg 1560
 cctgtcttct catggatggt gtccaaatgc caccaggaac tgttgctgca ggaggggggg 1620
 tccccaaaga ccactacaga tatgtacctg ctgattctgc agcattttct gctgcatgcc 1680
 accccccag actcagcttc ccaaggtctg ggaccagtc ttcttcgggg ccgctcccc 1740
 acctcctgc acctgggcag actggctctg tggggcctgg gcatgtgctg ctaegtgttc 1800
 tcagcccagc agctccaggc agcacaggtc agccctgatg acatttctct tggttctctg 1860
 gtgcgtgcca aaggtgtcgt gccagggagt acggcgcccc tggaaatcct tcacatcact 1920
 ttccagtgtt tctttgccgc gttctacctg gactcagtg ctgatgtgcc accagctttg 1980
 ctcagacacc tcttcaattg tggcaggcca ggcaactcac caatggccag gctcctgccc 2040
 acgatgtgca tccaggcctc ggagggaaag gacagcagcg tggcagcttt gctgcagaag 2100
 gccgagccgc acaaccttca gatcacagca gccttctctg cagggetggt gtcccgggag 2160
 cactggggcc tgetggctga gtgccagaca tctgagaagg ccttgetccg gcgccaggcc 2220
 tgtgcccgtt ggtgtctggc ccgcagcctc cgcaagcact tccactccat ccgcccagct 2280
 gcaccoggtg aggccaaag cgtgcatgcc atgccoggtt tcactctggt catcoggagc 2340
 ctgtacgaga tgcaggagga gggctggctt cggaaaggct cactggcctt gaatgtggg 2400
 cacctcaagt tgacattttg cagtgtgggc cccactgagt gtgctgcctt ggcctttgtg 2460
 ctgcagcacc tccggcggcc cgtggccctg cagctggact acaactctgt gggtgacatt 2520
 ggggtggagc agctgctgcc ttgccttggg gtctgcaagg ctctgtattt gogcgataac 2580
 aatctctcag accgaggcat ctgcaagctc attgaatgtg ctcttcaact cgagcaattg 2640
 cagaagttag ctctattcaa caacaaattg actgacggct gtgcacactc catggctaag 2700
 ctcttgcctt gcaggcagaa cttcttggca ttgaggctgg ggaataacta catcactgcc 2760
 gccggagccc aagtgtctggc cgaggggctc cgaggcaaca cctccttgcg gttcctggga 2820
 ttctggggca acagagtggg tgacgagggg gccaggccc tggctgaagc cttgggtgat 2880
 caccagagct tgagggtggc cagcctgggt gggaaacaaca ttggcagtgt gggtgcccaa 2940
 gccttggcac tgatgctggc aaagaacgtc atgctagaag aactctgcct ggaggagaac 3000
 catctccagg atgaagggtt atgttctctc gcagaaggac tgaagaaaaa ttcaagtttg 3060
 aaaatcctga agttgtccaa taactgcact acctacctag gggcagaagc cctcctgcag 3120

ES 2 486 340 T3

gcccttgaaa ggaatgacac catcctggaa gtctggctcc gagggaacac tttctctcta 3180
 gaggagggtg acaagctcgg ctgcagggac accagaactct tgctttgaag tctccgggag 3240
 gatgttcgtc tcagtttgtt tgtgagcagg ctgtgagttt gggccccaga ggctgggtga 3300
 catgtgttgg cagcctcttc aaaatgagcc ctgtcctgcc taaggctgaa cttgttttct 3360
 gggaacacca taggtcacct ttattctggc agaggagggg gcatcagtgc cctccaggat 3420
 agacttttcc caagcctact tttgccattg acttcttccc aagattcaat cccaggatgt 3480
 acaaggacag cccctcctcc atagtatggg actggcctct gctgatcctc ccaggcttcc 3540
 gtgtgggtca gtggggccca tggatgtgct tgtaactga gtgccttttg gtggagaggg 3600
 cgggcctctc acaaaagacc ccttaccact gctctgatga agaggagtac acagaacaca 3660
 taattcagga agcagctttc cccatgtctc gactcatcca tccaggccat tccccgtctc 3720
 tggttcctcc cctcctcctg gactcctgca cacgctcctt cctctgaggc tgaattcag 3780
 aatattagtg acctcagctt tgatatttca cttacagcac ccccaaccct ggcaccagg 3840
 gtgggaaggg ctacacctta gcctgccctc ctttccggtg ttaagacat ttttggagg 3900
 ggacaegtga cagccgtttg tcccccaaga cattctaggt ttgcaagaaa aatatgacca 3960
 cactccagct gggatcacat gtggactttt atttccagtg aatcagtta ctcttcagtt 4020
 aagcctttgg aaacagctcg actttaaaaa gctccaaatg cagctttaa aaattaatct 4080
 gggccagaat ttcaaaeggc ctcaactaggc ttctggttga tgctgtgaa ctgaactctg 4140
 acaacagact tctgaaatag acccacaaga ggcagttcca tttcatttgt gccagaatgc 4200
 tttaggatgt acagttatgg attgaaagt tacaggaaaa aaaattaggc cgttccttca 4260
 aagcaaatgt cttcctggat tattcaaaat gatgtatgt gaagcctttg taaattgtca 4320
 gatgctgtgc aatgttatt attttaaca ttatgatgtg tgaaaactgg ttaatattta 4380
 taggtcactt tgttttactg tcttaagttt atactcttat agacaacatg gccgtgaact 4440
 ttatgctgta aataatcaga ggggaataaa ctggtgagtc aaaac 4485

Figura 12

SEC ID N° 1

Secuencia de ADNc de Nod2

gtagacagat ccaggctcac cagtcctgtg ccactgggct tttggcgttc tgcacaaggc	60
ctaccgcag atgccatgcc tgctcccca gcctaatggg ctttgatggg ggaagagggt	120
ggttcagcct ctacagatga ggaggaaaga gcaagtgtcc tectcggaca ttctccgggt	180
tgtgaaatgt gctcgcagga ggcttttcag gcacagagga gccagctggg cgagctgctg	240
gtctcagggg ccctggaagg cttcgagagt gtcctggact ggctgctgtc ctgggaggtc	300
ctctcctggg aggactacga gggcttcac ctctcgggc agcctctctc ccacttggcc	360
aggcgccttc tggacaccgt ctggaataag ggtacttggg cctgtcagaa gctcatcgcg	420
gctgccaag aagcccaggc cgacagccag tcccccaagc tgcattggctg ctgggacccc	480
cactcgtctc acccagcccg agacctgcag agtcaccggc cagccattgt caggaggctc	540
cacagccatg tggagaacat gctggacctg gcatgggagc ggggtttcgt cagccagtat	600
gaatgtgatg aaatcaggtt gccgatcttc acaccgtccc agagggcaag aaggctgctt	660
gatcttgcca cggtgaaagc gaatggattg gctgccttcc ttctacaaca tgttcaggaa	720
ttaccagtcc cattggccct gcctttggaa gctgccacat gcaagaagta tatggccaag	780
ctgaggacca cgggtgtctg tcagtctcgc ttctcagta cctatgatgg agcagagacg	840
ctctgcctgg aggacatata cacagagaat gtctcggagg tctgggcaga tgtgggcatg	900
gctggacccc cgcagaagag cccagccacc ctgggcctgg aggagctctt cagcacccct	960
ggccacctca atgacgatgc ggacactgtg ctggtggtgg gtgaggcggg cagtggcaag	1020
agcacgctcc tgcagcggct gcacttgctg tgggctgcag ggcaagactt ccaggaattt	1080
ctctttgtct tccattcag ctgccggcag ctgcagtgea tggccaaacc actctctgtg	1140
eggactctac tctttgagca ctgctgttgg cctgatgttg gtcaagaaga catcttccag	1200
ttactccttg accaccctga ccgtgtctg ttaacctttg atggctttga cgagttcaag	1260
ttcaggttca cggatcgtga acgccactgc tccccgaccg accccacctc tgtccagacc	1320

ES 2 486 340 T3

ctgctcttca accttctgca gggcaacctg ctgaagaatg cccgcaaggt ggtgaccagc 1380
 cgtccggccg ctgtgtcggc gttcctcagg aagtacatcc gcaccgagtt caacctcaag 1440
 ggcttctctg aacagggcat cgagctgtac ctgaggaagc gccatcatga gcccggggtg 1500
 gcggaccgcc tcatccgcct gctccaagag acctcagccc tgcacggttt gtgccacctg 1560
 cctgtcttct catggatggg gtccaaatgc caccaggaac tgttgctgca ggaggggggg 1620
 tccccaaaga ccactacaga tatgtacctg ctgattctgc agcattttct gctgcatgcc 1680
 accccccag actcagcttc ccaaggtctg ggaccagtc ttcttcgggg ccgcctcccc 1740
 accctcctgc acctgggcag actggctctg tggggcctgg gcatgtgctg ctacgtgttc 1800
 tcagcccagc agctccaggg agcacaggtc agccctgatg acatttctct tggcttctg 1860
 gtgcgtgcca aaggtgtcgt gccagggagt acggcgcccc tggaaatcct tcacatcact 1920
 ttccagtgtc tctttgccgc gttctacctg gcactcagtg ctgatgtgcc accagctttg 1980
 ctcagacacc tcttcaattg tggcaggcca ggcaactcac caatggccag gctcctgccc 2040
 acgatgtgca tccaggcctc ggagggaaag gacagcagcg tggcagcttt gctgcagaag 2100
 gccgagccgc acaaccttca gatcacagca gccttctctg cagggtctgt gtcccgggag 2160
 cactggggcc tgctggctga gtgccagaca tctgagaagg ccctgctccg gcgccaggcc 2220
 tgtgcccgtt ggtgtctggc ccgagcctc cgcaagcact tccactccat cccgccagct 2280
 gcaccgggtg aggccaagag cgtgcatgcc atgccoggtt tcatctggct catccggagc 2340
 ctgtacgaga tgcaggagga gggctggct cggaaggctg cacgtggcct gaatgtggg 2400
 cacctcaagt tgacattttg cagtgtgggc cccactgagt gtgctgcctt ggcctttgtg 2460
 ctgcagcacc tccggcggcc cgtggccctg cagctggact acaactctgt gggtgacatt 2520
 ggcgtggagc agctgctgcc ttgccttggg gtctgcaagg ctctgtattt gcgcgataac 2580
 aatatctcag accgaggcat ctgcaagctc attgaatgtg ctcttcactg cgagcaattg 2640
 cagaagttag ctctattcaa caacaaattg actgacggct gtgcacactc catggctaag 2700
 ctcttgcac gcaggcagaa cttcttggca ttgaggtgg ggaataacta catcactgcc 2760
 gcgggagccc aagtgtctggc cgaggggctc cgaggcaaca cctccttga gttcctggga 2820
 ttctggggca acagagtggg tgacgagggg gcccgagccc tggctgaagc cttgggtgat 2880
 caccagagct tgaggtggct cagcctgggt gggaaacaaca ttggcagtg gggtgcccaa 2940
 gccttggcac tgatgctggc aaagaacgtc atgctagaag aactctgctt ggaggagaac 3000
 catctccagg atgaagggtg atgttctctc gcagaaggac tgaagaaaaa ttcaagttt 3060
 aaaatcctga agttgtccaa taactgcatc acctacctag gggcagaagc cctcctgcag 3120

gcccttgaaa ggaatgacac catcctggaa gtctggctcc gagggaacac tttctctcta 3180
 gaggaggttg acaagctcgg ctgcagggac accagactct tgctttgaag tctccgggag 3240
 gatgttcgto tcagtttggt tgtgagcagg ctgtgagttt gggcccaga ggctgggtga 3300
 catgtgttg cagcctcttc aaaatgagcc ctgtcctgcc taaggctgaa cttgttttct 3360
 gggaaacca taggtcacct ttattctggc agaggagga gcatcagtgc cctccaggat 3420
 agacttttcc caagcctact tttgccattg acttcttccc aagattcaat cccaggatgt 3480
 acaaggacag cccctcctcc atagtatggg actggcctct gctgatcctc ccaggcttcc 3540
 gtgtgggtca gtggggccca tggatgtgct tgtaactga gtgccttttg gtggagaggc 3600
 ccggcctctc acaaaagacc ccttaccact gctctgatga agaggagtac acagaacaca 3660
 taattcagga agcagctttc cccatgtctc gactcatcca tccaggccat tccccgtctc 3720
 tggttcctcc cctcctcctg gactcctgca caogctcctt cctctgagge tgaaattcag 3780
 aatattagtg acctcagctt tgatatttca cttacagdac ccccaaccct ggcaccagg 3840
 gtgggaaggg ctacacctta gctgcctc ctttcoggtg ttaagacat ttttgaagg 3900
 ggacacgtga cagcogtttg ttcccaga cattctaggt ttgcaagaaa aatatgacca 3960
 cactccagct gggatcacat gtggactttt atttccagtg aaatcagtta ctcttcagtt 4020
 aagcctttgg aaacagctcg actttaaaaa gctccaatg cagctttaa aaattaatct 4080
 gggccagaat ttcaaacggc ctactagge ttctgggtga tgcctgtgaa ctgaactctg 4140
 acaacagact tctgaaatag accacaaga ggcagttcca tttcatttgt gccagaatgc 4200
 tttaggatgt acagttatgg attgaaagt tacaggaaaa aaaattagge cgttccttca 4260
 aagcaaagt ctctctggat tattcaaat gatgtatgt gaagccttg taaattgtca 4320
 gatgetgtgc aaatgttatt attttaaca ttatgatgtg tgaaaactgg ttaatattta 4380
 taggtcactt tgttttactg tcttaagttt atactcttat agacaacatg gcogtgaact 4440
 ttatgctgta aataatcaga ggggaataaa ctgttgagtc aaaac 4485

