(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
22 de mayo de 2020 (22.05.2020)

WIPO | PCT

(10) Número de publicación internacional
WO 2020/099708 A2

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

(21) Número de la solicitud internacional:
PCT/ES2019/070777

(22) Fecha de presentación internacional:
13 de noviembre de 2019 (13.11.2019)

(25) Idioma de presentación: español

(26) Idioma de publicación: español

(30) Datos relativos a la prioridad:
P201831095 13 de noviembre de 2018 (13.11.2018) 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: **GAHETE ORTIZ, Manuel D.**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **LÓPEZ MIRANDA, José**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **PÉREZ MARTÍNEZ, Pablo**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **DELGADO LISTA, Javier**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **GARCÍA RIOS, Antonio**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **ALCALÁ DÍAZ, Juan Francisco**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **YUBERO SERRANO, Elena**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Avda. Menéndez Pidal s/n, 14004 Córdoba (ES). **CAMARGO GARCÍA, Antonio**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **DEL RIO MORENO, Mercedes**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **ALORS PEREZ, Emilia**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **LUQUE HUERTAS, Raúl M.**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **CASTAÑO FUENTES, Justo P.**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **VENTURA SOTO, Sebastián**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez Pidal, s/n, 14004 Córdoba (ES). **REYES PUPO, Oscar**; Edificio IMIBIC, Hospital Universitario Reina Sofía, Av. Menendez 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))*

WO 2020/099708 A2

(54) **Title:** METHOD FOR PREDICTING OR PROGNOSING THE DEVELOPMENT OF TYPE 2 DIABETES MELLITUS IN AN INDIVIDUAL

(54) **Título:** MÉTODO PARA PREDECIR O PRONOSTICAR EL DESARROLLO DE DIABETES MELLITUS TIPO 2 EN UN INDIVIDUO

(57) **Abstract:** The invention relates to biomarkers, a method, a diagnostic kit and uses for predicting or prognosing the development of type 2 diabetes mellitus in an individual.

(57) **Resumen:** Biomarcadores, método, kit de diagnóstico y usos para predecir o pronosticar el desarrollo de diabetes mellitus tipo 2 en un individuo.

MÉTODO PARA PREDECIR O PRONOSTICAR EL DESARROLLO DE DIABETES MELLITUS TIPO 2 EN UN INDIVIDUO

CAMPO DE LA INVENCIÓN

La presente invención se encuentra dentro del campo de la medicina y la endocrinología y está dirigida a un método para predecir o pronosticar el desarrollo de DMT2 en un individuo, a un kit o dispositivo de pronóstico y a sus usos.

ANTECEDENTES DE LA INVENCIÓN

La diabetes mellitus tipo 2 (DMT2) es un importante problema de salud en los países desarrollados, ya que su incidencia y prevalencia han aumentado en las últimas décadas, junto con el aumento de la obesidad, una alimentación poco saludable y el sedentarismo. Si bien los principales factores de riesgo de DMT2 están bien definidos (antecedentes familiares, IMC> 30 y/o desregulación del metabolismo de la insulina), es una afección heterogénea, cuya base fisiopatológica sigue siendo desconocida en gran medida.

La DMT2 está estrechamente asociada con la enfermedad cardiovascular (ECV), ya que, por ejemplo, los pacientes con infarto de miocardio y DMT2 tienen un mayor riesgo de desarrollar un nuevo evento cardiovascular que aquellos sin DMT2 (Martin-Timon I, et al. *Type 2 diabetes and cardiovascular disease: have all risk factors the same strength.* World J Diabetes. 2014). Por lo tanto, la identificación temprana de individuos con alto riesgo de desarrollo de DMT2, especialmente entre pacientes con ECV es fundamental para la prevención (*Blanco-Rojo R, et al. HDL Cholesterol efflux normalised to apoa-i is associated with future development of type 2 diabetes: from the CORDIOPREV trial. Sci Rep. 2017; Camargo A, et al. Postprandial endotoxemia may influence the development of type 2 diabetes mellitus: from the CORDIOPREV study. Clin Nutr., 2018*).

Tradicionalmente, esta estrategia se ha basado en biomarcadores como los parámetros relacionados con la homeostasis de la glucosa, incluida la hemoglobina glicosilada (HbA1c), los niveles de las diferentes clases de lípidos y diversos scores predictivos, como el índice de resistencia a la insulina o FINDRISC; sin embargo, estos parámetros y pruebas tienen limitaciones y no pueden predecir con precisión el riesgo individual de desarrollar DMT2.

Específicamente se utilizan dos estrategias de cribado en personas de riesgo:

1. El cribado oportunista, dentro del contexto de cribado de otros factores de riesgo cardiovascular, consiste en:
   - En personas a partir de 45 años, cribado de DMT2 cada 4 años mediante glucemia basal en plasma venoso (GBP).
   - En personas de riesgo, definidas por hipertensión, hiperlipemia, obesidad, diabetes gestacional o patología obstétrica, glucemia basal alterada o intolerancia a la glucosa, cribado de DMT2 anual mediante GBP.
2. El cribado en dos etapas mediante índices de resistencia a la insulina FINDRISC: consiste en realizar el FINDRISC cada 3-5 años en personas a partir de los 40 años y en personas entre 25-39 años si tienen factores de riesgo de DMT2. La actuación a seguir según la puntuación en FINDRISC es:
   - ≤ 14p: repetir FINDRISC a los 5 años.
   - ≥ 15p: realizar glucemia en ayunas o HbA1c:
     - a) Si no hay DMT2 ni prediabetes: FINDRISC cada 3 años.
     - b) Si Prediabetes: ofrecer medidas intensivas de modificación de estilo de vida y control anual (con GBP o HbA1c).
     - c) Si Diabetes: tratamiento de DMT2 y seguimiento.

En estudios recientes se ha demostrado que el desarrollo de DMT2 se ve afectado críticamente por la pérdida de flexibilidad fenotípica (es decir, la capacidad para hacer frente a factores estresantes manteniendo la homeostasis metabólica), y que la identificación de factores que puedan modular la plasticidad fenotípica que define la susceptibilidad individual para desarrollar DMT2 pueden tener un potencial predictivo (Stroeve JHM, et al. Phenotypic flexibility as a measure of health: the optimal nutritional stress response test. Genes Nutr., 2015).

En concreto, existen estudios que sugieren que bajo condiciones metabólicas adversas (obesidad, resistencia a la insulina, etc.) la maquinaria de *splicing* está desregulada en la mayoría de los tejidos, incluidas las células mononucleares de sangre periférica (PBMC); y que las desregulaciones en los procesos de *splicing* están asociadas con el desarrollo de varias patologías (Dlamini Z, et al. Abnormalities in alternative splicing in diabetes: therapeutic targets. J Mol Endocrinol. 2017.). Específicamente, 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 DMT2.

Por todo esto, podría ser de utilidad la caracterización del perfil molecular de la maquinaria de *splicing* en pacientes con riesgo de desarrollar DMT2, de forma que pudiera servir como biomarcador para predecir el desarrollo futuro de DMT2 en dichos pacientes.

DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Patrón de expresión basal de los componentes del spliceosoma y factores de splicing seleccionados en las PBMCs de los pacientes que desarrollarán DMT2 (*Incident-T2DM*) y controles (non-T2DM). a) Cambio (Fold-change) en los niveles de expresión entre pacientes que desarrollarán DMT2 y controles, representados en rojo (aumento) o azul (disminución). Los componentes del spliceosoma o factores de splicing significativamente alterados están resaltados en negrita. b) Niveles de expresión de ARNm [ajustados por un factor de normalización (NF) calculado a partir de los niveles de expresión de GAPDH y ACTB] de componentes del spliceosoma específicos (primera fila) y factores de splicing (segunda fila) en las PBMCs de los pacientes no diabéticos y controles. Los valores representan la media ± EEM. Los asteriscos indican valores que difieren significativamente de los pacientes control (t-test: *, $p<0.05$).

Figura 2. Patrón de expresión basal de los componentes del spliceosoma y factores de splicing seleccionados en las PBMCs de los pacientes que desarrollarán DMT2 (Incident-T2DM) y controles (non-T2DM) durante la fase postprandial. a) Cambio (Fold-change) en los niveles de expresión entre pacientes que desarrollarán DMT2 y controles, representados en rojo (aumento) o azul (disminución). Los componentes del spliceosoma o factores de splicing significativamente alterados están resaltados en negrita. b) Niveles de expresión de ARNm [ajustados por un factor de normalización (NF) calculado a partir de los niveles de expresión de GAPDH y ACTB] de componentes del spliceosoma y factores de splicing específicos en las PBMCs de los pacientes no diabéticos y controles. Los valores representan la media ± EEM. Los asteriscos indican valores que difieren significativamente de los pacientes no diabéticos (t-test: *, $p<0.05$).

Figura 3. Modelos predictivos basados en la expresión de los componentes del spliceosoma y los factores de splicing, generados mediante el algoritmo computacional Random Forest y análisis de curvas ROC. Los parámetros de las curvas ROC fueron calculados para los modelos predictivos generados por Random Forest, considerando la expresión de todos los elementos de la maquinaria de splicing medidos durante el ayuno (naranja), postprandio (azul) o la combinación de estos (verde) en los pacientes no diabéticos y diabéticos (a) o utilizando una selección de los componentes de la maquinaria de splicing más relevantes y discriminatorios (b). Específicamente, estos componentes seleccionados fueron RNU4ATAC, TIA1, KHDRBS1, SRSF10, PTBP1, RAVER1, RNU2, RNU5, SRSF9, U2AF2, RBM45, SRSF4, RBM3 en ayuno, RNU4ATAC, RNU6ATAC, RAVER1, SF3B1, SRSF3, NOVA1, SRM160, SRSF6, ESRP1, U2AF1 en el postprandio, y SRSF3, SRSF10, SRSF6, TRA2B, PTBP1, SF3B1 en ayuno, así como RNU4ATAC, RAVER1, RNU4ATAC, NOVA1, RNU4, RNU6, PRPF8 en el postprandio, para los análisis combinados. Las mismas curvas ROC fueron calculadas considerando únicamente aquellos pacientes que desarrollaron diabetes durante los dos primeros años de seguimiento (c y d). En este caso, la selección de los componentes de la maquinaria fueron RNUU4ATAC, PTBP1, TRA2A, RM17, RNU12, TIA1, SRSF5, RNU2 en ayuno, RNU4ATA, PTBP1, MAGOH, SRSF9, RAVER1, PRP8, PRPF40A, SRRM1, SRSF6, SNRNP200, TIA1, RNU2 en el postprandio, y RNU4ATAC, TRA2A, SRSF5, RBM17, SRSF10, SRSF3, RNU2 en ayuno y RNU4ATAC, RNU4, RNU6ATAC, MAGOH, RAVER1, PRPF40A, RBM4, U2AF2, SRSF10, RNU11 TRA2B, SND1 en el postprandio para los análisis combinados. Se determinaron también las curvas ROC de la hemoglobina glicosilada (HbA1c) y el FINDRISK. *HbA1c:* hemoglobina glicosilada; *AUC*: Área debajo de la curva; *TPR*: ratio de verdaderos positivos; *TNR*: ratio de verdaderos negativos; *p*: p valor.

Figura 4. Niveles de expresión de componentes clave del spliceosoma en las PBMCs de sujetos sanos tratados con suero de pacientes que desarrollarán DMT2 (*Incident-T2DM*) y controles (non-T2DM) de ayunas durante 24 horas (a) o postprandial durante 4 horas (b). Niveles de expresión de ARNm, ajustados por los niveles de expresión de ACTB, de componentes del spliceosoma y factores de splicing específicos. Los valores representan la media ± EEM. Los asteriscos indican valores que difieren significativamente de las PBMCs tratadas con suero de pacientes no diabéticos (t-test: *, $p<0.05$; ***, $p<0.001$).

