

(12) SOLICITUD INTERNACIONAL PUBLICADA EN VIRTUD DEL TRATADO DE COOPERACIÓN EN MATERIA DE PATENTES (PCT)

(19) Organización Mundial de la
Propiedad Intelectual
Oficina internacional



(43) Fecha de publicación internacional
13 de agosto de 2020 (13.08.2020)

WIPO | PCT

(10) Número de publicación internacional
WO 2020/161378 A2

(51) Clasificación internacional de patentes:
C12Q 1/68 (2018.01)

(21) Número de la solicitud internacional:
PCT/ES2020/070086

(22) Fecha de presentación internacional:
08 de febrero de 2020 (08.02.2020)

(25) Idioma de presentación: español

(26) Idioma de publicación: español

(30) Datos relativos a la prioridad:
P201930104 08 de febrero de 2019 (08.02.2019) ES

(71) Solicitantes: **UNIVERSIDAD DE CÓRDOBA** [ES/ES];
Avda. Medina Azahara, 5, 14005 Córdoba (ES). **SERVICIO ANDALUZ DE SALUD** [ES/ES]; Avda. de la Constitución, 18, 41071 Sevilla (ES).

(72) Inventores: **LUQUE HUERTAS, Raúl Miguel**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **GAHETE ORTIZ, Manuel David**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **CASTAÑO FUENTES, Justo P.**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **JIMÉNEZ VACAS, Juan Manuel**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **MORÁIS SARMENTO BORGES, André**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **LÓPEZ LÓPEZ, Fernando**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **REQUENA TAPIA, María José**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **GÓMEZ GÓMEZ, Enrique**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES). **CARRASCO VALIENTE, Julia**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menéndez Pidal, s/n, 14004 Córdoba (ES).

(81) Estados designados (a menos que se indique otra cosa, para toda clase de protección nacional admisible): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU,

RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Estados designados (a menos que se indique otra cosa, para toda clase de protección regional admisible): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), euroasiática (AM, AZ, BY, KG, KZ, RU, TJ, TM), europea (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publicada:

— sin informe de búsqueda internacional, será publicada nuevamente cuando se reciba dicho informe (Regla 48.2(g))

(54) Title: METHOD FOR DIAGNOSING OR PROGNOSING THE DEVELOPMENT OF PROSTATE CANCER IN AN INDIVIDUAL

(54) Título: MÉTODO PARA DIAGNOSTICAR O PRONOSTICAR EL DESARROLLO DE CÁNCER DE PRÓSTATA EN UN INDIVIDUO

(57) Abstract: The invention relates to biomarkers, a method, a diagnostic kit and uses for diagnosing, predicting or prognosing the development of prostate cancer (PCa) in an individual.

(57) Resumen: Biomarcadores, método, kit de diagnóstico y usos para diagnosticar, predecir o pronosticar el desarrollo de cáncer de próstata (CaP) en un individuo.



WO 2020/161378 A2

Método para diagnosticar o pronosticar el desarrollo de cáncer de próstata en un individuo

CAMPO DE LA INVENCION

5

La presente invención se encuentra dentro del campo de la medicina y en concreto de la oncología, y está dirigida a un método para el diagnóstico temprano y para predecir o pronosticar el desarrollo de adenocarcinomas prostáticos en un individuo, a un kit o dispositivo de diagnóstico o pronóstico.

10

ANTECEDENTES DE LA INVENCION

El cáncer de próstata (CaP) se considera uno de los problemas médicos más importantes a los que se enfrenta la población masculina. De hecho, en algunos países occidentales es la neoplasia sólida más frecuente, superando en número al cáncer de pulmón y colorrectal. Además, representa la quinta causa de muerte por cáncer en este colectivo en todo el mundo.

En general se estima que en 2018 hubo 3,91 millones de casos nuevos de cáncer (excluyendo el cáncer de piel no melanoma) y 1,93 millones de muertes por cáncer en Europa. Los tumores más comunes fueron los cánceres de mama (523,000 casos), seguidos de colorrectal (500,000), cáncer de pulmón (470,000) y próstata (450,000). Estos cuatro cánceres representan la mitad de la carga total de cáncer en Europa. Las tasas de incidencia del cáncer de próstata varían cinco veces (37-189 por 100,000), con las tasas más altas estimadas en el norte y el oeste de Europa, incluyendo Irlanda (189.3), Estonia (162.4), Noruega (157.3) y Francia (144.9). (Ferlay et al. *Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. European Journal of Cancer*, 2018-11-01).

Aunque los métodos de diagnóstico de esta patología han mejorado notablemente en los últimos años, debido principalmente al desarrollo de técnicas no invasivas (p.ej. análisis de los niveles sanguíneos del antígeno prostático específico [PSA] y/o PSA libre, así como de los niveles en orina del antígeno 3 del cáncer de próstata [PCA3]), la única herramienta con capacidad pronóstica, aunque limitada, sigue siendo la escala de Gleason, que es un sistema que se emplea para medir el grado de agresividad de un CaP, calculado a partir del análisis histológico de muestras tumorales. La escala de Gleason clasifica los tumores de próstata en 5 grupos: Grado 1 o puntuación de Gleason de 6 o menos (cáncer de bajo grado); Grado 2 o puntuación de

Gleason de 3 + 4 = 7 (cáncer de mediano grado); Grado 3 o puntuación de Gleason de 4 + 3 = 7 (cáncer de mediano grado); Grado 4 o puntuación de Gleason de 8 (cáncer de grado alto); grado 5 o puntuación de Gleason de 9 a 10 (cáncer de grado alto).

- 5 El marcador más usado en el diagnóstico del CaP y en la monitorización del tratamiento es el antígeno prostático específico (PSA), pero tiene sus limitaciones. En términos de marcador diagnóstico, el PSA presenta una especificidad baja (33%). Una mejora en el diagnóstico con marcadores de cribado más específicos, evitaría un gran número de biopsias de próstata innecesarias y constituiría, a su vez, una prueba diagnóstica menos invasiva. El PSA tiene,
10 además, poco valor como biomarcador de pronóstico.

- Por otra parte, la primera aproximación terapéutica para el tratamiento del CaP, cuando este se encuentra aún confinado en la próstata, consiste en la prostatectomía radical, es decir, extirpación de la glándula prostática y las vesículas seminales, eliminando el tumor. Sin embargo,
15 para el tratamiento de recidivas o de tumores con elevada agresividad en el momento del diagnóstico se procede a la terapia hormonal, también llamada terapia de deprivación androgénica (ADT), la cual consiste en el uso de agonistas y antagonistas de LHRH como la leuprolida (Lupron®) y el degarelix (Firmagon®), respectivamente, capaces de reducir los niveles sistémicos de andrógenos, claves en el desarrollo de la glándula prostática así como en la
20 fisiopatología del CaP, a través del bloqueo del eje hipotalámico-hipofisario-gonadal. Por último, los pacientes que desarrollan tumores que se vuelven resistentes a la ADT y que, por tanto, sufren el fenotipo más agresivo de esta enfermedad o CaP resistente a la castración (CPRC), se tratan con novedosos fármacos como son la abiraterona (Zytiga®), capaz de reducir la producción exotesticular e intratumoral de testosterona, y el anti-andrógeno enzalutamida
25 (Xtandi®), el cual antagoniza el receptor de andrógenos (AR), evitando así su señalización. Sin embargo, estos fármacos no son todo lo efectivos que se desearía, ya que un importante porcentaje de los casos se vuelven resistentes a los mismos, adquiriendo los pacientes, por tanto, un pronóstico devastador. Por tanto, es evidente que es necesario el desarrollo de estrategias terapéuticas para combatir el CaP de forma global y más efectiva, algo que pasa por el
30 descubrimiento de nuevas dianas terapéuticas.

- El CaP es extremadamente heterogéneo y puede presentarse como una enfermedad indolente o agresiva. Desafortunadamente, no disponemos de biomarcadores que predigan la progresión de la enfermedad en el momento del diagnóstico. El uso del PSA ha liderado el diagnóstico de
35 muchos cánceres potencialmente indolentes, y los tratamientos agresivos de estos han causado una significativa morbilidad sin beneficio clínico en muchos casos. Por el contrario, aproximadamente un 30% de los pacientes tratados con intención curativa van a experimentar

recaída a los 5 años de iniciado el tratamiento, de ahí la importancia de definir mejor los grupos de riesgo.

En este sentido, se ha propuesto recientemente que la desregulación del mecanismo de splicing podría representar una de las bases de la aparición de diversas patologías tumorales, debido a que la alteración de la expresión de numerosas variantes de splicing alternativo se asocia con el desarrollo y el aumento de la agresividad de estas patologías. En concreto, diversos estudios han demostrado el papel determinante que juegan ciertas variantes de splicing en CaP, entre las cuales destacan AR-v7 (variante del receptor de andrógenos caracterizada por no poseer dominio de unión a la testosterona, permaneciendo activo de forma constitutiva; es una de las principales causas del desarrollo del CPRC), así como SST5TMD4 (variante del receptor de somatostatina SST5, que posee solo 4 dominios transmembrana y se relaciona con una mayor agresividad tumoral y con presencia de metástasis; es una de las posibles causas de la falta de respuesta de algunos tumores de próstata al tratamiento con análogos de somatostatina), ls In1-ghrelina [variante de la ghrelina, asociada con la regulación de procesos tumorales claves en el CaP, tales como la proliferación y migración celular (Hormaechea-Agulla et al. *The oncogenic role of the In1-ghrelin splicing variant in prostate cancer aggressiveness. Molecular Cancer*, 2017)], el KLF6-SV1 [variante del gen *KLF6* que presenta efectos antagónicos con respecto a la variante canónica, incrementando la agresividad tumoral y disminuyendo la apoptosis (DiFeo et al. *The role of KLF6 and its splice variants in cancer therapy. Drug Resistance Updates*, 2009)] y OPNb y OPNc [variantes del gen *SPP1/OPN* que actúan promoviendo la progresión tumoral del CaP a través de la activación de la vía PI3K (Tilli et al. *Both osteopontin-c and osteopontin-b splicing isoforms exert pro-tumorigenic roles in prostate cancer cells. The Prostate*, 2012)].

El proceso de *splicing* es catalizado por los spliceosomas menor y mayor, que actúan sobre diferentes tipos de intrones. El spliceosoma es un complejo macromolecular, cuyo núcleo funcional está compuesto por varias subunidades de pequeñas ribonucleoproteínas nucleares (snRNPs), que interactúan dinámicamente para regular el proceso de splicing. Además, la actividad del spliceosoma está modulada por más de 300 factores de *splicing* (SFs) que reconocen específicamente ciertas secuencias en exones e intrones. Por lo tanto, la desregulación de la expresión y/o función de ciertos componentes del spliceosoma puede conducir a la alteración aberrante del proceso normal de *splicing*. De hecho, la función correcta de la maquinaria de *splicing*, es decir, los componentes del spliceosoma y los SFs, es un mecanismo esencial para mantener la homeostasis de todo el cuerpo. Por lo tanto, las desregulaciones del proceso de *splicing* pueden jugar un papel importante como factor patógeno y podrían servir como un marcador predictivo del desarrollo de CaP.

El *splicing* alternativo está emergiendo como un sello distintivo en el cáncer y otras patologías. En particular, en el CaP, la retención de intrones podría ser especialmente relevante, ya que los genes que codifican factores clave asociados con su desarrollo y progresión (por ejemplo, receptor de andrógenos, antígeno prostático específico o ghrelina) pueden experimentar este proceso, lo que genera variantes oncogénicas involucradas en el desarrollo, la agresividad y respuesta al tratamiento de CaP.

Por todo esto, podría ser de utilidad la caracterización del perfil molecular de la maquinaria de *splicing* en pacientes con sospecha de CaP, de forma que pudiera servir también como herramienta para predecir el desarrollo futuro de CaP en dichos pacientes.

DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Niveles de expresión de 26 genes que codifican componentes de la maquinaria de *splicing* y que se encuentran desregulados en CaP (n=84) con respecto a la zona no tumoral adyacente (n=84). Los niveles de ARNm se determinaron por PCR cuantitativa y se ajustaron con un factor de normalización calculado a partir de los genes constitutivos B-Tubulina y GAPDH. Los resultados se expresan como Z-Score. Los asteriscos (*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$) indican diferencias estadísticamente significativas. **B)** Curvas ROC que muestran la sensibilidad, especificidad y el área bajo la curva de distintas combinaciones de la expresión de hasta 26 componentes de la maquinaria de *splicing* y factores asociados para diferenciar entre muestras de CaP y sus respectivas regiones adyacentes no tumorales.

Figura 2. Relación de la expresión de 11 componentes de la maquinaria de *splicing* y factores asociados en la zona tumoral de muestras fijadas en parafina (n=84) con el desarrollo de recidivas de CaP. **A)** Comparación entre los niveles de expresión de componentes de la maquinaria de *splicing* en muestras de CaP derivadas de pacientes que desarrollan recidivas de CaP (n=23) y pacientes que no las desarrollan (n=57). Los niveles de ARNm se determinaron por PCR cuantitativa y se ajustaron con un factor de normalización calculado a partir de los genes constitutivos B-Tubulina y GAPDH. Los asteriscos (*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$) indican diferencias estadísticamente significativas. **B)** Curvas ROC que muestran la sensibilidad, especificidad y el área bajo la curva de distintas combinaciones de la expresión de 11 componentes de la maquinaria de *splicing* y factores asociados para diferenciar entre pacientes que desarrollarán recidivas de CaP y pacientes que no las desarrollarán.

Figura 3. Relación de la expresión de *snRNP200*, *SRSF3* y *SRRM1* con parámetros clínicos y moleculares de agresividad del CaP. Asociación entre los niveles de expresión de *snRNP200*,

SRSF3 y *SRRM1* en muestras frescas de CaP (n=42) y la presencia y tipo de metástasis **(A)**, así como una puntuación de Gleason mayor o igual a 8 **(B)**. El asterisco (*, $p < 0.05$) indica diferencias estadísticamente significativas. Correlación entre los niveles de expresión de snRNP200, *SRSF3* y *SRRM1* y el grado de Gleason **(C)**, expresión de PSA **(D)**, AR **(E)** y AR-v7 **(F)** y el ratio de la expresión de AR-v7 y AR **(G)**. Se indica el p-valor (p) y el coeficiente de la correlación de Pearson y Spearman (R; dependiendo si los datos siguen o no una distribución normal).

Figura 4. *SnRNP200*, *SRSF3* y *SRRM1* modulan la proliferación de células derivadas de CaP. Tasa de proliferación de las células 22Rv1 **(A)** y DU145 **(B)** a 24, 48 y 72 h, en respuesta al silenciamiento de snRNP200, *SRSF3* y *SRRM1* mediante siRNAs específicos. Los resultados se representan como porcentaje de proliferación con respecto a las células transfectadas con siRNA control o Scramble. Los asteriscos (*, $p < 0.05$; ***, $p < 0.001$) indican los valores que difieren significativamente del control (Scramble).

Figura 5. Efectos del silenciamiento de *snRNP200*, *SRSF3* y *SRRM1* sobre los niveles de expresión del Receptor de Andrógenos (AR), de la variante de splicing 7 del AR (AR-v7) y del ratio entre los mismos (AR-v7/AR) en la línea celular 22Rv1. Los asteriscos (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$) indican que los valores difieren

DESCRIPCIÓN DE LA INVENCION

La presente invención supone una herramienta para la identificación de pacientes con CaP. La creación de los modelos de evaluación de los biomarcadores aquí propuestos permite: 1) **discriminar entre individuos sanos e individuos con CaP**, con elevada sensibilidad y especificidad, 2) **estratificar clínicamente los individuos** en base a parámetros clínicos y la expresión de componentes de la maquinaria de splicing en células de tejido prostático; y 3) **facilitar el seguimiento clínico de estos individuos**, contribuyendo al desarrollo de una medicina personalizada.

La presente invención propone la **evaluación estandarizada de un panel de biomarcadores**, en este caso relacionados con la maquinaria de *splicing*, en tejido de individuos. Esta información permitiría, por ejemplo, el desarrollo de un chip o array para la estratificación de pacientes en diversos grupos de riesgo y la definición de sus necesidades específicas de seguimiento clínico y manejo terapéutico.

La desregulación del proceso de corte y splicing de moléculas de ARNm, o *splicing*, está emergiendo como un sello distintivo en diversas patologías. Específicamente en el CaP se ha observado la alteración significativa de algunos componentes del spliceosoma y factores de splicing.

5 El objetivo de la presente invención ha sido la identificación de ciertos componentes de la maquinaria de *splicing* desregulados en células de pacientes con CaP.

Los autores de la presente invención han identificado que la alteración en la expresión de ciertos
10 componentes del spliceosoma y factores de *splicing* están relacionados con el desarrollo de CaP, pudiendo asociarse al aumento de expresión de diferentes isoformas proteicas de carácter oncogénico, que contribuyen en el desarrollo y la progresión de cáncer de próstata.

Concretamente, los autores han determinado que la sobreexpresión de *RNU6*, *SF3B1*,
15 *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*; *SND1*, *SRSF2*, *SRSF3* y *SRRM1* se asocia a agresividad tumoral. Adicionalmente se han asociado a recidivas 11 componentes del spliceosoma y factores de splicing *snRNP200*, *ESRP1*, *ESRP2*, *RBM45*, *SND1*, *SRSF3*, *SRSF4*, *SRSF5*, *SRSF10*, *SRRM1* y *TIA1*. Además, *snRNP200*, *SRSF3* y *SRRM1* tienen gran potencial para su uso como diana terapéutica, ya que son los únicos factores alterados en todas las
20 comparativas y su modulación in vitro está asociada a la agresividad tumoral.

Por tanto, **un primer aspecto** de la invención se refiere al uso del producto de la expresión de los genes *RNU6*, *SF3B1*, *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *SRSF3*, *SRRM1*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1* o cualquiera de sus combinaciones, para
25 predecir o pronosticar el desarrollo de CaP en un individuo. Por tanto, se puede emplear el producto de expresión de uno o más genes para el diagnóstico, predecir la agresividad tumoral, capacidad de invasión, recidivas.

En una realización preferida de este primer aspecto, se emplea el uso del producto de expresión
30 de todos los genes simultáneamente.

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*RNU6*", (también llamado en literatura U6; *RNU6*, *RNU6A*, U6-1) se refiere al gen, factor de la maquinaria de *splicing*. En el contexto de la presente invención, "*RNU6*" se define también por
35 una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia del snRNP "*RNU6*", y que comprendería diversas variantes procedentes de:

- a) moléculas de ácido nucleico que comprende la SEQ ID NO: 1,
- b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

Gene ID: 26827

Localización del gen: 15q23

SEQ ID N° 1:

GTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCC
CCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTT

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término “*SF3B1*”, (también llamado en literatura *MDS*; *PRP10*; *Hsh155*; *PRPF10*; *SAP155*; *SF3b155*) se refiere al gen, que codifica la subunidad 1 del complejo de proteína del factor splicing 3b. El factor de empalme 3b, junto con el factor de empalme 3a y una unidad de ARN 12S, forma el complejo de ribonucleoproteínas nucleares pequeñas U2 (snRNP U2). El complejo de factor de empalme 3b/3a se une al pre-ARNm corriente arriba del sitio de la rama del intrón de manera independiente de la secuencia y puede anclar el snRNP U2 al pre-ARNm. El factor de empalme 3b también es un componente del spliceosoma menor tipo U12. El splicing alternativo resulta en múltiples variantes de transcripción que codifican diferentes isoformas.

En el contexto de la presente invención, “*SF3B1*” se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia del snRNP “*SF3B1*”, y que comprendería diversas variantes procedentes de:

- a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 3,
- b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 3. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SF3B1*. Entre otras posibles secuencias nucleotídicas que codifican *SF3B1* se encuentra, pero sin limitarse, la SEQ ID NO: 2.

Gene ID: 23451

Localización del gen: 2q33.1

SEQ ID N° 2:

```

10 AGAGTGCAGCCCCCAGCTATTTTTCTCCGTGGCGGGCGGCGACGAGCGGAAGTTCTTGGGA
   GCGCCAGTTCGGTCTGTGTGTTTCGAGTGGACAAAATGGCGAAGATCGCCAAGACTCACGA
   AGATATTGAAGCACAGATTCGAGAAATTCAAGGCAAGAAGGCAGCTCTTGATGAAGCTCAA
   GGAGTGGGCCTCGATTCTACAGGTTATTATGACCAGGAAATTTATGGTGGAAAGTGACAGCA
   GATTTGCTGGATACGTGACATCAATTGCTGCAACTGAACTTGAAGATGATGACGATGACTA
   TTCATCATCTACGAGTTTGCTTGGTCAGAAGAAGCCAGGATATCATGCCCTGTGGCATTG
15 CTTAATGATATACCACAGTCAACAGAACAGTATGATCCATTTGCTGAGCACAGACCTCCAAA
   GATTGCAGACCGGGAAGATGAATACAAAAGCATAGGCGGACCATGATAATTTCCCCAGA
   GCGTCTTGATCCTTTTGCAGATGGAGGGAAAACCCCTGATCCTAAAATGAATGCTAGGACT
   TACATGGATGTAATGCGAGAACAACACTTGAATAAAGAAGAACGAGAAATTAGGCAACAGC
   TAGCAGAAAAAGCTAAAGCTGGAGAACTAAAAGTCGTCAATGGAGCAGCAGCGTCCCAGC
20 CTCCATCAAAACGAAAACGGCGTTGGGATCAAACAGCTGATCAGACTCCTGGTGCCACTC
   CCAAAAACTATCAAGTTGGGATCAGGCAGAGACCCCTGGGCATACTCCTTCTTAAGATG
   GGATGAGACACCAGGTCGTGCAAAGGGAAGCGAGACTCCTGGAGCAACCCCAGGCTCAA
   AAATATGGGATCCTACACCTAGCCACACACCAGCGGGAGCTGCTACTCCTGGACGAGGTG
   ATACACCAGGCCATGCGACACCAGGCCATGGAGGCGCAACTTCCAGTGCTCGTAAAAACA
25 GATGGGATGAAACCCCCAAAACAGAGAGAGATACTCCTGGGCATGGAAGTGGATGGGCTG
   AGACTCCTCGAACAGATCGAGGTGGAGATTCTATTGGTGAAACACCGACTCCTGGAGCCA
   GTAAAAGAAAATCACGGTGGGATGAAACACCAGCTAGTCAGATGGGTGGAAGCACTCCAG
   TTCTGACCCCTGGAAAGACACCAATTGGCACACCAGCCATGAACATGGCTACCCCTACTCC
   AGGTCACATAATGAGTATGACTCCTGAACAGCTTCAGGCTTGGCGGTGGGAAAGAGAAATT
30 GATGAGAGAAATCGCCCACTTTCTGATGAGGAATTAGATGCTATGTTCCAGAAAGGATATA
   AGGTAATTCCTCCTCCAGCTGGTTATGTTCTTATTCGAACTCCAGCTCGAAAGCTGACAGC
   TACTCCAACACCTTTGGGTGGTATGACTGGTTTCCACATGCAAACCTGAAGATCGAACTATG
   AAAAGTGTTAATGACCAGCCATCTGGAAATCTTCCATTTTTAAACCTGATGATATTCAATAC
   TTTGATAAACTATTGGTTGATGTTGATGAATCAACACTTAGTCCAGAAGAGCAAAAAGAGAG
35 AAAATAATGAAGTTGCTTTTAAAAATTAAGAATGGAACACCACCAATGAGAAAGGCTGCAT
   TCGCTCAGATTACTGATAAAGCTCGTGAATTTGGAGCTGGTCCTTTGTTTAATCAGATTCTT
   CCTCTGCTGATGTCTCCTACACTTGAGGATCAAGAGCGTCATTTACTTGTGAAAGTTATTGA
   TAGGATACTGTACAACTTGATGACTTAGTTCGTCCATATGTGCATAAGATCCTCGTGGTCA
   TTGAACCGCTATTGATTGATGAAGATTACTATGCTAGAGTGGAAGGCCGAGAGATCATTTT
40 TAATTTGGCAAAGGCTGCTGGTCTGGCTACTATGATCTCTACCATGAGACCTGATATAGATA
   ACATGGATGAGTATGTCCGTAACACAACAGCTAGAGCTTTTGCTGTTGTAGCCTCTGCCCT
   GGGCATTCTTCTTTATTGCCCTTCTTAAAGCTGTGTGCAAAAGCAAGAAGTCCTGGCAA
   GCGAGACACACTGGTATTAAGATTGTACAACAGATAGCTATTCTTATGGGCTGTGCCATCTT
   GCCACATCTTAGAAGTTTAGTTGAAATCATTGAACATGGTCTTGTGGATGAGCAGCAGAAA