Figura 13

SEC ID N° 2

MGEEGGSASH DEEERASVLL GHSPGCEMCS QEAFQAQRSQ LVELLVSGSL -EGFESVLDWL
 LSWEVLSWED YEGFHLLGQP LSHLARRLLD TVWNKGTWAC QKLIAAAQEA QADSQSPKLIH
 GCWDPHSLHP ARDLQSHRPA IVRRLHSHVE NMLDLAWERG FVSQYECDEI RLPIFTSPQR
 ARLLDLATV KANGLAAFLI QHVQELPVPL ALPLEAATCK KYMAKLRTTV SAQSRFLSTY
 DGAETLCLED IYTENVLEVW ADVGMAGPPQ KSPATLGLLE LFSTPGHLND DADTVLVVGE
 AGSGKSTLLQ RLHLLWAAGQ DFQEFVFP FSCRQLQCMA KPLSVRTLLF EHCCWPDVGO
 EDIFQLLLDH PDRVLLTFDG FDEFKFRFTD RERHCSPTDP TSVQTLFNL LQGNLLKNAR
 KVVTSRPAAV SAFLRKYIRT EFNLKGFSEQ GIELYLRKRH HEPGVADRLI RLLQETSALH
 GLCHLPVFSW MVSKCHQELL LQEGGSPKTT TDMYLLILQH FLLHATPPDS ASQGLGPSLL
 RGRLPPTLLHL GRLALWGLGM CCYVFSQAQL QAAQVSPDDI SLGFLVRAKG VVPGSTAPLE
 FLHITFQCFE AAFYLALSAD VPPALLRHLF NCGRPGNSPM ARLLPTMCIQ ASEGKDSSVA
 ALLQKAEPHN LQITAAFLAG LLSREHWGLL AECQTSEKAL LRRQACARWC LARSLRKHFH
 SIPPAPGEA KSVHAMPFI WLIRSLYEMQ EERLARKAAR GLNVGHLKLT FCSVGPTECA
 ALAFVLQHLR RPVALQLDYN SVGDIGVEQL LPCLGVCKAL YLRDNNISDR GICKLIECAL
 HCEQLQKLAL FNNKLTGCA HSMAKLLACR QNFLALRLGN NYITAAGAQV LAEGLRGNTS
 LQFLGFWGNR VGDEGAQALA EALGDHQLR WLSLVGNNIG SVGAQALALM LAKNVMLEEL
 CLEENHLQDE GVCSLAEGK KNSSLKILKL SNNCITYLGA EALLQALERN DTILEVWLRG
 NTFSLEEVDK LGCRDTRLLL *

Figura 14

SEC ID N° 3

MCSQEAFQAA RSQVLVLLVS GSLEGFESVL DWLLSWEVLS WEDYEGFHLL GQPLSHLARR
LLDVTWVWNGT WACQKLIAAA QEAQADSQSP KLHGCWDPHS LHPARDLQSH RPAIVRRRLHS
HVENMLDLAW ERGFVSQYEC DEIRLPIFTP SQRARLLDL ATVKANGLAA FLLQHVQELP
VPLALPLEAA TCKKYMAKLR TTVSAQSRFL STYDGAETLC LEDIYTENVL EVWADVGMAG
PPQKSPATLG LEELFSTPGH LNDDADTVLV VGEAGSGKST LLQRLHLLWA AGQDFQEFLE
VPPFSCRQLQ CMAKPLSVRT LLFEHCCWPD VGQEDIFQLL LDHPDRVLLT FDGFDEFKFR
FTDRERHCSP TDPTSVQTLN FNLLQGNLLK NARKVVTSRP AAVSAFLRKY IRTEFNLKGF
SEQIELYLR KRHHPEGVAD RLIRLLQETS ALHGLCHLPV FSWMVSCKHQ ELLLOEGGSP
KTTTDMYLLI LQHFLHATP PDSASQGLGP SLLRGRLLPTL LHLGRLALWG LGMCCYVFS
A
QQLQAAQVSP DDISLGFLVR AKGVVPGSTA PLEFLHITFQ CFFAAFYLL SADVPPALLR
HLFNCGRPGN SPMARLLPTM CIQASEGKDS SVAALLQKAE PHNLQITAAF LAGLLSREHW
GLLAECQTSE KALLRRQACA RWCLARSLRK HFHSIPPAAP GEAKSVHAMP GFIWLIRSLY
EMQEERLARK AARGLNVGHL KLTFCVSGPT ECAALAFVLQ HLRRPVALQL DYNVSGDIGV
EQLLPCLGVC KALYLRDNNI SDRGICKLIE CALHCEQLQK LALFNNKLTG GCAHSMAKLL
ACRQNFLLR LGNNYITAAG AQVLAEGLRG NTSLQFLGFW GNRVGDGAQ ALAEALGDHQ
SLRWLSLVGN NIGSVGAQAL ALMLAKNVML EELCLEENHL QDEGVCSLAE GLKKNSSLKI
LKLSNNCITY LGAEALLQAL BRNDTILEVW LRGNTFSLEE VDKLGCRDTR LLL*

Figura 15

SEC ID N° 34

Secuencia AA de Nod2, mutante

MGEEGGSASH DEEERASVLL GHSPGCEMCS QEAFQAQRSQ LVELLVSGSL EGFESVLDWL
 LSWEVLSWED YEGFHLGQP LSHLARLLD TVWNKGTWAC QKLIAAAQEA QADSQSPKLH
 GCWDPHSLHP ARDLQSHRPA IVRRLHSHVE NMLDLAWERG FVSQYECDEI RLPIFTSPQR
 ARRLDLATV KANGLAAFL LQHVEQELPVPL ALPLEAATCK KYMAKLRRTTV SAQSRFLSTY
 DGAETLCLED IYTENVLEVW ADVGMAGPPQ KSPATLGLLE LFSTPGHLND DADTVLVVGE
 AGSGKSTLLQ RLHLLWAAGQ DFQEFLLVFP FSCRQLQCMA KPLSVRTLLF BHCCWPDVGQ
 EDIFQLLLDH PDRVLLTFDG FDEFKFRFTD RERHCSPTDP TSVQTLLENL LQGNLLKNAR
 KVVTSRPAAV SAFLRKYIRT EFNLKGFSQ GIELYLRKRH HEPGVADRLI RLLQETSALH
 GLCHLPVFSW MVSKCHQELL LQEGGSPKTT TDMYLLILQH FLLHATPPDS ASQGLGPSLL
 RGRLPTLLHL GRLALWGLGM CCYVFSQAQL QAAQVSPDDI SLGFLVRAKG VVPGSTAPLE
 FLHITFQCFE AAFYLALSAD VPPALLRHLF NCGRPGNSPM ARLLPTMCIQ ASEGKDSSVA
 ALLQKAEPHN LQITAAFLAG LLSREHWGLL AECQTSEKAL LRRQACARWC LARSLRKHFH
 SIPPAAPGEA KSVHAMPFI WLIRSLYEMQ EERLARKAAR GLNVGHLKLT FCSVGPTTECA
 ALAFVLQHLR RPVALQLDYN SVGDIGVEQL LPCLGVCKAL YLRDNNISDR GICKLIECAL
 HCEQLQKLAL FNNKLTGCA HSMAKLLACR QNFLALRLGN NYITAAGAQQV LAEGLRGNTS
 LQFLGFWGNR VGDEGAQALA EALGDHQSLR WLSLVGNNIG SVGAQALALM LAKNVMLEEL
 CLEENHLQDE GVCSLAEGLK KNSSLKILKL SNNCITYLGA EALLQAP*