DESCRIPCIÓN DE LA INVENCIÓN

La presente invención supone una herramienta para el **diagnóstico temprano** de pacientes con riesgo a desarrollar DMT2. La creación de los modelos de evaluación de los biomarcadores aquí propuestos permite: 1) **discriminar entre individuos sanos y pacientes con** alto riesgo de desarrollar DMT2, con elevada sensibilidad y especificidad, 2) **estratificar clínicamente los pacientes** en base a parámetros clínicos y la expresión de componentes de la maquinaria de splicing en células sanguíneas; y 3) **facilitar el seguimiento clínico de estos pacientes,** 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 células sanguíneas de pacientes. 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 empalme de moléculas de ARNm, o *splicing*, está emergiendo como un sello distintivo en diversas patologías. En la presente invención se realiza el análisis del patrón de expresión de ciertos componentes de maquinaria de *splicing*, incluidos ciertos componentes del spliceosoma mayor (n = 13) y menor (n = 4) y factores de splicing específicos (n = 28) en células sanguíneas de pacientes sin DMT2 pero con alto riesgo de desarrollar DMT2 (individuos que sufrieron un evento cardiovascular).

El objetivo de la presente invención ha sido la identificación de ciertos componentes de la maquinaria de *splicing* desregulados en células sanguíneas de pacientes con riesgo a desarrollar DMT2, así como su influencia en la actividad de la enfermedad.

Los autores de la presente invención han identificado que la alteración en la expresión de ciertos componentes del spliceosoma y factores de splicing están relacionados con el desarrollo de DMT2, precediendo a la instauración clínica de esta patología y que, por lo tanto, podrían servir como sensores y predictores tempranos de desarrollo de DMT2. Han identificado una huella molecular formada por ciertos de estos componentes que es capaz de predecir el desarrollo futuro de DMT2 en pacientes con alta precisión, que incluso supera la capacidad de los predictores clásicos de desarrollo de DMT2, tales como hemoglobina glicosilada, los índices de resistencia a la insulina o el FINDRISC.

Por tanto, **un primer aspecto** de la invención se refiere al uso del producto de la expresión de los genes *RNU2*, *RNU4*, *RNU4ATAC, RNU6, RNU6ATAC*, *RNU12*, *SNRNP200*, *ESRP1*, *SRSF1, SRSF5*, *SNRNP70* y/o *NOVA1*, o cualquiera de sus combinaciones, para predecir o pronosticar el desarrollo de DMT2 en un individuo.

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*RNU2*" (también llamado en literatura *U2*) se refiere al gen componente del spliceosoma mayor. El spliceosoma es un complejo de ribonucleoproteínas que cataliza la eliminación de intrones de precursores de ARNm nuclear. El spliceosoma está compuesto por cuatro pequeñas partículas de ribonucleoproteínas (las *snRNPs U1, U2 y U4 / U6*) y numerosas proteínas adicionales. El ARN nuclear pequeño U2 (snRNA) es un componente de ARN del snRNP *U2* que interactúa con la región 3 'del intrón en el sitio de la rama. Varios sitios en el genoma humano expresan snRNAs de U2. Este locus representa una matriz de genes snRNA *U2* ubicados en 17q21-22 que experimenta una evolución concertada para homogeneizar unidades de repetición dentro de la matriz. Cada repetición en esta matriz es de aproximadamente 6.1 kb de largo y contiene una sola copia del snRNA *U2*. En el contexto de la presente invención, "*RNU2*" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia del snRNP "*RNU2*", 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: 6066

Localización del gen: 17q21.31

SEQ ID NO: 1

ATCGCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAATATCT
GATACGTCCTCTATCCGAGGACAATATATTAAATGGATTTTTGGAGCAGGGAGATGGAATA
GGAGCTTGCTCCGTCCACTCCACGCATCGACCTGGTATTGCAGTACCTCCAGGAACGGTG
CACCC

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*", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que comprende la SEQ ID NO:2,

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: 26835

Localización del gen: 12q24.23

SEQ ID Nº 2:

AGCTTTGCGCAGTGGCAGTATCGTAGCCAATGAGGTCTATCCGAGGCGCGATTATTGCTAA TTGAAAACTTTTCCCAATACCCCGCCGTGACGACTTGCAATATAGTCGGCACTGGCAATTTT TGACAGTCTCTACGGAGACTG

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*RNU4ATAC*", (también llamado en literatura MOPD1, RFMN1, TALS, U4ATAC, RNU4ATAC) 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, "*RNU4ATAC*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia del snRNP "*RNU4*", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que comprende 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,

Gene ID: 100151683

Localización del gen: 2q14.2

SEQ ID Nº 3:

AACCATCCTTTTCTTGGGGTTGCGCTACTGTCCAATGAGCGCATAGTGAGGGCAGTACTGC
TAACGCCTGAACAACACACCCGCATCAACTAGAGCTTTTGCTTTATTTTGGTGCAATTTTTG
GAAAAAT

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*RNU6*", (también llamado en literatura RNU6, RNU6A, U6, U6-1) se refiere al gen, factor de la maquinaria de *splicing*. En el contexto de la presente invención, "*RNU6*" se define también por 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:4,

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 Nº 4:

GTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCC
CCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCCATATTTT

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*RNU6ATAC*", (también llamado en literatura RNU6ATAC1, U6ATAC) 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, "*RNU6ATAC*" se define también por una secuencia de nucleótidos o polinucleótidos, que constituye la secuencia del snRNP "*RNU6ATAC*", y que comprendería diversas variantes procedentes de:

a) moléculas de ácido nucleico que comprende la SEQ ID NO: 5,

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: 100151684

Localización del gen: 9q34.2

SEQ ID Nº 5:

GTGTTGTATGAAAGGAGAGAAGGTTAGCACTCCCCTTGACAAGGATGGAAGAGGCCCTCG
GGCCTGACAACACGCATACGGTTAAGGCATTGCCACCTACTTCGTGGCATCTAACCATCGT
TTTT

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 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),

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 Nº 6:

TGCCTTAAACTTATGAGTAAGGAAAATAACGATTCGGGGTGACGCCCGAATCCTCACTGCT
AATGTGAGACGAATTTTTGAGCGGGTAAAGGTCGCCCTCAAGGTGACCCGCCTACTTTGC
GGGATGCCTGGGAGTTGCGATCTGCCCG

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 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 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, "*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: 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,

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: 7. 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: 8.

Gene ID: 23020

Localización del gen: 2q11.2

SEQ ID Nº 7:

GAACCCGTTTCTTTTCCTTCCCCAGTGCGTCTTTCCTGCGTCGTTCCGGCGCGGCGGGAG
CAGAGATCTGCGGCCGTTTGCAGCTTGCGGTAGGGAGGCGTGGTGGTCTGAAGCCTCCG
AGCAGCCGCGGCCATGGCGGATGTAACCGCCCGTAGTCTGCAATACGAGTACAAGGCGA
ACTCGAATCTTGTGCTCCAAGCTGACCGTTCTCTCATTGACCGGACCCGCCGGGATGAAC
CCACAGGAGAGGTGCTGTCCCTTGTTGGGAAGCTGGAGGGCACCCGTATGGGAGACAAG
GCTCAACGGACCAAACCGCAGATGCAGGAGGAAAGAAGAGCCAAGCGAAGAAAGCGTGA
TGAGGACCGGCATGACATCAACAAGATGAAGGGTTATACTCTGCTGTCGGAGGGCATTGA
TGAGATGGTGGGCATCATCTACAAGCCCAAAACTAAAGAGACTCGGGAGACCTATGAGGT
GCTACTCAGCTTCATCCAGGCTGCTCTTGGGGACCAGCCACGTGATATCCTTTGTGGGGC
AGCTGATGAAGTTCTAGCTGTTCTAAAGAATGAAAAGCTGCGGGACAAGGAAAGGCGAAA
GGAGATTGACCTGCTGCTGGGTCAAACAGATGATACCAGATACCATGTGCTAGTGAACCTG
GGCAAAAAGATCACAGACTATGGTGGAGATAAGGAAATCCAAAATATGGATGACAACATTG
ATGAGACATACGGTGTGAATGTGCAGTTTGAGTCTGATGAGGAGGAAGGTGATGAAGACG
TATACGGGGAGGTTCGAGAAGAGGCATCTGATGATGACATGGAAGGGGACGAGGCTGTC
GTGCGCTGCACCCTCTCGGCTAATCTCGTAGCCTCAGGTGAACTGATGAGTTCCAAGAAG