```

GTTCCGACCATCAGTGCTTTGGCCATTGCTGCCTTGGCTGAAGCAGCAACTCCTTATGGTA
TCGAATCTTTTGATTCTGTGTTAAAGCCTTTATGGAAGGGTATCCGCCAACACAGAGGAAA
GGGTTTGGCTGCTTTCTTGAAGGCTATTGGGTATCTTATTCCTCTTATGGATGCAGAAATATG
5 CCAACTACTATACTAGAGAAGTGATGTTAATCCTTATTTCGAGAATTCCAGTCTCCTGATGAG
GAAATGAAAAAAATTGTGCTGAAGGTGGTAAAACAGTGTTGTGGGACAGATGGTGTAGAAG
CAAACATACATTAACACAGAGATTCTTCCTCCCTTTTTTAAACACTTCTGGCAGCACAGGATG
GCTTTGGATAGAAGAAATTACCGACAGTTAGTTGATACTACTGTGGAGTTGGCAAACAAAG
TAGGTGCAGCAGAAATTATATCCAGGATTGTGGATGATCTGAAAGATGAAGCCGAACAGTA
CAGAAAAATGGTGATGGAGACAATTGAGAAAATTATGGGTAATTTGGGAGCAGCAGATATT
10 GATCATAACTTGAAGAACAACCTGATTGATGGTATTCTTTATGCTTTCCAAGAACAGACTAC
AGAGGACTCAGTAATGTTGAACGGCTTTGGCACAGTGTTAATGCTCTTGGCAAACGAGTC
AAACCATACTTGCCTCAGATCTGTGGTACAGTTTTGTGGCGTTTAAATAACAAATCTGCTAA
AGTTAGGCAACAGGCAGCTGACTTGATTTCTCGAACTGCTGTTGTCATGAAGACTTGTCAA
GAGGAAAAATTGATGGGACACTTGGGTGTTGTATTGTATGAGTATTTGGGTGAAGAGTACC
15 CTGAAGTATTGGGCAGCATTCTTGGAGCACTGAAGGCCATTGTAAATGTCATAGGTATGCA
TAAGATGACTCCACCAATTAAGATCTGCTGCCTAGACTCACCCCCATCTTAAAGAACAGA
CATGAAAAAGTACAAGAGAATTGTATTGATCTTGTGGTTCGTATTGCTGACAGGGGAGCTG
AATATGTATCTGCAAGAGAGTGGATGAGGATTTGCTTTGAGCTTTTAGAGCTCTTAAAGCC
CACAAAAAGGCTATTCGTAGAGCCACAGTCAACACATTTGGTTATATTGCAAAGGCCATTG
20 GCCCTCATGATGTATTGGCTACACTTCTGAACAACCTCAAAGTTCAAGAAAGGCAGAACAG
AGTTTGTACCACTGTAGCAATAGCTATTGTTGCAGAAACATGTTCAACCCTTTACAGTACTCC
CTGCCCTTAATGAATGAATACAGAGTTCCTGAACTGAATGTTCAAAATGGAGTGTTAAATCG
CTTTCCTTCTTGTGTTGAATATATTGGTGAAATGGGAAAAGACTACATTTATGCCGTAACACC
GTTACTTGAAGATGCTTTAATGGATAGAGACCTTGTACACAGACAGACGGCTAGTGCAGTG
25 GTACAGCACATGTCACTTGGGGTTTATGGATTTGGTTGTGAAGATTCGCTGAATCACTTGTT
GAACTATGTATGGCCCAATGTATTTGAGACATCTCCTCATGTAATTCAGGCAGTTATGGGA
GCCCTAGAGGGCCTGAGAGTTGCTATTGGACCATGTAGAATGTTGCAATATTGTTTACAGG
GTCTGTTTCACCCAGCCCGGAAAGTCAGAGATGTATATTGGAAAATTTACAACTCCATCTAC
ATTGGTTCCCAGGACGCTCTCATAGCACATTACCCAAGAATCTACAACGATGATAAGAACA
30 CCTATATTCGTTATGAACTTGACTATATCTTATAATTTTATTGTTTATTTTGTGTTTAAATGCAC
AGCTACTTCACACCTTAAACTTGCTTTGATTTGGTGATGTAACTTTTAAACATTGCAGATCA
GTGTAGAACTGGTCATAGAGGAAGAGCTAGAAATCCAGTAGCATGATTTTTAAATAACCTGT
CTTTGTTTTTGTGTTAAACAGTAAATGCCAGTAGTGACCAAGAACACAGTGATTATATACA
CTATACTGGAGGGATTTTCATTTTAAATTCATCTTTATGAAGATTTAGAACTCATTCTTGTGT
35 TTAAAGGGAATGTTTAAATTGAGAAATAAACATTTGTGTACAAAATGCTAA

SEQ ID N° 3:

MAKIAKTHEDIEAQIREIQGKKAALDEAQGVGLDSTGYDQEIYGGSDSRFAGYVTSIAATELED
DDDDYSSSTSLLGQKKPGYHAPVALLNDIPQSTEQYDPFAEHRPPKIADREDEYKKHRRMTIIS
40 PERLDPFADGGKTPDPKMNARTYMDVMREQHLTKEEREIRQQLAEKAKAGELKVVNGAAASQ
PPSKRKRRWDQTADQTPGATPKKLSSWDQAETPGHTPSLRWDET PGRAKGSETPGATPGSK
IWDPTPSHTPAGAATPGRGDTPGHATPGHGGATSSARKNRWDET PKTERDTPGHGSGWAET
PRTDRGGDSIGETPTPGASKRKS RWDETPASQMGGSTPVLTPGKTPIGTPAMNMATPTPGHI
MSMTPEQLQAWRWEREIDERNRPLSDEELDAMFPEGYKVLPPPAGYVPIRTPARKLTATPTPL
45 GGMTGFHMQTEDRTMKSVNDQPSGNLPFLKPDDIQYFDKLLVDVDESTLSPEEQKERKIMKLL
LKIKNGTPPMRKAALRQITDKAREFGAGPLFNQILPLLMSPTLEDQERHLLVKVIDRILYKLLDLV
RPYVHKILVVIEPLLEDYARVEGREIISNLAKAAGLATMISTMRPDIDNMDEYVRNTTARAF

VVASALGIPSLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSLVEIIHGLVDEQQK
 VRTISALAI AALAE AATPYGIESFDSVLKPLWKIRQHRGKGLAAFLKAIGYLIPLMDAEYANYYT
 REVMLILIREFQSPDEEMKKIVLVVKQCCGTGVEANYIKTEILPPFFKHFQWHRMALDRRNY
 RQLVDTTVELANKVGAAEII SRIVDDLKDEAEQYRKMVMETIEKIMGNLGAADIDHKLEEQ LIDGI
 5 LYAFQEQTTEDSVMLNGFGTVVNALGKRVKPYLPQICGTVLWRLNNKSAKVRQQAADLISRTA
 VVMKTCQEEKLMGHLGVVLYEYLGE EYPEVLGSILGALKAI VNVIGMHKMT PPIKDLLPRLTPILK
 NRHEKVQENCIDL VGRIADRGAEYVSAREWMRICFELLELLKAHKKAIRRATVNTFGYIAKAIGP
 HDVLATLLNNLKVQERQNRVCTTVAIAIVAETCSPFTVLPALMNEYRVP ELNVQNGVLKSLSFLF
 EYIGEMGKDYIYAVTPLLEDALMDRDLVHRQTASAVVQHMSLGVYGF GCEDSL NHLLNYVWPN
 10 VFETSPHVIQAVMGALEGLRVAIGPCRMLQYCLQGLFHPARKVRDVYWKIYNSIYIGSQDALIAH
 YPRIYNDDKNTYIRYELDYIL

El término "*SNRNP200*" (también llamado en la literatura BRR2; RP33; HELIC2; ASCC3L1; U5-
 200KD) se refiere tanto al gen como a la proteína humana, subunidad 200 de la
 ribonucleoproteína nuclear pequeña U5. El splicing de pre-ARNm es catalizado por el
 15 spliceosoma, un complejo de subunidades de proteínas y ARN especializadas que elimina los
 intrones de un segmento de ARNm previo transcrito. El spliceosoma consiste en pequeñas
 proteínas de ARN nuclear (snRNPs) U1, U2, U4, U5 y U6, junto con aproximadamente 80
 proteínas conservadas. U5 snRNP contiene nueve proteínas específicas. Este gen codifica una
 de las proteínas específicas de snRNP U5. Esta proteína pertenece a la familia DEXH-box de
 20 ARN helicasas putativas. Es un componente central de los snRNPs U4 / U6-U5 y parece catalizar
 un desenrollado dependiente de ATP de los duplicados de ARN U4 / U6. Las mutaciones en este
 gen causan retinitis pigmentosa autosómica dominante. Alternativamente, se han encontrado
 variantes de transcripción empalmadas que codifican diferentes isoformas, pero no se ha
 determinado la naturaleza completa de estas variantes. En el contexto de la presente invención,
 25 "*SNRNP200*" se define también por una secuencia de nucleótidos o polinucleótido, que
 constituye la secuencia codificante de la proteína "*SNRNP200*", y que comprendería diversas
 variantes procedentes de:

- a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia
 aminoacídica de la SEQ ID NO: 5,
- 30 b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia
 polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del
 código genético,
- d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia
 35 aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la
 SEQ ID NO: 5. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad
 y las características estructurales de la proteína *SNRNP200*. Entre otras posibles secuencias
 nucleotídicas que codifican *SNRNP200* se encuentra, pero sin limitarse, la SEQ ID NO: 4.

Gene ID: 23020

Localización del gen: 2q11.2

SEQ ID N° 4:

```
5  GAACCCGTTTCTTTTCTTCCCCAGTGCGTCTTTCCTGCGTCGTTCCGGCGCGGCGGGAG
   CAGAGATCTGCGGCCGTTTGCAGCTTGCGGTAGGGAGGCGTGGTGGTCTGAAGCCTCCG
   AGCAGCCGCGGCCATGGCGGATGTAACCGCCCGTAGTCTGCAATACGAGTACAAGGCGA
   ACTCGAATCTTGTGCTCCAAGCTGACCGTTCTCTCATTGACCGGACCCGCCGGGATGAAC
   CCACAGGAGAGGTGCTGTCCCTTGTTGGGAAGCTGGAGGGCACCCGTATGGGAGACAAG
10  GCTCAACGGACCAAACCGCAGATGCAGGAGGAAAGAAGAGCCAAGCGAAGAAAGCGTGA
   TGAGGACCGGCATGACATCAACAAGATGAAGGGTTATACTCTGCTGTCTGGAGGGCATTGA
   TGAGATGGTGGGCATCATCTACAAGCCCAAACTAAAGAGACTCGGGAGACCTATGAGGT
   GCTACTCAGCTTCATCCAGGCTGCTCTTGGGGACCAGCCACGTGATATCCTTTGTGGGGC
   AGCTGATGAAGTTCTAGCTGTTCTAAAGAATGAAAAGCTGCGGGACAAGGAAAGGCGAAA
15  GGAGATTGACCTGCTGCTGGGTCAAACAGATGATACCAGATACCATGTGCTAGTGAACCTG
   GGCAAAAAGATCACAGACTATGGTGGAGATAAGGAAATCCAAAATATGGATGACAACATTG
   ATGAGACATACGGTGTGAATGTGCAGTTTGAGTCTGATGAGGAGGAAGGTGATGAAGACG
   TATACGGGGAGGTTTCGAGAAGAGGCATCTGATGATGACATGGAAGGGGACGAGGCTGTC
   GTGCGCTGCACCCTCTCGGCTAATCTCGTAGCCTCAGGTGAACTGATGAGTTCCAAGAAG
20  AAGGATTTGCACCCTCGGGATATTGATGCATTTTGGCTGCAGCGGCAGCTCAGTCGTTTCT
   ATGATGATGCCATCGTGTCTGCAGAAGAAGGCAGATGAAGTATTGGAGATTTTGAAGACGG
   CCAGTGATGATCGGGAATGTGAAAATCAGCTGGTTCTGCTGCTTGGTTTCAACACCTTTGA
   TTTCATTAAAGTGTTGCGGCAGCACAGGATGATGATTTTATACTGTACCTTGCTGGCCAGT
   GCACAAAGTGAAGCTGAAAAGGAAAGGATTATGGGAAAGATGGAAGCTGACCCAGAGCTA
25  TCCAAGTTCTCTACCAGCTTCATGAAACCGAGAAGGAGGATCTGATCCGAGAGGAAAGG
   TCCCGGAGAGAGCGAGTGCCTCAGTCTCGAATGGACACAGATCTGGAAACCATGGATCTC
   GACCAGGGTGGAGAGGCACTGGCTCCACGGCAGGTTCTGGACTTGGAGGACCTGGTTTTT
   ACCCAAGGGAGCCACTTTATGGCCAATAAACGCTGTCAGCTTCCTGATGGATCCTTCCGTC
   GCCAGCGTAAGGGCTATGAAGAGGTGCATGTGCCTGCTCTGAAGCCCAAGCCCTTTGGCT
30  CAGAAGAACAACCTGCTTCCAGTGGAAGGCTGCCAAAGTATGCCAGGCTGGGTTTGAGG
   GCTTCAAACACTGAATCGGATCCAGAGTAAGCTCTACCGTGCTGCCCTTGAGACGGATGA
   GAATCTGCTGCTGTGTGCTCCTACTGGTGCTGGGAAGACCAACGTGGCCCTGATGTGCAT
   GCTCCGAGAGATTGGGAAACACATAAACATGGACGGCACCATCAATGTGGATGACTTCAA
   GATTATCTACATTGCCCCCATGCGCTCCTTGGTGCAGGAGATGGTGGGCAGCTTTGGAAG
35  GCGCCTGGCCACTTATGGCATCACTGTTGCTGAACTGACTGGGGACCACCAGCTGTGCAA
   AGAAGAGATCAGTGCCACTCAGATCATCGTCTGCACCCCGAGAAGTGGGACATCATCAC
   CCGCAAGGGTGGTGAGCGCACCTACACCCAGCTGGTGCGGCTCATCATTCTGGATGAGAT
   TCATCTTCTCCACGATGACAGAGGTCCTGTCTTAGAAGCTTTAGTGGCCAGGGCCATCCGA
   AACATTGAGATGACCCAAGAGGATGTCCGACTCATTGGTCTCAGTGCCACCCTACCCAAC
40  ATGAAGATGTAGCCACCTTTCTACGTGTTGACCCTGCCAAGGGTCTCTTTTACTTTGACAAC
   AGCTTCCGTCCAGTGCCTCTGGAACAGACATATGTGGGTATCACAGAGAAAAAAGCTATCA
   AGCGTTTCCAGATCATGAATGAAATCGTCTATGAAAAAATCATGGAACATGCTGGAAAAAT
   CAGGTGCTGGTGTGTTGTCCACTCCCGGAAGGAGACTGGAAAGACAGCCAGGGCCATCCG
   GGACATGTGCCTAGAAAAGGACACTCTGGGTCTGTTTCTGAGGGAGGGGCTCAGCCTCCAC
45  AGAAGTCCTGCGAACAGAAGCTGAGCAGTGCAAGAACCTAGAGCTGAAGGATCTTCTGCC
   TTATGGCTTTGCTATTCATCACGCAGGCATGACCAGGGTTGACCGAACACTCGTGGAGGAT
```

CTTTTTGCTGATAAACATATTCAGGTTTTAGTTTTCCACAGCAACTCTAGCTTGGGGTGTGAA
TCTCCCTGCACATACAGTCATCATCAAAGGCACCCAGGTGTACAGTCCAGAGAAGGGGCG
TTGGACAGAACTGGGAGCACTGGACATTCTGCAGATGCTGGGACGTGCCGGAAGACCCCA
GTATGACACCAAGGGTGAAGGCATACTCATCACATCTCATGGGGAGCTACAGTACTACCTG
5 TCCCTCCTCAATCAACAACTTCCTATTGAAAGCCAGATGGTTTTCAAAGCTTCCTGACATGCT
CAATGCAGAAATCGTGCTAGGAAATGTCCAGAATGCCAAGGATGCGGTGAACTGGCTGGG
CTATGCCTACCTCTATATCCGAATGCTGCGATCCCCAACCCCTCTATGGCATCTCTCATGAT
GACCTCAAGGGAGATCCCCTGCTGGACCAGCGCCGACTAGATCTGGTTCATACAGCTGCC
CTGATGCTGGACAAGAACAATCTGGTCAAGTACGACAAGAAGACGGGCAACTTCCAGGTG
10 ACAGAACTGGGCCGTATAGCCAGCCACTACTACATCACCAATGATACAGTGCAGACTTACA
ACCAGCTGCTGAAGCCCACCCTGAGTGAGATTGAGCTTTTCAGGGTCTTCTCATTGTCTC
TGAGTTCAAGAACATCACAGTGAGAGAGGAGGAGAAGCTGGAGCTGCAGAAGTTGCTGGA
GAGGGTGCCTATCCCTGTAAAGGAGAGCATTGAGGAACCCAGTGCTAAGATCAACGTTCTT
CTGCAAGCCTTCATCTCACAGCTGAAATTGGAGGGCTTTGCACTGATGGCTGACATGGTGT
15 ATGTCACACAGTCGGCTGGCCGGTTGATGCGAGCGATATTTGAAATTGTCTGAACCGAG
GTTGGGCACAGCTTACAGACAAGACCCTGAACCTCTGCAAGATGATCGACAAACGCATGT
GGCAGTCCATGTGTCTCTGCGCCAGTTCCGGAAACTCCCTGAGGAAGTAGTGAAGAAGA
TTGAGAAGAAGAATTTCCCCTTTGAGCGTCTGTACGACCTGAATCATAATGAGATTGGGGA
GCTTATCCGCATGCCAAAGATGGGGAAGACCATCCACAAATATGTCCATCTGTTTCCCAAG
20 TTGGAGTTGTCAAGTGCACCTGCAGCCTATCACACGCTCCACCCTGAAGGTGGAGCTGACC
ATCACGCCAGACTTCCAGTGGGATGAAAAGGTGCATGGTTCATCCGAGGCTTTTTGGATT
TGGTGGAGGATGTGGACAGCGAGGTGATTCTGCACCATGAGTATTTTCTCCTCAAGGCCA
AGTACGCCCAGGACGAGCACCTCATTACATTCTTCGTGCCTGTCTTTGAACCGCTGCCCCC
TCAGTACTTCATCCGAGTGGTGTCTGACCGCTGGCTCTCTTGAGACCCAGCTGCCTGTC
25 TCCTTCCGGCACCTGATCTTGCCGGAGAAGTACCCCCCTCCAACCGAACTTTTGGACCTG
CAGCCCTTGCCCGTGTCTGCTCTGAGAAACAGTGCCTTTGAGAGTCTTTACCAAGATAAAT
TTCCTTTCTTCAATCCCATCCAGACCCAGGTGTTTAACTGTATACAACAGTGACGACAAC
GTGTTTGTGGGGGCCCCCACGGGCAGCGGGAAGACTATTTGTGCAGAGTTTGCCATCCTG
CGAATGCTGCTGCAGAGCTCGGAGGGGCGCTGTGTGTACATCACCCCCATGGAGGCCCT
30 GGCAGAGCAGGTATACATGGACTGGTACGAGAAGTTCCAGGACAGGCTCAACAAGAAGGT
GGTACTCCTGACAGGCGAGACCAGCACAGACCTGAAGCTGCTGGGCAAAGGGAACATTAT
CATCAGCACCCCTGAGAAGTGGGACATACTTTCCCGCGATGGAAGCAGCGCAAGAACGT
GCAGAACATCAACCTCTTCGTGGTGGATGAGGTCCACCTTATCGGGGGCGAGAATGGGCC
TGTCTTAGAAGTGATCTGCTCCCGAATGCGCTACATCTCCTCCAGATTGAGCGGCCCAT
35 CGCATTGTGGCACTCAGCTCTTCGCTCTCCAATGCCAAGGATGTGGCCCACTGGCTGGGC
TGCAGTGCCACCTCCACCTTCAACTTCCATCCCAATGTGCGTCCCGTCCCCTTGGAGCTGC
ACATCCAGGGCTTCAACATCAGCCATACACAAACCCGCCTGCTCTCCATGGCCAAGCCTGT
GTACCATGCTATCACCAAGCACTCGCCCAAGAAGCCTGTCATTGTCTTTGTGCCGTCTCGC
AAGCAGACCCGCCTCACTGCCATTGACATCCTCACCACCTGTGCAGCAGACATCCAACGG
40 CAGAGGTTCTTGCACTGCACCGAGAAGGATCTGATTCCGTACCTGGAGAAGCTAAGTGAC
AGCACGCTCAAGGAAACGCTGCTAAATGGGGTGGGCTACCTGCATGAGGGGCTCAGCCC
CATGGAGCGACGCCTGGTGGAGCAGCTCTTCAGCTCAGGGGCTATCCAGGTGGTGGTGG
CTTCTCGGAGTCTCTGCTGGGGCATGAACGTGGCTGCCACCTGGTAATCATCATGGATA
CCCAGTACTACAATGGCAAGATCCACGCCTATGTGGATTACCCCATCTATGACGTGCTTCA
45 GATGGTGGGCCACGCCAACCGCCCTTTGCAGGACGATGAGGGGCGCTGTGTGCATCATGT
GTCAGGGCTCCAAGAAGGATTTCTTCAAGAAGTTCTTATATGAGCCATTGCCAGTAGAATC
TCACCTGGACCACTGTATGCATGACCACTTCAATGCTGAGATCGTCACCAAGACCATTGAG
ACAAGCAGGATGCTGTGGACTACCTCACCTGGACCTTTCTGTACCGCCGCATGACACAG
AACCCCAATTACTACAACCTGCAGGGCATCTCCCATCGTCACTTGTCCGACCACTTGTGAG

AGCTGGTGGAGCAGACCCTGAGTGACCTGGAGCAGTCCAAGTGCATCAGCATCGAGGAC
GAGATGGACGTGGCGCCTCTGAACCTAGGCATGATCGCCGCCTACTATTACATCAACTACA
CCACCATTGAGCTCTTCAGCATGTCCCTCAATGCCAAGACCAAGGTGCGAGGGGCTTATCG
AGATCATCTCCAATGCAGCAGAGTATGAGAACATTCCCATCCGGCACCATGAAGACAATCT
5 CCTGAGGGCAGTTGGCTCAGAAGGTCCCCCACAAGCTGAATAACCCTAAGTTCAATGATCC
GCACGTCAAGACCAACCTGCTCCTGCAGGCTCACTTGTCTCGCATGCAGCTGAGTGCTGA
GTTGCAGTCAGATACGGAGGAAATCCTTAGTAAGGCAATCCGGCTCATCCAGGCCTGCGT
GGATGTCCTTTCCAGCAATGGGTGGCTCAGCCCTGCTCTGGCAGCTATGGAAGTGGCCCA
GATGGTCACCCAAGCCATGTGGTCCAAGGACTCATACCTGAAGCAGCTGCCACACTTCAC
10 CTCTGAGCATATCAAACGTTGCACAGACAAGGGAGTGGAGAGTGTTTTCGACATCATGGAG
ATGGAGGATGAAGAACGGAACGCGTTGCTTCAGCTGACTGACAGCCAGATTGCAGATGTG
GCTCGCTTTTGTAAACGCTACCCTAATATCGAACTATCTTATGAGGTGGTAGATAAGGACA
GCATCCGCAGTGGCGGGCCAGTTGTGGTGGTGGTGCAGCTGGAGCGAGAGGAGGAAGTC
ACAGGCCCTGTCATTGCGCCTCTCTTCCCGCAGAAACGTGAAGAGGGCTGGTGGGTGGTG
15 ATTGGAGATGCCAAGTCCAATAGCCTCATCTCCATCAAGAGGCTGACCTTGCAGCAGAAG
GCCAAGGTGAAGTTGGAAGTTTGTGGCCCCAGCCACTGGTGCCCACTACACTCTGTAC
TTCATGAGTGACGCTTACATGGGATGTGACCAGGAGTACAAATTCAGCGTGGATGTGAAAG
AAGCTGAGACAGACAGTGATTGAGATTGAGTCCTGAGGCATTTACTTTTGGGTAAAGGAGA
GTTGAGCCTGAATTAGGAATGTGTACATTGTAGGAATCCTGGTTGTGGGGACCAGGTCTGT
20 GGGCCTCAGGTCTGGCCAGCCAGGGCTGGTGTCTGTCCCCGCCTACCTCCACTTCCTTTCC
CTTGCTCACTCTGGATCCAGTGACAGCAGGTGTCATGGGTCAAGCATAAATCATATATAGC
ATTTTCAGGCATGTTCTGCTAGTTCTTTTGTAGTCTGACATTCTAATAAAATAATTTGTAGAA
ACCATTTGTCTTTGTAGTGATTCCAAATTAAGTTTCTTTCTCCAACCTGAGGGCACGGC
CAAAAAGATCTGGTTATTTTTTAGCCAGGAACGTGCTTGTTAATGAGTATGTCTGGAGGACA
25 GACCTGCTCATTAGGTGTGCTGTCCCTGTAGCCTCGTGAGTCAGCCAGAGGAGGGTAC
ATGCGACTGTGGCCTGGCCTCAGTGGTACCCACACATCAGCACTACCACAAGAACCAACA
CTGAGCCTCGGAAGCTAGATCACAGGTTAGGGGTTTCTCTAGATGGGGGTTCTGAAATTTG
CAGTGTCTGCTCCTGGGAGGCAGCACCAGAAAGGGCACTGAAATGTACTAGCTGGATGTG
ACCCAGTCTTAATAAACAGGTTTTCTAATCCAGAAAAAAAAAAAAA

30

SEQ ID N° 5:

MADV TARSLQY EYKANSNLVLQADRSLIDRTRRDEPTGEVLSLVGKLEGTRMGDKAQRTPKQ
MQEERRAKRRKRDEDRHDINKMKGTYLLSEGIDEMVGIIYKPKTKETRETYEVLLSFIQAALGD
QPRDILCGAADEVLA VLKNEKLRDKERRKEIDLLLQQTDDTRYHVLVNLGKKITDYGGDK EIQN
35 MDDNIDETYGVNVQFESDEEEGEDVYGEVREEASDDDMEGDEAVVRCTLSANLVASGELMS
SKKKDLHPRDIDAFWLQRQLSRFYDDAIVSQKKADEVLEILKTASDDRECENQLVLLLGFNFTD
FIKVLQRHRMMILYCTLLASAQSEAEKERIMGKMEADPELSKFLYQLHETEKEDLIREERSRRER
VRQSRMDTDLETMDLDQGG EALAPRQVLDLEDLVFTQGS HFMANKRCQLPDG SFRRQRKGY
EEVHVPALPKPKPGFSEEQLLPVEKLPKYAQAGFEGFKTLNRIQSKLYRAALETDENLLLCAPTG
40 AGKTNVALMCMLREIGKHINMDGTINVD DFKIIYIAPMRSLVQEMVGSFGKRLATYGITVAELTG
DHQLCKEEISATQII VCTPEKWDIITRKGGERTYTQLVRLIILDEIHLHDDRGPVLEALVARAIRNI
EMTQEDVRLIGLSATLPNYEDVATFLRVDPAKGLFYFDNSFRPVPLEQTYVGITEKKAIRFQIM
NEIVYEKIMEHAGKNQVLV FVHSRKETGKTARAI RDMCLEKDTLGLFLREGSASTEVLRTEAEQ
CKNLELKDLLPYGFAIHHAGMTRVDRTLVEDLFADKHIQVLVSTATLAWGVNLP AHTVIKGTQV
45 YSPEKGRWTEL GALDILQMLGRAGRPQYDTKGEGILITSHGELQYYLSLLNQQLPIESQMVSKL
PDMLNAEIVLGNVQNAKDAVNWLGYAYLYIRMLRSPTLYGISHDDLKGDPLLDQRRRLDLVHTAA
LMLDKNNLVKYDKKTGNFQVTELGRIASHYYITNDTVQTYNQLLKPTLSEIELFRVFSLSSEFKNI