FIGURA 16

Exón 11 de Nod2, de tipo silvestre

cagacatgag caggatgtgt ctaagggaca ggtgggcttc agtagactgg ctaactcctg

cagtctcttt aactggacag tttcaagagg aaaaccaaga atccttgaag ctcaccattg

tatcttcttt tccagGTTGT CCAATAACTG CATCACCTAC CTAGGGGCAG AAGCCCTCCT

L S N N C I T Y L G A E A L L

GCAGGCCCTT GAAAGGAATG ACACCATCCT GGAAGTCTGg taaggcccct gggcaggcct

Q A L E R N D T I L E V

gttttagctc tccgaacctc agtttttcta tctgtaaaat ggggtgacgg gagagaggaa

tggcagaatt ttgaggatcc cttctgatte tgacattcag tgagaatgat totgcatgtg

Exón 11 de Nod2, mutante

cagacatgag caggatgtgt ctaagggaca ggtgggcttc agtagactgg ctaactcctg

cagtctcttt aactggacag tttcaagagg aaaaccaaga atccttgaag ctcaccattg

tatcttcttt tccagGTTGT CCAATAACTG CATCACCTAC CTAGGGGCAG AAGCCCTCCT

L S N N C I T Y L G A E A L L

GCAGGCCCTT TGAAAGGAAT GACACCATCC TGGAAAGTCTG gtaaggcccc tgggcaggcc

Q A P *

tgtttagct ctccgaacct cagtttttct atctgtaaaa tggggtgacg ggagagagga

atggcagaat tttgaggatc ctttctgatt ctgacattca gtgagaatga ttctgcatgt

g

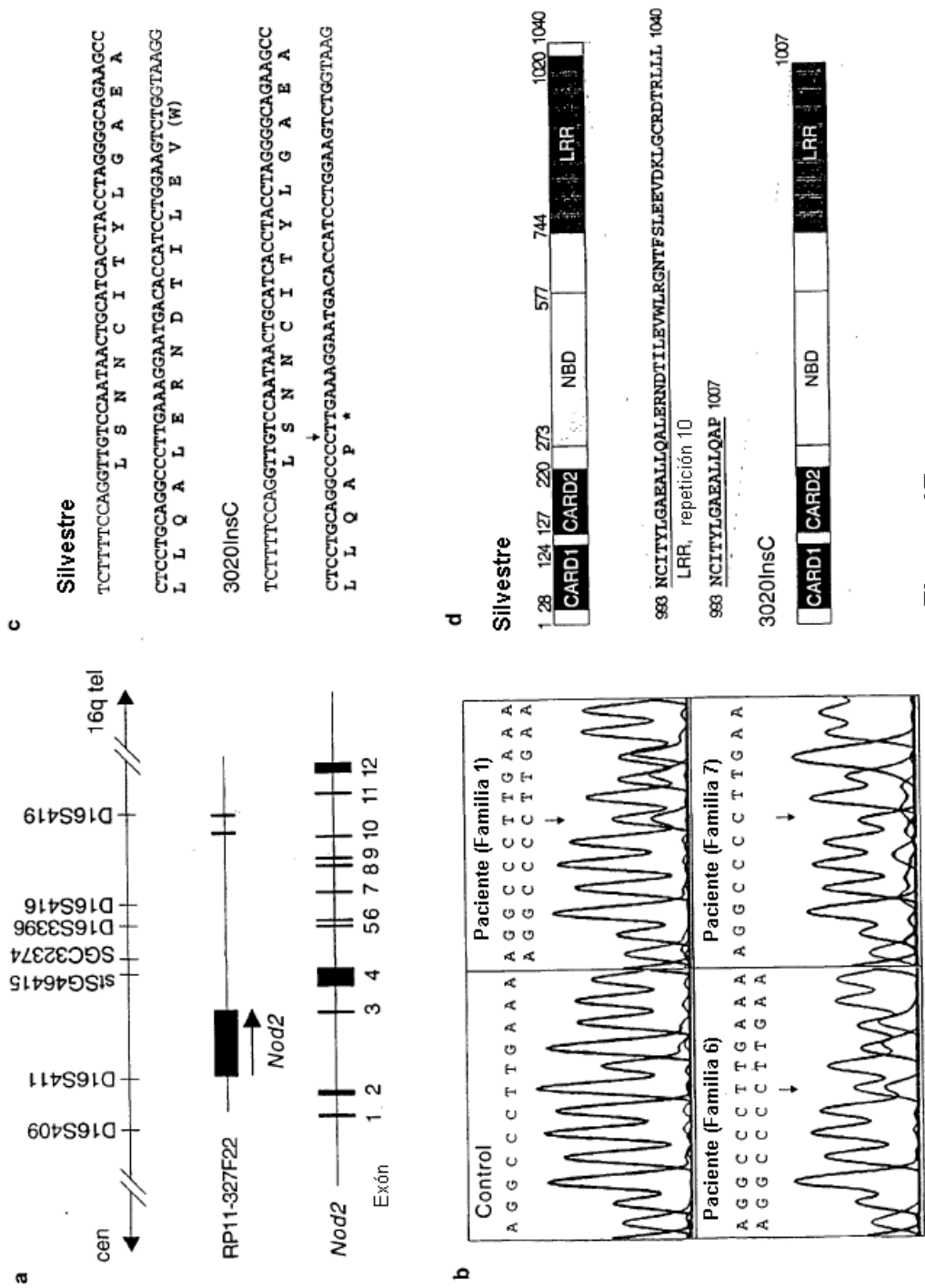


Figura 17

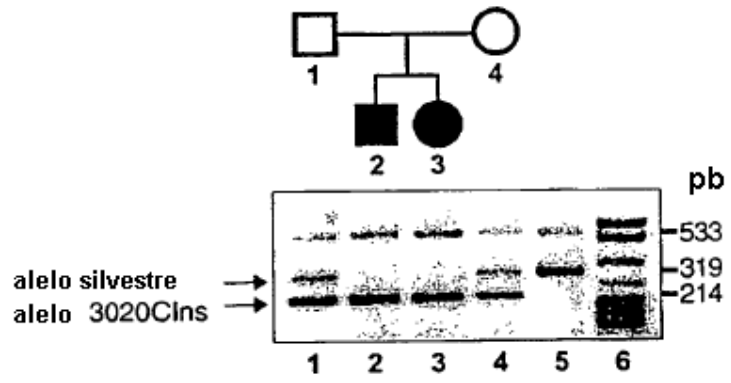


Figura 18

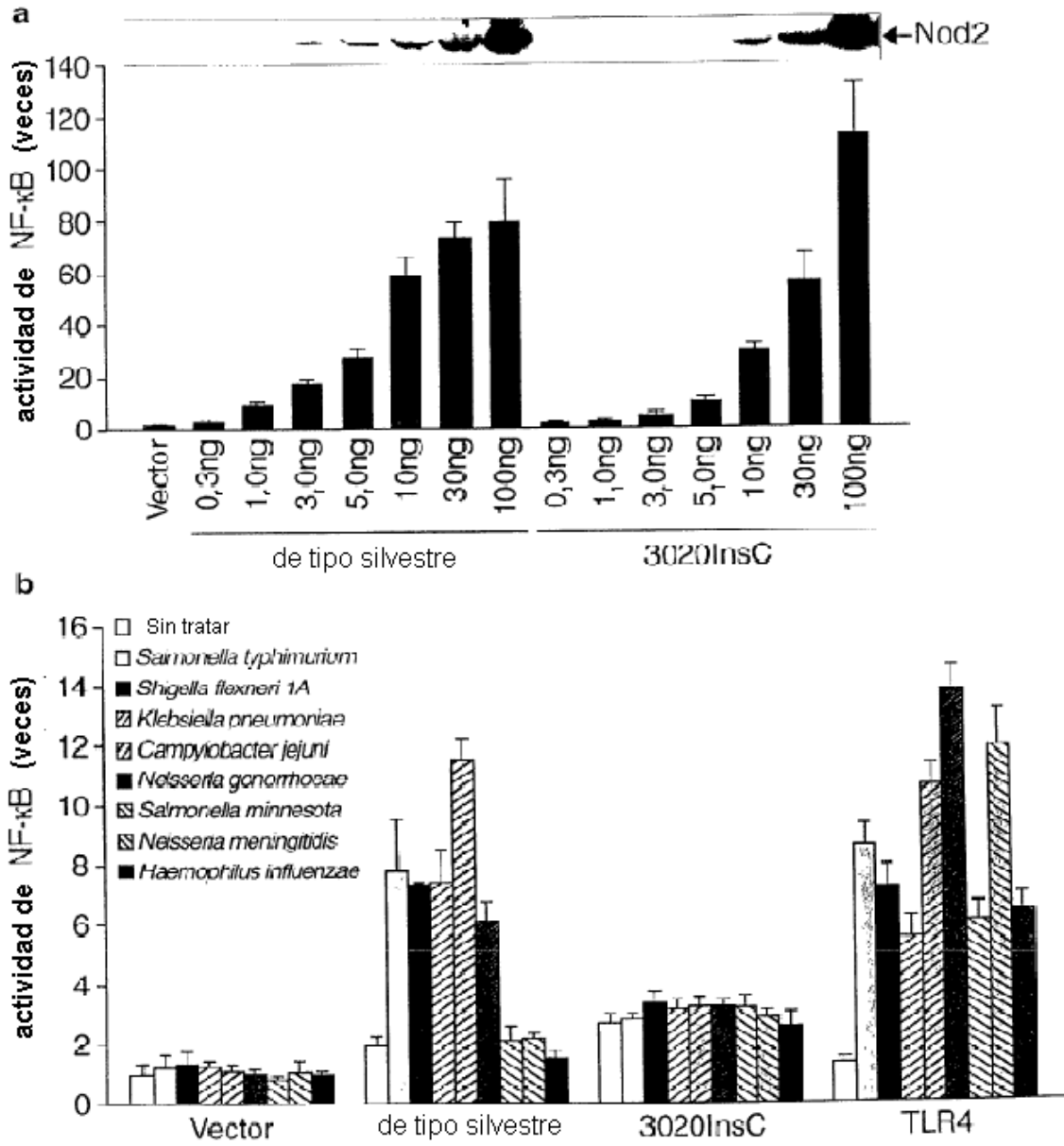


Figura 19

Figura 20

SEC ID N° 53

atgggggaagaggggtgggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctgggtcgagctgctgggtctcagggtcctcggaggcttcg
agagtgtcctggactggctgctgtcctgggaggtcctcctcctgggaggactacgagggttccacctcctgggaccagct
ctctcccacttggccagggcctctcggacacogtctggaataagggtacttgggctgtcagaagctcatcgcggctgc
ccaagaagccaggccgacagccagtcaccaagctgcatggctgctgggacccccactcgctccaccagccgagacc
tgcagagtcccgccagccatgtcaggaggtccacagccatgtggagaacatgctggacctggcaaggagcgggggt
ttcgtcagccagatgaaatggtgatgaaatcaggttgcagatcttcacaccgtccagagggcaagaaggctgcttgact
tgccacgggtgaaagcgaatggattggctgctcctctcaacaatggttcaggaattaccagtcctcctcagtaacct
tggaagctgcccacatgcaagaagtatatggccaagctgaggaccacgggtgctgctcagctcctcctcagtaacct
gatggagcagagacgctcctcctggaggacatatacacagagaatgtcctggagggtcgggagatgtgggcatggctgg
acccccgcagaagagccagccaccctgggctcggagagctctcagcaccctggccaccctcaatgacgatgggaca
ctgtgctggtgggtgggtgagggggcagtgggcaagagcacgctcctgcagcggctgcaacttgctgtgggtgcagggcaa
gacttcagggaattctctctgtcttcccatcagctgcccagctgcagtgcatggcacaaccactcctctgtgctggac
tctactcttgagcactgctgtggcctgatgtgggtcaagaagacatctccagttactccttgaccaccctgaccgtg
tctgttaacctttgatggctttgacagttcaagttcaggttcacggatcgtgaaagccactgctccccgaccgacccc
acctctgtccagaccctgctctcaacctctcgcagggaacctgctgaagaatgcccgaagggtggtagaccagccgtcc
ggcgcctgtgctgggcttctcaggaagtacatccgcaccaggttcaacctcaagggtctctcgaacagggcatcgagc
tgtacctgaggaagccacatcagcccggtggcggaccgctcatccgctgctccaagagaacctcagccctgac
ggtttgtagccacctgctctctcctatggatgggtgctcaaatgccaaccaggaactgtgctgcaaggaggggggtcccc
aaagaccactacagatagtacctgctgatctcgcagattttctgctgcatgccacccccccagactcagcttccaag
gtcgggaccagctctctcggggccgctccccacctcctgcacctgggagactggctctgtggggcctgggcatg
tgctgctacgtgtctcagcccagcagctccagggcagcacaggtcagccctgatgacattctcttggctcctcgtgctg
tgccaagggtgctgctgcccagggagtagcggcgcctggaaatccttcacatcacttccagtgctctcttgcggctct
acctggcactcagtgctgatgtgccaccagcttgctcagacacctctcaatgtggcagggccaggcaactcaccaatg
gccaggtcctgcccagatgtgcatccaggcctcggagggaaaggacagcagcgtggcagctttgctgcaagaaggccga
ggcgcacaaccttcagatcacagcagccttctggcagggctgtgttccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggcctgctcggcgccaggcctgtgccgctgggtgctggcccagcctccgcaagcactccac
tccatcccgcagctgcaccgggtgaggccaagagcgtgcatgccatgccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcacgtggcctgaaatgtgggcacctcaagtgacattttgcagtg
tggggcccactgagtgctgctcctggcctttgtgctgcagcacctccggcggccgctggcctgagctggactacaac
tctgtgggtgacattggcgtggagcagctgctgctccttgggtgctgcaaggctctgtatttgccgataacaatat
ctcagaccgagcactcgcaggctcattgaaatgtgctctcactgagcaattgcagaagttagctctattcaacaaca
aattgactgacggctgtgcacactccatggctaaagctccttgcatgcaggcagaactcttggcatgaggctggggaat
aactacatcactgcccggggagcccaagtgtgctggcggaggggtccgaggcaaacctccttgagttcctgggattctg
gggcaacagagtggtgacgagggggccaggccctggctgaaagcttgggtgatcaccagagcttgaggtggctcagcc
tgggtgggaaacaacattggcagtggtgggtgcccagccttggcactgatgctggcaagaacgtcatgctagaagaactc
tgctgggaggaaccatctccagatgaaggtgatgtctctcgcagaaggactgaagaaaaatccaagttgaaaaat
cctgaagttgtccaataactgcatcacctacctaggggcagaagcctcctgcaggcccttgaagggaatgaccacatcc
tggaaagtctggctccgagggaaacactttctctctagaggaggttgacaagctcggctgcagggaaccagactcttgctt
tga

Figura 21

SEC ID N° 54

atgggggaagaggggtggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgtga
aatgtgctcgagggaggttttcaggcacagaggagccagctggctcgagctgctggtctcagggctccctggaaggcttcg
agagtgtcctggactggctgctgctcctgggaggtcctcctcctgggaggactacgagggcttccacctcctggccagcct
ctctcccacttggccagggccttctggacaccgtctggaataaagggtacttgggctgtcagaagctcatcgcggtgc
ccaagaagcccaggccagccagc
tgagagtccaccggccagccattgtcaggaggctccacagccatgtggagaacatgtggacctggcatgggagcggggt
ttcgtcagccagatgaaatgtgagaaatcaggttgcagatcttcacaccgtcccagagggaagaaggctgcttgatct
tgccacggtgaaagcgaatggattgctgctcctcctctcaaacatggtcaggaattaccagctccatggccctgctt
tggagctgccacatgcaagaagtatatggccaagctgaggaccacgggtgctgctcagctcctcctcagctacat
gatggagcagagacgctctgctggaggacatatacacagagaatgtcctggaggctgggagatgtgggcatggctgg
atccccgcagaagagccagccaccctgggctggaggagctcttcagcaacctggccaacctcaatgacgatgggaca
ctgtgctgggtgggtgggtagggcgggagtggaagagcagctcctgcagcggctgcaacttgctgtagggctgcagggcaa
ggctccaggaaattctctttgtcttccattcagctgcccggcagctgcaatggccaaaacctctctgtgaggac
tctactcttgagcactgctgttggcctgatgttggcaagaagacatcttccagttactccttgaccacctgaccgtg
tccgttaacctttgatggctttgacagagttcaagttcaggttcacggatcgtgaacgccactgctccccgacggcccc
acctctgtccagaccctgctctcaaccttctgcagggaacctgctgagaatgcccgaagggtggtagaccagcctcc
ggcctgctgtgctggcgttctcaggaagtacatccgcaccgagttcaacctcaagggttctctgaaacagggcatcgagc
tgtacctgaggaagcccatcatgagccggggtggcggaccgctcatccgctgctccaagagacctcagccctgcac
ggtttgtgccacctgctgcttctcatggatggtgtccaaatgccaccaggaactgttgctgcaggagggggggtcccc
aaagaccactacagatagtacctgctgatctgcagcatttctgctgcatgccacccccagactcagcttcccag
gtctgggacccagctcttctcgggqccqctccccacctcctqacactqggcagactggctctgtgggqccctggqcatg
tgctgctacgtgttctcagccagcagctccagggcagcaggtcagccctgatgacatttctctggcttctctggtgog
tgccaaagggtgctgctgcccaggagtagggcgcctggaaattccttcacatcacttccagtgcttcttggcgcgttct
acctggcactcagtgctgatgtgccaccagctttgctcagacacctctcaattgtggcaggccaggcaactcaccaatg
gccaggctcctgcccacgatgtgcatccaggcctcggagggaaggacagcagcgtggcagctttgctgcagaaggccga
gcccacaaaccttcagatcacagcagccttctcggcagggtgttgtccgggagcactggggcctgctgggtgagtgcc
agacatctgagaaggccctgctccggcggcaggcctgtgcccgtgggtgtctggcccgcagcctccgcaagcacttccac
tccatcccggccagctgcaccgggtgaggccaagagcgtgcatgccaatggccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcaactggcctgaatgtgggacacctcaagttgacatttgcagtg
tgggcccactgagtgctgcccctggcctttgtgctgcagcacctccggcggcctggcctgagctggactacaac
tctgtgggtgacatggcgtggagcagctgctgcttgccttgggtgcttgcaaggctctgtatttggcggataacaatat
ctcagaccgaggcatctgcaagctcatgaaatgtgctcttccactgcagcaatgcaagaagttagctctattcaacaaca
aatgactgacggctgtgcacactccatggctaagctccttgcatgcaggcagaacttcttggcattgaggctggggaat
aactacatcactgcccggggagcccagtgctggccgaggggctccgaggcaacacctccttgcagttcctgggattctg
gggcaacagagtgggtgacagggggcccaaggcctggctgaaaccttgggtgatcaccagagcttgggtggctcagcc
tgggtggggaacaacatggcagtggtgggtgcccagccttggcactgatgctggcaagaacgtcatgctagaagaactc
tgcttggaggagaacctctcaggatgaagggtgatgttctctcgcagaaggactgaagaaaaattcaagttgaaaaat
cctgaagttgtccaataactgcatcacctacctaggggagaagccctcctgcaggcccttgaaaggaatgacaccatc
ctggaagtctggctcggagggaaccttctctctagaggagggttgacaagctcggctgcagggaaccagactcttgcct
ttga