AAGGATTTGCACCCTCGGGATATTGATGCATTTTGGCTGCAGCGGCAGCTCAGTCGTTTCT
ATGATGATGCCATCGTGTCGCAGAAGAAGGCAGATGAAGTATTGGAGATTTTGAAGACGG
CCAGTGATGATCGGGAATGTGAAAATCAGCTGGTTCTGCTGCTTGGTTTCAACACCTTTGA
TTTCATTAAAGTGTTGCGGCAGCACAGGATGATGATTTTATACTGTACCTTGCTGGCCAGT
GCACAAAGTGAAGCTGAAAAGGAAAGGATTATGGGAAAGATGGAAGCTGACCCAGAGCTA
TCCAAGTTCCTCTACCAGCTTCATGAAACCGAGAAGGAGGATCTGATCCGAGAGGAAAGG
TCCCGGAGAGAGCGAGTGCGTCAGTCTCGAATGGACACAGATCTGGAAACCATGGATCTC
GACCAGGGTGGAGAGGCACTGGCTCCACGGCAGGTTCTGGACTTGGAGGACCTGGTTTTT
ACCCAAGGGAGCCACTTTATGGCCAATAAACGCTGTCAGCTTCCTGATGGATCCTTCCGTC
GCCAGCGTAAGGGCTATGAAGAGGTGCATGTGCCTGCTCTGAAGCCCAAGCCCTTTGGCT
CAGAAGAACAACTGCTTCCAGTGGAAAAGCTGCCAAAGTATGCCCAGGCTGGGTTTGAGG
GCTTCAAAACACTGAATCGGATCCAGAGTAAGCTCTACCGTGCTGCCCTTGAGACGGATGA
GAATCTGCTGCTGTGTGCTCCTACTGGTGCTGGGAAGACCAACGTGGCCCTGATGTGCAT
GCTCCGAGAGATTGGGAAACACATAAACATGGACGGCACCATCAATGTGGATGACTTCAA
GATTATCTACATTGCCCCCATGCGCTCCTTGGTGCAGGAGATGGTGGGCAGCTTTGGAAA
GCGCCTGGCCACTTATGGCATCACTGTTGCTGAACTGACTGGGGACCACCAGCTGTGCAA
AGAAGAGATCAGTGCCACTCAGATCATCGTCTGCACCCCCGAGAAGTGGGACATCATCAC
CCGCAAGGGTGGTGAGCGCACCTACACCCAGCTGGTGCGGCTCATCATTCTGGATGAGAT
TCATCTTCTCCACGATGACAGAGGTCCTGTCTTAGAAGCTTTAGTGGCCAGGGCCATCCGA
AACATTGAGATGACCCAAGAGGATGTCCGACTCATTGGTCTCAGTGCCACCCTACCCAACT
ATGAAGATGTAGCCACCTTTCTACGTGTTGACCCTGCCAAGGGTCTCTTTTACTTTGACAAC
AGCTTCCGTCCAGTGCCTCTGGAACAGACATATGTGGGTATCACAGAGAAAAAAGCTATCA
AGCGTTTCCAGATCATGAATGAAATCGTCTATGAAAAAATCATGGAACATGCTGGAAAAAAT
CAGGTGCTGGTGTTTGTCCACTCCCGGAAGGAGACTGGAAAGACAGCCAGGGCCATCCG
GGACATGTGCCTAGAAAAGGACACTCTGGGTCTGTTTCTGAGGGAGGGCTCAGCCTCCAC
AGAAGTCCTGCGAACAGAAGCTGAGCAGTGCAAGAACCTAGAGCTGAAGGATCTTCTGCC
TTATGGCTTTGCTATTCATCACGCAGGCATGACCAGGGTTGACCGAACACTCGTGGAGGAT
CTTTTTGCTGATAAACATATTCAGGTTTTAGTTTCCACAGCAACTCTAGCTTGGGGTGTGAA
TCTCCCTGCACATACAGTCATCATCAAAGGCACCCAGGTGTACAGTCCAGAGAAGGGGCG
TTGGACAGAACTGGGAGCACTGGACATTCTGCAGATGCTGGGACGTGCCGGAAGACCCCA
GTATGACACCAAGGGTGAAGGCATACTCATCACATCTCATGGGGAGCTACAGTACTACCTG
TCCCTCCTCAATCAACAACTTCCTATTGAAAGCCAGATGGTTTCAAAGCTTCCTGACATGCT
CAATGCAGAAATCGTGCTAGGAAATGTCCAGAATGCCAAGGATGCGGTGAACTGGCTGGG
CTATGCCTACCTCTATATCCGAATGCTGCGATCCCCAACCCTCTATGGCATCTCTCATGAT
GACCTCAAGGGAGATCCCCTGCTGGACCAGCGCCGACTAGATCTGGTTCATACAGCTGCC
CTGATGCTGGACAAGAACAATCTGGTCAAGTACGACAAGAAGACGGGCAACTTCCAGGTG
ACAGAACTGGGCCGTATAGCCAGCCACTACTACATCACCAATGATACAGTGCAGACTTACA
ACCAGCTGCTGAAGCCCACCCTGAGTGAGATTGAGCTTTTCAGGGTCTTCTCATTGTCCTC
TGAGTTCAAGAACATCACAGTGAGAGAGGAGGAGAAGCTGGAGCTGCAGAAGTTGCTGGA
GAGGGTGCCTATCCCTGTAAAGGAGAGCATTGAGGAACCCAGTGCTAAGATCAACGTTCTT
CTGCAAGCCTTCATCTCACAGCTGAAATTGGAGGGCTTTGCACTGATGGCTGACATGGTGT
ATGTCACACAGTCGGCTGGCCGGTTGATGCGAGCGATATTTGAAATTGTCCTGAACCGAG
GTTGGGCACAGCTTACAGACAAGACCCTGAACCTCTGCAAGATGATCGACAAACGCATGT
GGCAGTCCATGTGTCCTCTGCGCCAGTTCCGGAAACTCCCTGAGGAAGTAGTGAAGAAGA
TTGAGAAGAAGAATTTCCCCTTTGAGCGTCTGTACGACCTGAATCATAATGAGATTGGGGA
GCTTATCCGCATGCCAAAGATGGGGAAGACCATCCACAAATATGTCCATCTGTTTCCCAAG
TTGGAGTTGTCAGTGCACCTGCAGCCTATCACACGCTCCACCCTGAAGGTGGAGCTGACC
ATCACGCCAGACTTCCAGTGGGATGAAAAGGTGCATGGTTCATCCGAGGCTTTTTGGATTC
TGGTGGAGGATGTGGACAGCGAGGTGATTCTGCACCATGAGTATTTTCTCCTCAAGGCCA

AGTACGCCCAGGACGAGCACCTCATTACATTCTTCGTGCCTGTCTTTGAACCGCTGCCCCC
TCAGTACTTCATCCGAGTGGTGTCTGACCGCTGGCTCTCTTGTGAGACCCAGCTGCCTGTC
TCCTTCCGGCACCTGATCTTGCCGGAGAAGTACCCCCCTCCAACCGAACTTTTGGACCTG
CAGCCCTTGCCCGTGTCTGCTCTGAGAAACAGTGCCTTTGAGAGTCTTTACCAAGATAAAT
TTCCTTTCTTCAATCCCATCCAGACCCAGGTGTTTAACACTGTATACAACAGTGACGACAAC
GTGTTTGTGGGGGCCCCCACGGGCAGCGGGAAGACTATTTGTGCAGAGTTTGCCATCCTG
CGAATGCTGCTGCAGAGCTCGGAGGGGCGCTGTGTGTACATCACCCCCATGGAGGCCCT
GGCAGAGCAGGTATACATGGACTGGTACGAGAAGTTCCAGGACAGGCTCAACAAGAAGGT
GGTACTCCTGACAGGCGAGACCAGCACAGACCTGAAGCTGCTGGGCAAAGGGAACATTAT
CATCAGCACCCCTGAGAAGTGGGACATACTTTCCCGGCGATGGAAGCAGCGCAAGAACGT
GCAGAACATCAACCTCTTCGTGGTGGATGAGGTCCACCTTATCGGGGGCGAGAATGGGCC
TGTCTTAGAAGTGATCTGCTCCCGAATGCGCTACATCTCCTCCCAGATTGAGCGGCCCATT
CGCATTGTGGCACTCAGCTCTTCGCTCTCCAATGCCAAGGATGTGGCCCACTGGCTGGGC
TGCAGTGCCACCTCCACCTTCAACTTCCATCCCAATGTGCGTCCCGTCCCCTTGGAGCTGC
ACATCCAGGGCTTCAACATCAGCCATACACAAACCCGCCTGCTCTCCATGGCCAAGCCTGT
GTACCATGCTATCACCAAGCACTCGCCCAAGAAGCCTGTCATTGTCTTTGTGCCGTCTCGC
AAGCAGACCCGCCTCACTGCCATTGACATCCTCACCACCTGTGCAGCAGACATCCAACGG
CAGAGGTTCTTGCACTGCACCGAGAAGGATCTGATTCCGTACCTGGAGAAGCTAAGTGAC
AGCACGCTCAAGGAAACGCTGCTAAATGGGGTGGGCTACCTGCATGAGGGGCTCAGCCC
CATGGAGCGACGCCTGGTGGAGCAGCTCTTCAGCTCAGGGGCTATCCAGGTGGTGGTGG
CTTCTCGGAGTCTCTGCTGGGGCATGAACGTGGCTGCCCACCTGGTAATCATCATGGATA
CCCAGTACTACAATGGCAAGATCCACGCCTATGTGGATTACCCCATCTATGACGTGCTTCA
GATGGTGGGCCACGCCAACCGCCCTTTGCAGGACGATGAGGGGCGCTGTGTCATCATGT
GTCAGGGCTCCAAGAAGGATTTCTTCAAGAAGTTCTTATATGAGCCATTGCCAGTAGAATC
TCACCTGGACCACTGTATGCATGACCACTTCAATGCTGAGATCGTCACCAAGACCATTGAG
AACAAGCAGGATGCTGTGGACTACCTCACCTGGACCTTTCTGTACCGCCGCATGACACAG
AACCCCAATTACTACAACCTGCAGGGCATCTCCCATCGTCACTTGTCGGACCACTTGTCAG
AGCTGGTGGAGCAGACCCTGAGTGACCTGGAGCAGTCCAAGTGCATCAGCATCGAGGAC
GAGATGGACGTGGCGCCTCTGAACCTAGGCATGATCGCCGCCTACTATTACATCAACTACA
CCACCATTGAGCTCTTCAGCATGTCCCTCAATGCCAAGACCAAGGTGCGAGGGCTTATCG
AGATCATCTCCAATGCAGCAGAGTATGAGAACATTCCCATCCGGCACCATGAAGACAATCT
CCTGAGGCAGTTGGCTCAGAAGGTCCCCCACAAGCTGAATAACCCTAAGTTCAATGATCC
GCACGTCAAGACCAACCTGCTCCTGCAGGCTCACTTGTCTCGCATGCAGCTGAGTGCTGA
GTTGCAGTCAGATACGGAGGAAATCCTTAGTAAGGCAATCCGGCTCATCCAGGCCTGCGT
GGATGTCCTTTCCAGCAATGGGTGGCTCAGCCCTGCTCTGGCAGCTATGGAACTGGCCCA
GATGGTCACCCAAGCCATGTGGTCCAAGGACTCATACCTGAAGCAGCTGCCACACTTCAC
CTCTGAGCATATCAAACGTTGCACAGACAAGGGAGTGGAGAGTGTTTTCGACATCATGGAG
ATGGAGGATGAAGAACGGAACGCGTTGCTTCAGCTGACTGACAGCCAGATTGCAGATGTG
GCTCGCTTTTGTAACCGCTACCCTAATATCGAACTATCTTATGAGGTGGTAGATAAGGACA
GCATCCGCAGTGGCGGGCCAGTTGTGGTGCTGGTGCAGCTGGAGCGAGAGGAGGAAGTC
ACAGGCCCTGTCATTGCGCCTCTCTTCCCGCAGAAACGTGAAGAGGGCTGGTGGGTGGTG
ATTGGAGATGCCAAGTCCAATAGCCTCATCTCCATCAAGAGGCTGACCTTGCAGCAGAAG
GCCAAGGTGAAGTTGGACTTTGTGGCCCCAGCCACTGGTGCCCACAACTACACTCTGTAC
TTCATGAGTGACGCTTACATGGGATGTGACCAGGAGTACAAATTCAGCGTGGATGTGAAAG
AAGCTGAGACAGACAGTGATTCAGATTGAGTCCTGAGGCATTTACTTTTGGGTAAAGGAGA
GTTGAGCCTGAATTAGGAATGTGTACATTGTAGGAATCCTGGTTGTGGGGACCAGGTCTGT
GGGCCTCAGGTCTGGCCAGCCAGGGCTGGTGCTGTCCCCGCCTACCTCCACTTCCTTTCC
CTTGCTCACTCTGGATCCAGTGACAGCAGGTGTCATGGGTCAAGCATAAATCATATATAGC
ATTTTCAGGCATGTTCCTGGTAGTTCTTTTGAGTCTGACATTCTAATAAAATAATTTGTAGAA

ACCATTTGTCTTTGTAGTGATTCCAAATTAAAAGTTTTCTTTCTCCAACCTGAGGGCACGGC
CAAAAAGATCTGGTTATTTTTTAGCCAGGAACGTGCTTGTTAATGAGTATGTCTGGAGGACA
GACCTGCTCATTAGGTGTGCTGTCCCCTGTAGCCTCGTGAGTCAGCCCAGAGGAGGGTAC
ATGCGACTGTGGCCTGGCCTCAGTGGTACCCACACATCAGCACTACCACAAGAACCAACA
CTGAGCCTCGGAAGCTAGATCACAGGTTAGGGGTTTCTCTAGATGGGGGTTCTGAAATTTG
CAGTGTCTGCTCCTGGGAGGCAGCACCAGAAAGGGCACTGAAATGTACTAGCTGGATGTG
ACCCAGTCTTAATAAACAGGTTTTCTAATCCAGAAAAAAAAAAAA

SEQ ID Nº 8:

MADVTARSLQYEYKANSNLVLQADRSLIDRTRRDEPTGEVLSLVGKLEGTRMGDKAQRTKPQ
MQEERRAKRRKRDEDRHDINKMKGYTLLSEGIDEMVGIIYKPKTKETRETYEVLLSFIQAALGD
QPRDILCGAADEVLAVLKNEKLRDKERRKEIDLLLGQTDDTRYHVLVNLGKKITDYGGDKEIQN
MDDNIDETYGVNVQFESDEEEGDEDVYGEVREEASDDDMEGDEAVVRCTLSANLVASGELMS
SKKKDLHPRDIDAFWLQRQLSRFYDDAIVSQKKADEVLEILKTASDDRECENQLVLLLGFNTFD
FIKVLRQHRMMILYCTLLASAQSEAEKERIMGKMEADPELSKFLYQLHETEKEDLIREERSRRER
VRQSRMDTDLETMDLDQGGEALAPRQVLDLEDLVFTQGSHFMANKRCQLPDGSFRRQRKGY
EEVHVPALKPKPFGSEEQLLPVEKLPKYAQAGFEGFKTLNRIQSKLYRAALETDENLLLCAPTG
AGKTNVALMCMLREIGKHINMDGTINVDDFKIIYIAPMRSLVQEMVGSFGKRLATYGITVAELTG
DHQLCKEEISATQIIVCTPEKWDIITRKGGERTYTQLVRLIILDEIHLLHDDRGPVLEALVARAIRNI
EMTQEDVRLIGLSATLPNYEDVATFLRVDPAKGLFYFDNSFRPVPLEQTYVGITEKKAIKRFQIM
NEIVYEKIMEHAGKNQVLVFVHSRKETGKTARAIRDMCLEKDTLGLFLREGSASTEVLRTEAEQ
CKNLELKDLLPYGFAIHHAGMTRVDRTLVEDLFADKHIQVLVSTATLAWGVNLPAHTVIIKGTQV
YSPEKGRWTELGALDILQMLGRAGRPQYDTKGEGILITSHGELQYYLSLLNQQLPIESQMVSKL
PDMLNAEIVLGNVQNAKDAVNWLGYAYLYIRMLRSPTLYGISHDDLKGDPLLDQRRLDLVHTAA
LMLDKNNLVKYDKKTGNFQVTELGRIASHYYITNDTVQTYNQLLKPTLSEIELFRVFSLSSEFKNI
TVREEEKLELQKLLERVPIPVKESIEEPSAKINVLLQAFISQLKLEGFALMADMVYVTQSAGRLM
RAIFEIVLNRGWAQLTDKTLNLCKMIDKRMWQSMCPLRQFRKLPEEVVKKIEKKNFPFERLYDL
NHNEIGELIRMPKMGKTIHKYVHLFPKLELSVHLQPITRSTLKVELTITPDFQWDEKVHGSSEAF
WILVEDVDSEVILHHEYFLLKAKYAQDEHLITFFVPVFEPLPPQYFIRVVSDRWLSCETQLPVSF
RHLILPEKYPPPTELLDLQPLPVSALRNSAFESLYQDKFPFFNPIQTQVFNTVYNSDDNVFVGAP
TGSGKTICAEFAILRMLLQSSEGRCVYITPMEALAEQVYMDWYEKFQDRLNKKVVLLTGETSTD
LKLLGKGNIIISTPEKWDILSRRWKQRKNVQNINLFVVDEVHLIGGENGPVLEVICSRMRYISSQI
ERPIRIVALSSSLSNAKDVAHWLGCSATSTFNFHPNVRPVPLELHIQGFNISHTQTRLLSMAKPV
YHAITKHSPKKPVIVFVPSRKQTRLTAIDILTTCAADIQRQRFLHCTEKDLIPYLEKLSDSTLKETLL
NGVGYLHEGLSPMERRLVEQLFSSGAIQVVVASRSLCWGMNVAAHLVIIMDTQYYNGKIHAYV
DYPIYDVLQMVGHANRPLQDDEGRCVIMCQGSKKDFFKKFLYEPLPVESHLDHCMHDHFNAEI
VTKTIENKQDAVDYLTWTFLYRRMTQNPNYYNLQGISHRHLSDHLSELVEQTLSDLEQSKCISIE
DEMDVAPLNLGMIAAYYYINYTTIELFSMSLNAKTKVRGLIEIISNAAEYENIPIRHHEDNLLRQLA
QKVPHKLNNPKFNDPHVKTNLLLQAHLSRMQLSAELQSDTEEILSKAIRLIQACVDVLSSNGWL
SPALAAMELAQMVTQAMWSKDSYLKQLPHFTSEHIKRCTDKGVESVFDIMEMEDEERNALLQL
TDSQIADVARFCNRYPNIELSYEVVDKDSIRSGGPVVVLVQLEREEEVTGPVIAPLFPQKREEG
WWVVIGDAKSNSLISIKRLTLQQKAKVKLDFVAPATGAHNYTLYFMSDAYMGCDQEYKFSVDV
KEAETDSDSD

El término "*ESRP1*" (también llamado en la literatura DFNB109, RBM35A, RMB35A) se refiere tanto al gen como a la proteína humana. ESPR1 es un regulador de *splicing* específico del tipo de célula 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: 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,

Gene ID: 54845

Localización del gen: 8q22.1

SEQ ID Nº 9:

ACAGGTGGCAGCTCTCGTCCCCTGAGAGCGGGCGAAGGCCAGGGTCCCACACTCGCAGC
GCTGCGACGGCGCTCGGGACCTCCCTCGTCCACTGCTTGAGTTCCAGAGGTGGGTGCTT
CCCTGTCCTGAACTTCAGAGTGCGGAGTCATAAATGGGTTCCGGCTGGCTTACTGCAGTA
GCCTCGCTCCTCCCCAGCCCCGGTAACTCCGAGCTACCCGTCCAGGCCCTCGGGCGTCG
CGGGGGCAGGGACTGGGCGCGGAACGAGGCAGGGAGGGACCTGGAAAAACCACCCAGA
TTGCATTGCAGTGGGCGAGGCCGCCTGGAGGAGCCGGTTCCCCCTAACCACCTCCCCGT
GGGGCTCTCGGTGCGCGGTTCCCAGGTGCTCAGCTCTGCTGGGCCCAGGAGGTGCCGCC
TCACAGGGACGCGGAACCCCGTGCGTGGCCCCCGCCGGGTGGAACAGATAGCGCGGGG
CGGTCCGGAGGCTCGTCGCCAAGCAGGTGACTCTTGCTGAAAAAGTGGTTGGAACACTTG
AGGAAACCCGGCCCCGCCTGTTCTTTCTAGGTCTTTGGAGTTTGGATTAATCATTTGTGTA
GCCCGTTTGGATAAACCGAAGACTTTATTAAATCAGCGCGTTTAACAGGAATTCCGCAGTA
GTATCCACATTAGAATCTTGAGTCTTGGAGTTGAACATATTCACACAGACTTGCCTTCTTCC
TGTTTAGTTTATGCCTTGTGTTCCGTTATTGGAACGCTAAGCTTGTGGGAGTTGTTTACATC
CTACTGCTCAAGGTCATCGCTAAGGTGTGATTTTTCACAAAAAGAATTTGCAACCTCCGGC
ATGAATGACTTAAGGGAAGTCTAATCCCGGTTTCTGATTTTTTTTTTTTTTTTAATTTAAAAGT
TAATCTTTCTGGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCG
AGGCGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCT
CTACTAAAAACACAAAAATTAGCCGGGCGGGGTGGCGCGCACCTGTAATCCCAGCTGCTC
GGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAGGCGGGGGGTTGCAGTGAGCCG
AGATCTGGCCATTGCACTCCAGCGTGGGCAACAGAGTGAGACTCCATCTCAAAAAAAAAG
GTTAATCTTTCCAACTAGATTTTCAAGGATGAGGATTTTGTTGTTGTTGTTGTTGTTCTC
AAATGTATTCCCAGGGCTTGGAACAGAGCCTGACATATACTAGGCACTCAACAAATATTTGT
TGAATGATTGTAATGAGTAACACCCATTTTTGCAGATCTTTGTCTTCTGAGCCTAGGGCATA
GGTCATCACTGCAGGGGTGAGATTGTCAAAATGGGAGTCTACAGGTAATTTAAGACTTAAA
TGTTTAAAGAGTATGTGCTCATTCTTCAACAAACTTACTTTTGTTAAATTAAAATGGTAAAAT
GTGGTGGAGGGGTTGGAATATATGTAATTCAAGACAGTTCTGAATACAAAAATGTTTTACTG
TCTATCACCACCATCTATAAATCTAATTCACTAAGGATAATCTGTGTAAGGTGGCTGGAAAG
AACCTTGAGGAGAGAGGCTTATTTAAGTATTGGCTCAGGACCACACCTAAAATTCTCAAAA

CGTTGAGATTCTGTTGTTTTGTTTTTAAGCGCCAGAGACCCAAGTTGAGGAACAGCCTATAA
AATAACTGGCCTGTACTCTTACATACATGAAAGCCATCAAAGACAAAGACTGAAGAAGAACT
TTTGCAGATTAAAGGACTTTAAGAGACATGATCCTGAACCAGAAAAAGTAAAAAATAAAAAA
ATAAAAAGATCTGTAGATTAGATAATAGTTTTATATCCATGTTAATTTCCCATTTGTTGTAATA
TTACTGTGGTTGGGCAAGAGAGTGTCTTTGTTGGTGGAAGTATCTACTGAAGTATTTAGGG
GTAAAGAGGAATCATGTCTGTAACTTATGCTCAAAAGGTTCAGACAAAATATATGTAAAGGG
AATGATAAAGCAAACGCAATAAAATGTTAACATTGGGGAATCTGTATAAGGGAATTATTAAC
TTTTCTGTAAGTCTAAAATTATTTCAAATTAAAAATAATTTTTTTAAGCCTC

El término "*SRSF1*" (también llamado en la literatura ASF; SF2; SFRS1; SF2p33; SRp30a) se refiere tanto al gen que codifica a un miembro de la familia de proteínas del factor de *splicing* rico en arginina / serina. La proteína codificada puede activar o reprimir el *splicing*, dependiendo de su estado de fosforilación y sus compañeros de interacción. Se han encontrado múltiples variantes de transcripción para este gen. Hay un pseudogen de este gen en el cromosoma 13. A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*SRSF1*", se refiere tanto al gen como a la proteína humana, factor de la maquinaria de splicing. En el contexto de la presente invención, "*SRSF1*" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "*SRSF1*", 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: 10,

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: 10. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SRSF1*. Entre otras posibles secuencias nucleotídicas que codifican *SRSF1* se encuentra, pero sin limitarse, la SEQ ID NO: 11.