TVREEEKLELQKLLERVPIPVKESIEEPSAKINVLLQAFISQLKLEGFALMADMVYVTQSAGRLM
 RAIFEIVLNRGWAQLTDKTLNLCKMIDKRMWQSMCPLRQFRKLPEEVVKKIEKKNFPFERLYDL
 NHNEIGELIRMPKMGKTIHKYVHLFPKLELSVHLQPITRSTLKVELTITPDFQWDEKVGHSSEAF
 WILVEDVDSEVILHHEYFLLKAKYAQDEHLITFFVPVFEPLPPQYFIRVWSDRWLSCETQLPVSF
 5 RHLILPEKYPPPTTELLDLQPLPVSA LRNSAFESLYQDKFPFFNPIQTQVFNTVYNSDDNVFVGAP
 TGSGKTICAEFAILRMLLSSEGRVCYITPMEALAEQVYMDWYEKFQDRNLNKKVLLTGETSTD
 LKLLGKGNIIISTPEKWDILSRRWKQRKNVQNINLFVVDEVHLIGGENGPVLEVICSRMRYISSQI
 ERPIRIVALSSSLNAKDV AHWLGCSATSTFNHFNVRPVPLELHIQGFNISHTQTRLLSMAKPV
 YHAITKHSPKKPVIVFVPSRKQTRLTAIDILTTCAADIQRQRFLHCTEKDLIPYLEKLSDSTLKETLL
 10 NGVGYLHEGLSPMERRLVEQLFSSGAIQVVVASRSLCWGMNVAAHLVIIMDTQYYNGKIHAYV
 DYPIYDVLQMVGHANRPLQDDEGRVCVIMCQGSKKDFFKKFLYEPLPVESHLDHCMHDHFNAEI
 VTKTIENKQDAVDYLTWTFLYRRMTQNPNYNQLQGISHRHLSDHLSERVEQTLSDLEQSKCISIE
 DEMDVAPLNLGMIAAYYYINYTTIELFSMSLNAKTKVRGLIEIISNAAEYENIPIRHHEDNLLRQLA
 QKVPHKLNNPKFNDPHVKTNLLLQAHL SRMQLSAELQSDTEEILSKAIRLIQACVDVLSSNGWL
 15 SPALAAMELAQMVTQAMWSKDSYKQLPHFTSEHIKRCCTDKGVESVFDIMEMEDEERNALLQL
 TDSQIADVARFCNRYPNIELSYEVVDKDSIRSGGPVVVLVQLEREEVEVTGPVIAPLFPQKREEG
 WWWVIGDAKSNSLISIKRLTLQQKAKVKLDFVAPATGAHNYTLYFMSDAYMGCDQEYKFSVDV
 KEAETDSDSD

20 El término "*RNU12*" (también llamado en la literatura *RNU12L*; *RNU12P*; *RNU12-1*; *dJ222E13.7*)
 se refiere al pequeño ARN nuclear (snRNP) codificado por este gen es parte del complejo de
 spliceosoma menor dependiente de U12. Además del ARN codificado, este complejo de
 ribonucleoproteínas consiste en snRNPs U11, U12, U5 y U6atac. En el contexto de la presente
 invención, "*RNU12*" se define también por una secuencia de nucleótidos o polinucleótidos, que
 25 constituye la secuencia del snRNP "*RNU12*", y que comprendería diversas variantes procedentes
 de:

a) Moléculas de ácido nucleico que comprende la SEQ ID NO: 6,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia
 polinucleotídica de a),

30 c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del
 código genético,

Gene ID: 267010

Localización del gen: 22q13.2

SEQ ID N° 6:

35 TGCCTTAACTTATGAGTAAGGAAAATAACGATTCGGGGTGACGCCCGAATCCTCACTGCT
 AATGTGAGACGAATTTTGTAGCGGGTAAAGGTGCGCCTCAAGGTGACCCGCCTACTTTGC
 GGGATGCCTGGGAGTTGCGATCTGCCCG

El término "*ESRP1*" (también llamado en la literatura RBM35A; RMB35A; DFN109) se refiere tanto al gen como a la proteína humana. *ESRP1* es un regulador de *splicing* epitelial. En el contexto de la presente invención, "*ESRP1*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*ESRP1*", y que comprendería diversas variantes procedentes de:

- a) moléculas de ácido nucleico que codifican un polipéptido que comprende la SEQ ID NO: 7,
- b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

Gene ID: 54845

Localización del gen: 8q22.1

SEQ ID N° 7:

```

ACAGGTGGCAGCTCTCGTCCCCTGAGAGCGGGCGAAGGCCAGGGTCCCACACTCGCAGC
GCTGCGACGGCGCTCGGGACCTCCCTCGTCCACTGCTTGAGTTCCAGAGGTGGGTGCTT
CCCTGTCCTGAACCTCAGAGTGCAGGAGTCATAAATGGGTTCCGGCTGGCTTACTGCAGTA
GCCTCGCTCCTCCCCAGCCCCGGTAACCTCCGAGCTACCCGTCCAGGCCCTCGGGCGTCCG
CGGGGGCAGGGACTGGGCGCGGAACGAGGCAGGGAGGGACCTGGAAAAACCAACCCAGA
TTGCATTGCAGTGGGCGAGGCCGCTTGGAGGAGCCGGTTCCCCCTAACCACTCCCCGT
GGGGCTCTCGGTGCGCGGTTCCCAGGTGCTCAGCTCTGCTGGGCCCAGGAGGTGCCGCC
TCACAGGGACGCGGAACCCCGTGCGTGGCCCCCGCCGGGTGGAACAGATAGCGCGGGG
CGGTCCGGAGGCTCGTCGCCAAGCAGGTGACTCTTGCTGAAAAAGTGGTTGGAACACTTG
AGGAAACCCGGCCCCGCTGTTCTTTCTAGGTCTTTGGAGTTTGGATTAATCATTTGTGTA
GCCCCGTTTGGATAAACCGAAGACTTTATTAATCAGCGCGTTTAACAGGAATTCCGCAGTA
GTATCCACATTAGAATCTTGAGTCTTGAGTGAACATATTCACACAGACTTGCCTTCTTCC
TGTTTAGTTTATGCCTTGTGTTCCGTTATTGGAACGCTAAGCTTGTGGGAGTTGTTTACATC
CTACTGCTCAAGGTCATCGCTAAGGTGTGATTTTTTCAAAAAAGAATTTGCAACCTCCGGC
ATGAATGACTTAAGGGAAGTCTAATCCCGGTTTCTGATTTTTTTTTTTTTTTTAAATTTAAAGT
TAATCTTTCTGGGCCGGGCGCGGTGGCTCACGCTGTAATCCAGCACTTTGGGAGGCCG
AGGCGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCT
CTACTAAAAACAAAAAATTAGCCGGGCGGGGTGGCGCGCACCTGTAATCCAGCTGCTC
GGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGGGGGTTGCAGTGAGCCG
AGATCTGGCCATTGCACTCCAGCGTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAG
GTTAATCTTTCCAACCTAGATTTTCAAGGATGAGGATTTTGTGTTGTTGTTGTTGTTGTTCTC
AAATGTATTCCCAGGGCTTGGAACAGAGCCTGACATATACTAGGCACTCAACAAATATTTGT
TGAATGATTGTAATGAGTAACACCCATTTTGCAGATCTTTGTCTTCTGAGCCTAGGGCATA
GGTCATCACTGCAGGGGTGAGATTGTCAAAATGGGAGTCTACAGGTAATTTAAGACTTAAA
TGTTTAAAGAGTATGTGCTCATTCTTCAACAACTTACTTTTGTAAATTTAAATGGTAAAT
GTGGTGGAGGGGTTGGAATATATGTAATTCAAGACAGTTCTGAATACAAAAATGTTTTACTG
TCTATCACCAACCATCTATAAATCTAATTCATAAGGATAATCTGTGTAAGGTGGCTGGAAAG
AACCTTGAGGAGAGAGGCTTATTTAAGTATTGGCTCAGGACCACACCTAAAATTCTCAAAA

```


CGTTGAGATTCTGTTGTTTTGTTTTTAAGCGCCAGAGACCCAAGTTGAGGAACAGCCTATAA
 AATAACTGGCCTGTACTCTTACATACATGAAAGCCATCAAAGACAAAGACTGAAGAAGAACT
 TTTGCAGATTAAAGGACTTTAAGAGACATGATCCTGAACCAGAAAAAGTAAAAAATAAAAAA
 ATAAAAAGATCTGTAGATTAGATAATAGTTTTATATCCATGTTAATTTCCCATTTGTTGTAATA
 5 TTAAGAGAGGAATCATGTCTGTAAGTGTCTTTGTTGGTGGGAAGTATCTACTGAAGTATTTAGGG
 GTAAAGAGGAATCATGTCTGTAAGTGTCTTTGTTGGTGGGAAGTATCTACTGAAGTATTTAGGG
 AATGATAAAGCAAACGCAATAAAATGTTAACATTGGGGAATCTGTATAAGGGAATTATTAAC
 TTTTCTGTAAGTCTAAAATTATTTCAAATTAAAAATAATTTTTTTTAAGCCTC

10 El término "*ESRP2*" (también llamado en la literatura *RBM35B*) se refiere tanto al gen como a la proteína humana. *ESRP2* es un regulador de *splicing* epitelial. En el contexto de la presente invención, "*ESRP2*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*ESRP2*", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 9,

b) moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

20 d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 9. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *ESRP2*. Entre otras posibles secuencias nucleotídicas que codifican *ESRP2* se encuentra, pero sin limitarse, la SEQ ID NO: 8.

25

Gene ID: 80004

Localización del gen: 16q22.1

SEQ ID N° 8:

AGCCGGGCGCGAGGTGCCGCCCCCGATACACGGTGTCCCGCCCAAGCTGATCCGCGTCT
 30 GCGGTTCGGTCGGTGCGTGCGTGCGCCTCGTCGGTCCGCGTGTCTGGCCGAGAGCCCCCT
 TCCTCTGCGGCCATGACTCCGCGCGCGCGCGCCCCCTCCCCGGGGCCCTGACCCCGC
 GGCCGACCCCGCGCGGACCCCTGCCCTGGCCCGGATCACTGGTCGTCTCTTCGGGG
 CTACGGCGGGTGCGCTGGGACGGGACCTGGGCTCGGACGAGACCGACTTAATCCTCCTA
 GTTTGGCAAGTGTTGAGCCGCGGAGCCGCCAGGTGGGGACGCTGCACAAATCGCTGGT
 35 TCGTGCCGAGGCGGCCGCACTGAGTACGCAGTGCCGCGAGGCGAGCGGCCTGAGCGCC
 GACAGCCTGGCGCGGGCAGAGCCGCTGGACAAGGTGCTGCAGCAGTTCTCACAGCTGGT

GAACGGGGATGTGGCTTTGCTGGGCGGGGGCCCCCTACATGCTCTGCACTGATGGGCAGC
AGCTATTGCGACAGGTCCTGCACCCCGAGGCCTCCAGGAAGAACCTGGTGCTCCCCGACA
TGTTCTTCTCCTTCTATGACCTCCGAAGAGAATTCCATATGCAGCATCCAAGCACCTGCCCT
GCCAGGGACCTCACTGTGGCCACCATGGCACAGGGTTTAGGACTGGAGACAGATGCCAC
5 AGAGGATGACTTTGGGGTCTGGGAAGTCAAGACAATGGTAGCTGTTATCCTCCATCTACTC
AAAGAGCCCAGCAGTCAATTGTTTTCTGAAGCCCGAGGTGATAAAGCAGAAATACGAGACG
GGGCCTTGCAAGGCTGATGTGGTGGACAGTGAGACTGTGGTACGGGCTCGTGGGTTGCC
GTGGCAGTCATCAGACCAGGACGTGGCTCGCTTCTTCAAAGGGCTCAACGTGGCCAGGG
GTGGTGTAGCACTCTGCCTCAACGCCCAGGGCCGCAGAAATGGCGAGGGCCCTCATCCGC
10 TTTGTGGACAGCGAGCAGCGGGACCTAGCGCTGCAGAGACACAAGCACCACATGGGCGT
CCGCTATATTGAGGTGTATAAAGCGACAGGGGAGGAGTTTGTAAAGATTGCAGGGGGCAC
ATCACTAGAGGTGGCTCGTTTCTTGTACGGGAAGACCAAGTGATCCTGCGGCTGCGGGG
ACTGCCCTTCTCGGCTGGGGCCAACGGACGTGCTTGGCTTCTTGGGGCCAGAGTGCCAG
TGACTGGGGGTACCGAGGGGCTGCTCTTTGTGCGCCATCCTGATGGCCGGCCGACTGGT
15 GATGCCTTCGCCCTCTTTGCTTGTGAGGAGCTGGCACAGGCTGCACTGCGCAGGCACAAG
GGCATGCTGGGTAAAGCGATACATTGAACTCTTCCGGAGCACTGCAGCCGAAGTGCAGCAG
GTCTTGAACCGCTATGCATCCGGCCCACTCCTTCTTACACTGACTGCCCCACTGCTGCCCA
TCCCCTTCCCACTGGCACCTGGGACTGGGAGGGACTGTGTACGCCTCCGAGGCCTGCC
TACACGGCCACCAATTGAAGACATCCTGAGCTTCTTGGGGGAGGCAGCAGCTGACATTCGG
20 CCCCACGGTGTACACATGGTGCTCAACCAGCAGGGCCGGCCATCGGGCGATGCCTTCATT
CAGATGACATCAGCAGAGCGAGCCCTAGCTGCTGCTCAGCGTTGCCATAAGAAGGTGATG
AAGGAGCGCTACGTGGAGGTGGTCCCCTGTTCCACAGAGGAGATGAGCCGAGTGCTGAT
GGGGGGCACCTTGGGCCGCAGTGGCATGTCCCCTCCACCCTGCAAGCTGCCCTGCCTCT
CACCACCTACCTACACCACCTTCCAAGCCACCCCAACGCTCATTCCCACGGAGACGGCAG
25 CTCTATACCCCTCTTCAGCACTGCTCCCAGCTGCCAGGGTGCTGCTGCCCCACCCCTG
TTGCCTACTATCCAGGGCCAGCCACTCAACTCTACCTGAACTACACAGCCTACTACCCAAG
CCCCCAGTCTCCCCCACCCTGTGGGCTACCTCACTACACCCTGCTGCCCTGGCCTC
TGCTCCCACCTCAGTGTTGTCCCAGTCAGGAGCCTTGGTCCGCATGCAGGGTGTCCATA
CACGGCTGGTATGAAGGATCTGCTCAGCGTCTTCCAGGCCTACCAGCTACCCGCTGATGA
30 CTACACCAGTCTGATGCCTGTTGGTGACCCACCTCGCACTGTGTTACAAGCCCCCAAGGA
ATGGGTGTGTTTGTAGGAGAGAAAGCCAGGAGGTAAGAGCCAGCTGATATCCTCGGCGAA
CATGTCTCTCCTGAGTCCAGAAGACCAGCACCTCAACCTGGTAGCTTCTTTCTGGCTTGT
CAAAGCTCTCAGAAGGTACCTAGAGGAGCCCAAGCCCCAGCTCCATCCTCCACTTATTCTG
CCTGTTTCCCCCAAAGACAATGGCTGGACCCTGCATGCAGGGCTGGGGGTGGAATGGGG
35 CTAACCAGCTCCTGATGGCCTGAGCCAGGCATCTTGACTGGCACCTGGAGAGCCCTTAAG
TCTGTCTTGGCTGTGGCCCATGCCGACAGATATCGTGGGGCTGACAGGTCCACGGCAGG
CTTGCTTTCTTTTATAAAATGGAAGCTCTGGTACCTTCAATGTATGACTCCTGGGAGAATCA
AGGGTCCATCTGAGCCTCTGAGTAAAGATCCCAATGTTCTACCTCTCCCTGTCCCTCTTGT
AGGGGATAGGGAGGCAGAGAGAGCCAGCCCCCTACCCTCAGAGTATCTGGACCTCAGAGA
40 CCATGTTGTGCCAGGGGTGGTCCCACCTAAAGATGCTAGCCCCTCTCCAGGTGGGCATAA
GGAGTAACAGATGGCAAAACCACAAACTATTTTGTATGGACTGTGCTGCAGTATCACCAGAA
GACATTAGGGGGCAGTAGGCCCCCACACAAAACCTTCAGGCTTGAATTTTAAAGGGGAGG
ACTTTCTGCCAACTTTTCTTGTATGCCTTGGGAAAGCCAGTTGCCCTGAACCCAGCAGACA
CCATGGAATGTCTTTGCACGCATTAAATGGTACAGAACTGAAGCCTCGGAAGCAATTTGG
45 AACTCGATCTTCTCTTCTTAAATGAAAAGTTATTGACCAAATGGACTTTTTTAAAGACACAG
GACCTTAACTTTGCCCCAAAGTGAGGGGCTCCACACCAACCCCAGGCGGAGGAACACTC
AGACAGATTAAGGATACTGTTGACCTGTCACTGTTTATTATTTTACGCACTAAAACCTGAGGAG
CCTCAACTGCTGGCTCTTCTTCCCTTTGTATTTGTGTAAGGAGCACTGCACTCCCATAAAAG
GTTTTAAATACAAAATGTACAAGAACACACAATTCCAAGTGCTGTAACATAACTGAGAAC

CAGTTCCTTTACTAAACATCCATTTTATAAAACACAAGGTTTCAATTTGAGCCCATCTGAGC
CTTAAAGATCCATTCTGAATACCAAAAACAGGGCTTCACAGCCAGGCCCAGAAGAGGTCTG
GTGATAATGGCTGGCCCTGGGTGGGGATAGTTTACACCCGGGCAGCAGCACCACACATGA
ACCCAAAGACATGTTCTTTTTAAAGCTGTTTTTCAGCCATGTTTCTCTGTGCATCTCCAGTAA
5 GCAGAAAGGCTACCCATTCCATTCTCAACCCAAGAGCTAGCACAGTTAGAGTAGGAGGGG
GTGCGTACTAGCACGTGCCCAGTTGCTCAGTGCTGCTAGTAGAAATTGATTTGCATAGTCC
AATGGATGTGTGCTTTAACACCACTATGTTGCACAAAAATTTAAGTCTTTATCTACAAAGCC
AAAAAATATTGACTCTTAACACCAAAGCTTTTACAAAGCTGATATAAAACTGCTTACATAGTA
TACAAAGCTCTATTTTAAAATTTAATGTTTATTTTAAATAGGAAAGCATTCTAGTGCTACAG
10 GCATCAAGGTATTTAGAAAGGCATCAAAATTTGGTGCATAGCAACTCTGGATCATAGAGGCT
TTTAATTCTTGAGGGCACAAGGGTATTGCAAGGGGAAGTCTTGCAATAACCTCATGTGAG
GCAGCTACACCAGGGGGCAAAGGGTAGGGGGCAAGTCTGGCCATTTCTCTGGAATGTCAAC
CGAGAGGCCAACTTCTTGCCCTCCAAAGGAAGCTGTGGGCGTCCAACCACAGGCCCAGA
AGCAAATCTCACTATCCACATTCTGTGACACAACCTCACTTTTCAAGAAAATTTGAAAACACC
15 CAGCTCCAAGAGGGGGCCACAGTGATCTGCTGAAGAATAGACCTTTTTCTCTAACTAACTTT
AGGGGAGAAATCTAAAATATATTTTGTTTTTTTTAAAGACCATACTTCCTCATCCTGAGTCAA
CAGCTCCTAACCTACTGACAACAGGTGGTGGGCAATAGGCCAGCTCCCTTTCCAGACCT
GAACTGGCCTTAGTAGAGAGAGTCCCTGGCATCAAGGGCCTGCAGTCTCCTGGAATAACAG
CATCAGGCCTTCAACCAGCACACTGTCTCATATATAGGCACTCAAAAAATCTGCTTCCTGA
20 ATCAATGGAAGTATATTATTTTCAATTTGAAAGCATACTGCAAACCTTCAGAGGCTCTGTAATGG
GCCCAGGAACAGGTAAAAAAAATCCCATTAGGGAGAGGGGTAAAGGGCAGGTTTTTTTTTTT
TTTAAAGTACCCAACCTCGGGTACCTCCTTTGGTAGACAGCAGGCTGCAGTTTCTGCTGGGC
CCTGCCTTGGACACCAGTCTCACCAGCAGCAGCACCAGGGTGTCTTTGGTGTGGAGGGTC
AGGTTATCACTTCTGGATGTGCTCAGAGCCTTCTGGGGAAAAGGCTGGTTTTTATGATCATA
25 CAGACAGGACAGGGTAGGCCAGAATCATTTAGGCTATAAATAGAATGGCAGATAAGCCTT
CCATGCCAGTGTGACTCTTAAACCTTCCCATCTGGAGTCTGTATTTTCTCTGCCCATGCT
GATGTACTGCTCATTGTGCTTTAGCTCTGCACACAGCCTGGCCTCTAGCTAGAGTTAGAA
CTGCAAAGCTGCCTAAACCACTGAGACTGGCTTCAAAGGTCACAGAACAGCCAAAGGCAG
ATGCTTCTGTTTGGAAAAGGTTCTACGACCAGATGACACAAACAGGAAGCGCACTCCACAG
30 CCTACTAGTTTTTATGATGGCAAGAAATGAAAGTGACCAAGCTAGAACCCACCATTTCAGA
GTCACCTTAAATGTAGCCATCTTACCCAGGGTCTGAGGGAAAAAGCAATTTCAACATGGA
GCCAGTATGGAGAGATTGCAAGATTACCTAGAAAGTTACCAAAGGCTATGCTCTGGAATG
TCTTTTTATATTCCAACACTTCAATGATGGAATGGTCTCTTCTCATTTTATCTAACAACAGTTT
AGTTTTATAAAGAGGTATGAGCTACATTTGGCTTTATTTCTATAAATCTATGAACGTCTTCA
35 GGACACATGCTGACTTTCTGAGGCATAGGTGTGGGCTTGGGATATTTTTTTCTTTTTCTTTT
TTTGGCTTCTAGAAGGTCTTATCAA
AAGTTCCTATTTTAAAAATTATCATAAAACTAAAGGGCCCTTCAAATGACCATAAAGCAGTA
CAAATCACCAATCCTAGCTAAGTAACCAGGTCACCTGCAAAGCAGCACCTCGGAAACAAG
AGAGCCTGCTGCATTCACTCAAAGCACCCATCATATAAAAGTATCCTAGTTAAATATAAA
40 AAGCAAAGTTCATTTCCCCAAAGCTCAAACAGCCTATGCTGCTCCTGTTCTTTTGCCAAATC
TACCCAAGGCCTGAGGTGGAGGAATGCTACTTTCCCCACAAATGGTGGGTCCCACTCCTG
GCCCCAGACCTGAAGGCTGCAGAGTTGGAACCCAAGGTCCAAGGAGAGAAACATGCTGA
GAAGTTGGTGGTGGACACAAGGCTGCAGTAAGCACTGTTAGAGCCCATCAGATCTTCCTAA
ATCCAGGGACTCAGGACTCGGGAGGGGAGCCTGCCTGCTCACCCCTGCTCCTTGGGAAA
45 CAGAAATCTGGGACATGTCTCTCCCAATTATCTCGTTCACTGAAAAACAAGAAAAGGGAGA
GTGATTAGTGACCTCTGAGTGATTAGCCATATGATGCCCAACGCTGAGCTCTACCTCAGAC
ATTTCCCTTCCCATGGCTTTCTGTAAAGATAAAGGCAGTGTCTGAAGCCTGCTCCTGAGCT
GAGTGGCATTAGAGATGGAAACATCGCTCCACAACTACTCTGGACCTTTTTTCTCTCTCA
TTCAGTTCATTTTGGGTGGGAAGGGGGCTTACCCTCACTCTAAAGGTATGTTTAAATAAGT

ATTCACCTTACTCTAACTTTTCTTTAATTCATATCTCTTTGAGAAAGTTTTTATTTTCTACAAT
GCCACCTCCTATGTAAATGGAATGAAAATAAATGCTTTGGTCAGAATAAA

SEQ ID NO: 9

5 MTPPPPPPPPPGPDPAADPAADPCPWPGSLVVLFGATAGALGRDLGSDETDLILLVWQVVEP
RSRQVGT LHKSLVRAEAAALSTQCREASGLSADSLARAEP LDKVLQQFSQLVNGDVALLGGG
PYMLCTDGGQLLRQVLHPEASRKNLVLPDMFFSFYDLRREFHMQHPSTCPARDLTVATMAQG
LGLETDATEDDFGVWEVKTMVAVILHLLKEPSSQLFSKPEVIKQKYETGPCKADVVDSETVVRA
RGLPWQSSDQDVARFFKGLNVARGGVALCLNAQGRRNGEALIRFVDSEQRDLALQRHKHHM
GVRYIEVYKATGEEFVKIAGGTSLEVARFLSREDQVILRLRGLPFSAGPTDVLGFLGPECPTVG
10 GTEGLLFVRHPDGRPTGDFAFALFACEELAQAALRRHKGMLGKRYIELFRSTAAEVQQVLNRYA
SGPLLPTLTAPLLPIPFPLAPGTGRDCVRLRGLPYTATIEDILSFLGEAAADIRPHGVHMLNQQ
GRPSGDAFIQMTSAERALAAQRCHKKVMKERYVEVPCSTEEMSRVLMGGTLGRSGMSPP
PCKLPCLSPPTYTTFQATPTLIPTETAALYPSSALLPAARVPAAPTVPAYYPGPATQLYLYNTAY
YPSPPVSPTTVGYLTPTAALASAPTSVLSQSGALVRMQGVPTAGMKDLLSVFQAYQLPADD
15 YTSLMPVGDPPRTVLQAPKEWWCL

El término "*MAGOH*" (también llamado en la literatura *MAGOH1*; *MAGOHA*) se refiere tanto al gen como a la proteína humana. *MAGOH* subunidad del complejo de unión exón. En el contexto de la presente invención, "*MAGOH*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*MAGOH*", y que comprendería diversas variantes procedentes de:

- a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 11,
- b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- 25 c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,
- d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la
- 30 SEQ ID NO: 11 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *MAGOH*. Entre otras posibles secuencias nucleotídicas que codifican *MAGOH* se encuentra, pero sin limitarse, la SEQ ID NO: 10.