Figura 22

SEC ID N° 55

MGEEGGSASH DEEERASVLL GHSPGCEMCS QEAFQAQRSQ LVELLVSGSL EGFESVLDWL
 60
 LSWEVLSWED YEGFHLLGQP LSHLARRLLD TVWNKGTWAC QKLIAAAQEA QADSQSPKLH
 120
 GCWDPHSLHP ARDLQSHRPA IVRRLHSHVE NMLDLAWERG FVSQYECDEI RLPIFTPSQR
 180
 ARRLDLATV KANGLAAFL L QHVQELPVPL ALPLEAATCK KYMAKLRRTV SAQSRLFSTY
 240
 DGAETLCLED IYTENVLEW ADVGMAGSPQ KSPATLGLLE LFSTPGHLND DADTVLVVGE
 300
 AGSGKSTLLQ RLHLLWAAGQ DFQEFLLVFP FSCRQLQCMA KPLSVRTLLF EHCCWPDVGO
 360
 EDIFQLLDH PDRVLLTFDG FDEFKFRFTD RERHCSPTDP TSVQTLLEFNL LQGNLLKNAR
 420
 KVVTSRPAAV SAFLRKYIRT EFNKGFSEQ GIELYLRKRH HEPGVADRLI RLLQETSALH
 480
 GLCHLPVFSW MVSKCHQELL LQEGGSPKTT TDMYLLILQH FLLHATPPDS ASQGLGPSLL
 540
 RGRLPTLLHL GRLALWGLGM CCYVFSQQQL QAAQVSPDDI SLGFLVRAKG VVPGSTAPLE
 600
 FLHITFQCFE AAFYLALSAD VPPALLRHLF NCGRPGNSPM ARLLPTMCIQ ASEGKDSSVA
 660
 ALLQKAEPHN LQITAAFLAG LLSREHWGLL AECQTSEKAL LRRQACARWC LARSLRKHFH
 720
 SIPPAAPGEA KSVHAMPFI WLIRSLYEMQ EERLARKAAR GLNVGHLKLT FCSVGPTECA
 780
 ALAFVLQHLR RPVALQLDYN SVGDIGVEQL LPCLGVCKAL YLRDNNISDR GICKLIECAL
 840
 HCEQLQKLAL FNNKLTGCA HSMAKLLACR QNFLALRLGN NYITAAGAQV LAEGLRGNTS
 900
 LQFLGFWGNR VGDEGAQALA EALGDHQSLR WLSLVGNNIG SVGAQALALM LAKNVMLEEL
 960
 CLEENHLQDE GVCSLAEGLK KNSSLKILKL SNNCITYLGA EALLQAP*
 1007

Figura 23

SEC ID N°56

atgggggaagaggggtgggtcagcctctcagcatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgtga
aatgtgctcgcaggaggcttttcaggcacaagaggagccagctgggtcgagctgctgggtctcagggccctgggaaggcttcg
agagtgctcctggactggctgctgtcctgggaggtcctcctcctgggaggactacgagggcttccacctcctgggocagcct
ctctcccacttggccaggcgcttctggacaccgtctggaataagggtacttgggcctgtcagaagctcatcgcggctgc
ccaagaagcccaggccagcagccagtcctcccaagctgcatggctgctgggacccccactcgtccaccagcccgagacc
tgagagtcaccggccagccattgtcaggaggtccacagccatgtggagaacatgctggacctggcatgggagcgggggt
ttcgtcagccagatgaaatgtgatgaaatcaggttgccgatcttcacaccgtcccagaggccaagaaggctgcttgatct
tggccaggtgaaagcaatggattggctgcttcttctacaacatgttcaggaattaccagtcctattggccctgctct
tggaaagctgccacatgcaagaagtatatggccaagctgaggaccacgggtgtctgctcagctcctcctcagtacctat
gatggagcagagacgctctgctcctggaggacataacacagagaatgtcctggaggtctgggcagatgtgggcatggctgg
acccccgcagaagagcccagccaccctgggctggaggagctcttcagcaccctggccacctcaatgacgatgcccgaca
ctgtgctgggtgggtgggtgagggcgggagtgggcaagagcagcctcctgcagcggctgcacttggctggtggctgcagggcaa
gacttccaggaattctctcttgtcttccattcagctgcccggcagctgcatggccaaaaccactctctgtgggac
tctactctttagcactgctgttggcctgatgttggccaagaagacatcttccagttactccttgaccaccctgaccgtg
tctctgttaacctttagtggctttagcaggttcaagttcaggttccaggtatcgtgaaacgacctgctcccgaaccgacccc
acctctgtccagaccctgctcttcaaccttctgcagggcaacctgctgaaagaatgcccgaagggtggtagccagcctcc
ggccgctgtgtcggcgttctcaggaagtacatccgcaccaggttcaacctcaagggttctctgaaacagggcatcgagc
tgtacctgaggaaagcctcatgagccggggtggcgggacgcctcatcgcctgctccaagagacctcagccctgcac
ggtttgtgccacctgctgtcttctcatggatgggtgtccaaatgccaccaggaactgttgcagggaggggggtcccc
aaagaccactacagatatgtacctgctgattctgcagcatttctgctgcatgccacccccccagactcagcttcccaag
gtctgggaccagctctctctcggggccgctcctccaccctcctgcacctgggcagactggctctgtggggcctgggcatg
tgcctgctacgtgtctcagcccagcagctccagggcagcaggtcagcctgatgacatttctcttggcttctctgggtg
tgccaaaagggtgtcgtgccaggagtagcggcggcctggaattccttccacatcaacttccagtgcttcttggcggcttct
acctggcactcagtgctgatgtgccaccagcttggctcagacaccttccaattgtggcaggccaggcaactcaccatg
gccaggctcctgccacagatgtgcatccaggcctcggagggaaggacagcagcgtggcagcttggctgcagaaggccga
gccgcacaaccttcagatcacagcagccttctggcagggtgtgttgcgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctctggcggccagcctgtgcccgtgggtctggcccgcagcctccgcaagcacttccac
tccatcccgcagctgcaccgggtgaggccaagagcgtgcatgccatgccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcacgtggcctgaatgttgggacacctcaagttagacatttgcagtg
tgggcccactgagtgctgctgcccctggccttggctgagcagcactccggcggccctggcccctgagctggactacaac
tctgtgggtgacattggcgtggagcagctgctgcttgccttgggtgtctgcaaggctctgtatttggcggataacaatat
ctcagaccgagcatctgcaagctcattgaaatgtgctcttcaactgcagcaaatgcaagaagttagctctattcaacaaca
aattgactgacggctgtgcacactccatggctaaagctccttgcagcagggcagaacttctggcattgaggctggggaat
aactacatcactgcccgggagcccaagtgtgctggcggagggtcctcagggcaaacacctccttgcagttcctgggattctg
gggcaacagagtggtgacgagggggccaggccctggctgaaagccttgggtgatcaccagagcttgagggtgctcagcc
tgggtggggaacaacatggcagtggtgggtgcccaagccttggcaetgatgctggcaagaacgtcatgctagaagaactc
tgcttggaggagaacctctccaggatgaaggtgtatgttctctcgcagaaggactgaaagaaaaatcaagttgaaaat
cctgaagttgtccaataactgcatcacctacctaggggcagaagcctcctgcaggcccttgaaggaaatgacacccatcc
tggaaagtctggctccgagggaacacttctctctagaggaggttgacaagctcggctgcaggggacaccagactcttctct
tga

Figura 24

SEC ID N° 57

MGEEGGSASHDEEERASVLLGHS PGCEMCSQEAFQAQRSQVLVELLVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLI AAAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRR LSHSHVENMLDLAWERG
FVSQYECDEIRLP IFTPSQRARRLLDLATVKANGLAAFL LQH VQELPVPLALPLEAATCKKYM AKLR TT VSAQSRFLSTY
DGAETLCLEDIY TENVLEVWADVGMAGPPQKSPATL GLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEF LFVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVQEDIFQLLLDHPDRVLLTFDGFDFKFRFTDRERHCSPTDP
TSVQTL LFNLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNLKGFSEQGI ELYLRKRHHEPGVADR LIRLLQETSALH
GLCHLPVFSW MVSKCHQELLQEGGSPKTTTDMYLLI LQHFLHATPPDSASQGLGPSLLRGR LPTLLHLGRLALWGLGM
CCYV FSAQQLQAAQVSPDDISLGFVRAKGVVPGSTABLEFLHITFQCFFAAFYLALSADVPPALLRHLFNCRPGNSPM
ARLLPTMCIQASEGKDSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLWRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGF IWLIRSLYEMQEERLARKAARGLNVGHLKLTFC SVGPT ECAALAFVLQHLRRFPVALQLDYN
SVGDIGVEQLLPCLGVCALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLT DGCASHMAKLLACRQNFALRLGN
NYITTAAGAQLAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSLRWLSLVGN NIGSVCAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSLKIKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEV DKLGCRDTRLLL
*

Figura 25

SEC ID N° 58

atgggggaagaggggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacatctcctgggttgtga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctgggtcgagctgctgggtcctcagggtccctggaaggctcgcg
agagtgtcctggactggctgctgtcctgggaggtcctcctcctgggaggactacgagggttccacctcctgggccagcct
ctctcccacttggccaggcgccttctggacacgctcggaaataagggctacttgggctgtcagaagctcatcgcggctgc
caaagaagcccaggccgacagccagtcccccaaagtcgcatggctgctgggacccccactcgcctccaccagcccagacc
tgcagagtcaccggccagccattgtcaggaggctccacagccatgtggagaacatgctggacctggccatgggagcgggggt
ttcgtcagccagatgaatgtgatgaaatcaggttgcagatcttcacaccgtcccagagggaagaaggctgcttgatct
tgccacggtgaaagcgaatggattggctgcctccttctacaacatgctcaggaattaccagtccattggccctgcctt
tggaaagctgcccacatgcaagaagtataggccaagctgaggaccacgggtgtcgtcagtcagtcgctcctcagtaacctat
gatggagcagagacgctctgcctggaggacatatacacagagaatgtcctggaggctcgggcagatgtgggcatggctgg
acccccgagaagagcccagccacctgggctggaggagctcttcagcaccctggccacctcaatgacgatgcccagaca
ctgtgctgggtgggtgggtgaggcgggagctggcaagagcagcctcctcagcggctgcaactgctgtgggtgagggcaa
gacttccaggaatttctcttgccttccattcagctgcccagctgcagtgcatggccaaaccactctctgtgaggac
tctactctttgagcactgctgttggcctgatgttggccaagaagacatcttccagtaactccttgaccacctgacgctg
tctgttaacctttgatggctttgacgagttcaagttcaggttcacggatcgtgaacgccactgctccccagccgacccc
acctctgtccagaccctgctcttcaaccttctgcagggaacctgctgaagaatgcccgaaggtggtagaccagccgctcc
ggcctgctgtcggcgttctcaggaagtacatccgaccgagttcaacctcaagggttctctgaacagggtcatcgagc
tgtacctgaggaagcgcctcatgagcccgggtggcggaccgctcatccgctgctccaagagacctcagccctgcac
ggtttgtgccacctgctctctcctcatggatgggtgtccaaatgccaccaggaactgttgctgcaggaggggggtcccc
aaagaccactacagatagtactyctgatcttgcagcatttctgctgcatgcccacccccagactcagcttcccaag
gtctgggacccagctcttctcggggccgctccccacctcctgcaactgggcagactggctctgtggggcctgggcatg
tgctgctacgtgttctcagcccagcagctccaggcagcacaggtcagccctgatgacatttctcttggttctcgtggtg
tgccaaaaggtgtcgtgcccaggagtaaggcgcctcctggaattccttccacatcacttccagtgctctcttgcccgttct
acctggcactcagtgctgatgtgccaccagcttgcctcagacacctctcaattgtggcaggccaggcaactcaccaatg
gccaggctcctgcccagatgtgcatccaggcctcggagggaagggaagcagcagcgtggcagctttgctgcagaaggccga
ggccgacaaacctcagatcacagcagccttctggcagggtgtgtgtcccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctccggcgcagggcctgtgcccgtggtgtctggcccagcctccgcaagcacttccac
tccatcccgcagctgcaccgggtgaggccaagagcgtgcatgccaatgcccgggttcacttggtcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcagtgggcctgaaatgtgggcacctcaagttgacattttgcagtg
tgggcccactgagtgctgcccctggccttctgtgctgcagcaactccggcggccctggcctgcagctggactacaac
tctgtgggtgacattggcgtggagcagctgctgccttgccttgggtgtctgcaaggctctgtatttgcgcgataacaat
ctcagaccgaggcatctgcaagctcattgaaatgtgctcttcaactgcgagcaatgcaagaagttagctctatcaacaaca
aattgactgacggctgtgcaactccatggctaagctccttgcctgcaaggcagaactcttggcatggagctggggaat
aactacatcactgcccgggagcccaagtgctggccgaggggtcctcagggcaacacctccttgcagttcctgggattctg
ggcgaacagagtggtgacgagggggcccaggccctggctgaaagccttgggtgatcaccagagcttgaggtggctcagcc
tgggtggggaacaaacattggcagtggtgggtgcccagccttggcactgatgctggcaagaacgctcatgctagaagaactc
tgctggaggagaacctctccaggatgaaggtgtatgttctcgcagaaggactgaagaaaaattcaagtttgaat
cctgaagtgtccaataactgcatcactacctaggggcagaagccctcctgcaggccttgaaaggaatgacaccatcc
tggaaagtctggctccgagggaacatttctctctagaggaggttgacaagctcggctgcagggacaccagactcttgctt
tga