Gene ID: 6426

Localización del gen: 17q22

SEQ ID Nº 10:

GTTGCGCTCCCGCGCGTGCGTGTTGGGATCGAATCGCTGTTTCCTTCCGCTTCTCTTCCTC
TGTCTCCCCCCCATATCCGTGCGCCGAGCTGATAAAGGCGCCATTTTGGAGGGGCCGCG
GGAGACGTGGTGCCGCTGCGGGCTCGCTCTGCCGTGCGCTAGGCTTGGTGGGAAGGCCT
GTTCTCGAGTCCGCGCTTTTCGTCACCGCCATGTCGGGAGGTGGTGTGATTCGTGGCCCC
GCAGGGAACAACGATTGCCGCATCTACGTGGGTAACTTACCTCCAGACATCCGAACCAAG
GACATTGAGGACGTGTTCTACAAATACGGCGCTATCCGCGACATCGACCTCAAGAATCGC
CGCGGGGGACCGCCCTTCGCCTTCGTTGAGTTCGAGGACCCGCGAGACGCGGAAGACGC
GGTGTATGGTCGCGACGGCTATGATTACGATGGGTACCGTCTGCGGGTGGAGTTTCCTCG
AAGCGGCCGTGGAACAGGCCGAGGCGGCGGCGGGGGTGGAGGTGGCGGAGCTCCCCG
AGGTCGCTATGGCCCCCCATCCAGGCGGTCTGAAAACAGAGTGGTTGTCTCTGGACTGCC
TCCAAGTGGAAGTTGGCAGGATTTAAAGGATCACATGCGTGAAGCAGGTGATGTATGTTAT
GCTGATGTTTACCGAGATGGCACTGGTGTCGTGGAGTTTGTACGGAAAGAAGATATGACCT
ATGCAGTTCGAAAACTGGATAACACTAAGTTTAGATCTCATGAGGGAGAAACTGCCTACAT
CCGGGTTAAAGTTGATGGGCCCAGAAGTCCAAGTTATGGAAGATCTCGATCTCGAAGCCG
TAGTCGTAGCAGAAGCCGTAGCAGAAGCAACAGCAGGAGTCGCAGTTACTCCCCAAGGAG
AAGCAGAGGATCACCACGCTATTCTCCCCGTCATAGCAGATCTCGCTCTCGTACATAAGAT
GATTGGTGACACTTTTTGTAGAACCCATGTTGTATACAGTTTTCCTTTATTCAGTACAATCTT
TTCATTTTTTAATTCAAACTGTTTTGTTCAGAATGGGCTAAAGTGTTGAATTGCATTCTTGTA
ATATCCCCTTGCTCCTAACATCTACATTCCCTTCGTGTCTTTGATAAATTGTATTTTAAGTGA
TGTCATAGACAGGATTGTTTAAATTTAGTTAACTCCATACTCTTCAGACTGTGATATTGTGTA
AATGTCTATCTGCCCTGGTTTGTGTGAACTGGGATGTTGGGGGTGTTTGTGGTTATCTTAC
CTGGGGAAGTTCTTATGTTTATCTTGCTTTTCATGTGTCTTTCTGTAGACATATCTGAAGAG
ATGGATTAAGAATGCTTTGGATTAAGGATTGTGGAGCACATTTCAATCATTTTAGGATTGTC
AAAAGGAGGATTGAGGAGGATCAGATCAATAATGGAGGCAATGGTATGACTCCAAGTGCTA
TTGTCACAGATGAAATTGGCAGTATTGACCTTATACTAAAAGGCAGGGGTTAAAAATGATTA
TATACATTTTCCTTAAAACACTTGCAAACATTTTATTCAGTTGTCTTTAGCTACAATTGCTTTG
CTTTTTAAACCTTGGCAATTGTGGCAAAATTATATTGCCCATTTTGTAGCAACTTATTTTGCT
CCCTTCCCCCCATTTTTGTTTTAATAGGGACTAATGTGGGAAGAACTGGCTAATTTGTCACA
GTGCTTAGTTACAACTGTTAATGTGTGACCTGCTGTTGGTGTACATGTGGGTACAGGGTGT
TTTTAAATCCAACAAGATAGAGTATAATATCAATACTGCTAAATCTGCATGTCCTCTGTGTGA
CTGATAGAGCGTTGCTATTTCATTTTTTTAAGACAAAATGAAAGCAAAATATAGAGTTCCAAT
GTATTGGTGTAGATAATCTAGTTGGGAATACTTTTAAGTCTCACCTTCCCCTTTAAACTAATA
TTCATAATTGGTTCATATGTTTAAAAGACTTTAATTTACAAATTAAATTGCAAATGGGAGCAT
TAGATTTAGTTTTAGACTTAGGTGGGTAGCAATGCCAGTAAACTTAAATTACGTAACTTCTT
GCAACCACGAAACCTGTAATACGCTGTACAGTAACAAGTGTTGGCATTATCAGTTGAACTG
TAAATACAAAATGCTTCTTCCAATTAGTCTCTATGATGATTAAGTTTCTAAAATTTATCTGAAC
ACCATTCAGAAACTTGTTTTGGGGAATTTGATAGTTATTGATGTGCATCTGTTAAACTGATG
ACAGACATAACTCATCATTCCCCAGAAACCTTTTTTGATTACAGTATCTAACATTTTGCCTCC
TCTTTTTTGGTTTTGCTGGTTATAAAGGTTTGGATTGGAGAGGGCTCACTGGATCCCAATCC
TTGGAGCTGGATCATTGGATTCAAATCATAATGTGGATAGGATAGGGAGGATGAATTACCA
GGATTCATGGAGCGGGATCAGATTACCAGGAACATAGGAGTGGATTCCTGCCCCAACCAA
ACCGCATTCGTGTGGATTTTTTTATTCAACTTAATTGGCTATTCCAAAGATTTTTTTTTTCCTA
TTTTTGACGATTGGAGCCCTTAAGATGCACGATGGAATTGTGTTTTGCGTTTTTTGGTAAAA
GGAGCAAAGCGAGGACCTGGAGATAAACGCTGGAGCAATCTCCTTGGAAGGATTCAGCAC
GAGTAGATGGTAAACATTTAAAGGGGAAAGGGGGGGTTTGTTTAAAATAGTAAATCAGTAA
GTCACTTCTAAATTTAAAGAAAACAAAATTGGAGTTGAAGAATAAGTAGGTTTCCAATTGGC
TATTGCCGTTTTCTTTGAAAAAATAAACATTTTTTAAAAAACTATGCATGGTTGTCCTTTTTCC
TCTTCATGTAAGATTCTAACTGGGTCTATCAGTTAATCTTTAAATTGTTAAGTAAGATAAGAT
TTTGACTCTTGTGTTAATGTGTTAGCAAATTAAAAGTTCTTAAAAGGCAATCTAATGGTATTA

GCCATCTTTTATTGTTAATTGTAAAAGTCTTCAGGGGAAAGCAAAAGGGGAGAATAAGGCA
TTTGTGTATGTAACTTGGTAAATGACGGTGGGGGATGGATCTAGCATCTGAAAGATAAGCT
TCTCTACTTTGTTATAAAGTGGTTAAAAAACTATAGATGCTGCTTATTTTCTGGTGGTCATAG
ACAACATAGGCTTTTGTGCAAAATTGGTTGATGGCTACTAATGTTCACTTGGAGATAGCTTT
TGATATTCTCAATGAAACTCATCTCAAAAAAAGGTAAGTATTAAATGTTAACATCAGCACAGA
TGTATTAGAACTGTTTTTTGTTTTTGAGACAGAGTCTCGCTCTGTTCTCCAGACTGGAGGGC
AGTGGTGTGATCTAGGCTCACTGCAACCTCCACCCCTGGATTTGAGTGATTCTCGTGCCTC
AGTCTCCCAAGTAGCTGAGACTACAAGTGTGTGCCACCCTTGCCCGGCTAATTTTGTATTT
TTAGTAGAGATGTGGTTTCTCTGTGTTACCCAGGCTGGTCCCAAAACTCCTGGCCTCAAGT
GATCTGCCTGCCTTGGCCTCCCAAAGTGTTAGGATTACAGGTGTGAGCCACCATGCTCAG
CCTGTAGAACTTTTAACCCAAGTCTCATTTCTTTTTTGAAAGGGAAGAGTGCACAAGATTAA
CTGCTTCTTTGGATGAATCATTGTTAATAAAAAGCTGGGCATTTAGAATTTTGCCTTATAAGC
CCTTCTCCAACCATAAGATTATTTTGTACCAAAAACTTTGGTGTTCTCTACCAAAGCAGTTAA
AAACTTTTAGCCTGCTACTTCTTGTATTTGTCTACTGACAGCCCCTTGGTACTATTTAGGTT
GGGGGAGGGGACCTAAAATAAATAGACTTTAACATTTCCCTTGGGTGCTAATCATAGTTGG
AAGTTGAATTTAAGGTGATTATTTGGGTGACAATTAAAAACCTAAGGAAAACCAGAAATCTT
GGTAGTGGAAGAAATGTGTAAGGTCACCCCAATCGGTAGATTTTAATGAACGTTGTGGAAT
GTTGGGAAGAGGGGATGTTAAGTTGAATGCAGAATTTCACTAAGTACTTAGTGTAAGTTTAA
GGATGTAGCTCTTTTTATCTAAGAATTCAATGTAATGGCCAAAAGGCAGATTTACTGTTTAA
AAATTTGAATAATTTTACATGACATTCTTGAAATTCTAAGAAGTTTTATGTGTAGAACATTTTA
AAAATTCATCAGATTATTAAAGGGAAAATAAATGATTAATGATAATTTTGGAAGGTTAATGTG
AGCTAGACTTAAGTAAACTTTGGTTCATTTGTGTTCATTGAATGTTTTGGAAATGACCAAAAA
ATGTAAATGGCCTTCACTCAAGTTTGAGTGTTTAAAGTTGAAAGATGTGCTCTACTAAAAGT
TATAGTAATTCTAACCTCACATTGAAATGAGACAGTATTCCTTGTTATACAGGCTGAATTTGA
AGATTAGAGAGGATCTAATGTTTACTTAGGTAAAGGGGCATAGGTTTTGTAGTTAAGATGAC
CAGACAGCTAAAAGCTGTGATGGGAAGTATGGACTGCTCCTATTTATAGTCTCAGAAAATG
GACCTTTAGGTCTCTATCCGTATTGGCAATTATTAGAAGAAAGTTACACCCTTTTGAAACTA
CAAAAGCTGTCTTGGAATTTCCCCCTCTTCTCCCTATTTATGTCCCCTTAGAATATTTTAGG
GAGCCTATAATTATTTTCTAACCAAGGAAAAACTTAAGTCTCTTTAAGAAGCAATTACTTTTC
ATAACATCAGATTGAATAACCCACCTTGCTGTTCAGCCCACATCCTACTGGAAACAAAAGGT
AAGAAACCCATTTTCTGGTTCTTGATTGTTTGGGTCTGAATTTTGTTTTTAAAACTAAGCTAA
GTTTAATGTTTTTTTAAAATGCTGTTTGGAATATGAATAGATTCCCCGTAAAATGATTTTTCCT
AAGTTTTATGCTTTAGTAAATATTCAGTGCTCACGTCTGTGCATCATAGTGCTTGCGTTTAAT
ATGATTTATTGTAGAATCTCAACTTTTCTTGGTGTTTGTTGTCTTTGAAACATTGTCTTGGTC
ATTAGGGCTGGTGTTTTCACATTTCTGTGGTCAAGGTGGATTTCTTATGTGTGCCTTTTTGC
TTACTTTGTATATGAATTTTGTAATTTAAATTGCAAGTAAGTTATATATATGTATTTACCATAA
ATAGTATTAAAAGATGAGAAACTGTTAGACTGAAGTTCTGTTGTAACATAACCATTATTTCCA
TCACAGTATGAAGACTGCAAACGCAGAAAACAGATTACAGTCTCTTATCCATTTTTTGAAAT
CCAAAAACTACGAAAACAAAAGATTTTCTGTTGTTGAGCTAATTAAATGTGAACCCTGACCA
GAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID Nº 11:

MSGGGVIRGPAGNNDCRIYVGNLPPDIRTKDIEDVFYKYGAIRDIDLKNRRGGPPFAFVEFEDP
RDAEDAVYGRDGYDYDGYRLRVEFPRSGRGTGRGGGGGGGGGAPRGRYGPPSRRSENRV
VVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVVEFVRKEDMTYAVRKLDNTKFRSHE
GETAYIRVKVDGPRSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSR
SRT

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 este gen es un miembro de la familia rica en serina / arginina (SR) de factores de *splicing* pre-ARNm, que constituyen parte del spliceosoma. Cada uno de estos factores contiene un RRM (*RNA Recognition Motif*) para unir ARN y un dominio RS para unir otras proteínas. El dominio RS es rico en residuos de serina y arginina y facilita la interacción entre diferentes factores de *splicing* SR. Además de ser crítico para el *splicing* de ARNm, también se ha demostrado que las proteínas SR están implicadas en la exportación de ARNm desde el núcleo y en la traducción. El corte y *splicing* alternativo da como resultado múltiples variantes de transcripción.

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*SRSF5*", se refiere tanto al gen como a la proteína humana, factor de la maquinaria de splicing. En el contexto de la presente invención, "*SRSF5*" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "*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: 12,

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: 12. 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: 13.