Gene ID: 4116

35 Localización del gen: 1p32.3

SEQ ID N° 10

AGCCACTTCCGGGCGAAGAGGTTAGCGAGAAGGAACCCCTCCACGCAGCGACGACGGGA
 CCTGACGTCAGTTCGCGCGCCGCCGCAAACGTCAGTGTCTGGGCGCAGACGGCGGCAGT
 GCGGCTTGCTCTTGGAAGTTCAGGCTCGGTTGTCTTTTGGGAGCCATGGAGAGTGACTTTT
 5 ATCTGCGTTACTACGTGGGGCACAAGGGCAAGTTCGGCCACGAGTTCCTGGAGTTTGAGT
 TTCGACCGGACGGGAAGTTAAGATATGCCAACACAGCAATTACAAGAATGATGTCATGAT
 CAGAAAAGAGGCTTATGTACATAAAAGCGTGATGGAGGAACTGAAGAGAATAATTGACGAC
 AGTGAAATTACCAAAGAGGATGATGCATTGTGGCCTCCTCCTGACCGAGTGGGCCGGCAG
 GAGCTTGAAATCGTCATTGGAGATGAACACATTTCTTTTACAACATCAAAAATTGGTTCCCT
 10 TATTGATGTCAATCAATCCAAGGATCCAGAAGGCTTACGAGTATTTTATTATCTTGTCCAGG
 ACCTGAAGTGTGGTCTTCAGTCTTATTGGATTACACTTCAAGATTAAACCAATCTAGACT
 GAATATTGGTGTGGACATGGGGGGTGGGTGGGAGTAGAAAATTTTGTGTATATCAGGGCA
 GTATTTTTTTATGAACTATAAATGATTGTCTTTAATAAATATGTGATAAAATCCAATTTTATTA
 TTTTATAAAGACCTGAACATGTGAAAAAAAAAAAAAAAAAAAA

15

SEQ ID N° 11

MESDFYLRYVGHKGKFGHEFLEFEFRPDGKLRYANNSNYKNDVMIRKEAYVHKSVMEELKRI
 IDDSEITKEDDALWPPDRVGRQELEIVIGDEHISFTTSKIGSLIDVNQSKDPEGLRVFYLVQDL
 KCLVFSLIGLHFKIKPI

20

El término "*SND1*" (también llamado en la literatura *p100*; *TDRD11*; *Tudor-SN*) se refiere tanto al gen como a la proteína humana. El gen *SND1* codifica un coactivador transcripcional que interactúa con el dominio ácido del antígeno nuclear 2 del virus de Epstein-Barr (EBNA 2), un activador transcripcional que se requiere para la transformación de los linfocitos B. También se cree que esta proteína es esencial para el crecimiento celular normal. En el contexto de la presente invención, "*SND1*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*SND1*", y que comprendería diversas variantes procedentes de:

25

a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 13,

30

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

35

d) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 13 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad

y las características estructurales de la proteína *SND1*. Entre otras posibles secuencias nucleotídicas que codifican *SND1* se encuentra, pero sin limitarse, la SEQ ID NO: 12.

Gene ID: 27044

5 Localización del gen: 7q32.1

SEQ ID N° 12

ATATTTAGCCCATCCTTGCAAATCAACTCTTTCAGCGCTTCCTGGAACCTACCACCGTTTTG
GGGACCAAGAGGCGGGGCTCCGCGAGCAGGGTGCCTATTGGCCTGAGGGCCGGCGGGC
TCTGGGCGCTAATTGGGCGGAGGGCGGTGGCGCTCTGGCACGCTTGCGCGGCGAGTAGA
10 ACGTGTGGCGGCGGCGGAGATCGCGTCTCTTTCGCTCCGTGTCCCGCTGCTGCTCCTGT
GAGCGCCCGGCGAGTCCGTCCCGTCCACCGTCCGCAGCTGGTAGCCAGCCTGCCCCCTCG
CCTCGACTCCCTTTACCAACACCGACACCCACATTGACACCTCCAGTCCGGCCAGCCGC
TCCACTCGTTGCCTTTGCATCTCCACACATGGCGTCCCTCCGCGCAGAGCGGCGGCTCCTC
CGGGGGACCCGCGGTCCCCACCGTGCAGCGGGGCATCATCAAGATGGTCTCTCAGGGT
15 GCGCCATCATTGTCCGAGGTCAGCCTCGTGGTGGGCCTCCTCCTGAGCGGCAGATCAACC
TCAGCAACATTCGTGCTGGAAATCTTGCTCGCCGGGCAGCCGCCACACAACCTGATGCAA
AGGATACCCCTGATGAGCCCTGGGCATTTCCAGCTCGAGAGTTCTTTCGAAAGAAGCTGA
TTGGGAAGGAAGTCTGTTTCACGATAGAAAACAAGACTCCCCAGGGGCGAGAGTATGGCA
TGATCTACCTTGGAAAAGATACCAATGGGGAAAACATTGCAGAATCACTGGTTGCAGAGGG
20 CTTAGCCACCCGGAGAGAAGGCATGAGAGCTAATAATCCTGAGCAGAACCGGCTTTTCAGA
ATGTGAAGAACAAGCAAAGGCAGCCAAGAAAGGGATGTGGAGTGAGGGGAACGGTTCACA
TACTATCCGGGATCTCAAGTATACCATTTGAAAACCCAAGGCACTTTGTGGACTCACACCAC
CAGAAGCCTGTTAATGCTATCATCGAGCATGTGCGGGACGGCAGTGTGGTCAGGGCCCTG
CTCCTCCCAGATTACTACCTGGTTACAGTCATGCTGTGAGGCATCAAGTGCCCAACTTTTC
25 GACGGGAAGCAGATGGCAGTGAACTCCAGAGCCTTTTGTGTCAGAAGCCAAATTTTTCAC
TGAGTCGCGACTGCTTCAGAGAGATGTTTCAGATCATTCTGGAGAGCTGCCACAACCAGAA
CATTCTGGGTACCATCCTTCATCCAAATGGCAACATCACAGAGCTCCTCCTGAAGGAAGGT
TTCGCACGCTGTGTGACTGGTCGATTGCAGTTTACACCCGGGGCGCAGAAAAGCTGAGG
GCGGCAGAGAGGTTTGCCAAAGAGCGCAGGCTGAGAATATGGAGAGACTATGTGGCTCC
30 CACAGCTAATTTGGACCAAAAGGACAAGCAGTTTGTGTTGCAAGGTGATGCAGGTTCTGAAT
GCTGATGCCATTGTTGTGAAGCTGAACTCAGGCGATTACAAGACGATTCACCTGTCCAGCA
TCCGACCACCGAGGCTGGAGGGGGAGAACACCCAGGATAAGAACAAGAACTGCGTCCC
CTGTATGACATTCTTACATGTTTGAGGCCCGGGAATTTCTTCGAAAAAGCTTATTGGGAA
GAAGGTCAATGTGACGGTGGACTACATTAGACCAGCCAGCCAGCCACAGAGACAGTGCC
35 TGCCTTTTCAGAGCGTACCTGTGCCACTGTCACCATTTGGAGGAATAAACATTGCTGAGGCT
CTTGTCAGCAAAGGTCTAGCCACAGTGATCAGATACCGGCAGGATGATGACCAGAGATCA
TCACACTACGATGAACTGCTTGCTGCAGAGGCCAGAGCTATTAAGAATGGCAAAGGATTGC
ATAGCAAGAAGGAAGTGCCTATCCACCGTGTTGCAGATATATCTGGGGATACCCAAAAAGC
AAAGCAGTTCCTGCCTTTTCTTCAGCGGGCAGGTCGTTCTGAAGCTGTGGTGGAATACGTC
40 TTCAGTGGTTCTCGTCTCAAACCTATTTGCCAAAGGAACTTGCCTTATCACCTTCTTGCT
TGCAGGCATTGAATGCCCCAGAGGAGCCCGAAACCTCCCAGGCTTGGTGCAGGAAGGAG
AGCCCTTCAGCGAGGAAGCTACACTTTTACCAAGGAACTGGTGTGTCAGCGAGAGGTGG
AGGTGGAGGTGGAGAGCATGGACAAGGCCGGCAACTTTATCGGCTGGCTGCACATCGAC
GGTGCCAACCTGTCCGTCTGCTGGTGGAGCACGCGCTCTCCAAGGTCCACTTCACCGCC
45 GAACGCAGCTCCTACTACAAGTCCCTGCTGTCTGCCGAGGAGGCCGCAAAGCAGAAGAAA

GAGAAGGTCTGGGCCCCACTATGAGGAGCAGCCCGTGGAGGAGGTGATGCCAGTGCTGGA
 GGAGAAGGAGCGATCTGCTAGCTACAAGCCCGTGTTTGTGACTGAGATCACTGATGACCT
 GCACTTCTACGTGCAGGATGTGGAGACCGGCACCCAGTTGGAGAAGCTGATGGAGAACAT
 GCGCAATGACATTGCCAGTCACCCCCCTGTAGAGGGCTCCTATGCCCCCGCAGGGGAG
 5 AGTTCTGCATTGCCAAATTTGTAGATGGAGAATGGTACCGTGCCCGAGTAGAGAAAGTCGA
 GTCTCCTGCCAAAATACATGTCTTCTACATTGACTACGGCAACAGAGAGGTCTGCCATCC
 ACCCGCCTGGGTACCCTATCACCTGCCTTCAGCACTCGGGTGCTGCCAGCTCAAGCCACG
 GAGTATGCCTTCGCCTTCATCCAGGTGCCCCAAGATGATGATGCCCGCACGGACGCCGTG
 GACAGCGTAGTTCGGGATATCCAGAACACTCAGTGCCTGCTCAACGTGGAACACCTGAGT
 10 GCCGGCTGCCCCCATGTCACCCTGCAGTTTGCAGATTCCAAGGGCGATGTGGGGCTGGG
 CTTGGTGAAGGAAGGGCTGGTCATGGTGGAGGTGCGCAAGGAGAAACAGTTCCAGAAAGT
 GATCACAGAATACCTGAATGCCCAAGAGTCAGCCAAGAGCGCCAGGCTGAACCTGTGGCG
 CTATGGAGACTTTTCGAGCTGATGATGCAGACGAATTTGGCTACAGCCGCTAAGGAGGGGA
 TCGGGTTTGGCCCCCAGCCCCGCTCACGCCAGTCCCTCTTCTCTGCCGGGAGGGTGTTT
 15 TCAACTCCAAACCCCAGAGAGGGGTTGTAGATTGGGTCCAGCTTTGCTTCAGTGTGTGGAA
 ATGTCTCGTGGGGTGGCATCGGGGCTGCGGGGTGGGGACCCCAAGGCTTTCTGGGGCAG
 ACCCTTGTCTCTGGGATGATGGGCACTGCTATCCACAGTCTCTGCCAGTTGGTTTTATTT
 GGAGGTTTGTGGGCTTTTTTTAAAAAAGTCCTCAAATCAGGAAGAAACATCAAAGAC
 TATGTCCTAGTGGAGGGAGTAATCCTAACACCCAGGCTGGCCGCCAGCTGGCACCTGCCT
 20 CTATCCCAGACTGCCCTCGTCCCAGCTCTCTGTCCAACGTTGATTATGTGATTTTTCTGAT
 ACGTCCATTCTCAAATGCCAGTGTGTTACATCTTCGCTCTGGCCAGCCATTCTGTATTTA
 AAGCTTTTGAGGCCCAATAAAATAGTACGTGCTGTCTGCAGCCCTTATTGATCACAAAAAA
 AAA

SEQ ID N° 13

MASSAQSGGSSGGPAVPTVQRGIIKMLVSGCAIIVRGQPRGGPPPERQINLSNIRAGNLARRAA
 ATQPDADKTPDEPWAFPAREFLRKKLIGKEVCFTIENKTPQGREYGMILGKDTNGENIAESLV
 AEGLATRREGMRANNPEQNRLSECEEQAKAAKKGMWSEGNGSHTIRDLKYTIENPRHFVDSH
 HQKPVNAIIEHVRDGSVVRALLLPDYLLVTVMLSGIKCPTFRREADGSETPEPFAAEAKFFTESR
 LLQRDVQIILESCHNQNILGTILHPNGNITELLKEGFARCDWSIAVYTRGAELRAAERFAKER
 30 RLRIWRDYPAPTANLDQKDKQFVAKVMQVLNADAIVVKLNSGDYKTIHLSSIRPPRLEGENTQD
 KNKKLRPLYDIPYMFEAREFLRKKLIGKKVNVTVDIYPASPATETVPAFSERTCATVTIGGINIAE
 ALVSKGLATVIRYRQDDQRRSSHYDELLAAEARAIKNGKGLHSKKEVPIHRVADISGDTQKAKQ
 FLPFLQRAGRSEAVVEYVFSGSRLKLYLPKETCLITFLLAGIECPRGARNLPGLVQEGEPFSEEA
 TLFTKELVLQREVEVEVESMDKAGNFIGWLHIDGANLSVLLVEHALSKVHFTAERSSYYKSLLSA
 35 EEAAKQKKEKVWAHYEEQPVEEVMPVLEEKERSASYKPVFVTEITDDLHFYVQDVETGTQLEK
 LMENMRNDIASHPPEGSYAPRRGEFCIAKFVDGEWYRARVEKVESPAKIHVFYIDYGNREVL
 PSTRLGTLSPAFSTRVLPAQATEYAFAFIQVPQDDDARTDAVDSVVRDIQNTQCLLNVEHLSAG
 CPHVTLQFADSKGDVGLGLVKEGLVMVEVRKEKQFQKVITEYLNAQESAKSARLNLWRYGDF
 RADDADFEFGYSR

40

El término "*SRSF2*" (también llamado en la literatura *SC35*; *PR264*; *SC-35*; *SFRS2*; *SFRS2A*; *SRp30b*) se refiere tanto al gen como a la proteína humana. La proteína codificada por el gen *SRSF2* es un miembro de la familia rica en serina / arginina (SR) de factores de empalme de pre-ARNm, que constituyen parte del espliceosoma. Cada uno de estos factores contiene un motivo

45 de reconocimiento de ARN (RRM) para unirse al ARN y un dominio RS para unirse a otras

proteínas. El dominio RS es rico en residuos de serina y arginina y facilita la interacción entre diferentes factores de splicing de SR. Además de ser crítico para el splicing del ARNm, también se ha demostrado que las proteínas SR están involucradas en la exportación del ARNm desde el núcleo y en la traducción. Se han encontrado dos variantes de transcripción que codifican la misma proteína y una variante de transcripción no codificante para este gen. En el contexto de la presente invención, “*SRSF2*” se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia “*SRSF2*”, y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 15,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 15 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF2*. Entre otras posibles secuencias nucleotídicas que codifican *SRSF2* se encuentra, pero sin limitarse, la SEQ ID NO: 14.

Gene ID: 6427

Localización del gen: 17q25.1

SEQ ID N° 14

AGAAGGTTTTCATTTCCGGGTGGCGCGGGCGCCATTTTGTGAGGAGCGATATAAACGGGCG
CAGAGGCCGGCTGCCCCGCCAGTTGTTACTCAGGTGCGCTAGCCTGCGGAGCCCCGTCCG
TGCTGTTCTGCGGCAAGGCCTTTCCAGTGTCCCCACGCGGAAGGCAACTGCCTGAGAGG
CGCGGCGTTCGCACCGCCCAGAGCTGAGGAAGCCGGCGCCAGTTCGCGGGGCTCCGGGC
CGCCACTCAGAGCTATGAGCTACGGCCGCCCCCTCCCGATGTGGAGGGTATGACCTCC
CTCAAGGTGGACAACCTGACCTACCGCACCTCGCCCCGACACGCTGAGGCGCGTCTTCGA
GAAGTACGGGCGCGTTCGGCGACGTGTACATCCCGCGGGACCGCTACACCAAGGAGTCCC
GCGGCTTCGCCTTCGTTTCGCTTTACGACAAGCGCGACGCTGAGGACGCTATGGATGCCA
TGGACGGGGCCGTGCTGGACGGCCGCGAGCTGCGGGTGCAAATGGCGCGCTACGGCCG
CCCCCGGACTCACACCACAGCCGCCGGGGACCGCCACCCCGCAGGTACGGGGGGCGGT
GGCTACGGACGCCGGAGCCGCAGCCCTAGGCGGCGTTCGCCGACGCCGATCCCGGAGTC
GGAGCCGTTCCAGGTCTCGCAGCCGATCTCGCTACAGCCGCTCGAAGTCTCGGTCCCGC
ACTCGTTCTCGATCTCGGTTCGACCTCCAAGTCCAGATCCGCACGAAGGTCCAAGTCCAAG
TCCTCGTTCGGTCTCCAGATCTCGTTCGCGGTCCAGGTCCCGGTCTCGGTCCAGGAGTCCT

CCCCCAGTGTCCAAGAGGGAATCCAAATCCAGGTCGCGATCGAAGAGTCCCCCAAGTCT
 CCTGAAGAGGAAGGAGCGGTGTCCTCTTAAGAAAATGATGTATCGGCAAGCAGTGTAAC
 GGAGGACTTGGGGAAAAAGGACCACATAGTCCATCGAAGAAGAGTCCTTGGAAACAAGCAA
 CTGGCTATTGAAAAGGTTATTTTGTAACTTTGTCTAACTTTTTACTTGTTTAAGCTTTGCCT
 5 CAGTTGGCAAACCTTCATTTTATGTGCCATTTTGTGCTGTTATTCAAATTTCTTGTAATTTAG
 TGAGGTGAACGACTTCAGATTTTCATTATTGGATTTGGATATTTGAGGTAAAATTTCAATTTGT
 TATATAGTGCTGACTTTTTTTGTTTGAAATTAAACAGATTGGTAACCTAATTTGTGGCCTCCT
 GACTTTTAAGGAAAACGTGTGCAGCCATTACACACAGCCTAAAGCTGTCAAGAGATTGACT
 CGGCATTGCCTTCATTCCTTAAAATTAACCTACAAAAGTTGGTGTAAATTTGTATATGTT
 10 ATTTACCTTCAGATCTAAATGGTAATCTGAACCCAAATTTGTATAAAGACTTTTCAGGTGAAA
 AGACTTGATTTTTTTGAAAGGATTGTTTATCAAACACAATTCTAATCTCTTCTCTTATGTATTTT
 TGTGCACTAGGCGCAGTTGTGTAGCAGTTGAGTAATGCTGGTTAGCTGTAAAGGTGGCGT
 GTTGCAGTGCAGAGTGCTTGGCTGTTTCTCCTGTTTTCTCCCGATTGCTCCTGTGTAAAGATG
 CCTTGTCGTGCAGAAACAAATGGCTGTCCAGTTTATTAATAATGCCTGACAACTGCACTTCCA
 15 GTCACCCGGGCCTTGCATATAAATAACGGAGCATACAGTGAGCACATCTAGCTGATGATAA
 ATACACCTTTTTTTCCCTCTTCCCCCTAAAAATGGTAAATCTGATCATATCTACATGTATGAA
 CTTAACATGGAAAATGTTAAGGAAGCAAATGGTTGTAATTTGTAAGTACTTATAACATGGT
 GTATCTTTTTTGCTTATGAATATTCTGTATTATAACCATTGTTTCTGTAGTTTAATTAACATT
 TTCTTGGTGTTAGCTTTTCTCAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 20 AA

SEQ ID N° 15

MSYGRPPPDVEGMTSLKVDNLTYRTSPDTLRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRF
 HDKRDAEDAMDAMDGAVLDGRELRVQMARYGRPPDSHHSRRGPPPRRYGGGGYGRRSRS
 25 PRRRRRSRSRSRSRSRSRSRSRYSRKSRSRTRSRSRSTSKSRSAARRSKSKSSSVSRSRSR
 SRSRSRSRSPPPVSKRESKSRSRSPPKSPPEEGAVSS

El término "*SRSF3*" (también llamado en la literatura *SFRS3*; *SRp20*) se refiere tanto al gen como
 a la proteína humana. La proteína codificada por el gen *SRSF2* es un miembro de la familia rica
 30 en serina / arginina (SR) de factores de empalme de pre-ARNm, que constituyen parte del
 espliceosoma. Cada uno de estos factores contiene un motivo de reconocimiento de ARN (RRM)
 para unirse al ARN y un dominio RS para unirse a otras proteínas. El dominio RS es rico en
 residuos de serina y arginina y facilita la interacción entre diferentes factores de splicing de SR.
 Además de ser crítico para el splicing del ARNm, también se ha demostrado que las proteínas
 35 SR están involucradas en la exportación del ARNm desde el núcleo y en la traducción. Se han
 encontrado dos variantes de transcripción que codifican la misma proteína y una variante de
 transcripción no codificante para este gen. En el contexto de la presente invención, "*SRSF3*" se
 define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia
 "*SRSF3*", y que comprendería diversas variantes procedentes de:

40 a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia
 aminoacídica de la SEQ ID NO: 17,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

5 d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 17 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF3*. Entre otras posibles secuencias nucleotídicas que codifican *SRSF3* se encuentra, pero sin limitarse, la SEQ ID NO: 16.