SNP	Alelo	Secuencia de SNP*	Polimorfismo AA	Localización AA	Localización ADnc
SNP4	1	GGCAGATGTGGGCATGGCTGGACCC; SEC ID N° 70	P	P 268 S	nt 802
	2	GGCAGATGTGGGCATGGCTGGAICC; SEC ID N° 71	S		
SNP20	1	AGACATCTGAGAAGGCCCTGCTCCGG; SEC ID N° 72	R	R 702 W	nt 2104
	2	AGACATCTGAGAAGGCCCTGCTCIGG SEC ID N° 73	W		
SNP7	1	CTGCAGCACCTCCGGGGCCCGTG; SEC ID N° 74	V	V 793 M	nt 2377
	2	CTGCAGCACCTCCGGGGCCCATG; SEC ID N° 75	M		
SNP18	1	TTGCAGAAAGTTAGCTCTATTCAAC; SEC ID N° 76	N	N 852 S	nt 2555
	2	TTGCAGAAAGTTAGCTCTATTCAAG; SEC ID N° 77	S		
SNP25	1	ACTGACGGCTGTGCACACTCCATG; SEC ID N° 78	M	M 863 V	nt 2587
	2	ACTGACGGCTGTGCACACTCCGTG; SEC ID N° 79	V		
SNP17	1	TGCAGTTCCTGGGATTCTGGGGC; SEC ID N° 80	G	G 908 R	nt 2722
	2	TGCAGTTCCTGGGATTCTGGGGC; SEC ID N° 81	R		
SNP23	1	CACTGATGCTGGCAAAGAACGTC; SEC ID N° 82	V	V 955 I	nt 2863
	2	CACTGATGCTGGCAAAGAACATC; SEC ID N° 83	I		
3020C Ins	1	GGGCAGAAAGCCCTCCTGCAGGCCCT; SEC ID N° 90	de tipo silvestre		nt 3020
	2	GGGCAGAAAGCCCTCCTGCAGGCCCT; SEC ID N° 91	mut.cambio de marco	Δ33	
		*base mutada subrayada			
		* Los n° de nucleótido/aminoácido designan las posiciones en Nod2a según		Ogura et al. J. Biol. Chem. 276:4812 [2001]	
		AA = aminoácido			
		SNP= polimorfismo de un solo nucleótido			

Figura 26

snp4	snp20	snp7	snp18	snp17	snp23	3020insC	comentarios
1	1	1	1	1	1	1	común
2	1	1	1	1	1	1	aproximadamente 10% frecuencia de alelo en los controles
2	2	1	1	1	1	1	aproximadamente 4% frecuencia de alelo en los controles
2	1	1	1	2	1	1	aproximadamente 1-2% frecuencia de alelo en los controles
2	1	1	1	1	1	2	aproximadamente 3-4% frecuencia de alelo en los controles
1	1	1	1	1	1	1	muy raro: 4/400 pacientes heterocigotos; en el alelo 1 de snp4
1	1	2	1	1	1	1	muy raro: brazos no funcionan
2	1	1	2	1	1	1	muy raro: brazos funcionan
1	1	1	1	1	2	1	muy raro: brazos funcionan

Figura 27

Figura 28

SEC ID N° 59

MGEEGGSASHDEERASVLLGHS PGCEMCSQEAFAQRSQVVELLVSGSLEGFESVLDWLLSWEVLSWEDYEGPHLLGQP
LSHLARRLLDTVWNKGTWACQKLI AAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKKYMAKLRRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLLELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFLFVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVCQEDIFQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEOGIELYLRKRHHEPGVADRLLIRLLOETSALH
GLCHLPVFSWVMVSKCHQELLQEGGSPKTTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLEPTLLHLGRLALWGLGM
CCYVFSQQLOAAQVSPDDISLGFLVRAKGVVPGSTAPLEFLHITFQCFFAAYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKDSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPFAAPGEAKSVHAMPGFIWLRSLYEMQEERLARKAARGLNVGHLKLTFCVSGPTECAALAFVLQHLRFPVALQLDYN
SVGDIGVEQLLPCLVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTGCAHSMAKLLACRQNFALRLGN
NYITAACAQVLAEGLRGNTSLQFLGFWRNRVGEDEGAQALAEALGDHQSLRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSLKILKLSNNCITYLGAELLOALERNDTILEVWLRGNTFSLEEVDKLGCRRDTRLLL
*

Figura 29

SEC ID N° 60

atgggggaagaggggtgggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgtga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctggctcgagctgctggctcagggctccctggaaaggcttcg
agagtgtcctggactggctgctgtcctgggaggtcctcctcctgggaggaactacgagggcttccacctcctgggcccagcct
ctctccacttggccaggcgcctctggacacgctctggaataagggtacttgggctgtcagaagctcatcgcggctgc
ccaagaagcccaggccgacagccagctcccccaagctgcatggctgctgggacccccactcgctccaccagcccgagacc
tgcagagtaccggccagccatgtcaggaggctcccacagccatgtggagaaca t gctggacctggcatgggagcgggggt
ttcgtcagccagatgaaatgtgatgaaatcaggttgccgatcttcacaccgtcccagagggcaagaaggctgcttgatct
tgccacgggtgaaagcgaatggatggctgccttctctacaacatggttcaggaaat taccagctccat tggccctgcctt
tggaaagctgccacatgcaagaagtata tggccaagctgaggaccagggctgctgctcagctcctcctcagtaacctat
gatggagcagagacgctctgctcggaggacatatacacagagaatgtcctggaggctcgggcagatgtgggcatggctgg
atccccgcagaagagcccagccacctgggctggaggagctctcagcaccctggccacctcaatgacgatgcggaca
ctgtgctgggtgggtgggtgagggcgggagctggcaagagcagcctcctgcagcggctgcacttgctgtgggctgcagggcaa
gacttccaggaatttctcttggcttcccattcagctgcccggcagctgcagtgcatggccaaacctctctgtgcccagc
tctactcttgagcactgctgttggcctgatgttggtaagaagacatcttccagtaactccttgaccacctgacogtg
tctgttaacctt t gatggctt gacaggtt caagttcaggttcaaggatcgtgaacgccactgctccccgaccgacccc
acctctgtccagaccctgctcttcaaccttctgcagggcaacctgctgaagaatgcccgcgaaggtgggtgaccagccctcc
ggccgctgtgtcggcgttctcaggaagtacatccgcaccaggttcaacctcaagggttctctgaacagggcatcgagc
tgaacctgaggaagcgcacatcatgagcccgggtggcgaccgctcctcctcctgctccaagagacctcagccctgcac
gggttgtgcccacctgcctgtctctcatggatgggtgtccaaatgccaccaggaactgttgcctgcaggaggggggtcccc
aaagaccactacagatattgtaacctgctgattctgcagcatcttctgctgcatgcccacccccagactcagcttccccag
gtctgggaccctcagctcttctcggggccgcttccccacctcctgcacctgggcagactggctctgtggggcctgggcatg
tgctgctacgtgttctcagcccagcagctccaggcagcacaggtcagccctgatgacatttctctggcttccctgggtgctg
tgccaaaggtgtcgtgcccaggagtagcggcggcctggaaatccttcacatcacttccagtgctctcttgccgcttct
acctggcactcagtgctgatgtgccaccagctt t gctcagacaccttcaat t g t g g c a g g c c a g g c a a c t c a c c a a t g
gccaggtcctgcccacgatgtgcatccaggcctcggagggaaaaggacagcagcgtggcagctt t g c t g c a g a a g g c c g a
gcccacaaaccttcagatcacagcagccttctggcagggtgtgtgccgggagcactggggcctgctggctgagtgcc
agacatctgagaagggcctgctccggcgcagggcctgtgcccgtgggtgtctggcccgcagcctccgcaagcacttccac
tccatccccagctgcaccgggtgagggcaagagcgtgcatgccaatgcccgggttcatctggctcatccggagcctgta
cgagatgcaggagggagcggctggctcggaaaggctgcacgtggcctgaaatgttgggcaacctcaagttagcatttgcagtg
tgggccccactgagtgctgcccctggcctt t g t g c t g c a g c a c c t c c g g c g c c g t g g c c c t g c a g c t g g a c t a c a a c
tctgtgggtgacatggcgtggagcagctgctgccttgccttgggtgtctgcaaggctctgtatttgccgataaacaata
ctcagaccgaggtcatctgcaagctcat t g a a t g t g c t c t t c a c t g c g a g c a a t t g c a g a a g t t a g c t c a t t c a a c a c a
aat t g a c t g a c g g c t g t g c a c a c t c c a t g g c t a a g c t c e t t g c a t g c a g g c a g a a c t t e t t g g c a t t g a g g c t g g g g a a t
aactacatcactgcccgggagcccaagtgtggccgaggggtccgagggcaacacctcttgcagttcctgggattctg
gggcaacagagtggtgacgagggggccaggccctggctgaagccttgggtgatcaccagagcttgaggtggctcagcc
tgggtggggaacaacattggcagtggtgggtgccccagccttggcactgatgctggcaagaacgtcatgctagaagaactc
tgctggaggagaacctctccaggatgaaggtgatgttctctcgcagaaggactgaagaaaaat t c a a g t t g a a a a t
cctgaagt t g t c a a t a a c t g c a t c a c c t a c c t a g g g g c a g a a g c c c t c t g c a g g c c c t g a a a g g a a t g a c a c c a t c c
tggaaagtctggctccgagggaaaccttctctctagaggaggttgacaagctcggctgcagggacaccagactcttgcct
tga

Figura 30

SEC ID N° 61

MGEEGGSASHDEEERASVLLGHSPGCEMCSQEAFAQARSQVLELLVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKYMAKLRRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGSPQKSPATLGLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFVFPFSCRQLQCMAPLSVRTLLFEHCWPDVQEDIFQLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLFNLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEQGIELYLRKRHHEPGVADRILRLLQETSALH
GLCHLPVFSWMVSKCHQELLQEGGSPKTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLLPTLLHLGRLLWGLGM
CCYVFAQQLQAAQVSPDDISLGFVRAKGVVPGSTAPLEFLHITFCFFAAFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKSSVAALLOKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGFIWLRSLYEMQEERLARKAARGLNVGHLKLTFCVSGPTECAALAFVLQHLRRPVALQLDYN
SVGDIGVEQLLPCLVCKALYLRDNNISDRGICKLIECALHCEQLQKALFNKLTGCAHSMKLLACRQNFALRLGN
NYITAGAQLAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSRLWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSLKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRTRELLL
*

Figura 31

SEC ID N° 62

atgggggaagaggggtgggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgtga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctggtcgagctgctggctcagggtccctggaaggcttcg
agagtgtcctggactggctgctgctcctgggaggtcctcctcctgggaggactacgagggctccacctcctgggcccagcct
ctctcccacttggccaggcgccttctggacaccgtctggaataagggctactgggctctcagaagctcatcgcggctgc
ccaagaagcccaggccgacagccagtcccccaaagctgcatggctgctgggaacccactcgtccaccagcccgagacc
tgcagagtcaacggccagccattgtcaggaggctccacagccatgtggagaacatgctggacctggcatgggagcgggggt
ttcgtcagccagtatgaatgtgatgaaatcaggttgcogattctcacaccgtcccagagggaagaaggctgcttgatct
tgccacgggtgaaagcgaatggatggctgcctccttctacaacatgttcaggaattaccagtcctatggccctgcctt
tggaaagctgccacatgcaagaagtatatggccaagctgaggaccacgggtgtctgctcagtcctgctcctcagtaacctat
gatggagcagagagcctctgctcctggaggacatatacacagagaatgtcctggaggctcgggcagatgtgggcatggctgg
acccccgcagaagagcccagccaccctgggctggaggagctcttcagcaccctcggccacctcaatgacgatgggaca
cgtgtcctgggtgggtgaggcgggagtggaagagcagctcctgcagcggctgcacttgcctggtggctgcaggggcaa
gactccagggaatttctcttgtcttccatcagctgcccagctgcagtgcatggccaaaccactctctgtgcccagc
tctactcttggagcactgctgttggcctgatgttggctcaagaagacatctccagttactccttgaccacctgaccgtg
tctgttaacctttgatggctttgacaggttcaagttcaggttccaggatcgtgaaagccactgctcccccgaacgacccc
acctctgtccagacctgctcttcaaccttctgcagggcaacctgctgaagaatgcccgaagggtggtagaccagcgtcc
ggccgctgtgtcggcgttctcaggaagtacatccgcaccgagttcaacctcaagggtctctgaaacagggcatcgagc
tgtacctgaggaagcgcacatcatgagccgggtggcggaccgctctccgctgctccaagagacctcagccctgcac
ggtttgtgccaacctgctctctctcatggatgggtgtccaaatgcccaccaggaactgttgcctgcaggaggggggtccc
aaagaccactacagatatgtacctgctgattctgcagcatttctgctgcatgcccacccccagactcagcttcccag
gtctgggaccagctctctctcggggcgcctccccacctctcgcacctgggcagactggctctgtggggcctgggcatg
tgctgctacgtgttctcagcccagcagctccaggcagcagaggtcagccctgatgacatttctcttggcttctgggtgcg
tgccaaaggtgtcgtgcccaggagtagcggcgcctggaaatcctctcacatcacttccagtgctctcttgcgcgctct
acctggcactcagtgctgatgtgccaccagcttgcctcagacacctcttcaatgtggcaggccaggcaactcaccatg
gcccaggtctctgcccacgatgtgcatccaggcctcggagggaaggacagcagcgtggcagcttgcctgcagaaggccga
gcccacacaccttcagatcacagcagcctcctggcagggtgtgtcccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctccggcgcagggcctgtgcccgtgggtgtctggccgcagcctcccgaagcactccac
tccatcccgcagctgcacgggtgagggcaagagcgtgcatgcatgcccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggtgcacgtggcctgaaatgtgggcacctcaagttgacatttgcagtg
tgggccccactgagtgctgcccctggccttggctgctcagcacctccggcggccatggccctgcagctggactacaac
tctgtgggtgacatggcgtggagcagctgctgcccctggctgggtgtctgcaaggctctgtatttggcggataaacaat
ctcagaccagggcatctgcaagctcatgaaatgtgctctcactgagagcaatgcagaagttagctctattcaacaaca
aattgactgacggctgtgcacactccatggctaagctccttgcatgagggcagaacttcttggcattgaggctggggaat
aactacatcactgcccggggagccccagtgctggccgaggggctccgaggcaacacctctctgcagttcctgggattctg
gggcaacagagtggtgacgagggggcccaggccctggctgaaagccttgggtgatalcaccagagcttgagggtggctcagcc
tgggtggggaacaacattggcagtggtgggtgcccagccttggcactgatgctggcaaaagagcctcatgctagaagaactc
tgccctggaggagaacctctccaggatgaaaggtgtatgttctctcgcagaaggactgaagaaaaattcaagttgaaaat
cctgaagttgtccaaactgcatcacctacctaaggggcagaagcctcctgcagggccttgaagggaatgacacccatcc
tggaaagtctggctccgagggaacacttctcctctagaggaggtgacaagctcggctgcagggaacaccagactcttgcct
tga