Gene ID: 6430

Localización del gen: 14q24.1

SEQ ID Nº 12:

GGAAGTGACGTAGGACGCGCCCTCCATTTTGTGGAGCGCCAGAGCTGCTAAGTGCGTCAG
TTGTGGAGTGGCGTAGACGAGTTAAGTCCTGGTCTGCGTGGAGGTCGACGACTCCGTCGC
AGACTACGGACCTGTCTGGGTCTCAGCCGCCAAAGACCCCGTCCGGTAGGAAGTACTAGC
CGGACATCATGAGTGGCTGTCGGGTATTCATCGGGAGACTAAATCCAGCGGCCAGGGAGA
AGGACGTGGAAAGATTCTTCAAGGGATATGGACGGATAAGAGATATTGATCTGAAAAGAGG
CTTTGGTTTTGTGGAATTTGAGGATCCAAGGGATGCAGATGATGCTGTGTATGAGCTTGAT
GGAAAAGAACTCTGTAGTGAAAGGGTTACTATTGAACATGCTAGGGCTCGGTCACGAGGT
GGAAGAGGTAGAGGACGATACTCTGACCGTTTTAGTAGTCGCAGACCTCGAAATGATAGA
CGAAATGCTCCACCTGTAAGAACAGAAAATCGTCTTATAGTTGAGAATTTATCCTCAAGAGT
CAGCTGGCAGGATCTCAAAGATTTCATGAGACAAGCTGGGGAAGTAACGTTTGCGGATGC
ACACCGACCTAAATTAAATGAAGGGGTGGTTGAGTTTGCCTCTTATGGTGACTTAAAGAAT
GCTATTGAAAAACTTTCTGGAAAGGAAATAAATGGGAGAAAAATAAAATTAATTGAAGGCAG
CAAAAGGCACAGTAGGTCAAGAAGCAGGTCTCGATCCCGGACCAGAAGTTCCTCTAGGTC
TCGTAGCCGATCCCGTTCCCGTAGTCGCAAATCTTACAGCCGGTCAAGAAGCAGGAGCAG
GAGCCGGAGCCGGAGCAAGTCCCGTTCTGTTAGTAGGTCTCCCGTGCCTGAGAAGAGCC
AGAAACGTGGTTCTTCAAGTAGATCTAAGTCTCCAGCATCTGTGGATCGCCAGAGGTCCCG
GTCCCGATCAAGGTCCAGATCAGTTGACAGTGGCAATTAAACTGTAAATAACTTGCCCTGG
GGGCCTTTTTTTAAAAAACAAAAACCACAAAAATTCCCAAACCATACTTGCTAAAAATTCTG
GTAAGTATGTGCTTTTCTGTGGGGGTGGGATTTGGAAGGGGGGTTGGGTTGGGCTGGATA
TCTTTGTAGATGTGGACCACCAAGGGGTTGTTGAAAACTAATTGTATTAAATGTCTTTTGAT
AAGCCTTCTGCTCACATTTTTGTGAATGTCTGAAGTATATAGTTTGTGTATATTGACAGAGC
TCTTTTATAACTAAAGCAAATTTAATTTTTTTGTACTAGAAAAAAATTTGAACATTTTAGTTCT
TGGTTATAAAAATGTTAATTCAGAATTAGTTTAATGCCTTAATTAAACTAATTAATAGCTTTGG
ACACTTAAAAGAGCTCTAAATTTGCTTGTACATAAAGGCTTAATTTGTTTTCCTTGTTAGGGT
CAAGGGTGTCCTCCACTCTTTAACAGCTGCTGGACAGACACATTAGAGCAGCTGTTTGTTA
TTGATAATAAAATATTATAAAACTA

SEQ ID Nº 13

:

MSGCRVFIGRLNPAAREKDVERFFKGYGRIRDIDLKRGFGFVEFEDPRDADDAVYELDGKELC
SERVTIEHARARSRGGRGRGRYSDRFSSRRPRNDRRNAPPVRTENRLIVENLSSRVSWQDLK
DFMRQAGEVTFADAHRPKLNEGVVEFASYGDLKNAIEKLSGKEINGRKIKLIEGSKRHSRSRSR
SRSRTRSSSRSRSRSRSRSRKSYSRSRSRSRSRSRSKSRSVSRSPVPEKSQKRGSSSRSKS
PASVDRQRSRSRSRSRSVDSGN

El término "NOVA1" se refiere al gen que codifica una proteína de unión a ARN neuronal específica, un miembro de la familia Nova de antígenos de enfermedad paraneoplásica, que es

reconocida e inhibida por los anticuerpos paraneoplásicos. Estos anticuerpos se encuentran en los sueros de pacientes con ataxia opsoclono paraneoplásica, cáncer de mama y cáncer de pulmón de células pequeñas.

A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*NOVA1*", se refiere tanto al gen como a la proteína humana, factor de la maquinaria de splicing. En el contexto de la presente invención, "*NOVA1*" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "*NOVA1*", 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: 14,

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: 14. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *NOVA1*. Entre otras posibles secuencias nucleotídicas que codifican *NOVA1* se encuentra, pero sin limitarse, la SEQ ID NO: 15.

Gene ID: 4857

Localización del gen: 14q12

SEQ ID Nº 14:

CTCTCCCTTCTCCACTCTCTCCCCCTGTCTCCTTTCTTCTTCTTCTTTCACCCTCCGTCTCTC
ACACCCCCTCCATTCCCCTGTCTCCTTTCTGACACTGCACTGCAGCTGCTCCTCAGCCCTG
CCCCCTCCCCAGTGAGAACAAACCAGCAACATTGCTTTTTTTCCTAAAGAGATTTATATTGA
TCCGATTAAAAAAAAAAAACCTTAAGAAACCCCAAACGCAAAAAAAAAAAAAAAAAAAAAG
AAAAAAGAAAAGAAAAAGCCAAAACAAAAGGGAGAACCTTCTCCCGGTAGCAGCGGCAGG
AACTGCAAACATGATGGCGGCAGCTCCCATCCAGCAGAACGGGACCCACACTGGGGTTCC
CATAGACCTGGACCCGCCGGACTCGCGGAAAAGGCCGCTGGAAGCCCCCCCTGAAGCCG
GCAGCACCAAGAGGACCAATACGGGCGAAGACGGCCAGTATTTTCTAAAGGTTCTCATAC
CTAGTTATGCTGCTGGATCTATAATTGGGAAGGGAGGACAGACAATTGTTCAGTTGCAAAA
AGAAACTGGAGCCACCATCAAGCTGTCTAAGTCCAAAGATTTTTACCCAGGTACTACTGAG
CGAGTGTGCTTGATCCAGGGAACGGTTGAAGCACTGAATGCAGTTCATGGATTCATTGCAG
AAAAAATTCGAGAAATGCCCCAAAATGTGGCCAAGACAGAACCAGTCAGCATTCTACAACC
CCAGACCACCGTTAATCCAGATCGCATCAAACAAGTAAAGATTATAGTTCCCAACAGCACA
GCAGGTCTGATAATAGGGAAGGGAGGTGCTACTGTGAAGGCTGTAATGGAGCAGTCAGGG
GCTTGGGTGCAGCTTTCCCAGAAACCTGATGGGATCAACTTGCAAGAGAGGGTTGTCACT
GTGAGTGGAGAACCTGAACAAAACCGAAAAGCTGTTGAACTTATCATCCAGAAGATACAAG

AGGATCCACAAAGTGGCAGCTGTCTCAATATCAGTTATGCCAATGTGACAGGTCCAGTGGC
AAATTCCAATCCAACCGGATCTCCTTATGCAAACACTGCTGAAGTGTTACCAACTGCTGCA
GCAGCTGCAGGGCTATTAGGACATGCTAACCTTGCTGGCGTTGCAGCCTTTCCAGCAGTTT
TATCTGGCTTCACAGGCAATGACCTGGTGGCCATCACCTCTGCACTTAATACATTAGCCAG
CTATGGATATAATCTCAACACTTTAGGTTTAGGTCTCAGTCAAGCAGCAGCAACAGGGGCT
TTGGCTGCAGCAGCTGCCAGTGCCAACCCAGCAGCAGCAGCAGCCAATTTATTGGCCACC
TATGCCAGTGAAGCCTCAGCCAGTGGCAGCACAGCTGGTGGTACGGCGGGGACATTTGC
ATTAGGTAGCCTGGCTGCTGCTACTGCTGCAACCAATGGATATTTTGGAGCTGCTTCTCCC
CTAGCTGCCAGTGCCATTCTAGGAACAGAAAAGTCCACAGATGGATCCAAGGATGTAGTTG
AAATAGCAGTGCCAGAAAACTTAGTTGGTGCAATACTTGGCAAAGGAGGGAAAACATTAGT
GGAATACCAGGAGTTGACTGGTGCAAGGATACAGATCTCCAAAAAAGGAGAATTCGTACCT
GGCACAAGGAATCGGAAGGTAACCATTACTGGAACACCAGCTGCAACACAGGCTGCTCAA
TATTTAATTACACAAAGGATCACATATGAGCAAGGAGTTCGGGCTGCCAATCCTCAGAAAG
TGGGTTGAGTGCCCCAGTTACACATCAGATTGTTTTAACCCCTCCTTTACCCCATTTTCAAG
AAGGATGTACTGTACTTTGCAGAAGTGAAGTTTTTCTGTTATTAATATATAATTATGCAAATG
AATGCGACTATGTTGACAATGTGTATATGTAAATAATATGTGTTTTACCAGATGTTTCATAGA
AAGAATTTTTTCTTGATCTGTTTTGTTCTCTATACTTTGCTTGTGTATATTTGTCAGAGGTGTT
TCTAGTGTAAGATTTAAGCCTGCCATTTTACCAGCATTATTGTAGTTTAATGATTGAATGTAG
ACAGGGATATGCGTATAGTTTTCAGTATTAGTTCTAGATAACACTAAATTAACTACTGTTAG
GTTGAGTATGGTGGGGTCAGTGACCTAAAATGGAGTGAGGCCAAAGCACTGTCCTGTAAG
TCTTACTTCCTGCTTAGGGCACAGTGAAGTAGGAAACAATATTTTGAAAATAAGTTTTAAATT
TAAAATGATCAAAAAGCAATATAGTTGCATAAAAGCACTGTAAAATATTTAAAAGGTTAAAAC
TGTGGAAAATTATATTGGTAAGTTTACAGATCAATAAAAGCACCTGTTCTCCATCTGAACTA
GACAATGGAAATAATGCTGCATGCTGGCCATGGCCCATTCTTCATCATTTGTAAGTTCAACA
AAAGTTCTCACATGGAGTCCCACCTCTTCAGAGGTTTGTACATTTGTTTTTAAGCACTGAAT
TCACTACTGATCCCATCGCCTGGCCAGTAGAACAGTCATTACTCCATTAACATCCTCACTGT
TTAGACACATAACTGTGGTACAGTGTATTGGAAATTTTATAAACAAAAGTGAAAGTGCCAAC
AAATTATTGATAGCTGATAATGTTTCATTATCTGCAACTGCTTGATAAGTATGTTGCATTTTA
AGAGCTTATAATTGTGTATAATTTGTTAACACTAGAAACCTATTAGTATTGTGAATGTAGATT
TTACTGTGAAGCTATCTGTGATTTAGCTGTTTGCTCCCATGATGGAGTCTTTGCAGCATGGC
GCTAGCAGCCAATGCAGTTTCTAATACTCGGTAATTTGCATGTTTTGTGGAGCATTTTTATG
TCACCAACCAGACAGTATTTCCTGCATGCTTATTTAGAAGAGGCAGCTTATCTTGAGAGGTA
GTGTTATCTACCTTTGTCAGGCTTTTTGACAGGTCATTTCAGAGTAAGCCTTTGTTCCCAAG
ACCCAACAACTGTCACCCTCTTCTGTACCTCTCCTGAGTGCCAACTGTCCAGGCCATTTGA
CACACCATCTGTTAACCTCTGAGTTTGCCCACTCAAGGCCACTCATAGGGGCATCCATCCC
CAAGCACCTCCTCATGCTGTGCATGCAGTCTTAAATTCAATGGACAAAAATAAAATGCTGG
CTACCTCTGGATCATCTGGCTGAGCAACTGAATTACAAAAGAGAATTACTTCCATCTCAACT
TCAACCCATTGATTACGTCCATCCTAGCAAGCTAAATGGCATCCCAGCTGCTCCTTTCTGT
GCAACCAATTAAAGAACAATGAGTGTGATGCTCCATGTCTGAATTTCGTCCAGCCTCTCTCT
GAACTGTGATCTTTGTCCTCATGAACTTTCCCTTTTGTTCATTGAACTATATGGACTCTTCAT
TTCATATTGATTTACTGTGCAATTTACTTTTGGACATTGAGAACTTGAAATTATTTCCTGATC
CCTTCCCCTTCCACTATTAATAATTCATTTCTGTCAAACTGTAAGAGTAGACTCATTTTTTTTT
TTTTAGTTTTTAACATTGGACTGTTATTTCATTTAGAGTTCTCTATCTCTAAATATTTATTTAG
AGAATGATTTTAAAAGGGAATGATATGCTTGTTTAAATGAAAGAGAAAAGCTGTAGTAAACT
GTGTTAATTGGTAATGACTATTTATCGTCGATACTCTGTAGCTGTGTAAGTTTTGACAAATA
GTGTATCTCGTGGAATCAGTGGTTAGCATTGCCGCTATTATATTTACTCATTTTATCATTATA
AATGTGCTTAGTTCATCATGTAGCATCA