10

Gene ID: 6428

Localización del gen: 6p21.31-p21.2

SEQ ID N°16

15 ATGTGATTCCAGTATGGCGTATAAATAAAGGCGAGGAGAAGGCGGTGGTCCGCCATTTTCG
TGGACGCCGGGTGAGTGAGAGAGTTGGTTGGTGGTGGGCCGAGGAAAGCGGGAAGACT
CATCGGAGCGTGTGGATTTGAGCCGCCGCATTTTTTAACCCTAGATCTCGAAATGCATCGT
GATTCCTGTCCATTGGACTGTAAGGTTTATGTAGGCAATCTTGGAACAATGGCAACAAGA
CGGAATTGGAACGGGCTTTTGGCTACTATGGACCACTCCGAAGTGTGTGGGTGCTAGAA
20 ACCCACC CGGCTTTGCTTTTGTGAATTTGAAGATCCCCGAGATGCAGCTGATGCAGTCCG
AGAGCTAGATGGAAGAACACTATGTGGCTGCCGTGTAAGAGTGGAACTGTGCAATGGTGA
AAAAAGAAGTAGAAATCGTGGCCACCTCCCTCTTGGGGTCGTCGCCCTCGAGATGATTAT
CGTAGGAGGAGTCCTCCACCTCGTCGCAGATCTCCAAGAAGGAGAAGCTTCTCTCGCAGC
CGGAGCAGGTCCCTTTCTAGAGATAGGAGAAGAGAGAGATCGCTGTCTCGGGAGAGAAAT
25 CACAAGCCGTCCCGATCCTTCTCTAGGTCTCGTAGTCGATCTAGGTCAAATGAAAGGAAAT
AGAAGACAGTTTGCAAGAGAAGTGGTGTACAGGAAATTACTTCATTTGACAGGAGTATGTA
CAGAAAATTCAAGTTTTGTTTGAGACTTCATAAGCTTGGTGCATTTTTTAAGATGTTTTAGCTG
TTCAAATCTGTTTGTCTCTTGAACAGTGACACAAAGGTGTAATTCTCTATGGTTTGAAATG
GATCATACGAGGCATGTAATACCAAGAATTGTTACTTTACAATGTTCCCTTAAGCAAAATTG
AATTTGCTTTGAACTTTTAGTTATGCACAGACTGATAATAAACCTCTAAACCTGCCCAGCGG
30 AAGTGTGTTTTTTTTTAAATTTAAATACAGAAACAACCTGGCAAAAATTGAACTAAGATTTACT
TTTTTTTCCATAGCTGGGATATAGGCTGCAGCTATAGTTGAACAAGCAGTCTTTAAAACTG
CTGTGAAACACAGGCCATCAGGGAAAACGAAATGCTGCACTATTAAATTAGAGGTTTTTGA
AAAATCCAACCTCTCATCCTGGGCAGAGGTTGCCTAGTTGGTATAGAATGTTAAGTTTCAAGA
AAGTTTACCTTTGCTTTAGGTCATAAGTTCCCTATTTGATTGCTGTATATGGATACATGGCTG
35 TTCGTGACATTCTTTATGTGCAAATTTGTGATTTCAAAAATGTCCTGCCAGTTTAAGGGTAC
ATTGTAGAGCCGAACCTTTGAGTTACTGTGCAAGATTTTTTTTTCATGCTGTCATTTGTAATAT
GTTTTGTGAGAATCCTTGGGATTAAAGTTTTGGTTACAAATTGTTCTTTAACTTGAAAGCCTG
TTTTTCCTTGCAAACCTCAAATCTGTGAGCTTGGTACCAAGTCCAGGTATAACATTCCTATTG
GAAGCCATACTTATATTTTCTTGTAAGTGCTTTTGAATTAATAAAATATTAGCATAATTGTGT
40 ATAGTCAGTTGAACCCACTGTTACCATTGTTCTTATCCCATGGGAAGCAGTTGGTTACACGA
TTCTTATTTTATAAGAAACAGCTGAGAGGCACTATGGATTAGTCTTCTGAAGTGAAGGAAAT

ATAGATGTCACCTAAGTGATAGTTAACCCATTTTTTTTTTTTTTTAGGCATAGAAGCCAGTTCA
 GGGTCCATAATATTTAGTGACCAACATTTTAAAGTATAGCAGCAACCTGGTTCTTAAACACA
 AAGTAAGTTGCCCATTAACAAATGGCTTTTATCTTTAGCATGAAAACCTTCCACAGGTCTAA
 AAATTGCTTCCATTTTATAATTTGAGGTGTTGCATGGGAATTCTAAGCTGATCCATCATGAT
 5 GTAAAAGTTTACAATATGGTTCAAATGTAAACAGTGCAGAATTGAATATGGAGGCATGCATAA
 CCTTCCTCTTAGAAAAATGGCAGGTGTTGTAATTTCAAATTTTTGTGCAATTAGATTAAATCAT
 AATGCAACAGTCTTGTGGCTTAGTTTCCTTAAATATGCTCTTAAGATAGTTTTGGATTATGC
 GGTATTGACTGTCTTAAATATGAAAGATAAGTCATTGCATTAAGAGTTCAAGCTAAATGGAT
 ACATTAAGATACAGTTCCTAAACAAATTATTTATTTCCACCTTTTACACAAATCTTGGGAGAA
 10 TTGGTGTGTCAGGAAGGTTTCAGCCTTAAAGTTAAGCATGTTCAAGAAAGACACTTTTCAGACA
 GCCTGTTTTATTTACTGAGAGCTCTGGCATTGGGCATTTTGTCTTTAGTATTCTCACATAGCC
 TACAACCTGTTCTGTTAGGAACAAGCCAAGATAGCTAAGAAGTTACGGAAGCCATGCAAT
 ATGTCAATTACATTGCTTTTTTATAAATGGACAACCTTATGTGGGCATTTTTTGGGCAGTATATGA
 CTTCTTGCAGGCTCTTAAGTTGGAAGGGTTTCTGTAAAAAGGACAGTTACGGGTATAATAT
 15 GGCTAAGAGAAATAATACAGAACCTGAAATAAAAGTAACTTCACCAGGGAGTTATCCTGAC
 TTAGGTGACATACAGTCCCAGTTGGCAGTTACTTTAACCAGGAGCCCTAGCATAACCTCAA
 GACTCTTAGAACTTTAGAACATGGAAATTGTTAAAAGCTTTGAATTGCACATTTAGATTGTG
 GCTATTAGGAATTTTTTAAAGTTGTAATTCCTAAAATAACAGGTCACACTAACCGGTCTTGTC
 CATATTCAGGGTTGAGTGTAAAAATGAAAGGAATCACAAAACCTGAAGTTAGACCTGTGGTTT
 20 AAGGATCAAGGTGTTCACTAGAAAGTCACTGTTACTAATACTTGATGACAATGTAAAGAAACA
 TTAAGGATTATATTTGATTGTTTTCAAAGACACTGGCTGGATGTGGTGGCTCACACTTGTA
 TTCCTAGCACTTTGGGAGGCTAGGGCGGGTGGGAGGATCTCTTGAAGCCAGGAGTTGGA
 GACCAGTCTGCCCAACATGGCAAGATTCACATTTCTATTAAAAGGAATATTTAGTTTGA

SEQ ID N°17

25
 MHRDSCPLDCKVYVGNLGNNGNKTLELAFGYGPLRSVWVARNPPGFVFEFEDPRDAAD
 AVRELDGRTLGCGRVRVELSNGEKRSRNRGPPPSWGRRPRDDYRRRSPPRRRSRPRRSFS
 RSRSRSLSRDRRRERSLSRERNHKPSRSFSRSRSRSRNERK

El término "*SRRM1*" (también llamado en la literatura *160-KD*; *POP101*; *SRM160*) se refiere tanto
 30 al gen como a la proteína humana. En el contexto de la presente invención, "*SRRM1*" se define
 también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia
 "*SRRM1*", y que comprendería diversas variantes procedentes de:

- a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 19,
- 35 b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,
- d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la
- 40

SEQ ID NO: 19 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRRM1*. Entre otras posibles secuencias nucleotídicas que codifican *SRRM1* se encuentra, pero sin limitarse, la SEQ ID NO: 18.

5 Gene ID: 10250

Localización del gen: 1p36.11

SEQ ID N° 18

GGAGTTTAGGGCCTGACAGAAGCCCGCCCCCGCTGGCGCTCGTGCGCACGCGTGCGG
 GCTCTCGGCGCACTGAGCAGGCGCGGCCTCGTGTGCGCCGGAGGGGGCGGGCGCAACG
 10 ACGCGCGCTGCGTCCCGGCGCTCGGCTTTCCCTCCGCCGGTCCCGCCCTCCGTGCGGGC
 GGCGCGGTGTACCCTGGGATAGGGAGCGATCTCCGAGCGAGGCGGCAAGATGGACGCG
 GGATTTTTCCGCGGAACAAGTGCAGAACAGGATAATCGGTTCAGCAACAAACAGAAGAAAC
 TACTGAAGCAGCTGAAATTTGCAGAATGCCTAGAAAAAAGGTGGACATGAGCAAAGTAAA
 TTTGGAGGTTATAAAGCCTTGGATAACAAAAAGAGTAACGGAAATCCTTGGGTTTGAAGAT
 15 GATGTTGTGATTGAGTTTATATTCAACCAGCTGGAAGTGAAGAATCCAGACTCCAAAATGAT
 GCAAATCAACCTGACTGGATTTTTGAATGGAAAAAATGCTCGAGAATTTATGGGAGAACTGT
 GGCCCCTGCTGCTAAGTGCACAAGAAAACATCGCGGGAATCCCTTCTGCTTTCTAGAACT
 GAAGAAAGAAGAAATAAAACAAAGACAGATTGAACAAGAAAAACTGGCATCTATGAAAAAG
 CAAGATGAAGACAAAGATAAAAGAGATAAGGAAGAAAAAGAAAGCAGCAGAGAAAAAAGG
 20 GAGCGGTCTCGTAGCCCAAGAAGACGCAAATCCAGATCTCCTTCCCTAGAAAGACGATCTT
 CCCCTGTCAGGAGAGAGAGAAAGCGCAGTCATTCTCGATCTCCCCGTCACAGAACCAAGA
 GCCGGAGTCCTTCCCTGCTCCAGAAAAGAAGGAAAAAACTCCAGAGCTCCAGAACCTT
 CAGTGAAAGTAAAAGAACCTTCAGTACAAGAGGCTACTTCTACTAGTGACATTCTGAAAGTT
 CCCAAACCTGAACCTATACCAGAGCCTAAAGAACCTTCTCCGGAAAAAAATCCAAAAAAG
 25 AAAAGGAGAAGGAGAAGACCCGACCACGATCTCGGTACGCTCCAAATCAAGATCCCGGA
 CGCGGTCCCGCTCTCCTTCTCACACTCGACCTAGACGGCGCCATAGATCCCGATCAAGAT
 CGTATTCACCTAGAAGGCGGCCAAGCCCAAGAAGGCGGCCATCTCCTCGAAGAAGAACTC
 CGCCAAGAAGAATGCCTCCTCCACCAAGGCATAGAAGGAGTAGATCTCCAGTAAGACGAA
 GAAGACGTTTCGTCAGCATCCTTGTCTGGGAGTAGCTCATCATCCTCTTCATCTCGTTACG
 30 GTCACCACCAAGAAGCCTCCCAAGAGGACATCCAGCCCCCTCGGAAAACTCGTAGGTT
 ATCTCCTTCAGCAAGTCCTCCAAGGCGAAGGCACAGGCCATCACCTCCTGCAACTCCACC
 ACCCAAACTCGGCATTCCCCTACACCCAGCAGTCAAACCGTACAAGAAAAAGTCGTGTT
 TCTGTGTCTCCAGGGAGAACTTCAGGTAAAGTGACAAAACATAAAGGTACTGAGAAAAGAG
 AATCCCCTTCACCAGCACCGAAGCCTAGAAAAGTAGAGTTATCTGAATCGGAAGAAGATAA
 35 AGGTGGCAAAATGGCTGCAGCAGATTCTGTGCAGCAGAGACGCCAATACAGACGACAAAA
 CCAGCAGTCTTCATCTGACTCTGGCTCCTCCTCCTCCTCAGAAGATGAACGACCCAAGAGA
 TCCCATGTGAAGAATGGTGAGGTTGGCAGGCGGCGGAGACATTCCCCTTCCCGGAGTGCT
 TCTCCATCACCACGAAAGCGCCAAAAAGAGACTTCCCCTCGTGGTAGACGGAGGAGAAGT
 CCATCCCCACCACCCACCAGAAGGCGACGGTCTCCTTCTCCCGCCCCCTCCTCCTCGACGG
 40 CGCAGGACTCCACACCACCACCACGACGAAGGACTCCTTCTCCTCCCCACGTGCGGCGC
 TCACCTTCTCCTAGAAGATACTCTCCTCCAATACAGAGGAGATACTCTCCTTCTCCACCTCC
 AAAGAGAAGAACGGCTTCACCTCCTCCCCCTCCTAAACGAAGAGCATCACCATCTCCACCA
 CCAAAGCGGCGGGTCTCCATTCTCCACCTCCCAAACAAAGAAGCTCCCCAGTCACCAAG
 AGACGTTACCTTCATTATCATCCAAGCATAGGAAAGGGTCTTCCCCAAGCCGCTCTACCC

GGGAGGCCCGATCACCACAACCAAAACAAACGGCATTGCGCCCTCACCACGGCCTCGAGCTC
CTCAGACCTCCTCAAGTCTCCACCCGTTTGAAGAGGAGCGTCGTCATCACCCCAAAGAA
GGCAGTCCCCGTCTCCAAGTACTAGGCCATTAGGAGAGTCTCCAGGACTCCGGAACCTA
AAAAGATAAAAAAGGCTGCTTCCCCAAGCCCACAGTCTGTAAGAAGGGTCTCATCCTCCCG
5 ATCTGTCTCCGGGTCTCCTGAGCCAGCAGCTAAAAAGCCCCCAGCACCTCCATCCCCCGT
CCAGTCTCAGTCACCGTCTACAACTGGTCACCAGCTGTACCGGTCAAAAAGGCCAAAAG
CCCAACACCGAGCCCATCACCGCCAAGAAATTCAGATCAGGAAGGAGGTGGAAAGAAAAA
GAAGAAAAAGAAGGACAAGAAACACAAAAAGGATAAGAAGCACAAAGACAAAAAACAC
AAGAAGGAAAAGGCTGTGGCTGCAGCTGCTGCAGCTGCTGTGACCCCTGCAGCCATTGCA
10 GCTGCCACAACCACATTAGCACAGGAAGAGCCAGTGGCAGCGCCAGAGCCGAAGAAGGA
GACTGAAAGTGAAGCTGAAGATAACCTTGATGATTTAGAAAAGCACCTGCGTGAAAAGGCC
CTGAGATCAATGAGGAAGGCCCAAGTGTCCCCACAGTCTTAGGGGGAAATGTTTGTTATGA
TGTAATTTTATTTGGTTTGTACGCAGTTCAATTTCAAATTTGCTAAAATGTGTTTGAGCTTT
AGACTATAACATTTGTTGTAATAATTGCTAGGTTGAAGTTCAACATGTAAAAAAGGGGGCA
15 TGGATTTACATTGCAAAAGGTGTCCACAGTGTATTAGTGACATTCTTTCATTGACAGCTGAC
ATAATTCATTGAGTGAAATATTTTAAGCCAAAAAATTCCTTTTTTAAAAAAGGGGGTTTA
AATACTGTTGGCATTTTTATGGTTCCTTTAAATGCCCTAGCTATTCCCAGAGGGGTTTTTT
GTTTGTTTTTTTGGTTTTGATTTTCTTTTTGTTTTCTTTCTTCTTATTTTTTTCATTTGAG
TCTTAGCTCCCATTTAAGTTATGCTTCTGACCTTGATGGTCTGTAAGCTTGCCCAGAAATA
20 AGACCACTGTTTTGAACTACCACAAAAGTATAAATGAATATTTTAATGCCACAATCTTTCCTG
TTGCCTGTGGAGTCTCTGCTGAAATGAATCAGGATTCGAGCTCTAGGATGAGACAGAAAAT
GAAAGCATGTTGTTTGCCAGGACACTGTGGGTTTATATTGATGTGTAACAAGTTGATTTGGA
ACACTGGACTCTCATTCTGTTATTCTGGTTTTGTTTTTTTTGTTTTGTTTTTTTTCTTTTGTA
AGGCAATGAGCTAGTCCCAGAAAGGATCCTTCAGTTACATACAATTTGTTTAATGAAATGTC
25 ATGGCTCTGTTCAATTTTTGTCTTGTCTTCCAATTGGTATATACACTTTCAGAGCCTCTT
GTATTTGGAAGGCTGGAAGGGCCCAGACTTTGGAATAGTGTCTTGGTTTCACTGTTTTGT
TTTGATTTTTTTTTGTTTTGATTTTTTTAACTAAAGCTATATAAAGCTTGTTGATTAAACA
GAATAAATTTCTAAATTTAAAAATTTTGCCACTCTTAAAAA

SEQ ID N° 19

30
MDAGFFRGTSAEQDNRFSENKQKKLLKQLKFAECLEKKVDMSKVNLEVIKPWITKRVTEILGFED
DVVIEFIFNQLEVKNPDSKMMQINLTGFLNGKNAREFMGELWPLLLSAQENIAGIPSAFLELKKE
EIKQRQIEQEKLASMKKQDEDDKDKRDKEEKESREKRERSRSPRRRKSRSPSPRRRSSPVRR
ERKRSHSRSPRHRTKSRSPSPAPEKKEKTPELPEPSVKVKEPSVQEATSTSDILKVPKPEPIPE
35 PKEPSPEKNSKKEKEKEKTRPRSRSRSKSRSRTRSRSPSHTRPRRRHRSRSPSPRRRSP
RRRPSPRRRTPPRRMPPPPRHRRSRSPVRRRRRSSASLSGSSSSSSSSSRSPPKKPPKRT
SSPPRKTRRLSPSASPPRRRHRSPPATPPPKTRHSPTPQQSNRTRKSRVSVSPGRTSGKVT
KHKGTEKRESPSPAPKPRKVELSESEEDKGGKMAAADSVQQRQYRRQNQQSSSDSGSSSS
SEDERPKRSHVKNGEVGRRRRHSPSRASPSPRKRQKETSPRGRRRRSPSPPPTRRRRSPS
40 PAPPRRRRRTPTPPRRRTTPSPPPRRRSPSPRRYSPPIQRRYSPSPPPKRRRTASPPPPPKRRA
SPSPPPKRRVSHSPPPKQRSSPVTKRRSPSLSSKHKRGSSPSRSTREARSPQPNKRHSPSPR
PRAPQTSSSPPPVRRGASSSPQRRQSPSPSTRPIRRVSRTPPEPKIKKAASPSPPQSVRRVSS
RSVSGSPEPAKKPPAPPSPVQSQSPSTNWSPAVPVKKAKSPTPSPSPPRNSDQEGGGKKK
KKKKDKKKKKDKKKKKKKKKKKEKAVAAAAAAVTPAAIAAATTTLAQEEPVAAPKPKETESE
45 AEDNLDDLEKHLREKALRSMRKAQVSPQS

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término “*RNU4*”, (también llamado en literatura *U4*; *U4BL*; *RNU4A*; *RNU4B2*) se refiere al gen, factor de la maquinaria de *splicing*. En el contexto de la presente invención, “*RNU4*” se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia del snRNP “*RNU4*”,
 5 y que comprendería diversas variantes procedentes de:

a) Moléculas de ácido nucleico que comprende la SEQ ID NO:20,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

10 c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

Gene ID: 26835

Localización del gen: 12q24.23

SEQ ID N° 20:

15 AGCTTTGCGCAGTGGCAGTATCGTAGCCAATGAGGTCTATCCGAGGCGCGATTATTGCTAA
 TTGAAACTTTTTCCCAATACCCCGCCGTGACGACTTGCAATATAGTCGGCACTGGCAATTTT
 TGACAGTCTCTACGGAGACTG

El término “*RBM45*” (también llamado en la literatura *DRB1*; *RB-1*) se refiere tanto al gen como a la proteína humana. El gen *RBM45* codifica un miembro de la familia de proteínas de unión al
 20 ARN del tipo de reconocimiento de ARN (RRM). Esta proteína exhibe unión preferencial al ARN poli (C). En el contexto de la presente invención, “*RBM45*” se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia “*RBM45*”, y que comprendería diversas variantes procedentes de:

25 a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 22,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

30 d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 22 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad

y las características estructurales de la proteína *RBM45*. Entre otras posibles secuencias nucleotídicas que codifican *RBM45* se encuentra, pero sin limitarse, la SEQ ID NO: 21.

Gene ID: 129831

5 Localización del gen: 2q31.2

SEQ ID N° 21

GACAGCGCCTGGCGCCGGACTCCTCTTTCTCCCGGAAGCGGAGCACCGAGCCGGCAAAG
 GCTTGGGTGTGAGACAGCAGCGGTGGCAGACACCGCAGAAGCAAAGAGCAGTGAGGCTC
 CTGCATTTCGGGTGGAGCACCATGGACGAAGCTGGCAGCTCTGCGAGCGGCGGGGGCTTC
 10 CGCCCGGGCGTGGACAGCCTGGACGAACCGCCCAACAGCCGCATCTTCCTTGTGATCAG
 CAAGTACACACCTGAGTCGGTGCTGAGGGAGCGCTTCTCGCCTTTTGGCGACATCCAGGA
 CATCTGGGTGGTGCGGGACAAGCACACCAAGGAGTCCAAGGGCATTGCTTTCGTCAAGTT
 CGCCCGCAGCTCACAGGCCTGCAGGGCCATGGAGGAGATGCATGGCCAGTGCCTCGGCC
 CCAACGACACCAAGCCCATCAAGGTGCGGGTGCCCGGGTGGGGTGCCCTCGGGGGAA
 15 GGAGTGGGCCTCTTGGACCTCCCTTCACCTGCTGCTCTGCCGGGGTGAGGGAGGAGTGG
 AACATCCGTTCCCGGCAGCTGACCAGACAGCACCTGGCTTGGGGACAGGGGCTGCTTTGA
 GGAGAGGCTCCCTCCACGAAGCTGCGCCGGGTTCAGGGGAAGGGGACGGGATGCGG
 AGTGTTCGCCCCACGAGAGCGGACCCGGCAGCCGCGGCCACCGGTGACAGAACAGCAG
 TGGCTGGGCTTCCCTCTCCCTCGCCCTACTAATCCACCAGCTCTTGAGTTCACTGCCAA
 20 ACTCGTCCTAGCCTACTTTGAGTACTTTTCTGACCTTCCTCTCTCAGATGCCAGATTTCCA
 TATATAAAAAACGGAAAACTCTACTTTTCCTCAGGTTTTTCATTGCTCAGTCCCGATCATCT
 GGAAGTCACCGAGATGTTGAAGATGAAGAACTTACAAGAATCTTTGTTATGATACCAAAGTC
 CTACACAGAAGAAGATCTGCGGGAAAAATTTAAGGTGTATGGAGATATCGAGTATTGCAGC
 ATTATTAAGAATAAAGTGACTGGAGAAAGTAAAGGTTTGGGCTACGTACGATACTTAAAACC
 25 ATCACAAGCTGCCCCAAGCAATAGAAAAGTGTGATCGAAGTTTTAGAGCAATCTTGGCTGAA
 CCTAAAAATAAAGCATCTGAATCCTCTGAACAAGATTATTATAGTAATATGAGGCAAGAAGC
 TTTGGGACATGAACCTAGAGTAAATATGTTTCCATTTGAACAACAATCTGAATTTTCAAGTTT
 TGACAAGAATGATAGCCGAGGCCAGGAAGCAATCTCCAAACGCTTGTGAGTTGTATCAAGA
 GTTCCTTTCACTGAAGAACAGCTTTTTCAGCATTTTTATAGTACCAGGATTGGAATATTG
 30 TGAAGTTCAACGAGATCCTTATTCAAATTATGGTCATGGAGTGGTTCAGTATTTTAATGTAG
 CATCAGCTATTTATGCAAAAACAAATTACATGGATTTTCACTACCTCCTGGGAACCGAATA
 GGTGTTTCCTTCATTGATGATGGAAGTAATGCAACAGATCTCCTTAGAAAAATGGCAACACA
 GATGGTAGCTGCACAGCTTGCATCAATGGTGTGGAATAACCCAAGTCAGCAACAATTTATG
 CAATTTGGAGGAAGCTCTGGATCACAGTTGCCTCAAATCCAGACAGATGTTGTACTTCCAT
 35 CATGCAAAAAAAGCTCCTGCTGAAACTCCTGTGAAAGAAAGACTTTTTATTGTGTTTAAT
 CCTCATCCTTTACCTTTAGACGTATTAGAAGATATATTCTGTCGTTTTGGTAACCTGATCGAA
 GTTTACCTTGTGTCAGGAAAAAATGTGGGGTATGCCAAGTATGCCGATAGAATAAGTGCTA
 ATGATGCCATTGCCACTCTACATGGAAAGATTCTGAATGGGGTGAGACTTAAAGTTATGCT
 GGCAGATTCGCCAAGAGAAGAATCTAACAACGGCAAAGAACTTACTGATTCTTGAGGTGA
 40 TGATTTTTCATACTTGTGTTGTGTTCTTTCTACAGAACAAGACTAAATAATGACATAATCCTC
 AGCTGACTGACTGAAAATGTGACTGGACGCATTCCCTGTGGACAGTTGACAGCTTTTTTTTT
 TTCCATATACCTGATAGTCTGTGTACAGCATTGTTTTGTCTGGGAAGCAGGGATTGCTGAC
 ATGTATTTTTGAATCCATACATTAATGCTAAAACGAATATAGTAGTTGTTCTTAGAGCAATA
 TGTGTACGTGTAGCAGAAATAAAGTTTTCTTTGCTTAACTAAA

SEQ ID N° 22

5 MDEAGSSASGGGFRPGVDSLDEPPNSRIFLVISKYTPESVLRERFSPFGDIQDIWVVRDKHTKE
 SKGIAFVKFARSSQACRAMEEMHGQCLGPNDTKPIKVRVPGSGCPRGKEWASWTSLLLHLLCR
 GEGGVEHPFPAADQTAPGLGTGAALRRGSPPTKLRRVAGKGTGCGVFAPRERTRQPRPPGD
 RTAVAGLPLSLALTNPPALEFTAKLVLAYFEYFSDLPLSDAQISIIYKKRKNSTFPQVFIAQSRSSG
 SHRDVEDEELTRIFVMIPKSYTEEDLREKFKVYGDIEYCSIKNKVTGESKGLGYVRYLKPSQAA
 QAIENCDRSFRAILAEPKNKASESSEQDYYSNMRQEALGHEPRVNMFPFEQQSEFSSFDKND
 10 SRGQEAISKRLSVSRVPFTEEQQLFSIFDIVPGLEYCEVQRDPYSNYGHGVVQYFNVASAIYAK
 YKLHGFQYPPGNRIGVSFIDGGSNATDLLRKMATQMVAQQLASMVWNNPSQQQFMQFGGSS
 GSQLPQIQTDVVLPSCKKKAPAETPVKERLFIVFNPHPLPLDVLEDIFCRFGNLIIEVYLVSGKNV
 GYAKYADRISANDAIATLHGKILNGVRLKVMLADSPREESNKRQRTY

15 El término "*SRSF4*" (también llamado en la literatura *SFRS4*; *SRP75*) se refiere tanto al gen como a la proteína humana. El gen *SRSF4* codifica un miembro de la familia del factor de splicing rico en arginina /serina. En el contexto de la presente invención, "*SRSF4*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*SRSF4*", y que comprendería diversas variantes procedentes de:

20 a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 24,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

25 c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 24 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF4*. Entre otras posibles secuencias
 30 nucleotídicas que codifican *SRSF4* se encuentra, pero sin limitarse, la SEQ ID NO: 23.