Figura 32

SEC ID N° 63

MGEEGGSASHDEERASVLLGHS PGCEMCSQEAFQAQRSQVLELLVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLDVTWNGTWCQKLI AAQEAQADSQSFKLHGCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKKYMALRRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFOEFLFVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVGOEDIFQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEGGIELYLKRHHPEPGVADRLIRLLQETSALH
GLCHLPVFSWVMKCHQELLLQEGGSPKTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLPTLLHLGRLALWGLGM
CCYVFAQQLQAAQVSPDDISLGFVLRAGVVPGSTAPLEFLHITFQCFFAAFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGF IWLIRSLYEMQEERLARKAARGLNVGHKLTFCSVGPTECAALAFVLQHLRPMALQLDYN
SVGDIGVEQLLPCLGVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTGCAHSMALLACRQNFALRLGN
NYITAGAQLVAEGLRGNTSLQPLGFWGNRVGDEGAQALAEALGDHQLRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSKILKLSNNCITYLGAELLLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
*

Figura 33

SEC ID N° 64

atgggggaagagggtggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgtga
aatgtgctcgcaggaggctttcaggcacagaggagccagctggtcgagctgctgggtctcagggtccctggaaggctcgcg
agagtgctcctggactggctgctgctcctgggaggtcctcctcctgggaggactacgagggtctccacctcctgggccagcct
ctctcccacttggccaggcctctcctggacaccgtctggaataaagggtacttgggctgtcagaagctcatcgcggtgc
ccaagaagcccaggccgacagccagctcccccaagctgcatggctgctgggacccccactcgctccaccagccccgagacc
tgcagagtcaccggccagccattgtcaggaggctccacagccatgtggagaaacatgctggacctggcatgggagcggggt
ttcgtcagccagatgaaatgtgatgaaatcaggttggcagatcttcacaccgtcccagagggaagaaggctgcttgatct
tgccacgggtgaaagcgaatggattggctgctcctcctacaacatgttcaggaattaccagctcccattggccctgctt
tggaaagctggccacatgcaagaagtatatggccaagctgaggaccaggtgtctgctcagctcctcctcagtacctat
gatggagcagagacgctcctgctcctggaggacatatacacagagaatgtcctggaggctgggagatgtgggcatggctgg
acccccgcagaagagcccagccaccctgggctggaggagctcttcagcaccctggccacctcaatgacgatggggaca
ctgtgctggtggtgggtgaggcggcagtgccaagagcagctcctcagcagcggctgcaacttgcctgtgggctgcagggcaa
gacttccaggaaattctcttcttctcctcctcagctgcggcagctgcagtgcatggccaaacctctctgtgctgggac
tctactcttfgagcactgctgttggcctgatgttggtaagaagacatcttccagttactccttgaccacctgaccctg
tctgttaacctttagtggtttgacagagttcaagttcaggttcacggatcgtgaacgccactgctcccagaccgaccccc
acctctgtccagaccctgctctcaaccttctcagggcaacctgctgaagaatgcccgcaagggtggtgaccagcgtccc
ggcctgctgtcggcgttctcaggaagtacatccgcaccaggttcaacctcaagggttctctgaacagggcacgcagc
tgtacctgaggaagcggccatcatgagccggggtggcggacggcctcatccgctgctccaagagacctcagccctgcac
ggttgtgcccacctgctgtctctcatggatgggtgtccaaatgccaccaggaactgttgcctgcaggaggggggtcccc
aaagaccactacagatattgtacctgctgattctgcagcattttctgctgcatgccacccccagactcagcttcccagg
gtctgggaccagctctctcctggggccgctccccacctcctgcacctgggcagactggctcctgtggggcctgggcatg
tgctgctacgtgttctcagcccagcagctccaggcagcacaggtcagccctgatgacatttctctggctcctgggtgctg
tgccaaagggtgctgctgcccaggagtagggcggcctggaaattccttcacatcacttccagtgcttcttgccgcttct
acctggcactcagtgctgatgtgcccaccagcttggctcagacacctctcaattgtggcaggccaggcaactcaccatg
gccagggtcctgcccacgatgtgcatccaggcctcggagggaaggacagcagcgtggcagcttgcctgcagaaggccga
gccgcacaaccttcagatcacagcagccttctggcagggtgtgtcccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctccggcggccaggcctgtgcccctgggtgtctggcccgcagcctccgcaagcacttccac
tccatcccgcagctgcaccgggtgaggccaagagcgtgcatgccatgccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcacgtggcctgaaatgttgggcaacctcaagttgacatttgcagtg
tgggccccactgagtgctgcccctggccttgtgctgcagcaccctccggcggccctggcctgcagctggactacaac
tctgtgggtgacattggcgtggagcagctgctgcccctggcttgggtgtctgcaaggctctgtatttgcgcgataacaat
ctcagaccgaggcatctgcaagctcatgaaatgtgctctcactgcagcaaatgcaagaagttagctctattcagcaaca
aattgactgacggctgtgcaactccatggctaaagctccttgcctgcatgcaaggcagaacttcttggcattgaggctggggaat
aactacatcactgcccggggagcccaagtgtggcggagggtccgaggcaaacacctccttgcagttcctgggattctg
gggcaacagagtggtgagcagggggccaggccctggctgaaagccttgggtgatcaccagagcttgagggtggctcagcc
tgggtggggaacaacattggcagtggtgggtgcccaagccttggcactgatgctggcaagaacgtcatgctagaagaactc
tgctggaggagaaccatctccaggatgaaaggtgtatgtctctcgcagaaggactgaagaaaaattcaagtttgaanaat
cctgaagttgtccaaataactgcatcacctacctaggggcagaagccctcctgcaggcccttgaaggaaatgacacctcc
tggaaagtctggctccgagggaaaccttctctctagaggaggttgacaagctcggctgcagggacaccagactcctgctt
tga

Figura 34

SEC ID N° 65

MGEEGGSASHDEEERASVLLGHSPGCEMCSQEAFQAQRSQLVELLVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLHGCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVP LALPLEAATCKKYMAKLRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFVFPFPPSCRQLQCMAPLSVRTLLFEHCCWPDVQEDIFQLLLDHPPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLFFNLLOGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNLKGFSEQGIELYLRKRHHEPGVADR LIRLLQETSALH
GLCHLPVFSWVSKCHQELLQEGGSPKTTTDMYLLILOHFL LHATPPDSASQGLGPSLLRGR LPTLLHLGR LALWGLGM
CCYVFSAAQLQAAQVSPDDISLGFVRAKGVVPGSTAPLEFLHITFQCFFAAFY LALSADVPPALLRHLFNGCRPGNSPM
ARLLPTMCIQASEGKDSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPCEAKSVHAMPGF IWLIRSLYEMQEERLARKAARGLNVGHLKLTFC SVGPTECAALAFVLQHLRRPVALQLDYN
SVGDIGVEQLLPCLGVCKALYL RDNNISDRGICKLIECALHCEQLQK LALFSNKLT DGCASHMAKLLACRQNF LALRLGN
NYITAAGAVLAEGLRCNTSLQFLGFWGNRVGDEGAQALAEALGDHQS LRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSLKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
*

Figura 35

SEC ID N° 66

atgggggaagaggggtggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcggacattctccgggttgga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctgggtcgagctgctgggtctcagggtccctggaaggcttcg
agagtgtcctggactggctgctgtcctgggaggtcctctcctgggaggactacgagggttccacctcctgggcccagcct
ctctcccacttggcccaggcgcctcttgacaccgtctggaataaggggtacttgggctgtcagaagctcatcgcggctgc
ccaagaagcccaggccgacagccagtcacccaagctgcatggctgctgggacccccactcgcctccacccagcccagagacc
tgcagagtccacggcccagccatttgtcaggaggctccacagccatgtggagaacatgctggacctggcatgggagcgggt
ttcgtcagccagtatgaaatgtgatgaaatcaggttgcgatctcaccagctccagaggccaaggctgctgatct
tgccacggtgaaagcgaatggatggctgctcctctctacaacatggttcaggaaattaccagctccattggccctgctt
tggaaagtgcacatgcaagaagtatatggccaagctgaggaccacgggtctgctcagctcgcctcctcagtaacctat
gatggagcagagacgctctgctggaggacatatacacagagaatgtcctggaggcttgggcagatgtgggcatggctgg
acccccgagaagagcccagccacccctgggctggaggagctctcagcacccttggccacctcaatgacgatgoggaaca
ctgtgctgggtgggtgaggcgggagtggaagagcagctcctgcagcggctgcaacttgcctgtgggtgagggcaaa
gacttccaggaatttctcttctgtcttccattcagctgcccggcagctgcaatggccaaaccactctctgtgcccga
tctactctttgagcactgctgttggcctgatgttggtaagaagacatctccagttactccttgaccacctgaccgtg
tctgttaaccttctgatggcttggcagagttcaagttcaggttccaggatcgtgaacgccactgctccccgaccgacccc
acctctgtccagaccctgctcttcaaccttctcaggggcaacctgctgaagaatgcccgaaggtgggtgaccagcctcc
ggcgcctgtgtcggcctcctcaggaagtacatccgacccagttcaacctcaagggttctctgaaacagggcatcgagc
tgtacctgaggaagcccatcatgagccccgggtggcgggacggcctcatccgctgctccaagagacctcagcctgcac
ggtttgtgccacctgctgctctctcatggatgggtgtccaaatgcccaccaggaactgttgcctgaggggggggtcccc
aaagaccatcacagatagtacctgctgatctgcaagcatttctgctgcatgcccacccccagactcagcttcccag
gtctgggacccagctctctcggggccgctccccacctcctgcacctgggcagactggctctgtggggcctgggcatg
tgctgctacgtgtctcagcccagcagctccaggcagcacaggtcagccctgatgacatttctctgggttctcgggtgctg
tgccaaaggtgtcgtgcccaggagtagcggcggcctggaaatccttccatcactttccagtgcttcttggcggcttct
acctggcactcagtgctgatgtgccaccagcttgcctcagacacctctcaattgtggcaggccaggcaactcaccatg
gccaggctcctgcccacgatgtgcatccaggctcggagggaagggacagcagctggcagcttgcctgcagaaggccga
ggcgcacacacctcagatcacagcagccttctggcagggctgtgtccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctccggcgcagcctgtgcccgtgggtgtctggcccagcctccgcaagcattccac
tccatccccgccagctgcaccgggtgaggccaagagcgtgcatgccaatgcccgggttcaatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcggaaaggctgcacgtggcctgaatgttgggacacctcaagtgcatttctgagtg
tgggccccactgagtgctgctgcccctggccttctgtgctgcaagcactccggcggcccgtggccctgcaagctggactacaac
tctgtgggtgacatggcgtggagcagctgctgcttgccttgggtgtctgcaaggctctgtatttgcgcatacaat
ctcagaccgaggcatctgcaagctcattgaatgtgctctcactcggagcaatgcaagaattagctctattcaacaaca
aattgactgacggctgtgcaacactccatggctaaagctccttgcctgcaagcagaacttctggcattgaggctggggaat
aactacatcactgcccgggagcccaggctgctggccgaggggctccgaggcaacacctcctgcaagttcctgggattctg
gggcaacagagtggtgacgagggggccaggccctggctgaagccttgggtgatacaccagagcttgagggtggctcagcc
tgggtggggaacaacatggcagtggtgggtgcccagccttggcactgatgctggcacaagaacatcatgtagaagaactc
tgccggaggagaaccaatctccaggatgaaggtgtatgttctctcgcagaaggactgaagaaaaatcaagttgaaaaat
cctgaagttgtccaataactgcatcacctacctaggggcagaagccctcctgcaagcccttgaagggaatgacacctcc
tggaaagtctggctccgagggaacacttctctctagaggaggttgacaagctcggctgagggaacaccagactcttgcctt
tga