SEQ ID Nº 15

MMAAAPIQQNGTHTGVPIDLDPPDSRKRPLEAPPEAGSTKRTNTGEDGQYFLKVLIPSYAAGSII
GKGGQTIVQLQKETGATIKLSKSKDFYPGTTERVCLIQGTVEALNAVHGFIAEKIREMPQNVAKT
EPVSILQPQTTVNPDRIKQVKIIVPNSTAGLIIGKGGATVKAVMEQSGAWVQLSQKPDGINLQER
VVTVSGEPEQNRKAVELIIQKIQEDPQSGSCLNISYANVTGPVANSNPTGSPYANTAEVLPTAA
AAAGLLGHANLAGVAAFPAVLSGFTGNDLVAITSALNTLASYGYNLNTLGLGLSQAAATGALAA
AAASANPAAAAANLLATYASEASASGSTAGGTAGTFALGSLAAATAATNGYFGAASPLAASAIL

GTEKSTDGSKDVVEIAVPENLVGAILGKGGKTLVEYQELTGARIQISKKGEFVPGTRNRKVTITG
TPAATQAAQYLITQRITYEQGVRAANPQKVG

El término "*SNRNP70*" (también llamado en la literatura RPU1; Snp1; U1AP; U170K; U1RNP; RNPU1Z; SNRP70; U1-70K) se refiere tanto al gen como a la proteína humana. A no ser que se indique lo contrario, cuando se utiliza en el presente documento el término "*SNRNP70*", se refiere tanto al gen como a la proteína humana, factor de la maquinaria de splicing. En el contexto de la presente invención, "*SNRNP70*" se define también por una secuencia de nucleótidos o polinucleótido, que constituye la secuencia codificante de la proteína "*SNRNP70*", 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: 16,

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: 16. en las que el polipéptido codificado por dichos ácidos nucleicos posee la actividad y las características estructurales de la proteína *SNRNP70*. Entre otras posibles secuencias nucleotídicas que codifican *SNRNP70 se* encuentra, pero sin limitarse, la SEQ ID NO: 17.

Gene ID: 6625

Localización del gen: 19q13.33

SEQ ID Nº 16:

ATGCGCGAAAGATCCGGGAAGTGTGATCGGAAAGGGGCGGAGCATCAGGGAGAGTCTCG
CGGACATAGGCGGTTCGGCGCGGAAAGCGGGAGGTGGAGGGGCGGCTTGGGGCAAGCG
CGCGCGCGCAGTGCAGAAGCCAGCCCCCCGCGGCTGAGGTACTCAAGGTGCCCAAAGGC
GGGGTAGTGACCTCGCGCGTGCGCTGTGCCCGCGGCAGCGCCGGGTCCTAGTGTGTGG
GTTGTTGTTGGCACCGCACGGCGCGTGCGCAGTGAGGACGGCGGAGGGATTTGCGGCCG
GGACCCACCCCCTGCTCCAGTCGCTATCGGAGGCCGCGCGGGTGGCTGAGCAGCGGCCT
GGTGCGCTCGCTTAGCGGGCGACGGAATCAGACGGACGTGGACGCCCCCGGAGTGGAA
GCCGAAGCAGGAGTTGTTGTTGCTGAGGGGCTGCCGCAGCCGCCGCGAGCCTCCGGACA
GACGCCAGAGCGAGGAGGGCGCTACGCGACTTGGCAAGATGACCCAGTTCCTGCCGCCC
AACCTTCTGGCCCTCTTTGCCCCCCGTGACCCTATTCCATACCTGCCACCCCTGGAGAAAC
TGCCACATGAAAAACACCACAATCAACCTTATTGTGGCATTGCGCCGTACATTCGAGAGTT

TGAGGACCCTCGAGATGCCCCTCCTCCAACTCGTGCTGAAACCCGAGAGGAGCGCATGGA
GAGGAAAAGACGGGAAAAGATTGAGCGGCGACAGCAAGAAGTGGAGACAGAGCTTAAAAT
GTGGGACCCTCACAATGATCCCAATGCTCAGGGGGATGCCTTCAAGACTCTCTTCGTGGC
GAGAGTGAATTATGACACAACAGAATCCAAGCTCCGGAGAGAGTTTGAGGTGTACGGACC
TATCAAAAGAATACACATGGTCTACAGTAAGCGGTCAGGAAAGCCCCGTGGCTATGCCTTC
ATCGAGTACGAACACGAGCGAGACATGCACTCCGCTTACAAACACGCAGATGGCAAGAAG
ATTGATGGCAGGAGGGTCCTTGTGGACGTGGAGAGGGGCCGAACCGTGAAGGGCTGGAG
GCCCCGGCGGCTAGGAGGAGGCCTCGGTGGTACCAGAAGAGGAGGGGCTGATGTGAAC
ATCCGGCATTCAGGCCGCGATGACACCTCCCGCTACGATGAGAGGCCCGGCCCCTCCCC
GCTTCCGCACAGGGACCGGGACCGGGACCGTGAGCGGGAGCGCAGAGAGCGGAGCCGG
GAGCGAGACAAGGAGCGAGAACGGCGACGCTCCCGCTCCCGGGACCGGCGGAGGCGCT
CACGGAGTCGCGACAAGGAGGAGCGGAGGCGCTCCAGGGAGCGGAGCAAGGACAAGGA
CCGGGACCGGAAGCGGCGAAGCAGCCGGAGTCGGGAGCGGGCCCGGCGGGAGCGGGA
GCGCAAGGAGGAGCTGCGTGGCGGCGGTGGCGACATGGCGGAGCCCTCCGAGGCGGGT
GACGCGCCCCCTGATGATGGGCCTCCAGGGGAGCTCGGGCCTGACGGCCCTGACGGTC
CAGAGGAAAAGGGCCGGGATCGTGACCGGGAGCGACGGCGGAGCCACCGGAGCGAGCG
CGAGCGGCGCCGGGACCGGGATCGTGACCGTGACCGTGACCGCGAGCACAAACGGGGG
GAGCGGGGCAGTGAGCGGGGCAGGGATGAGGCCCGAGGTGGGGGCGGTGGCCAGGAC
AACGGGCTGGAGGGTCTGGGCAACGACAGCCGAGACATGTACATGGAGTCTGAGGGCGG
CGACGGCTACCTGGCTCCGGAGAATGGGTATTTGATGGAGGCTGCGCCGGAGTGAAGAG
GTCGTCCTCTCCATCTGCTGTGTTTGGACGCGTTCCTGCCCAGCCCCTTGCTGTCATCCCC
TCCCCCAACCTTGGCCACTTGAGTTTGTCCTCCAAGGGTAGGTGTCTCATTTGTTCTGGCC
CCTTGGATTTAAAAATAAAATTAATTTCCTGTTGATAGTGGGCAAAAAAAAAAAAAAAAAA

SEQ ID Nº 17:

MTQFLPPNLLALFAPRDPIPYLPPLEKLPHEKHHNQPYCGIAPYIREFEDPRDAPPPTRAETREE
RMERKRREKIERRQQEVETELKMWDPHNDPNAQGDAFKTLFVARVNYDTTESKLRREFEVYG
PIKRIHMVYSKRSGKPRGYAFIEYEHERDMHSAYKHADGKKIDGRRVLVDVERGRTVKGWRPR
RLGGGLGGTRRGGADVNIRHSGRDDTSRYDERPGPSPLPHRDRDRDRERERRERSRERDKE
RERRRSRSRDRRRRSRSRDKEERRRSRERSKDKDRDRKRRSSRSRERARRERERKEELRG
GGGDMAEPSEAGDAPPDDGPPGELGPDGPDGPEEKGRDRDRERRRSHRSERERRRDRDR
DRDRDREHKRGERGSERGRDEARGGGGGQDNGLEGLGNDSRDMYMESEGGDGYLAPENG
YLMEAAPE

Los autores de la presente invención han demostrado que los niveles de expresión durante el ayuno y/o el post-prandio de los genes *RNU2, RNU4, RNU6, RNU12, RNU4ATAC, RNU6ATAC, SNRNP200, ESRP1, NOVA1, SRSF1* y *SRSF5* están significativamente reducidos y el de *SNRNP70* significativamente aumentado en PBMCs de pacientes que desarrollarán DMT2 en comparación con pacientes que no desarrollarán DMT2 (Figuras 1 y 2).

Para ello, se ha medido el patrón de expresión de elementos seleccionados de la maquinaria de *splicing* en PBMC de pacientes sin DMT2 pero que han sufrido un evento cardiovascular y que, por lo tanto, presentan alto riesgo de desarrollar DMT2, a los que hizo seguimiento durante 5 años. Establecieron que los niveles de expresión en PBMCs de ciertos componentes de la maquinaria de splicing estaban alteraros significativamente en pacientes que desarrollaron DMT2

durante los cinco años de seguimiento (incident-T2DM) y se asociaron con el desarrollo de DMT2. De hecho, las huellas moleculares comprendidas por los niveles en ayuno y, especialmente, post-prandiales de algunos de estos componentes de la maquinaria de *splicing* fueron capaces de diferenciar entre los pacientes que posteriormente desarrollaron DMT2 (incident-T2DM) de aquellos pacientes que no desarrollaron la enfermedad durante los cinco años de seguimiento con un AUC> 0,8 después del análisis de validación cruzada, que es significativamente más alto que los procedimientos estándar actuales (HbA1c y FINDRISC mostraron AUC <0,66 en nuestra cohorte). Además, los pacientes con bajos niveles de expresión de componentes especificos de la maquinaria de *splicing* en PBMC, incluidos *RNU2, RNU4, RNU6, RNU12, RNU4ATAC, RNU6ATAC, SNRNP200, ESRP1, NOVA1, SRSF1* y/o *SRSF5*, tienen un mayor riesgo (OR> 4) de desarrollo de DMT2 en comparación con aquellos pacientes con altos niveles de PBMC de estos componentes y/o los que presentan niveles aumentados de *SNRNP70.*

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 DMT2 en un individuo, de ahora en adelante primer método de la invención, que comprende los siguientes pasos:

a) Determinación de los niveles del producto de expresión de los genes *RNU2, RNU4, RNU6, RNU12, RNU4ATAC, RNU6ATAC, SNRNP200, ESRP1, NOVA1, SRSF1, SRSF y/o SNRNP70* 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.