Gene ID: 6429

Localización del gen: 1p35.3

SEQ ID N° 23

GAAGCCAGCTGGCTGCGGAAGGCGGTGCAGGAACCCCTTTGGCGAGCTCGCGCGAGGA
CGTGCCGCAGCGCCGCCTGTGCATCACCTTGCCCCGCCGCACGTGGCGCGCCCCCGG
GCAAGCTCCTCCCCAGTCCTGGCCGAGCGGGAAGCGAACGCGCGCTCTCTCCGGTTAC
CTTCCGCCTCGGGGAAGGGCGGGGGAAGTGAAGCGCGCGCTCGCGAGCTGCGTGCTC
5 GCGTACGTGCGCGGGGTTGCGGTGCGTCCGTCCCGCCGCCCGCCGTTGCCGCCGCCG
CCGCTGCCGCCGTGCTCTCGCTTTGCCCGCCGCCGCCTAAGGGGGGCTGGGGCCGGGG
CCAGCCATCACTGCCGTTGCCGGGATGCCGCGGGTGTACATCGGCCGCCTGAGCTACCA
GGCCCCGGGAGCGCGATGTGGAGCGCTTCTTTAAGGGCTACGGGAAGATCCTGGAGGTGG
ATCTGAAGAACGGATATGGTTTTGTGGAGTTTGATGATCTGCGTGATGCAGATGATGCTGT
10 TTATGAACTGAATGGCAAAGACCTTTGTGGTGAGCGAGTAATTGTTGAGCATGCCCGCGGC
CCACGGCGAGATGGCAGTTACGGTTCTGGACGCAGTGATATGGTTATAGAAGAAGTGGC
CGAGATAAATATGGCCCTCCTACTCGCACAGAGTACAGACTTATTGTGGAGAATTTGTCAA
GTCGGTGCAGCTGGCAAGACCTAAAGGATTATATGCGTCAGGCAGGAGAAGTGAATTATG
CAGATGCTCACAAGGGACGCAAAAATGAAGGGGTGATTGAATTTGTATCTTATTCTGATAT
15 GAAAAGAGCTTTGAAAAAGTTGGATGGAAGTCAATGGGAGAAAAATCAGATTAGTT
GAAGACAAGCCAGGTTCCAGACGACGCCGGTCTACTCCAGAAGCCGGAGTCATTCAAGG
TCTCGCTCTCGAAGCAGACATTCCCGTAAGAGCAGAAGCCGAAGTGGCAGCAGCAAAAGC
AGTCATTCTAAGAGTAGATCTCGGTCCAGGTCGGGCTCCCGCTCCCGGAGCAAGAGCCGG
AGCCGGAGCCAGAGTCGGAGCCGGAGCAAGAAAGAGAAAAAGCAGGAGCCCCAGCAAGGA
20 AAAGAGCCGCAGCCGCAGCCATAGCGCTGGCAAGAGCCGCAGCAAGAGCAAAAGACCAAG
CTGAAGAGAAGATCCAAAACAATGACAATGTCGGGAAACCCAAGAGCCGGAGTCCTAGCA
GGCATAAAAGTAAGAGCAAAAGTCGGAGCAGGAGTCAGGAGAGGAGAGTGGAGGAGGAG
AAGCGAGGGAGTGTGAGCAGGGGCAGGAGCCAGGAGAAGAGCCTCCGCCAGAGTCGGA
GCCGGAGCAGGAGCAAAAGGGGGCAGCAGGAGCCGGAGCAGGAGCCGCAGCAAGAGCAA
25 GGACAAGAGGAAGGGCAGGAAGAGAAGCAGAGAGGAGAGCCGCAGTCGCAGTCGCAGC
CGCAGCAAGAGTGAGAGGAGCAGAAAGCGAGGCAGCAAGCGAGACAGCAAGGCGGGCA
GCAGCAAGAAGAAGAAGAAGGAAGACACTGACCGCTCCAGTCCAGATCTCCATCCCGCT
CCGTGTCAAAGGAGCGGGAACATGCCAAGTCTGAATCCAGCCAGAGGGAAGGTCGAGGA
GAGAGTGAGAATGCTGGCACCAATCAGGAGACCCGGTCCAGGTGCGATCCAATTCAAA
30 TCGAAACCAAACCTTCCATCAGAATCACGCTCCAGATCAAAGTCAGCTTCAAAAACCCGAT
CTCGGTCCAAGTCTAGATCCAGGTCTGCTTCCAGATCGCCCTCCCGATCTAGATCTAGGTC
CCACTCAAGGTCCTAACTGGCTATGGCCACAGCTGGAAGTACCCGAGAAGTCTTTTGTACA
TGTTTGGTAGCCGTAGCACAAAGTGATTGGAGTAGAACATGTCACTGCTGTACATTTTTAACT
CCCCTAATGGTGTGTCTATAATTGTTAAATCTAAGTGCTTCTCTCAGTAAAGCCTCCTGGC
35 ACCAGGCCTTCTGCTCGACTGAAAAAATTTTCTCTTTGAAAATCCCCTTTTACTCATGGC
CCACAGTAGAATATCCAAAACGCCTTGGCTTTCAGGCCTGGCCTTTCCTACAGGGAGCTCA
GTAACCTGGACGGCTCTAAGGCTGGAATGACCACATAGGTAGGTATGGTGAGTTCAACCA
TTTTTGCTCTTGAATTGATGCCCTTCGATGTATGCCATTTAGTGAAAGTGCTAAGTCTTAAG
TTTCCTACCACTTTGGTTTCATATTTTTGGACTTAACAAAGTTGTGAATAGCACAGTCGAGG
40 AAAATTGATACCTGCAGTAACCCATAGGAAATAAACTGTAGAGTTCCATATTCTGGTATTGT
GATTATATTGTTTTATATTAATAAAGAAAAGAAAAGAAATTTTTTTAATTTTATTTTTCCCGT
CTTGCAAAGTATAGTGACCCCTGTTTCCATTAAATTTGAATAAAGACTATTTTTGCTTGACAA
AATTTCAAAAAA

SEQ ID N° 24

45 MPRVYIGRLSYQARERDVERFFKGYGKILEVDLKNYGFVEFDDL RDADDAVYELNGKDL CGE
RVIVEHARGPRRDGSYSGRSYGYRRSGRDKYGPPTRT EYRLIVENLSSRCSWQDLKDYM
RQAGEVTYADAHKGRKNEGVI EFVSYS DMKRALEKLDGTEVN GRKIRLVEDKPGSRRRRSYS
RSRSHSRSRSRHSRKSRSRSGSSKSSH SKSRSRSRSGSRSRSKSRSRSQSRSRSKKEKS

RSPSKEKSRSRSHSAGKSRSKSKDQAEKIQNNDNVGKPKSRSPSRHKS KSKSRSRSQERRV
 EEEKRGSVSRGRSQEKSLRQSRSRSRSKGGSRSRSRSRSKSKDKRKGRKRSREESR SRSR
 RSKSERSRKRGSKRDSKAGSSKKKKKEDTDRSQSRSPSRSVSKEREHAKSESSQREGRGES
 ENAGTNQETR SRSRSNSKSKPNLPSESRSRSKSASKTRSRSKSRSRASRSPSRSRSRSHSR
 S

El término "*SRSF5*" (también llamado en la literatura *HRS*; *SFRS5*; *SRP40*) se refiere tanto al gen como a la proteína humana. La proteína codificada por el gen *SRSF4* es un miembro de la familia rica en serina / arginina (SR) de factores de splicing de pre-ARNm, que constituyen parte del espliceosoma. Cada uno de estos factores contiene un motivo de reconocimiento de ARN (RRM) para unirse al ARN y un dominio RS para unirse a otras proteínas. El dominio RS es rico en residuos de serina y arginina y facilita la interacción entre diferentes factores de empalme de SR. En el contexto de la presente invención, "*SRSF5*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*SRSF5*", y que comprendería diversas variantes procedentes de:

a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 26,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 26 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF5*. Entre otras posibles secuencias nucleotídicas que codifican *SRSF5* se encuentra, pero sin limitarse, la SEQ ID NO: 25.

Gene ID: 6430

Localización del gen: 14q24.1

SEQ ID N° 25

GGAAGTGACGTAGGACGCGCCCTCCATTTTGTGGAGCGCCAGAGCTGCTAAGTGCGTCAG
 TTGTGGAGTGGCGTAGACGAGTTAAGTCCTGGTCTGCGTGGAGGTCGACGACTCCGTCGC
 AGACTACGGACCTGTCTGGGTCTCAGCCGCCAAAGACCCCGTCCGGTAGGAAGTACTAGC
 CGGACATCATGAGTGGCTGTCTGGGTATTCATCGGGAGACTAAATCCAGCGGCCAGGGAGA
 AGGACGTGGAAAGATTCTTCAAGGGATATGGACGGATAAGAGATATTGATCTGAAAAGAGG

CTTTGGTTTTGTGGAATTTGAGGATCCAAGGGATGCAGATGATGCTGTGTATGAGCTTGAT
 GGAAAAGAACTCTGTAGTGAAAGGGTTACTATTGAACATGCTAGGGCTCGGTCACGAGGT
 GGAAGAGGTAGAGGACGATACTCTGACCGTTTTAGTAGTCGCAGACCTCGAAATGATAGA
 CGAAATGCTCCACCTGTAAGAACAGAAAATCGTCTTATAGTTGAGAATTTATCCTCAAGAGT
 5 CAGCTGGCAGGATCTCAAAGATTTTCATGAGACAAGCTGGGGAAAGTAACGTTTGCGGATGC
 ACACCGACCTAAATTAATGAAGGGGTGGTTGAGTTTGCCTCTTATGGTGACTTAAAGAAT
 GCTATTGAAAACTTTCTGGAAAGGAAATAAATGGGAGAAAAATAAAATTAATTGAAGGCAG
 CAAAAGGCACAGTAGGTCAAGAAGCAGGTCTCGATCCCGGACCAGAAGTTCCTCTAGGTC
 TCGTAGCCGATCCCGTTCCCGTAGTCGCAAATCTTACAGCCGGTCAAGAAGCAGGAGCAG
 10 GAGCCGGAGCCGGAGCAAGTCCCGTTCTGTTAGTAGGTCTCCCGTGCCCTGAGAAGAGCC
 AGAAACGTGGTTCTTCAAGTAGATCTAAGTCTCCAGCATCTGTGGATCGCCAGAGGTCCCG
 GTCCCGATCAAGGTCCAGATCAGTTGACAGTGGCAATTAACTGTAAATAACTTGCCCTGG
 GGGCCTTTTTTTAAAAACAAAAACCACAAAAATTCCCAAACCATACTTGCTAAAAATTCTG
 GTAAGTATGTGCTTTTCTGTGGGGGTGGGATTTGGAAGGGGGGTGGGTTGGGCTGGATA
 15 TCTTTGTAGATGTGGACCACCAAGGGGTGTTGAAACTAATTGTATTAAATGTCTTTTGAT
 AAGCCTTCTGCTCACATTTTTGTGAATGTCTGAAGTATATAGTTTGTGTATATTGACAGAGC
 TCTTTTATAACTAAAGCAAATTTAATTTTTTTGTACTAGAAAAAAATTTGAACATTTTAGTTCT
 TGGTTATAAAAAATGTTAATTCAGAATTAGTTTAATGCCTTAATTAACTAATTAATAGCTTTGG
 AACTTAAAAGAGCTCTAAATTTGCTTGTACATAAAGGCTTAATTTGTTTTCCTTGTTAGGGT
 20 CAAGGGTGTCTCCACTCTTTAACAGCTGCTGGACAGACACATTAGAGCAGCTGTTTGTTA
 TTGATAATAAAATATTATAAAACTA

SEQ ID N° 26

MSGCRVFIGRLNPAAREKDVERFFKGYGRIRDIDLKRGFGFVEFEDPRDADDAVYELDGKELC
 25 SERVITIEHARARSRGGRGRGRYSDRFSSRRPRNDRRNAPPVVRTENRLIVENLSSRVSWQDLK
 DFMRQAGEVTFADAHRPKLNEGVEFASYGDLKNAIEKLSGKEINGRKIKLIEGSKRHSRHSR
 SRSRTRSSSRSRSRSRSRSRKSYSRSRSRSRSRSRKSRSVSRSPVPEKSQKRGSSSRSKS
 PASVDRQRSRSRSRSRVDSGN

30 El término "*SRSF10*" (también llamado en la literatura *NSSR*; *TASR*; *SRp38*; *TASR1*; *TASR2*;
FUSIP1; *FUSIP2*; *SFRS13*; *SRp40*; *SFRS13A*; *PPP1R149*) se refiere tanto al gen como a la
 proteína humana. El gen *SRSF10* es un miembro de la familia rica en serina / arginina (SR) que
 están involucradas en el empalme de ARN constitutivo y regulado. Los miembros de esta familia
 se caracterizan por los motivos N-terminal RNP1 y RNP2, que son necesarios para la unión al
 35 ARN, y múltiples repeticiones C-terminal SR / RS, que son importantes en la asociación
 mediadora con otras proteínas celulares. En el contexto de la presente invención, "*SRSF10*" se
 define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia
 "*SRSF10*", y que comprendería diversas variantes procedentes de:

a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia
 40 aminoacídica de la SEQ ID NO: 28,

b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),

c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,

5 d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 28 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF10*. Entre otras posibles secuencias nucleotídicas que codifican *SRSF10* se encuentra, pero sin limitarse, la SEQ ID NO: 27.

10

Gene ID: 10772

Localización del gen: 14p36.11

SEQ ID N° 27

15 GGACGCCTGGTGTTCCTCGCGGGAAAGAACCGTTGCGCCCGACTTTGCGCGGAGGTAGGA
GGGGGAAGTGGAGGCGGGAGTGAAGTCTCGCGAGAAGAGTCGGTTGCCGTAGCAGAGC
CCTCTAGCTGTGTGTGTCTGAGGCTCGGCCCGCTGAGCCGCGGACGGTTTGCTGAGCCC
GTTAGTGCGCCCGGCCGAGACACGCCGCCGCGCATGTCCCGCTACCTGCGTCCCCCAAC
ACGTCTCTGTTCTGTCAGGAACGTGGCCGACGACACCAGGTCTGAAGACTTGCGGCGTGAA
20 TTTGGTCGTTATGGTCCTATAGTTGATGTGTATGTTCCACTTGATTTCTACACTCGCCGTCC
AAGAGGATTTGCTTATGTTCAATTTGAGGATGTTCTGTATGCTGAAGACGCTTTACATAATT
TGGACAGAAAGTGGATTTGTGGACGGCAGATTGAAATACAGTTTGCCAGGGGGATCGAA
AGACACCAAATCAGATGAAAGCCAAGGAAGGGAGGAATGTGTACAGTTCTTCACGCTATGA
TGATTATGACAGATACAGACGTTCTAGAAGCCGAAGTTATGAAAGGAGGAGATCAAGAAGT
CGGTCTTTTGATTACAACATAGAAGATCGTATAGTCCTAGAAAACCAAACCTGCAGCTGGAA
25 TACCCAGTACAGTTCTGCTTACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAACCAA
AGAAGGGCAGTTCAAGCGACCAAAGGGTGGGTGGAAGGTGCTGCAGTATGAATACTGTAC
GAATATTTTGACTCTGGTCTGAAAAGATAAAAGAATGTTATCGAAAACCTACATGGAATAATT
GAAGTCCCTTCAAGTTTGAAAGTAAGCATTTTAGGACAAATAAAAGGAAATTCAACTTTGTA
CTTGTGGAACTAATCCCTAAATATGAATAGGTTTATATTGATTCATGGGTAAACAGGTCCAT
30 AATAAATTATTGGAACTAGGATGTCTGAATATCAAGGAAGACAGCCATAGTCTCTTACAGT
GCCTCTGTTGGTCTGTCTCAAACCTGAATTGGGTGGGAAAAGGTATGGTCCAATATAAAAGT
TCCATTTTGGCATTATTGGCAAATCTTGCCCTTTGTTTATTTTGGTGCCAGTGTTTTCTGCTT
AATCATTTGCTTTGTTGGCATCTGTGTTTATTTACTTGTACACCACATGCAGTTTACATCTGT
CTTAATACTCCTTCCCAGGTAAATTCCAATTATTTGACATCCAGCTAAGAGGGGCCATC
35 TCTTCTCACCTCTTTCCTAGTCAGTATATTGAGCAAATATTTATTGAGCCCTTACTGTGGGC
AAATCATTGTACTGGATAATTGAGAAAAATAGATAATTCCCTTATTGAGTAAATGTCTACTGA
GCACAATCTAGTGAATCATTACAGTATGGCCTCATTGTTTTGTTTGAGGTGTGTTATTCATA
ACAATATTTTACACCATTCGTATCAATGTAATTATAGAACAATATACGATCAAGGATAAGT
AATTGTGTGGTTATCTGCCATTTAAAGTATCCAGTATTTGATCACATTATTATAAATAATGA
40 AAAAATGATTTAATCTGTAATAAACTGGTTTATTGTGCAGTGAATATACTAGAGTTAT
AATAAATTGTTTACTCTGCCTCACCAAACACATGCTAGGATATAACCCCCAAAATAAGTATTT

AACTTTGCATTAGGTATAAAGGAGACTGGGTGCTATAATTAGATTATTTTGAGGCAGACAGA
GAGCTGTTATCCTAACTGATTTAGTATGTTCTGTAATTGAGAAAATGTTACCAAATTATACT
TTTTAGTGATTTACATGTACATTTTATAGGGGACATGTTCTGTGTATAGCGAATAAATAACTT
TTATAGTATCACAAAATGTTTTGGATTCTTAGTTTCTTATTAGGATATGATGTTTCTTTACCT
5 ATACATTGATTTTCATCACATTTGATTTTTGTCATCTTTTTGCCACATACTTAAAATTTTCCAT
ATAAAATTAATAAAAAATCTGGAAGTGGAATTTAAGCTTTTAATTTGTAGCAATTTTGGAAAT
ATAGCAGAGGTATATTTTCAAGGGAGGTGTGAAATGGAGGTATTATTGTTTGTATTTTATTTT
AGAAATTTTGTGTTGAACCTGCCCAAGATCACACAGCAAGTGTGAGATCCTAGGTTACAGCC
CATAATTTTTCACCTTGTTTCATCTTTTGTACCATGTTTGTGTTAATCCAATGGGGAAGTG
10 TTTTTATAGCTTTATTTTCATGGAGATAGTAACATTGATATAATGGGATCTAGAATTTACTTA
GAGTCAAAGTGACTAGATTTAATTACAAAAAGTATTTTGTGTATACCGTTATACACTCATT
ACATAGGGGTTCTTATGGACAGAGGTTCTATGTATTTTAAAAGAATATGTAGAGGCTATAG
ACTTTTTTGCCCTCTACCTCCCTTCTCCATTTCTTTTTGAAGTTAAGACATTTAAGGGGGAG
AATTAAGGTGCATATTGAGTTGGGGTTTTGTTTAAAAAAATTAGGTAAATACATATATAACT
15 ACAATAACAACAGTTGGTAGTTCTCTAAAATATTTAATTATGGTAAATCAGAAAAGGCCCTTT
AATTTTTGCATCTGTTGAATGAATTGCTTTAGGATCTCTTCTGTGTTGTTTCTCTCTATACT
GAATGCTGTCAATATTATTGAGCACTTACTGTGCCGAACCATGTGCTTTTCAAATATTTCTAT
TTAATCCTAAAAACACTCCTGGGAAATGGGTATTATTCCCATTTTACAGATGAGAAAATTGT
CCTTTGTAGAATACCCATCATGTGATAGCTTCTTGCATATTATGCTATTCAATTTCTCACAACC
20 TTACAAGGTAAATGGATAATGTTCCCTGGATGAGAGATTAACCGTACTCGAAGATCTTATGGT
TGTGTGGCTCAGAGACTCTAAGCATGGCTTTCCAAATCTCATTCTCTTGTAGAGATTTTCCA
TTGTATTGCATTTTGACCATGTTCTGTTTACAAAACCTAAGATTATATTCTGAAGTATCTGT
GACTTGAGTGATAGCTCAGGAATAATCTGAGGGCATTCTATTAGGAAGGAGGACTAGCAAA
GCCTATAGTTAGCCTATCCTAGGCATGAACTATTCTCATTTTAGATAGGATGACTTGATAAA
25 TAGGATGAAGTGAGCAACCTGTTTGTCTTCCGTTATTGCTATCTACAAGACATTCTGCTAGT
CACTGTGGAAAAGTTTCCTTCTCTTGGGTTTTTCTCTTCCCCTTCTCTTAACAGTGTTG
GTTATCTTTATTATGTTTCTGTGTATTTGTAACAGCTGTTTCACCCCTGTATTGCATAATGGG
GATTTTTCTTGAATAAAAAAGTGAGGCATGTAAATTCATTGAGAACTAACCGTAGTTTGTAA
AGAAGTAAAAGGCCTTTGTATCTAGTTTTTGCAGATCACATTTCTTTGAATGCTGGAAAGT
30 GAAATAAACTGGTCTAGGAGCCAAAAATAAATCTTGGGTTTTTTAAGATTTTGTCTTTAGAT
TTGTAACCACCTGGGTACAACTTTTCTTAATTTGTGTAAGAGGTTTTAGAGATGCACTCTG
TCCCTCTTGCTTTACAAATTGCCTACAAGAATGCAACCAAAATACTAGAACTGAACTTAC
ACAAAACCTTTTCAAGTTTTTTCTGTGTTGTAGTTCTAGGGAAGAGGATAGTTTATTAGTATT
TATTTTAAGCTTAAGGATAGGTTTTGGACCTGGATTCCATCATAGTATAACTGCCATCGAAA
35 AGAAGACAAATTGAAATTACAGCATTTACCTAAGGTTACAGTATGGCTGACTTCAGCAGC
CTTACATGGTAAGTTCTCTGTACCAGCTATTTTTGGTTCTCGCAATTGCATCGACTTGGGAA
ATTCCTATCACCATTCTCTATATAGGCTAACCAAACTTTTATCAGTACCTTT

SEQ ID N° 28

40

MSRYLRPPNTSLFVRNVADDTRSEDLRREFGRYGPVVDVYVPLDFYTRRPRGFAYVQFEDVRD
AEDALHNLDRKWICGRQIEIQFAQGDRKTPNQMKAKEGRNVYSSRYDDYDRYRRSRRSRYE
RRRSRSRSFDYNYRRSYSPRKPNCWNTQYSSAYYTSRKI

El término "*TIA1*" (también llamado en la literatura *TIA-1*, *WDM*) se refiere tanto al gen como a la proteína humana. El producto codificado por el gen *TIA1* es un miembro de una familia de proteínas de unión a ARN y posee actividad nucleolítica contra células diana de linfocitos citotóxicos (CTL). Se ha sugerido que esta proteína puede estar involucrada en la inducción de la apoptosis, ya que reconoce de manera preferente los homopolímeros poli (A) e induce la fragmentación del ADN en objetivos CTL. En el contexto de la presente invención, "*TIA1*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia "*TIA1*", y que comprendería diversas variantes procedentes de:

- a) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica de la SEQ ID NO: 30,
- b) Moléculas de ácido nucleico cuya cadena complementaria hibrida con la secuencia polinucleotídica de a),
- c) Moléculas de ácido nucleico cuya secuencia difiere de a) y/o b) debido a la degeneración del código genético,
- d) Moléculas de ácido nucleico que codifican un polipéptido que comprende la secuencia aminoacídica con una identidad de al menos un 80%, un 90%, un 95%, un 98% o un 99% con la SEQ ID NO: 30 en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *TIA1*. Entre otras posibles secuencias nucleotídicas que codifican *TIA1* se encuentra, pero sin limitarse, la SEQ ID NO: 29.