Figura 36

SEC ID N° 67

MGEEGGSASHDEEBERASVLLGHSPGCEMCSQEAFOAQRSQVLELLVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLHGCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKKYMAKLRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFVFPFSCRQLQCMAPLSVRTLLFEHCWPDVGGQEDIFQLLLDHFDRVLLTFDGGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEQGIELYLRKRHHEPGVADRRLIRLLQETSALH
GLCHLPVFSWVSKCHQELLQEGGSPKTTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLPTLLHLGRLALWGLGM
CCYVFAQQQAQVSPDDISLGFLVRAKGVVPGSTAPLEFLHITFQCFFAAFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGFIVLIRSLYEMQEERLARKAARGLNVGHKLKLTFCVGPTECAALAFVLQHLRRFVALQLDYN
SVGDIGVEQLLPCLVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTDGCAHSMAKLLACRQNFALRLGN
NYITAAGAQLVAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSLRWLSLVGNNIGSVGAQALALMLAKNIMLEEL
CLEENHLQDEGVCSLAEGLKKNSSKILKLSNNCITYLGAEALQALERNDTILEVWLRGNTFSLEVDKLGCRDTRLLL
*

Figura 37

SEC ID N° 68

atgggggaagagggtgggttcagcctctcacgatgaggaggaagagcaagtgtcctcctcggacatctcctcgggttgtga
aatgtgctcgcaggaggctttcaggcacagaggagccagctggctcgagctgctggctcagggctccctggaaggcttcg
agagtgctcctggactggctgctgtcctgggaggtcctcctcctgggaggactacgagggcttccacctcctgggcccagcct
ctctcccacttggccaggcccttctggacaccgtctggaataagggtacttgggcctgtcagaagctcatcgcggctgc
ccaagaagcccaggccgacagccagctcccccaagctgcatggctgctgggacccccactcgtccaccagcccgagacc
tgcagagtccaccggccagccatgtcaggaggctccacagccatgtggagaacatgctggacctggcatgggagcgggggt
ttcgtcagccagatgaatgtgatgaaatcaggttgcctgactctcaccacctcccagagggaagaaggctgcttgatct
tgccacgggtgaaagcgaatggattggctgccttctctacaacatgttcaggaattaccagttccattggccctgctt
tggaaagctgcccacatgcaagaagtaatatggccaagctgaggaccaggtgtctgctcagctctcctcagttacctat
gatggagcagagacgctcctcctggaggacatatacacagagaatgtcctggaggctcctgggagatgtgggcatggctgg
acccccgcagaagagcccagccaccctgggcttggaggagctctcagcacccttggccacctcaatgacgatgcccga
ctgtgctgggtgggtgaggcggcagtggaagagcacgctcctgcagcggctgcaactgctgtgggctgcaaggca
gacttccaggaaattctcttcttcttccattcagctgcccggcagctgcaatggccaaaccactctctgtgcccagc
tctactcttggagcactgctgttggcctgatgttggctcaagaagacatcttccagttactccttgaccacctgacctg
tctgttaacctttgatggcttggacagttcaagttcaggttccaggatcgtgaaacgcaactgctcccgcaccgaccc
acctctgtccagacctgctcttcaacctctcagaggcaacctgctgaaagaatgcccgcagggtggtagccagcctcc
ggcctgctgtgctggccttctcaggaagtacatccgcaccaggttcaacctcaagggtctctctgaaacagggatcagc
tgtacctgaggaagcgcacatcatgagcccgggtggcggaccgctctcctcctgctccaagagacctcagccctgca
ggttgtgcccacctgacctgtcttctcatggatgggtgtccaaatgcccaccaggaactgtgctgcaaggagggggtccc
aaagaccactacagatagtacctgctgatctctgcagcatcttctgctgcatgcccacccccagactcagcttcccag
gtctgggaccagctctctcctggggccgctcccacctcctgcaactgggagactggctctgtggggcctgggcatg
tgcctgctacgtgttctcagcccagcagctccaggcagcacaggtcagccctgatgacatttctctggcttcttgggtg
tgccaaagggtgctgctgcccaggagtagcggcctcctggaattccttccatcacttccagtgctcttctgcccgttct
acctggcactcagtgctgatgctgcccaccagcttggctcagacacctctcaattgtggcaggccaggcaactcaccatg
gccaggctcctgcccacgatgtgcatccaggctcggagggaaggacagcagcgtggcagcttctgctcagaaggccga
gcccacaaacctcagatcacagcagccttctggcagggtgtgttctccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctcgggcccaggcctgtgcccgtgggtgtctggcccagcctccgcaagcacttccac
tccatcccgcagctgcaaccgggtgaggccaagagcgtgcatgcccagcgggttcatctggctcatccggagcctgta
cgagatgcaaggagcggctggctcggaaagctgcaaggtggcctgaaatgttgggcacctcaagttgacatttggcagtg
tgggccccactgagtgctgctgcccctggccttggctgctgcaagcactccggcggccctggcctgcaagtggaactaca
tctgtgggtgacatggcgtggagcagctgctgcttgccttgggtgctgcaaggctctgtatttggcggatgacaat
ctcagaccaggcactgcaagctcatgaaatgtgctcttctcactgcagcaattgcaagaagttagctctattcaacaaca
aattgactgacggctgtgcaactcctggctaaagctccttgcattgcaaggcagaacttcttggcattgaggctggggaat
aactacatcactgcccggggagcccaagtgctggcggaggggctccaggccaacacctccttgcagttcctgggattctg
gggcaacagagtggtgacgagggggcccaggccctggctgaagccttgggtgatcaccagagctgaggtggctcagcc
tgggtggggaacaacatggcagtggtggctgcccaggccttggcactgatgctggcaagaacgtcatgctagaagaactc
tgctggaggagaacctctccaggatgaaggtgatgttctctcgcagaaggactgaagaaaaatcaagtttgaata
cctgaagttgtccaataactgcatcacctacctaggggcagaagcctcctgcagggccttgaagggaatgacacctcc
tggaaagtctggctccgaggggaacacttctctctagaggaggtgacaagctcggctgcaaggacaccagactcttctt
tga

Figura 38

SEC ID N° 69

MGEEGGSASHDEEERASVLLGHSPGCEMCSQEAFQAQRSQVLVSVGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLHGCDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKKYMALRRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFVFPFSCRQLQCMAPLSVRTLLFEHCWPDVGGEDI FQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNLKGFSEQGIELYLRKRHHPEPGVADRILIRLLQETSALH
GLCHLPVFSWVSKCHQELLQEGGSPKTTTMYLLIQHFLLHATPPDSASQGLGPSLLRGRLLPTLLHLGRLLAWGLGM
CCYVFAQQQLAAQVSPDDISLGFVRAKGVVPGSTAPLEFLHITFQCFFAIFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKDSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGFIWLRSLYEMQEERLARKAARGLNVGHKLKTFCCSVGPTCAALAFVLQHLRRPVVALQLDYN
SVGDIGVEQLLPCLVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTGCAHSVAKLLACRQNFALRLGN
NYITAAGAQLAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSLRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLBENHLQDEGVCSLAEGLKKNSSLKILKLSNNCITYLGAEALIQALERNDTILEVWLRGNTFSLEEVDKLGCRTRLLL
*

Figura 39

SEC ID N° 84

atgggggaagaggggtgggttcagcctctcacgatgaggaggaaagagcaagtgctcctcctcgggacattctccgggttgatga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctggctcgagctgctgggtcagggtccctggaaggettccg
agagtgctcctggactggctgctgctcctgggaggctcctcctcctgggaggactacgagggtccacacctcctgggcccagcct
ctctcccacttggccaggcgcctctctggacaccgctctggaat aagggtacttgggctgtcagaagctcatcgcggtgc
ccaagaagcccaggccagcagccagtcacccaagctgcatggctgctgggacccccactcgcctccaccagcccgagacc
tgcagagtcaccggccagccattgtcaggaggctccacagccatgtggagaacatgctggacctgggcatgggagcgggggt
ttcgtcagccagtatgaaatgtgatgaaatcaggttgcctgctctcacaccgtcccagagggcaagaaggtgcttgatct
tgccacgggtgaaagcgaatggatgggtgctcctcctctacaacatggtcagggaat taccagtcaccttggcctgctct
tggaaagctgccacatgcaagaagtata tggccaagctgaggaccaggtgctgctcagtcctcctcagtaacctat
gatggagcagagacgctctgctcctggaggacatatacacagagaatgtcctggagggtctgggacagatgtgggca tggctgg
atccccgcagaagagcccagccaccctgggctggaggagctctcagcaccctggccacctcaatgacgatggggaca
ctgtgctgggtgggtgggtgaggcggcagtggaagagcagctcctcagcagcggctgcaacttgctggtgggtcagggcaa
gacttccaggaatttctctttgtcttcccatcagctgcccggcagctgcaatgagccaaaaccactctctgtgcccggac
tctactctttgagcactgctgttggcctgatgttggccaagaagacatcttccagttactccttgaccacctgaccgtg
tccctgttaacctttgatggctttgacgagttcaagttcaggttcacggatcgtgaaagccactgctccccgaccgacccc
acctctgtccagaccctgctcttcaacctcttgcagggaacctgctgaaagtgcggcaaggtgggtgaccagccgtcc
ggccgctgtgctggcgttctcaggaagtacatccgcaaccgagttcaacctcaagggttctctgaaacagggcatcgagc
tgtacctgaggaagcgcctcatgagccgggggtggcggaccgctcatccgctgctccaagagacctcagccctgcac
ggtttgtgcccactgctgctctcctatggatgggtgctccaaatgcccaccaggaactgtgctgagggaggggggtcccc
aaagaccactacagataatgtaacctgctgattctgcagcatttctctgctgcatgcccacccccagactcagcttcccaag
gtctgggacccagctctctcctggggccgctccccaccctcctgcaacctgggcagactggctctgtggggctgggcatg
tgctgctacgtgttctcagcccagcagctccagggcagcacaggctcagccctgatgacatttctcttggcttctcgtgctg
tgccaaaggtgctgctgcccagggtacggcgccccctggaat tctctcacatcactttccagtgctctcttggccgcttct
acctggcactcagtgctgatgtgcccaccagcttctcagacacctcttcaattgtggcagggccaggaactcaccatg
gccaggctcctgcccacgatgtgcatccaggcctcggagggaaggacagcagcgtggcagctt tctgctgcaagaaggccga
gcccacacacctcagatcacagcagccttctcggcagggtgctgtgctccgggagcactggggcctgctggctgagtgcc
agacatctgagaaggccctgctcggcggccaggcctgtgcccggctgggtgcttggcccgcagcctccgcaagcacttccac
tccatcccggcagctgcaaccgggtgaggcccaagagcgtgcatgcccagccgggttcatctggctcatccggagcctgta
cgagatgcaggagggagcggctggctcgggaaggtgcacgtggcctgaatgttgggcaacctcaagttgacatttgcagtg
tgggccccactgagtgctgcccctggcctt tgtgctgcaagcactccggcggccctggcctgcaagctggactacaac
tctgtgggtgacattggcgtggagcagctgctgcttgccttgggtgctgcaaggctctgatat tggcggataaacaatat
ctcagaccgaggcatctgcaagctcat tgaatgtgctcttcaactgagcaat tgcagaagttagctctattcaacaaca
aattgactgacggctgtgcacactccatggctaaagctccttgcatgcaaggcagaactcttggcattgaggctggggaat
aactacatcactgcccgggagcccaagtgtgcccaggggctccgaggcaaacacctccttgcaattcctgggttctctg
gcycaacagagtggggtgacgagggggcccaggccctggctgaaagccttgggtgat caecagagcttgaggtggctcagcc
tgggtggggaacaacattggcagtggtgggtgcccaggccttggcactgatgctggcaaaagaactcatgctagaagaactc
tgctggaggagaacca tctccaggatgaaagtgatgttctctcgcagaaggactgaaagaaaattcaagtttgaanaat
cctgaaagttgtccaataactgcatcacctacctaggggcagaagccctcctgagggccttgaagggaatgacaccatcc
tggaaagctggctccgagggaaacacttctctctagaggaggttgacaagctcggctgaggggacaccagactcttggct
tga

Figura 40

SEC ID N° 85

MGEEGGSASHDEERASVLLGHS PGCEMCSQEAFQAQRSQVVELLVSGSLEGFESVLDWLLSWEVLSWEDYEGPHLLGQP
LSHLARRLLDTVWNKGTWACQKL TAAAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRRLHSHVENMLDLAWERG
FVSQYECDEIRLP IFTPSQRARRLLDLATVKANGLAAFLLQHVQELPVPLALPLEAATCKKYMAKLRRTTVSAQSRFLSTY
DGAETLCLEDIYTENVLEVWADVGMAGSPQKSPATLGLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAQO
DFQEFLLVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVGQEDI FQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEQGI ELYLRKRHHEPGVADRLIRLLQETSALH
GLCHLPVFSWMVSKCHQELLQEGGSPKTTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLEPTLLHLGRLLWGLGM
CCYVFSQQQLAAQVSPDDISLGFVRAKGVVPGSTAPLEFLHTTFQCFFAAFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARS LRKHFH
SIPPAAPGEAKSVHAMPGFIWLRSLYEMQEERLARKAARGLNVGHKLTFCSVGPTECAALAFVLOHLRRPVALQLDYN
SVGDIGVEQLLPCLGVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTDCCAHSMAKLLACRQNFALRLGN
NYITAAGAVLAEGLRGNTSLQFLGFWRNRVGDGAQALAEALGDHQSRLRWLSLVGNNGISVGAQALALMLAKNVMLEEL
CLEBNHLQDEGVCSLAEGLKKNSSKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
*