Gene ID: 7072

Localización del gen: 2p13.3

SEQ ID N° 29

```

CTGCGCAGTCTGAGGCGTCCCTCGCGTCGGGTGGAAGAGCGGCTGCGCAACAAGTGAAC
CTATGGATTAAGGTGTAGCCGGGCGGCCCGGAGGTTGAGAGTGCAAGATTATTCTGCGCC
TGCGCGGCCCGGATACGTCTGGTAACCTTTCCAGAGATGTGACAGAAGCTCTAATTCTGCA
ACTCTTTAGCCAGATTGGACCTTGTA AAAACTGCAAAATGATTATGGATACAGCTGGAAATG
ATCCCTATTGTTTTGTGGAGTTTCATGAGCATCGTCATGCAGCTGCAGCATTAGCTGCTATG
AATGGACGGAAGATAATGGGTAAGGAAGTCAAAGTGAATTGGGCAACAACCCCTAGCAGT
CAAAGAAAGATACAAGCAATCATTTCCATGTCTTTGTTGGTGATCTCAGCCCAGAAATTAC
AACTGAAGATATAAAAGCTGCTTTTGCACCATTTGGAAGAATATCAGATGCCCGAGTGGTA
AAAGACATGGCAACAGGAAAGTCTAAGGGATATGGCTTTGTCTCCTTTTTCAACAAATGGG
ATGCTGAAACGCCATTCAACAGATGGGTGGCCAGTGGCTTGGTGGAAGACAAATCAGAA
CTAACTGGGCAACCCGAAAGCCTCCCGCTCCAAAGAGTACATATGAGTCAAATACCAAACA
GCTATCATATGATGAGGTTGTAAATCAGTCTAGTCCAAGCAACTGTACTGTATACTGTGGAG
GTGTTACTTCTGGGCTAACAGAACAACTAATGCGTCAGACTTTTTTACCATTGGAACAATA
ATGGAAATTCGAGTCTTTCCAGATAAAGGATATTCATTTGTTCTGGTTCAATTCCCATGAAAG

```

TGCAGCACATGCAATTGTTTCTGTTAATGGTACTACCATTGAAGGTCATGTTGTGAAATGCT
ATTGGGGCAAAGAACTCTTGATATGATAAATCCCGTGCAACAGCAGAATCAAATTGGATAT
CCCCAACCTTATGGCCAGTGGGGCCAGTGGTATGGAAATGCACAACAAATTGGCCAGTAT
ATGCCTAATGTTTGGCAAGTTCCTGCATATGGAATGTATGGCCAGGCATGGAACCAGCAA
5 GGATTTAATCAGACACAGTCTTCTGCACCATGGATGGGACCAAATTATGGAGTGCAACCGC
CTCAAGGGCAAAATGGCAGCATGTTGCCCAATCAGCCTTCTGGGTATCGAGTGGCAGGGT
ATGAAACCCAGTGAATAAGGACTCCAGAATCTAAAGCCAGTGGCTTGAGGCTACAGGGAG
TGTAATAAGCCGTTGTTTACTTAAAGATTTATCAAATCAGTCAGTGCAAATGTCAGATACA
ATGTATTTATTTAAAAGATTCATTTTTAATCATGAAATTACTTATCATCCACATTGTTTTAAAA
10 AGAAACAAGATGCTGGATGTCTGCCAATTTTTGCCTTCATTACCTTTTTTGATAAAGTTTCTC
AGATCCTTGTTTCAAACACAAATGCAGGGATTGCTGCCACTTTTTAACTATTAAGAGGCAGA
AAATTGCACAATATTGAACTTTTTTCCACTGAAGTAGTGTGCAGTTCTAGTTTGCATTCTGA
TATGATTTAAACATGTAATATAAAGATGTTAAAAAACCACAACTGTGCAGAGTCTAG
AAGTTGTTTGCATCTTCAGCTTGTGCACAATTCTGTTTTAGGTTAAAAAAGGCATTGTTTG
15 AGCTGTCCCATCTCCACTGTTATCCCTTTGGGGTTTTTAAATATAAATTATTAGTTTACATCA
TTTTTGATCTACATCTTTTTTCACAAATTTGTCTTGCCTTATTAAAGTTCTGTAAAATATACT
TAAATGGAAAAATGATGTTTATTAGATTGAAAACTTTTCTCAGATGGATTGATAATTGCAT
TCATCTTGTGTTTTATATGAGAAGGTGCCTCAAGAATTTCTGTTGGATTTGTTTAAAAGGAT
TTTTATCTTTCGTGATAAACTTTGCTGTGTACCAGGAACATAAAAAACAAAACTTGTTACTA
20 AAGAAAATATCTGAAATGTGATAAGTTCTTATGCCATGTTAATTTTCATGTGTCAACTTCAACA
TTTACATGTATTATTTTATTATGTAAAATGTTTTAGCAATTTAATATTTTGCACAGTTAGCAAA
CTTTGTATGTCATTTCTTCAAGGCATCATGCAGAGTTGACATGAGATTTATAAGGTTTTAA
GTTGTTTGCATGTGAAAATCAAATACATACTTTGGTAGTCTTTGAATACAAAGTCATCTGCTC
TTGTTTTTCAAGAATTTTGAGACACAAAGTTGTATGTAAAGGAATATATTAATTTGCCGTTTT
25 CTAGGTAGATTTGCTCAAAAAGAGTGAATCAACTTAATATGTACAAATGATAGCTGTGAAAC
TGTAAGATATCTTTGTGTGAGGCTTGGAGTTCATTGTGACCTCCAAATTTTGCCTGAAGGAC
CAGCTGGGCAAAGCATCTTTTAAATGTTTCAAGAGGCCAAAGATAAACAAAAAACCCTTA
AAATCCTACCTCTTTAAACAGCCTTCAGATAAGAGAATCCTCAGTGCAATCATTATTTTGATT
CGTTTGGTACCTGTTTTTCTGGAGTTCCCGATTTTATTATTTGGGGTGGCTCCAAGCATT
30 AGAGGTTTAATCTTTGATGGCATTGTTCTAGTTTTGAAATTTCTAGTATATTTTCAAGTCTCT
TAGAAGACTTGTGTGGGAAGTTTCACTTTGTTTTCAGTGAAGATCACAAACCTCCTTCTTCC
TTTACTCAAGAGGAAAGGTCCCAGTATACATATTTGAATGGTTGATGGTTTTCAAGACCTTC
AGGGAGCTCCCTGCATTTTACCTAGAAACAGAAAAGGCCCGCAAATCTTAAGTTTCTTG
CCTGCATTTCCCGGGTAGGGGCAATGACTCCAAGCTGGTCTCTAAGCCAATACCCTTATA
35 AACCAGAGCCCAGGAAAGACAGCTCGAGTGTATAATTCTCTGGAGCTCAATTCTATGCAGT
TGTGCTGATATTTTATTAAAGTCACTGTGTATTTTTAAGTGTGATACATTAAGTGCCTTTA
TGGAAGATGAGTAAATTTTTTAAATACTTGGAAATTTTATTTCTTGTAACTTCTACAGATC
AGGGCATGCAACCAAAAGCAGCTTAAATGAAATATTTTAAATAAATATCAGGAAGCTATT
TTTAGATTTCTTCTGGCTTATGTTTCTACTTTAGGACCCTCATTGTTCTCTATTAAAAAAT
40 TATTTCTGTGCATCTCATGGACTGCAGGGTAAATTATTTGGGCATAAATAATTTAAATAGTT
TTCTTTTCAATTTTGAATATCTCCAGTAATAACAGTTTTTATTATCCAGCATATTGGCTTATTGCA
CAAATCTTAAATGTACATTGACTACTTTCTGAGAAGAAAGTGGTATCAGTACTCATGATGA
AAAGGTTACTACTGAACAAATTCACATTTTCAAGAACACCTCTATCTTTGGTTTAAATCTTACT
CTTAGTTTTTCCGTCTAAAAATCATACTGGTATTAGTATCAGGTAAGGAAATTAAGTTTTTA
45 AAATGGTTTCAATCTCTGCAATATGCAAAATTTAGATTTTACTTTCTGGTACTGTAAAGAACC
TGAAGTGATTTACACTTAATGGGTGATTAATCCAGTATTCTTTACCCTGAATGTTTGGATATT
AAAGTTCCTTTATGTTTTCTATAACCTGTGGGATCTTCTTGCAGTGATTATTGTGTGTGAGAT
TTTTTTTCTTTTGGTCTATCCATATTGTTATATTCAGGATTTTTTTTTTAAATCTTATTC
CAGAATCAGTGGTTTATATTGGGTACTGTTTAAACACCAAATGGAATTGGCATTCTGCAGAT

TTAATTAATTATGAAACCAGGGTCTCATTTTCCTTGCTGATACTTGTTGAAAATGAGATTCAC
 ATTCTAGTCTTTATTTTCCTCCTGTTTTGTCCCTGTGCTTGACATCTTCCTTTTATTTGTGTG
 TTATAGTTCTATTCCATTTGAGAAGGCAGTTGGTAAGAACTAGATTGCATGTACAAAGACAG
 GTTTACTAAGTGCTGTACAGTGGTCCTGAGGTTACAGTTGAATTAGAAAAACGAAATGTACT
 5 TACAGGAAATAAGAAAGCAAACCTTTCAAATGAGAGTGATGATTTCTTTAAAAAAATCAGT
 TTTTTTCTCTCAAATAATGTTCTTTATTTTACGAAATCGTCAATCTTAAGCATGAGCAGGGAT
 AAACAACTCCTAGAAGGAAGCAATTCATTCTTCCTGGATTTTCTCTGTTGTTAAATCACAAA
 AATGATAGTCCCCAATCGTTTCTTTATAGGAGGTTATTACATTTTATTACAGTCACTGCATTT
 TGAAGTGTGTTTGAATGTACATCCAAAATGATGAGTTTCAATTTAAGAGCCTTA
 10 ATAAAATGTGTGAGTGTGTCTCAATTGAA

SEQ ID N° 30

MIMDTAGNDPYCFVEFHEHRHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSNHFHV
 GDLSPEITTEDIKAAFAPFGRISDARVVKDMATGKSKGYGFVSFFNKWDAENAIQQMGGQWL
 15 GRQIRTNWATRKPAPKSTYESNTKQLSYDEVVNQSSPSNCTVYCGGVTSGLTEQLMRQTFS
 PFGQIMEIRVFPDKGYSFVRFNHESAAHAIVSVNGTTIEGHVVKCYWGKETLDMINPVQQQN
 QIGYPQPYGQWGQWYGNAQQIGQYMPNGWQVPAYGMYGQAWNQQGFNQTQSSAPWMGP
 NYGVQPPQGQNGSMLPNQPSGYRVAGYETQ

Métodos de la invención

20 Por tanto, **otro aspecto** de la invención se refiere a un método de obtención de datos útiles para diagnosticar, predecir o pronosticar el desarrollo de CaP en un individuo, de ahora en adelante primer método de la invención, que comprende los siguientes pasos:

25 a) Determinar los niveles del producto de expresión de los genes *RNU6*, *SF3B1*, *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGO1*, *SND1*, *SRSF2*, *SRSF3*, *SRRM1*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1* o cualquiera de sus combinaciones, en una muestra biológica ex vivo de dicho individuo,

b) Comparar los niveles obtenidos en el paso anterior con los estándares.

30 En una realización preferida de este aspecto de la invención se determinan los niveles del producto de expresión de todos los genes simultáneamente.

En la presente invención "**muestra biológica**", tal como aquí se utiliza, se refiere, pero no se limita, a células, tejidos y/o fluidos biológicos de un organismo, obtenidos mediante cualquier método conocido por un experto en la materia; es decir, cualquiera seleccionada del grupo de fluidos corporales o células, o de un extracto celular o el sobrenadante del cultivo celular de
 35 dichas células, y/o biopsia líquida (sangre, suero, plasma, orina,...). Preferiblemente la muestra es un tejido.

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

5 Los niveles de expresión de los genes van a dar un determinado perfil de expresión génica. El término "nivel", "nivel de expresión", también denominado "cantidad producto génico" se refiere al material bioquímico, ya sea ARN o proteína, resultado de la expresión de un gen. Algunas veces se usa una medida de la cantidad de producto génico para inferir qué tan activo es un gen. Se entiende por "perfil de expresión génica" el perfil génico obtenido tras la cuantificación del

10 ARNm y/o de proteína producida por los genes de interés o biomarcadores, es decir, por los genes empleados como marcadores biológicos en la presente invención, en una muestra biológica aislada. El perfil de expresión de los genes se realiza, preferiblemente, determinando el nivel de ARNm derivado de su transcripción, previa extracción del ARN total presente en la muestra biológica aislada, lo cual puede realizarse mediante protocolos conocidos en el estado

15 de la técnica. La determinación del nivel de ARNm derivado de la transcripción de los genes empleados como marcadores biológicos en la presente invención, puede realizarse, por ejemplo, aunque sin limitarnos, mediante amplificación por reacción en cadena de la polimerasa (PCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa (RT-PCR), RT-PCR cuantitativa (qPCR), retrotranscripción en combinación con la reacción en cadena de la

20 ligasa (RT-LCR), o cualquier otro método de amplificación de ácidos nucleicos; análisis en serie de la expresión génica (SAGE, SuperSAGE); chips de ADN elaborados con oligonucleótidos depositados por cualquier mecanismo; microarrays de ADN elaborados con oligonucleótidos sintetizados in situ mediante fotolitografía o por cualquier otro mecanismo; hibridación in situ utilizando sondas específicas marcadas con cualquier método de mareaje; mediante geles de

25 electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante resonancia magnética nuclear o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio. El perfil de expresión génica también podría obtenerse mediante la detección y/o cuantificación de las proteínas producto de la

30 traducción del ARNm derivado de la transcripción de los genes empleados como marcadores biológicos en la presente invención, mediante por ejemplo, pero sin limitarnos, inmunodetección por western blot. La detección cuantitativa de la expresión de los genes empleados como marcadores biológicos en la presente invención puede realizarse más preferiblemente mediante PCR en tiempo real (RT-PCR ó RTqPCR). La detección en tiempo real de los productos

35 amplificados puede llevarse a cabo mediante la utilización de moléculas fluorescentes que se intercalan en el ADN de cadena doble o mediante hibridación con diferentes tipos de sondas.

Así pues, la detección de los niveles de proteínas o del nivel de expresión de los genes puede hacerse por cualquiera de las técnicas conocidas por el experto en la materia.

Los pasos (b) y/o (c), del método descrito anteriormente pueden ser total o parcialmente automatizados, por ejemplo, por medio de un equipo robótico sensor para la detección de la presencia en el paso (b) o la clasificación computarizada en el paso (c).

En una realización preferida de este aspecto de la invención, la detección del paso (b) se realiza mediante PCR, preferiblemente RT-PCR.

En el caso de que se detecte el nivel o los niveles de proteínas puede hacerse, al igual que en el caso de los genes, por cualquiera de las técnicas conocidas por el experto en la materia. Así, en otra realización preferida de este aspecto de la invención, la detección de los niveles de expresión se realiza por una técnica inmunológica.

En una realización más preferida, las técnicas inmunológicas están basadas en reacciones de precipitación, de aglutinación, inmunomarcación, radioinmunoanálisis y técnicas radioinmunométricas, preferentemente ELISA (*Enzyme Linked Immunoabsorbent Assay*), o Western blot o en cualquiera de sus combinaciones.

Por tanto, **otro aspecto** de la invención se refiere a un método para diagnosticar, predecir o pronosticar el desarrollo de CaP en un individuo, de ahora en adelante segundo método de la invención, que comprende los pasos (a) y (b) del primer método de la invención, y además comprende:

(c) incluir a los individuos que presentan una mayor expresión de los factores *RNU6*, *SF3B1*, *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGO1*; *SND1*, *SRSF2*, *SRSF3*, *SRRM1*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1*, o cualquiera de sus combinaciones, respecto a una muestra de referencia, en el grupo de individuos con mayor probabilidad de desarrollar CaP.

Una "muestra de referencia", como se usa aquí, significa una muestra obtenida de un grupo de sujetos sanos que no tiene un estado de enfermedad o fenotipo particular

Los niveles de referencia adecuada expresión de los genes puede ser determinada mediante la medición de los niveles de expresión de dichos genes en varios temas adecuados, y esos niveles de referencia se puede ajustar a las poblaciones específicas (por ejemplo, un nivel de referencia puede estar relacionada con la edad, por lo que las comparaciones se puede hacer entre los niveles de expresión en las muestras de los sujetos de una cierta edad y niveles de referencia para una enfermedad particular, el fenotipo, o falta de ella en un determinado grupo de edad). En una realización preferida, la muestra de referencia se obtiene de tejido prostático adyacente

no tumoral. El experto en la técnica apreciará que el tipo de muestra de referencia puede variar dependiendo del método específico a realizar.

El perfil de expresión de los genes en la muestra de referencia de preferencia puede ser generado a partir de una población de dos o más personas. La población, por ejemplo, pueden
5 contener 3, 4, 5, 10, 15, 20, 30, 40, 50 o más personas. Además, el perfil de expresión de los genes en la muestra de referencia y en la muestra de la persona que va a ser diagnosticada de acuerdo con los métodos de la presente invención pueden ser generados a partir de la misma persona, siempre que los perfiles sean analizados y el perfil de referencia son generados a partir
10 de las muestras biológicas tomadas en diferentes momentos y se comparan entre sí. Por ejemplo, una muestra de un individuo puede obtener en el inicio de un período de estudio. Un perfil de marcador biológico de referencia de esta muestra se puede comparar con los perfiles de biomarcadores generados a partir de las muestras posteriores de la misma persona. En una realización preferida, la muestra de referencia es un conjunto de muestras de varios individuos.

Una vez que los niveles de expresión de los genes marcadores en relación con los valores de
15 referencia para dichos genes se han determinado, es necesario identificar si existen alteraciones en la expresión de dichos genes (aumento o disminución de la expresión). La expresión de un gen se considera aumentada en una muestra de la materia objeto de estudio cuando los niveles de incremento con respecto a la muestra de referencia son al menos de un 5%, por lo menos 10%, por lo menos 15%, por lo menos el 20%, al menos un 25%, por lo menos 30%, por lo menos
20 el 35%, por lo menos el 40%, por lo menos 45%, por lo menos el 50%, por lo menos el 55%, por lo menos el 60%, por lo menos el 65%, por lo menos el 70%, por lo menos el 75%, por lo menos el 80%, por lo menos el 85%, por lo menos el 90%, por lo menos el 95%, por lo menos 100%, por lo menos 110%, por lo menos 120%, por lo menos 130%, por lo menos 140%, por lo menos 150%, o más. Del mismo modo, la expresión de un gen se considerada disminuida cuando
25 sus niveles disminuyen con respecto a la muestra de referencia en al menos un 5%, por lo menos 10%, por lo menos 15%, por lo menos el 20%, por lo menos el 25%, al menos un 30%, por lo menos el 35%, por lo menos el 40%, por lo menos 45%, por lo menos el 50%, por lo menos el 55%, por lo menos el 60%, por lo menos el 65%, por lo menos 70%, por lo menos el 75%, por lo menos el 80%, por lo menos el 85%, por lo menos el 90%, por lo menos el 95%, por lo menos
30 100% (es decir, ausente).

KIT O DISPOSITIVO DE LA INVENCION

Otro **aspecto** de la invención se refiere a un kit o dispositivo, de ahora en adelante kit o dispositivo de la invención, que comprende los elementos necesarios para detectar los niveles de expresión
35 de snRNP200, SRSF3 y/o SRRM1, o cualquiera de sus combinaciones, y donde:

- Los cebadores o primers son secuencias de polinucleótidos de entre 10 y 30 pares de bases, más preferiblemente de entre 15 y 25 pares de bases, aún más preferiblemente de entre 18 y 22 pares de bases, y aún mucho más preferiblemente de alrededor de 20 pares de bases, que presentan una identidad de al menos un 80%, más preferiblemente de al menos un 90%, aún más preferiblemente de al menos un 95%, aún mucho más preferiblemente de al menos un 98%, y particularmente de un 100%, con un fragmento de las secuencias complementarias a la SEQ ID N° 1 (*RNU6*), SEQ ID N°: 2 (*SFR3B1*), SEQ ID N° 4 (*SNRNP200*), SEQ ID N° 6 (*RNU12*), SEQ ID N° 7 (*ESRP1*), SEQ ID N° 8 (*ESRP2*), SEQ ID N° 10 (*MAGOH*), SEQ ID N° 12 (*SND1*), SEQ ID N° 14 (*SRSF2*), SEQ ID N° 16 (*SRSF3*), SEQ ID N° 18 (*SRRM1*), SEQ ID N° 20 (*RNU4*), SEQ ID N° 21 (*RBM45*), SEQ ID N° 23 (*SRSF4*), SEQ ID N° 25 (*SRSF5*), SEQ ID N° 27 (*SRSF10*), SEQ ID N° 29 (*TIA1*),

- Las sondas son secuencias de polinucleótidos de entre 80 y 1100 pares de bases, más preferiblemente de entre 100 y 1000 pares de bases, y aún más preferiblemente de entre 200 y 500 pares de bases, que presentan una identidad de al menos un 80%, más preferiblemente de al menos un 90%, aún más preferiblemente de al menos un 95%, aún mucho más preferiblemente de al menos un 98%, y particularmente de un 100%, con un fragmento de las secuencias complementarias a la SEQ ID N° 1 (*RNU6*), SEQ ID N°: 2 (*SFR3B1*), SEQ ID N° 4 (*SNRNP200*), SEQ ID N° 6 (*RNU12*), SEQ ID N° 7 (*ESRP1*), SEQ ID N° 8 (*ESRP2*), SEQ ID N° 10 (*MAGOH*), SEQ ID N° 12 (*SND1*), SEQ ID N° 14 (*SRSF2*), SEQ ID N° 16 (*SRSF3*), SEQ ID N° 18 (*SRRM1*), SEQ ID N° 20 (*RNU4*), SEQ ID N° 21 (*RBM45*), SEQ ID N° 23 (*SRSF4*), SEQ ID N° 25 (*SRSF5*), SEQ ID N° 27 (*SRSF10*), SEQ ID N° 29 (*TIA1*),

- Los anticuerpos son capaces de unirse específicamente a una región formada por cualquiera de las secuencias aminoacídicas SEQ ID N°: 3 (*SFR3B1*), SEQ ID N° 5 (*SNRNP200*), SEQ ID N° 9 (*ESRP2*), SEQ ID N° 11 (*MAGOH*), SEQ ID N° 13 (*SND1*), SEQ ID N° 15 (*SRSF2*), SEQ ID N° 17 (*SRSF3*), SEQ ID N° 19 (*SRRM1*), SEQ ID N° 22 (*RBM45*), SEQ ID N° 24 (*SRSF4*), SEQ ID N° 26 (*SRSF5*), SEQ ID N° 28 (*SRSF10*), SEQ ID N° 30 (*TIA1*),

En otra realización preferida de este aspecto de la invención, los cebadores, sondas o anticuerpos están modificados o marcados, por ejemplo, pero sin limitarnos, mediante un marcaje radiactivo o inmunológico. Así, preferiblemente, los oligonucleótidos presentan modificaciones en alguno de sus nucleótidos, como por ejemplo, pero sin limitarnos a, nucleótidos que tengan alguno de sus átomos con un isótopo radiactivo, normalmente o nucleótidos marcados inmunológicamente, como por ejemplo con una molécula de digoxigenina, y/o inmovilizadas en una membrana. Varias posibilidades son conocidas en el estado de la técnica.

AUTOMATIZACIÓN DEL MÉTODO DE LA INVENCION IMPLEMENTÁNDOLO EN UN PROGRAMA DE ORDENADOR

Otro aspecto de la invención se refiere a un programa de ordenador que comprende instrucciones para realizar el procedimiento de acuerdo con cualquiera de los métodos de la invención.

En particular, la invención abarca programas de ordenador dispuestos sobre o dentro de una portadora. La portadora puede ser cualquier entidad o dispositivo capaz de soportar el programa. Como variante, la portadora podría ser un circuito integrado en el que va incluido el programa y que se haya adaptado para ejecutar, o para ser utilizado en la ejecución de los procesos correspondientes.

Por ejemplo, los programas podrían estar incorporados en un medio de almacenamiento, como una memoria ROM, una memoria CD ROM o una memoria ROM de semiconductor, una memoria USB, o un soporte de grabación magnética, por ejemplo, un disco flexible o un disco duro. Alternativamente, los programas podrían estar soportados en una señal portadora transmisible; por ejemplo, podría tratarse de una señal eléctrica u óptica que podría transportarse a través de cable eléctrico u óptico, por radio o por cualesquiera otros medios.

La invención se extiende también a programas de ordenador adaptados para que cualquier medio de procesamiento pueda llevar a la práctica los métodos de la invención. Los programas de ordenador también abarcan aplicaciones en la nube basadas en dicho procedimiento.

Otros aspectos de la invención se refieren al medio de almacenamiento legible y a la señal transmisible que comprende instrucciones de programa necesarias para la ejecución del método de invención por un ordenador.

EJEMPLOS DE LA INVENCION.

Para los estudios de los niveles de expresión de varios componentes de la maquinaria de splicing, se analizaron 2 cohortes independientes: **1)** Cohorte de 84 muestras de tejido prostático fijadas con formalina y embebidas en parafina (FFPE) de pacientes con CaP. Cada muestra presento tejido tumoral y tejido adyacente no tumoral, usado como control en el estudio. Los tejidos tumorales y no tumorales fueron separados por patólogos expertos. **2)** Cohorte de 42 muestras de tejido prostático obtenidas a partir de pacientes con alto riesgo de padecer CaP (PSA > 20,0 y/o resultado de tacto rectal positivo) mediante una biopsia prostática por punción con aguja gruesa.

RESULTADOS

Se ha demostrado que la expresión de 26 componentes de la maquinaria de splicing y factores asociados a la misma se encuentran desregulada en CaP. En concreto 25 de ellos se encuentran sobreexpresados, mientras que solo la expresión de *SRRM4* disminuye en muestras de tejido tumoral prostático con respecto a la zona adyacente no tumoral (Fig 1A). De hecho, la expresión de estos 26 genes es capaz de diferenciar entre muestras tumorales y sus respectivas regiones adyacentes no tumorales con un área bajo la curva de 0,877 (Fig 1B).

Además, algunos de estos factores se encontraron relacionados con parámetros clínicos de agresividad tumoral. Concretamente, la expresión de *RNU6*, *SF3B1*, *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *SRSF3* y *SRRM1* se encontró correlacionada positivamente con el grado de Gleason de las muestras (Tabla 1). Por otra parte, la expresión de *RNU4*, *snRNP200*, *ESRP2*, *SRSF3*, *SRRM1*, y *TIA1* se encontró correlacionada positivamente con el estadio T. Asimismo, la elevada expresión de una serie de factores se asoció con la presencia de invasión perineural y/o linfovascular, así como con CaP significativo desde un punto de vista clínico, es decir, CaP con un grado de Gleason mayor o igual a 7 (Tabla 1). Adicionalmente, la elevada expresión de 11 componentes del spliceosoma y factores de splicing (*snRNP200*, *ESRP1*, *ESRP2*, *RBM45*, *SND1*, *SRSF3*, *SRSF4*, *SRSF5*, *SRSF10*, *SRRM1*, *TIA1*) se asoció con el desarrollo posterior de recidivas de CaP (Fig 2A), sugiriendo un posible papel de estos factores (individuales o en conjunto) como marcadores de pronóstico tumoral del CaP. De hecho, los niveles de expresión combinados de estos 11 genes dibujaron una curva ROC con un área bajo la curva de 0.806 para diferenciar entre muestras de pacientes que recidivarán de la enfermedad y pacientes que no lo harán (Fig 2B).

Cabe destacar que *snRNP200*, *SRSF3* y *SRRM1*, fueron los únicos factores que se encontraron alterados significativa y consistentemente en todas las comparativas realizadas en este estudio (Fig 1A, Fig 2B, Tabla 1), indicando la elevada importancia de estos tres factores en la fisiopatología del CaP y definiendo, por tanto, una posible utilidad como novedosas dianas terapéuticas para esta patología. Por tanto, estos tres factores fueron los seleccionados para estudios más profundos.

En una cohorte independiente de biopsias de CaP significativo (n=42), se observó que un elevado nivel de *snRNP200*, *SRSF3* y *SRRM1* se asoció con la presencia de metástasis en el momento del diagnóstico, y en concreto, con metástasis de alta carga (Fig 3A), es decir, aquellas más agresivas. Además, mayores niveles de expresión de *snRNP200* (pero no de *SRRM1*, ni *SRSF3*) se asociaron significativamente con un nivel de Gleason mayor o igual que 8 (Fig 3B), aunque estos niveles de expresión no llegaron a correlacionar significativamente con el Grado Gleason (tendencia con $p=0.074$ y $R=0.272$ en el caso de *snRNP200*) (Fig 3C). Pese a esto, y aunque los niveles de expresión de *SRRM1* y *SRSF3* correlacionaron positivamente con los

niveles de expresión del marcador de diagnóstico "PSA", no se observó esta correlación con el factor *snRNP200*, poniendo de manifiesto la ya conocida limitada capacidad del PSA como marcador de agresividad y pronóstico del CaP (Fig 3D). Por último, se observó que tanto los niveles de expresión de *snRNP200*, como *SRSF3* y *SRRM1* correlacionaron positivamente con los niveles de expresión de la variante de splicing AR-v7 (Fig 3E) pero no con los niveles de la variante canónica AR (Fig 3F). De hecho, esto se reafirmó con las evidentes, significativas y pronunciadas correlaciones positivas observadas entre los niveles de expresión de estos tres factores de splicing y el ratio de expresión AR-v7/AR (Fig 3G). Estos resultados sugieren de nuevo el elevado potencial de estos tres factores (*snRNP200*, *SRRM1* y *SRSF3*) como dianas terapéuticas para el tratamiento del CaP y el CaP resistente a la castración (CPRC).

	Correlaciones		Asociaciones		
	Gleason Score	Estadio T	Invasión perineural	Invasión linfovascular	CaP Sig
RNU4	0,122	0,239*	9664718823 ± 3162530673 **	7730140948 ± 3078119236 *	7526218231 ± 3298861131 *
RNU6	0,235*	0,075	230034826 ± 7 2616606 **	119362920 ± 102513311	163738961 ± 75084561 *
PRPF40A	0,184	0,214	1801231 ± 878336 *	1669010 ± 855998	1789300 ± 832618 *
SF3B1	0,312**	0,121	9003875 ± 3487752 *	10853954 ± 4783529 **	5808930 ± 3941225
U2AF2	0,09	0,083	714287 ± 348778 *	807889 ± 481128	347329 ± 370880
snRNP200	0,264*	0,247*	2681185105 ± 1175955224 *	2812777484 ± 1275566115 *	2125357393 ± 1042911480 *
RNU12	0,321**	0,178	395441 ± 132486 **	133816 ± 149579	208236 ± 139628
ESRP1	0,319**	0,166	1658118 ± 916550	2151271 ± 1084569	1874015 ± 936328 *
ESRP2	0,281*	0,254*	1953298 ± 1179570	3964985 ± 1266549 **	2941102 ± 1300974 *
KHDRBS1	0,128	0,215	8816079 ± 4355945 *	7047264 ± 5942380	10326770 ± 4988242 *
MAGOH	0,331**	0,173	572836 ± 376052	1036475 ± 489737 *	772768 ± 395132
NOVA1	0,131	0,211	1375554 ± 1240422	2891594 ± 1397232 *	1968158 ± 1417671
PTB	0,189	0,197	462842 ± 504081	707443 ± 685520	923496 ± 547502
RBM3	0,141	0,139	1831860 ± 1133513	2948642 ± 1333148 *	2051801 ± 1189264
RBM45	0,145	0,159	2144099 ± 1086705	2444946 ± 1430495	1613514 ± 1237865
SND1	0,236*	0,167	6829797 ± 3127408 *	8116316 ± 3698131 *	7483290 ± 3461207 *
SRSF2	0,284*	0,161	903780 ± 409990 *	1099682 ± 533886 *	914182 ± 460892
SRSF3	0,289*	0,274*	13698858 ± 5050576 **	12653631 ± 5712504 *	10070038 ± 5031330 *
SRSF4	0,139	0,169	2131612 ± 857251 *	3422072 ± 1068338 **	941268 ± 902489
SRSF5	0,055	0,129	105348741 ± 37924790 **	141185003 ± 48026387 **	35477260 ± 37643655
SRSF9	0,088	0,211	33397890 ± 12917648 *	28501453 ± 15538733	26539564 ± 14468693
SRSF10	0,057	0,218	10135257 ± 4077596 *	10622878 ± 5604567	4248356 ± 4293783
SRRM1	0,211*	0,224*	9522292 ± 4314271 *	9444197 ± 4007860 *	9478531 ± 4760093 *
SRRM4	-0,154	0,107	53981 ± 33971	-2375 ± 34689	47458 ± 36610
TIA1	0,114	0,246*	5965156 ± 2459924 *	6334929 ± 2989097	5493989 ± 2630409*
TRA2A	0,141	0,191	27361791 ± 12693441 *	39100412 ± 15368002	29010940 ± 15801444

Leyenda del grado de correlación (Coeficiente R)



Tabla 1. Numerosos componentes de la maquinaria de splicing y factores asociados correlacionan positivamente o se asocian con parámetros clínicos de agresividad tumoral característicos del CaP en la zona tumoral de muestras fijadas en parafina (n=84). El color dentro de cada cuadro de correlaciones indica el grado de correlación (R de Spearman), que también se encuentra detallado numéricamente en el interior de cada cuadro. En el interior de los cuadros de las asociaciones clínicas se indica la diferencia entre las medias de los grupos ± error. Los asteriscos (*, p<0.05; **, p<0.01; ***, p<0.001) indican diferencias estadísticamente significativas.

REIVINDICACIONES

- 1.- Uso del producto de expresión de los genes *snRNP200*, *SRSF3* y *SRRM1* para diagnosticar, predecir o pronosticar el desarrollo de Cáncer de Próstata (CaP) en un individuo.
- 2.- El uso simultáneo del producto de expresión de los genes *snRNP200*, *SRSF3* y *SRRM1* según la reivindicación anterior, y además de los genes *RNU6*, *SF3B1*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1*
- 3.- Un método *in vitro* de obtención de datos útiles para diagnosticar, predecir o pronosticar el desarrollo de Cáncer de Próstata (CaP) en un individuo, que comprende:
 - a) Determinar en una muestra biológica aislada de dicho individuo los niveles del producto de expresión de los genes *snRNP200*, *SRSF3* y/o *SRRM1*, y más preferiblemente también de los genes *RNU6*, *SF3B1*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1*, según se define en las reivindicaciones 1-2, y
 - b) Comparar los niveles obtenidos en el paso anterior en relación a la cantidad de expresión detectada para dichos genes en una muestra referencia de respuesta conocida.
- 4.- Un método *in vitro* para predecir o pronosticar el desarrollo de Cáncer de Próstata (CaP) en un individuo, que comprende los pasos (a) y (b) según la reivindicación 2, y además comprende incluir en el grupo de individuos con mayor riesgo de desarrollo de CaP aquellos pacientes con elevados niveles de expresión de los factores de splicing *RNU6*, *SF3B1*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1*, en comparación con una muestra de referencia.
- 5.- El método según cualquiera de las reivindicaciones 2-3 donde la muestra biológica aislada es un tejido fresco, tejido embebido en parafina o ARN extraído de un tejido de un individuo.
- 6.- El método según cualquiera de las reivindicaciones 2-4 donde la detección del producto de expresión de los genes se realiza mediante cualquier técnica de expresión génica como RT-PCR y preferiblemente por PCR cuantitativa (qPCR).
- 7.- El método según cualquiera de las reivindicaciones 2-4 donde el producto de expresión son proteínas, preferiblemente *SF3B1*, *SNRNP200*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*,

SRSF3, SRRM1, RBM45, SRSF4, SRSF5, SRSF10 y/o TIA1, o cualquiera de sus combinaciones.

8.- El método según la reivindicación anterior donde las proteínas se identifican por un método inmunohistoquímico, preferiblemente seleccionado de entre Western Blot, espectrometría de masa, fluorescencia, absorción y/o luminiscencia, y más preferiblemente se utiliza Western Blot.

9.- El método según la reivindicación anterior donde al individuo se le incluye en el grupo de individuos con mayor probabilidad de desarrollar CaP, cuando estos componentes de maquinaria de splicing específicos, incluyendo *RNU6*, *SF3B1*, *snRNP200*, *RNU12*, *ESRP1*, *ESRP2*, *MAGOH*, *SND1*, *SRSF2*, *SRSF3*, *SRRM1*, *RNU4*, *RBM45*, *SRSF4*, *SRSF5*, *SRSF10*, y/o *TIA1*, o cualquiera de sus combinaciones, presentan altos niveles de expresión por debajo del tercil inferior de expresión, con un AUC de 0.902 (sensibilidad del 85% y especificidad del 82%).

10.-Un kit o dispositivo que comprende los elementos necesarios para cuantificar los niveles de expresión de los genes según se definen en las reivindicaciones 1-9, y donde:

- los cebadores o primers son secuencias de polinucleótidos de entre 10 y 30 pares de bases, más preferiblemente de entre 15 y 25 pares de bases, aún más preferiblemente de entre 18 y 22 pares de bases, y aún mucho más preferiblemente de alrededor de 20 pares de bases, que presentan una identidad de al menos un 80%, más preferiblemente de al menos un 90%, aún más preferiblemente de al menos un 95%, aún mucho más preferiblemente de al menos un 98%, y particularmente de un 100%, con un fragmento de las secuencias complementarias a la SEQ ID N° 1 (*RNU6*), SEQ ID N°: 2 (*SFR3B1*), SEQ ID N° 4 (*SNRNP200*), SEQ ID N° 6 (*RNU12*), SEQ ID N° 7 (*ESRP1*), SEQ ID N° 8 (*ESRP2*), SEQ ID N° 10 (*MAGOH*), SEQ ID N° 12 (*SND1*), SEQ ID N° 14 (*SRSF2*), SEQ ID N° 16 (*SRSF3*), SEQ ID N° 18 (*SRRM1*), SEQ ID N° 20 (*RNU4*), SEQ ID N° 21 (*RBM45*), SEQ ID N° 23 (*SRSF4*), SEQ ID N° 25 (*SRSF5*), SEQ ID N° 27 (*SRSF10*), SEQ ID N° 29 (*TIA1*),

- Las sondas son secuencias de polinucleótidos de entre 80 y 1100 pares de bases, más preferiblemente de entre 100 y 1000 pares de bases, y aún más preferiblemente de entre 200 y 500 pares de bases, que presentan una identidad de al menos un 80%, más preferiblemente de al menos un 90%, aún más preferiblemente de al menos un 95%, aún mucho más preferiblemente de al menos un 98%, y particularmente de un 100%, con un fragmento de las secuencias complementarias

a la SEQ ID N° 1 (*RNU6*), SEQ ID N°: 2 (*SFR3B1*), SEQ ID N° 4 (*SNRNP200*), SEQ ID N° 6 (*RNU12*), SEQ ID N° 7 (*ESRP1*), SEQ ID N° 8 (*ESRP2*), SEQ ID N° 10 (*MAGOH*), SEQ ID N° 12 (*SND1*), SEQ ID N° 14 (*SRSF2*), SEQ ID N° 16 (*SRSF3*), SEQ ID N° 18 (*SRRM1*), SEQ ID N° 20 (*RNU4*), SEQ ID N° 21 (*RBM45*), SEQ ID N° 23 (*SRSF4*), SEQ ID N° 25 (*SRSF5*), SEQ ID N° 27 (*SRSF10*), SEQ ID N° 29 (*TIA1*),

- Los anticuerpos son capaces de unirse específicamente a una región formada por cualquiera de las secuencias aminoacídicas SEQ ID N°: 3 (*SFR3B1*), SEQ ID N° 5 (*SNRNP200*), SEQ ID N° 9 (*ESRP2*), SEQ ID N° 11 (*MAGOH*), SEQ ID N° 13 (*SND1*), SEQ ID N° 15 (*SRSF2*), SEQ ID N° 17 (*SRSF3*), SEQ ID N° 19 (*SRRM1*), SEQ ID N° 22 (*RBM45*), SEQ ID N° 24 (*SRSF4*), SEQ ID N° 26 (*SRSF5*), SEQ ID N° 28 (*SRSF10*), SEQ ID N° 30 (*TIA1*).

11.- El uso del kit o dispositivo según la reivindicación anterior para diagnosticar, predecir o pronosticar el desarrollo de Cáncer de Próstata (CaP) en un individuo.

12.- Un programa de ordenador que comprende instrucciones de programa para hacer que un ordenador lleve a la práctica el procedimiento de acuerdo con el método según cualquiera de las reivindicaciones 3 a 5.

13.- Un medio de almacenamiento legible por un ordenador que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos del método según cualquiera de las reivindicaciones 3 a 9.

14.- Una señal transmisible que comprende instrucciones de programa capaces de hacer que un ordenador lleve a cabo los pasos del método según cualquiera de las reivindicaciones 3 a 9

Figure 1.

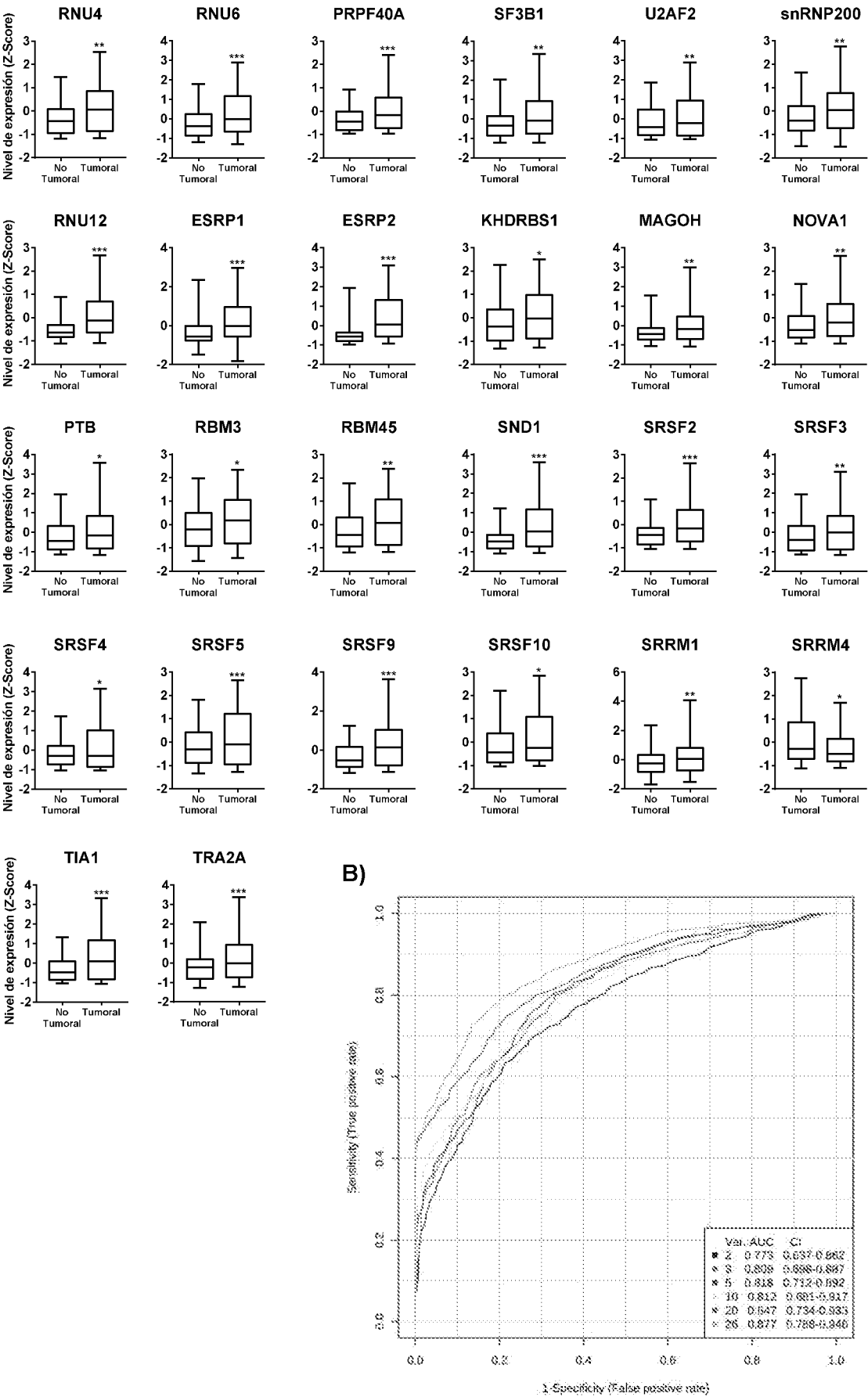
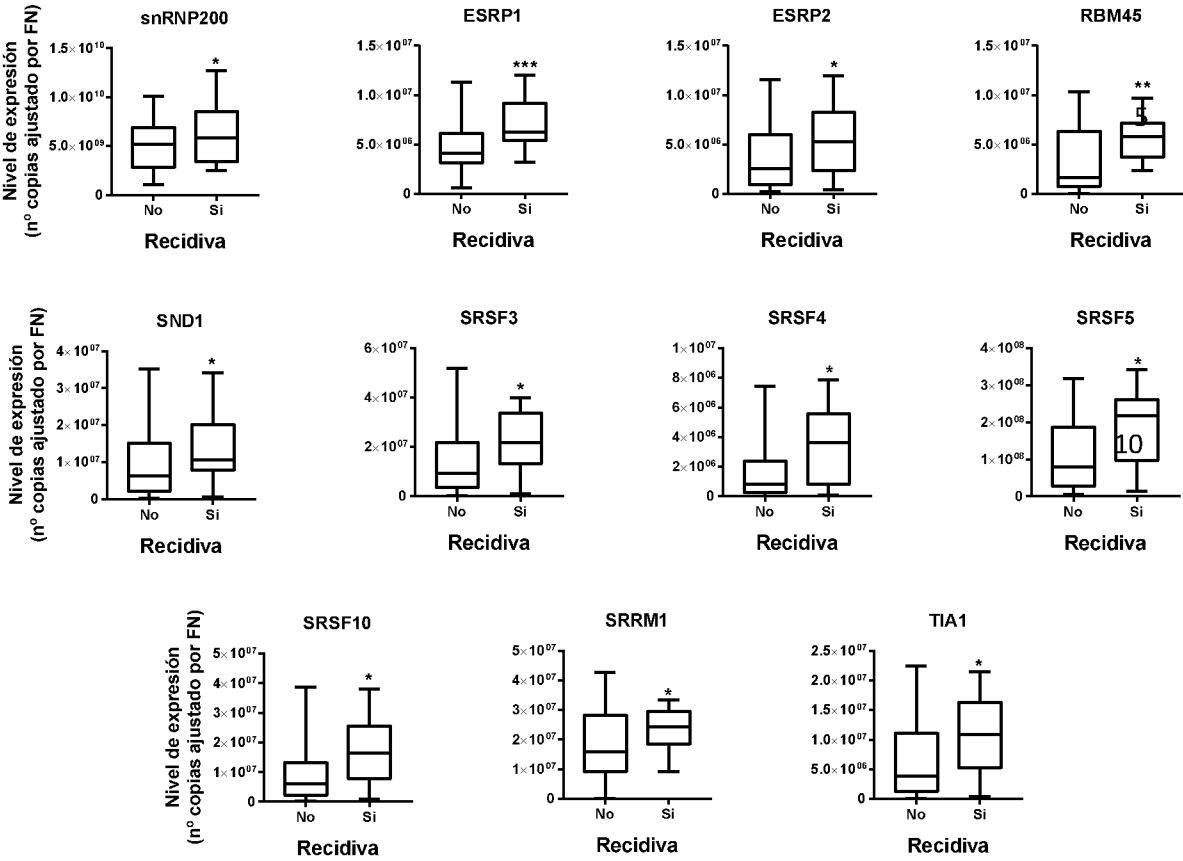


Figura 2

A)



B)

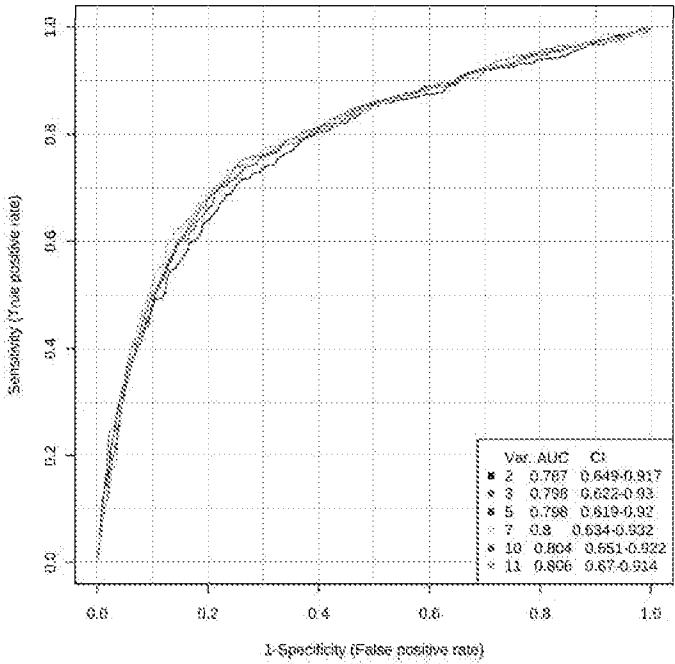


Figura 3

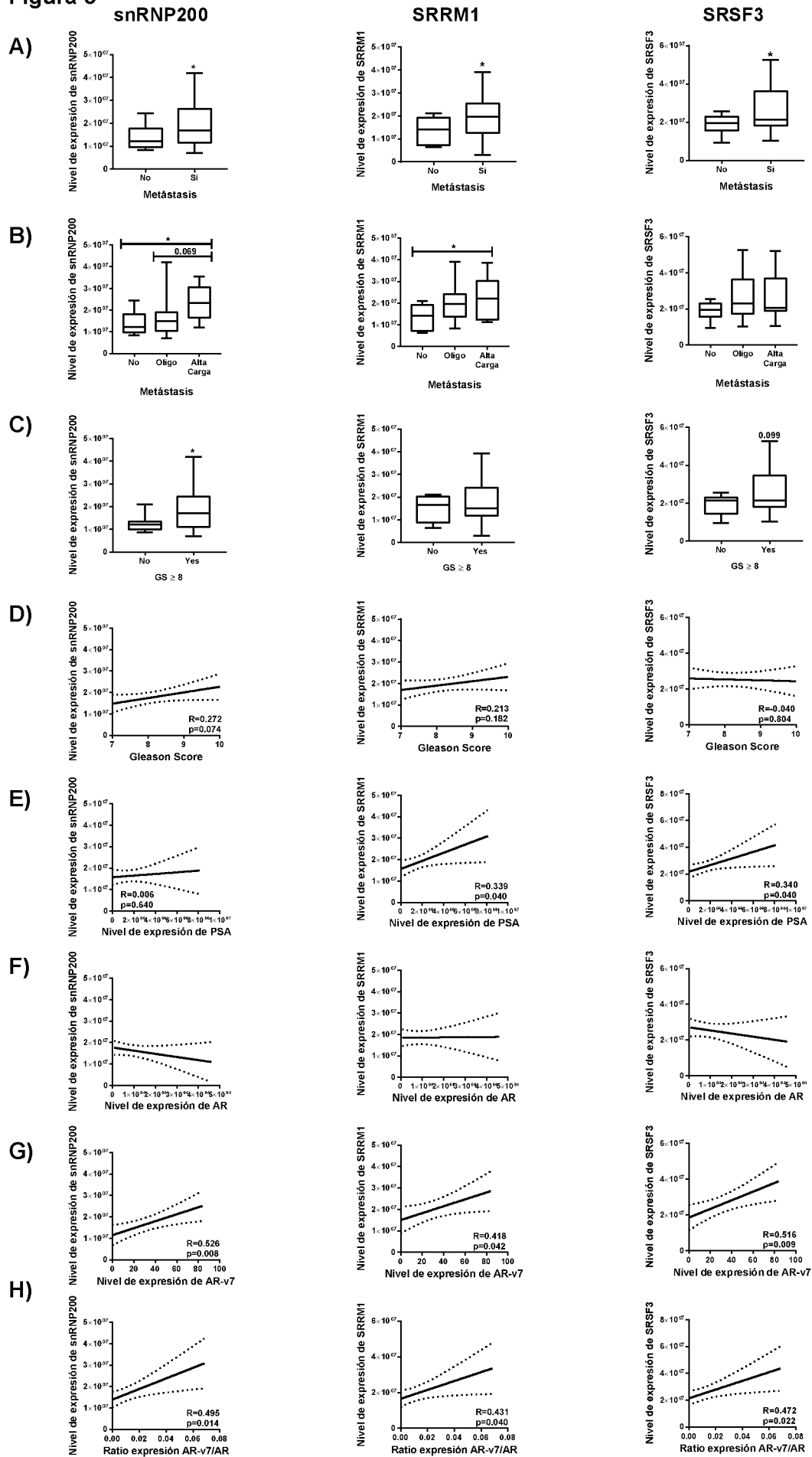


Figura 4

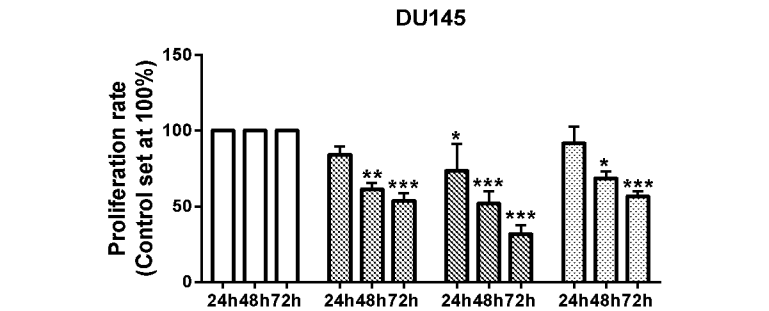
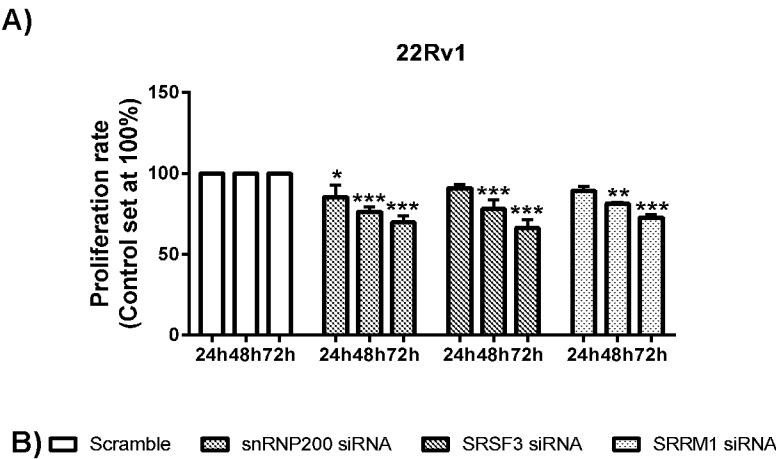


Figura 5