Figura 41

SEC ID N° 86

atgggggaagaggggtgggttcagcctctcacgatgaggaggaagagcaagtgtcctcctcggacattctccgggttgga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctgggtcgagctgctgggtctcagggtccctggaaggttcg
agagtgtcctggactggctgctgctcctgggaggtcctcctcctgggaggactacgagggtctccacctcctcctggccagcct
ctctcccacttggccaggcgcctctctggacaccgtctggaataaggggtacttgggctgtcagaagctcatcgggctgc
ccaagaagcccaggccgacagccagctccccaagctgcatgggtgctgggacccccactcgtccaccagcccagagacc
tgcagagtcaocggccagccattgtcaggaggtccacagccatgtggagaacatgctggacctggcatgggagcgggggt
ttcgtcagccagatgaatgtgatgaaatcaggttgcgctctccacaccgtcccagagggcaagaaggctgcttgatct
tgccacgggtgaaagcgaatggatggctgcctcctctcaaacatgttcaggaattaccagtcccatggccctgcct
tgggaagctgccacatgcaagaagtatatggccaagctgaggaccaggtgtcgtcagctcctcctcagtaacctat
gatggagcagagacgctctgcctggaggacatatacacagagaatgtcctggaggctctgggagatgtgggcatggctgg
atccccgcagaagagcccagccacctgggctggaggagctctcagcaccctggccacctcaatgacgatgcccaca
ctgtgctgggtgggtgggtgaggcgggagtggaagagcagcctcctgcagcggctgcacttgtctgggtgaggggcaa
gactccaggaatttctcttctctccattcagctgcccagctgcagtgcatggccaaacctctctgtgctggac
tctactcttggagcactgctgtggcctgatgtggctcaagaagacatctccagttactccttgaccacctgaccgtg
tccgttaacctttgatggctttgacaggtcaagttcaggttcacggatcgtgaacgccactgctcccagccgaccccc
acctgtgctcagaccctgctctcacccttctgcagggcaacctgctgaagaatgcccgaaggtggtgaccagcctcc
ggcgcgtgtgtcggcgttctcaggaagtacaaccgacccaggttcaacctcaagggcttctctgaaacagggcatcgagc
tgtacctgaggaagcgcctcatgagccccgggtggcgggacgcctcatccgctgctccaagagacctcagccctgacc
ggtttgtgccacctgctgtctctcatggatgggtgtccaaatgccaccaggaactgttgtgaggggggggggtcccc
aaagaccctacagatagtacctgctgatctgcagcatttctgctgcatgccacccccccagactcagcttcccag
gtctgggaccagctctctcctggggcgcctccccacctcctgcacctgggagactgggtctgtggggcctgggcatg
tgctgctacgtgtctcagcccagcagctccaggcagcaggtcagccctgatgacatttctctggcttctcctgggtgcg
tgccaaaggtgtcgtgccaggagtagggcgccttggaaatccttcaatcacttccagtgctctcttggcgcgttct
acctggcactcagtgctgatgtgccaccagcttgcctcagacacctctcaatgtggcaggccaggcaactcaccaatg
gccaggctcctgccacgatgtgcatccaggcctcggagggaagggacagcagcgtggcagcttctgctgcaagggccga
ggcgcacaaccttcagatcacagcagccttctggcagggtgtgtcccgggagcactggggcctgctggctgagtgcc
agacatctgagaagggcctgctccggcgcagggcctgtgcccgtgggtgtctggcccagcctccgcaagcacttccac
tccatcccggcagctgcaccgggtgagggcaagagcgtgcatgcccgggttcatctggctcatccggagcctgta
cgagatgagggagggagcggctggctcggaaagctgcacgtggcctgaaatgttgggacacctcaagtgcacatttgcagtg
tgggccccactgagtgctgctgcccctggccttctgctgagcaccctccggcggccctgggcctgagctggactacaac
tctgtgggtgacatggcgtggagcagctgctgcttgccttgcttgctgcaaggctctgtatttggcgcataacaatat
ctcagaccgaggcatctgcaagctcatgaaatgtgctcttcaactgcgagcaattgcagaagttagctctattcagcaaca
aattgactgacggctgtgacacactccatggctaaagctccttgcatgagggcagaactccttggcatgaggctggggaa
aactacatcactgcccggggagcccagtgctggcggaggggtccgaggcaaacacctcctgagctcctgggattctg
gggcaacagagtggtgacagggggccccaggcctggctgaaagccttgggtgatcaccagagcttgaggtgggtcagcc
tgggtggggaacaacatggcagtggtgggtgccaaagccttggcactgatgctggcaagaaagctcatgtagaagaactc
tgctggagggaacacatctccagatgaagggtgatgtctctcgcagaaggactgaagaaaaatccaagttgaaat
cctgaagtgttccaaataactgcatcactacctaggggcagaagcctcctgcaggccttgaaggaatgacaccatcc
tggaaagtctggctccgagggaacacttctctctagaggaggttgacaagctcggctgcagggacaccagactcttgctt
tga

Figura 42

SEC ID N° 87

MGEEGGSASHDEEERASVLLGHS PGCEMCSQEA FQAQRSQVLELVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLDVTWNKGTWACQKLI AAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRRLLSHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAFLQHVOELPVPLALPLEAATCKKYMAKLRRTVSAQSRFLSTY
DGAETLCLEDIYTENVEVWADVGMAGSPQKSPATLGLLEELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFLFVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVGQEDIFQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQGNLLKNARKVVTSRPAAVSAFLRKYIRTEFNKGFSEGGIELYLRKRHHEPGVADRLIRLLQBTSAH
GLCHLPVFSWVSKCHQELLLQEGGSPKTTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRLPTLLHLGRLALWGLGM
CCYVFAQQQLQAQVSPDDISLGFVRAKGVVPGSTAPLEFLHITFQCFFAAFYLALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLRRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGFIWLRSLYEMQEERLARKAARGLNVGHLKLTFC SVGTECAALAFVLQHLRRPVALQLDYN
SVGDIGVEQLPCLGVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNSKLTGCAHSMAKLLACRQNFIALRLGN
NYITAAGAQLAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSLRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGLKKNSSKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
*

Figura 43

SEC ID N° 88

atgggggaagagggtgggttcagcctctcacgatgaggaggaaagagcaagtgtcctcctcgggacattctccgggttgtga
aatgtgctcgcaggaggcttttcaggcacagaggagccagctgggtcgagctgcttgggtcctcagggtccctgggaaggcttcg
agagtgtcctggactggctgctgtcctgggaggctcctcctcgggaggactacgagggttccacctcctgggcccagcct
ctctcccacttggccaggcgccttctggacaccgtctggaaataagggtacttgggctgtcagaagctcatcgcggtgc
ccaagaagcccaggccgacagccagctcccccaagctgcatggctgctgggacccccactcgtccaccagcccgagacc
tgcagagtcaccggccagccattgtcaggaggctccacagccatgtggagaacatgctggacctggcctgggagcgggt
ttcgtcagccagatgaatgtgatgaaatcaggttgcgatcttcacaccgtcccagagggaagaaggctgctgatct
tgccacggtgaaagcgaatggattggctgctcctcctacaaatgttcagggaattaccagctccattggccctgctt
gggaagctgccacatgcaagaagtatatggccaagctgaggaccaggtgtctgctcagctcctcctcagtaacctat
gatggagcagagacgctctgctggaggacataacacagagaatgtcctggaggctcgggacagatgtgggcatggctgg
atccccgcagaagagccagccaccctgggctggaggagctcttcagcaccctggccacctcaatgacgatgcccagaca
ctgtgctgggtgggtgggtgagggggcagtggaagagcagctcctcgcagcggctgcaacttgtgaggctgcaaggcaca
gacttccaggaatttctctttgtcttccattcagctgcccggcagctgcaagtgcattggcacaaccactctctgtgaggac
tctactcttgagcactgctgttggcctgatgttggtcaagaagacatctccagttactccttgaccacctgacctg
tctgttaacctttgatggctttgacaggttcaagttcaggttcacggatcgtgaacgccactgctcccgcagccagccc
acctctgtccagacctgctctcaaccttctgcagggaacctgctgaagaatgcccgaagggtggtagaccagccgtcc
ggccgctgtgctggcgcttccctcaggaagtacatccgcaccgagttcaacctcaagggtctcttgaacagggcatcgagc
tgtacctgaggaagcccatcatgagccggggtggcggaccgctcctcgcctgctccaagagacctcagccctgcac
ggtttgtgcccacctgctgtctctcctatggatgggtgtccaaatgccaccaggaactgttgcctgaggggggggtccc
aaagaccactacagatagtacctgctgatctgcagcatttctgctgcatgccacccccagactcagcttcccag
gtctgggaccagctctctcctggggccgctccccacctcctgcacctgggacagactggctctgtgggctgggcatg
tgctgctacgtgtctcagccagcagctccaggcagcacaggctcagccctgatgacatttctcttggcttccctgggtgcg
tgccaaagggtgctgctcagggagtaaggccctggaaattccttcacatcacttccaggtgctcttggcggcttct
acctggcactcagtgctgatgtgccaaccagcttctcagacacctctcaattgtggcaggccagggcaactcacaatg
gccaggtcctgcccagatgtgcatccaggcctcggagggaaggacagcagcgtggcagcttggctgcagaaggccga
gcccacacaaccttcagatcacagcagccttctggcagggtgtgtgtcccgggagcactgggctgctggctgagtgcc
agacatctgagaaggccctgctctggcgcaggcctgtgcccgttgggtgtctggcccgcagcctccgcaagcacttccac
tccatcccggcagctgcaccgggtgaggccaagagcgtgcatgccatgccgggttcatctggctcatccggagcctgta
cgagatgcaggaggagcggctggctcgggaaggctgcacgtggcctgaatgttgggacctcaagtgacatttgcagtg
tgggcccactgagtgctgcccctggccttgtgctgcagcactccggcggccctggccctgcagctggactacaac
tctgtgggtgacatggcgtggagcagctgctgcttgccttgggtgtctgcaaggctctgtatctgcccgataaacaat
ctcagaccgaggcatctgcaagctcatgaaatgtgctctcactgcgagcaattgcagaagttagctctattcaacaaca
aattgactgacggctgtgcacactccatggctaaagctccttgcattgcaggcagaacttcttggcattgaggctggggaat
aactacatcactgccgcccggagcccaggtgctggccgagggtcctcagggcaaacctccttgcagttcctgggatctg
gggcaacagagtggtgacgagggggcccaggcctggctgaagccttgggtgatcaccagagcttgagggtggctcagcc
tgggtggggaacaacattggcagtggtgggtgcccagccttggcactgatgtcggcaaaagacgtcagctagaagaactc
tgctggaggagaacctctccaggtgaaggtgtatgttctctcgcagaaggactgaagaaaaattcaagtttgaaaat
cctgaagttgtccaataactgcatcactacctaggggcagaagcctcctgcaggcccttgaagggaatgacaccatcc
tggaaagtctggctccgaggggaacacttctctctagaggagggtgacaagctcggctgcagggaacaccagactcttgcct
tga

Figura 44

SEC ID N° 89

MGEEGGSASHDEERASVLLGHSPGCEMCSQEAFAQRSQLVVSGSLEGFESVLDWLLSWEVLSWEDYEGFHLLGQP
LSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLGHCWDPHSLHPARDLQSHRPAIVRRRLHSHVENMLDLAWERG
FVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLQHVQELPVPLALPLEAATCKKYMAKLRTTVAQSRFLSTY
DGAETLCLEDIYTENVLEWADVGMAGSPQKSPATLGLLELFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQ
DFQEFLVFPFSCRQLQCMAPLSVRTLLFEHCCWPDVQGEDIFQLLDHDPDRVLLTFDGFDEFKFRFTDRERHCSPTDP
TSVQTLLENLLQCNLLKNARKVVTSRPAAVSAFLRKYIRTEFNLKGFSEQGIELYLRKRHHEPGVADRILRLQETSALH
GLCHLPVFSWVSKCHQELLQEGGSPKTTDMYLLILQHFLHATPPDSASQGLGPSLLRGRPTLLHLGRLLALWGLGM
CCYVFSQQQLQAAQVSPDDISLGFLVRAKGVVPGSTAPLEFLHITFCQFFAAFYLAALSADVPPALLRHLFNCGRPGNSPM
ARLLPTMCIQASEGKDSVAALLOKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKALLWRQACARWCLARSLRKHFH
SIPPAAPGEAKSVHAMPGFIWLIRSLYEMQEERLARKAARGLNVGHKLKTFCSVGPTECAALAFVLQHLRRPVALQLDYN
SVGDIGVEQLLPCLGVCKALYLRDNNISDRGICKLIECALHCEQLQKLALFNNKLTGCAHSMAKLLACRQNFALRLGN
NYITAAGAQLVLAEGLRGNTSLQFLGFWGNRVGDEGAQALAEALGDHQSRLRWLSLVGNNIGSVGAQALALMLAKNVMLEEL
CLEENHLQDEGVCSLAEGKKNSSLKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL
*