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 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.

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

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 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 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 electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante resonancia magnética nuclear o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio. El perfil de expresión génica también podría obtenerse mediante la detección y/o cuantificación de las proteínas producto de la traducción del ARNm derivado de la transcripción de los genes 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 o RTqPCR). La detección en tiempo real de los productos amplificados puede llevarse a cabo mediante la utilización de moléculas fluorescentes que se intercalan en el ADN de cadena doble o mediante hibridación con diferentes tipos de sondas.

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 (Enzime Linked ImmunoadSorbent 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 DMT2 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 menor expresión de los factores *RNU2, RNU4, RNU6, RNU12, RNU4ATAC, RNU6ATAC, SNRNP200, ESRP1, NOVA1, SRSF1* y/o *SRSF5,* o cualquiera de sus combinaciones, respecto a una muestra de referencia, en el grupo de individuos con mayor probabilidad de desarrollar DMT2.

(d) incluir a los individuos que presentan una mayor expresión del factor *SNRNP70* respecto a una muestra de referencia, en el grupo de individuos con mayor probabilidad de desarrollar DMT2.

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 varios sujetos sanos o pacientes sin historia previa de DMT2. 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 generados a partir de una población de dos o más personas. La población, por ejemplo, pueden 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 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 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 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 menos por lo menos 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 1 10 %, 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 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 100% (es decir, ausente).

KIT O DISPOSITIVO DE LA INVENCIÓN

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 de *RNU2, RNU4, RNU6, RNU12, RNU4ATAC, RNU6ATAC, SNRNP200, ESRP1, NOVA1, SRSF1, SRSF5* y/o *SNRNP70*, 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 Nº 1 (*RNU2*), SEQ ID Nº: 2 (*RNU4*), SEQ ID Nº 3 (*RNU4ATAC),* SEQ ID Nº 4 (*RNU6*), SEQ

ID Nº 5 (*RNU6ATAC),* SEQ ID Nº 6 (*RNU12*), SEQ ID Nº 7 (*SNRNP200*), SEQ ID Nº 9 (*ESRP1*), SEQ ID Nº 10 (*SRSF1*), SEQ ID Nº 12 (*SRSF5*), SEQ ID Nº 14 (*NOVA1*), y/o SEQ ID Nº 16 (*SNRNF70*),

- 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 Nº 1 (*RNU2*), SEQ ID Nº: 2 (*RNU4*), SEQ ID Nº 3 (*RNU4ATAC),* SEQ ID Nº 4 (*RNU6*), SEQ ID Nº 5 (*RNU6ATAC),* SEQ ID Nº 6 (*RNU12*), SEQ ID Nº 7 (*SNRNP200*), SEQ ID Nº 9 (*ESRP1*), SEQ ID Nº 10 (*SRSF1*), SEQ ID Nº 12 (*SRSF5*), SEQ ID Nº 14 (*NOVA1*), y/o SEQ ID Nº 16 (*SNRNF70*),

- los anticuerpos son capaces de unirse específicamente a una región formada por cualquiera de las secuencias aminoacídicas SEQ ID Nº 8 (*SNRNP200*), SEQ ID Nº 11 (*SRSF1*), SEQ ID Nº 13 (*SRSF5*), SEQ ID Nº 15 (*NOVA1*), y/o SEQ ID Nº 17 (*SNRNF70*),

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 ^{32}P o tritio, 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 INVENCIÓN 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 INVENCIÓN.

Este estudio se ha desarrollado en el contexto del estudio CORDIOPREV, que es un ensayo prospectivo, aleatorizado, controlado y de intervención en el que un total de 1002 pacientes que han sufrido un evento cardiovascular y sin eventos clínicos en los últimos seis meses son seguidos anualmente para evaluar el posible desarrollo de DMT2. Los criterios de inclusión y exclusión y los factores de riesgo cardiovascular de los pacientes se han publicado previamente (Delgado-Lista, CORonary Diet Intervention with Olive oil and cardiovascular PREVention 562 study (the CORDIOPREV study): Rationale, methods, and baseline characteristics: A clinical trial comparing the efficacy of a Mediterranean diet rich in olive oil versus a low-fat diet on cardiovascular disease in coronary patients. Am Heart J. 2016.). En concreto, se seleccionaron los pacientes de 20 a 75 años de edad con enfermedad coronaria crónica establecida pero sin eventos clínicos en los últimos seis meses, sin otras enfermedades graves y con una esperanza de vida de al menos cinco años, con la intención de realizar un estudio de seguimiento a largo plazo. Además del tratamiento convencional para la enfermedad coronaria, los sujetos fueron aleatorizados en dos modelos dietéticos diferentes (dietas mediterráneas o bajas en grasas). Específicamente, 462 de esos pacientes no presentaron DMT2 (en el momento de la inclusión en el estudio) y, de ellos, 107 pacientes desarrollaron DMT2 (*incident-T2DM*) después de un seguimiento medio de 60 meses, según todos los criterios de diagnóstico de la Asociación Americana de Diabetes (ADA), evaluados sobre la base de las pruebas de tolerancia a la glucosa realizadas cada año. Se diagnosticó DMT2 si uno o más de los siguientes criterios estaban presentes en los sujetos del estudio:

concentración de glucosa en plasma en ayunas (FPG) ≥126 mg / dl,

glucosa en ayunas alterada (IFG);

FPG ≥200 mg / dl después de 2 h de prueba oral de glucosa (OGTT),

Tolerancia a la glucosa alterada (IGT);

Hemoglobina glucosilada (HbA1c) ≥6.5%.

Específicamente, entre los 107 pacientes que desarrollaron DMT2 (*incident-T2DM*), 43 sujetos fueron diagnosticados durante el primer año de seguimiento, 24 durante el segundo año, 11 durante el tercer año, 19 durante el cuarto año y 10 durante el quinto año.

Los 355 sujetos restantes no desarrollaron DMT2 durante el periodo de estudio. Todos los casos incidentes de DMT2 (n = 107) se incluyeron junto con 108 controles seleccionados y seleccionados al azar, que no desarrollaron DMT2 durante el seguimiento (*non-T2DM*)

La selección aleatoria de los sujetos no DMT2 se realizó mediante el muestreo estratificado computacional de los 462 sujetos *non-T2DM* del estudio CORDIOPREV de acuerdo con las siguientes variables clínicas, antropométricas y bioquímicas: dieta, edad, sexo, glucemia en ayunas, índice de masa corporal, Colesterol LDL y colesterol HDL. Para implementar este tipo de muestreo, la población objetivo se dividió primero en estratos separados y luego, las muestras se seleccionaron al azar dentro de cada estrato mediante un muestreo automático simple utilizando el software R.

MATERIALES Y METODOS

Para determinar dinámicamente el estado metabólico del paciente se implementaron en todos los pacientes:

La prueba de tolerancia oral a la glucosa (OGTT), (75 g de dextrosa monohidrato en 250 ml de agua, NUTER. TEC GLUCOSA, Subra, Toulouse, Francia) a las 8:00 am, y se recogieron muestras de plasma a 0, 30, 60, 90 y 120 min para determinar la glucosa en plasma y niveles de insulina;

y las pruebas de comidas ricas en grasa, a las 8:00 am se realizó una prueba de comida rica en grasa (0.7 g de grasa y 5 mg de colesterol por kg de peso corporal con 12% de ácidos grasos saturados, 10% de ácidos grasos poliinsaturados, 43% de ácidos grasos monoinsaturados, 10% de proteínas y 25% de carbohidratos). Se recogieron muestras de sangre antes de la comida y después de 4 h.

Para aislar PBMCs se recolectó la sangre venosa de los pacientes (12 h en ayunas durante la noche) en tubos que contenían EDTA durante la prueba de comida rica en grasas, a las 0 y 4 h.

Se aisló el ARN total de PBMC utilizando el kit de ARN Direct-zol (Zymo Research, Irvine, CA, EE. UU.). Se determinó la cantidad de ARN recuperado y su calidad se evaluó mediante el espectrofotómetro NanoDrop2000 (Thermo Fisher). Específicamente, todas las muestras de ARN pasaron los controles de calidad, siendo las relaciones de absorbancia 260/280 y 230/260 entre 1.8-2.0. Un µg de ARN se transcribió de forma inversa a ADNc utilizando cebadores hexaméricos aleatorios con el *kit First Strand Synthesis* (Thermo Fisher).

Se extrajeron PBMC de n = 7 sujetos sanos y luego se cultivaron en medio RPMI sin suero (Lonza). Se sembraron 500,000 células / pocillo de cada sujeto en placas de 12 pocillos de fijación ultra baja (Corning Costar, Sigma) y se trataron por duplicado con un 10% de ayuno de línea de base y suero postprandial derivado de pacientes de control e *incident-T2DM*. Después de los períodos de incubación (4 y 24 h), las PBMC se centrifugaron y el ARN se extrajo y se aisló utilizando el reactivo TRI (Sigma).

Para determinar la expresión de 48 transcripciones en 48 muestras, simultáneamente, se implementó un Array Dinámico 48.48 basado en tecnología microfluídica (Fluidigm, San Francisco, CA, EE. UU.). Se diseñaron específicamente cebadores específicos para transcritos humanos que incluyen componentes del espliceosoma principal (n = 13) y secundario (n = 4), los SF asociados (n = 28) y tres genes *housekeeping* (Tabla 1).

El panel de componentes de la maquinaria de *splicing* se seleccionó sobre la base de dos criterios principales:

1. la relevancia de los componentes del spleciosoma dado en el proceso de *splicing* (como los componentes del núcleo del spleciosoma) y
2. la implicación demostrada en la regulación del *splicing*

Las variantes ya han demostrado estar implicadas en la fisiopatología de la DMT2 (como es el caso de los 28 factores de *splicing* seleccionados en este estudio).

	Gen	Número acceso	Secuencia primer *Sense (Se)*	Secuencia primer *Antisense (As)*	Tamaño de producto	Posición nucleotídica
Componentes del spliceosoma	PRPF40A	NM_017892.3	GCTCGGAAGATGAAACGAAA	TGTCCTCAAATGCTGGCTCT	130	Se 2459; As 2288
	PRPF8	NM_006445.3	TGCCCACTACAACCGAGAA	AGGCCCGTCCTTCAGGTA	139	Se 2373; As 2511
	RBM22	NM_018047.2	CTCTGGGTTCCAACACCTACA	GGCACAGATTTTGCATTCCT	137	Se 133; As 269
	RNU11	NR_004407.1	AAGGGCTTCTGTCGTGAGTG	CCAGCTGCCCAAATACCA	108	Se 4; As 111
	RNU12	NR_029422.1	ATAACGATTCGGGGTGACG	CAGGCATCCCGCAAAGTA	106	Se 26; As 149
	RNU2	NR_002716.3	CTCGGCCTTTTGGCTAAGAT	TATTCCATCTCCCTGCTCCA	116	Se 8; As 123
	RNU4	NR_003925.1	TCGTAGCCAATGAGGTCTATCC	AAAATTGCCAGTGCCGACTA	103	Se 21; As 132
	RNU4atac	NR_023343.1	GTTGCGCTACTGTCCAATGA	CAAAAATTGCACCAAAATAA	85	Se 19; As 103
	RNU6	NR_004394.1	CGCTTCGGCAGCACATATA	AAAATATGGAACGCTTCACGAA	101	Se 6; As 106
	RNU6atac	NR_023344.1	TGAAAGGAGAGAAGGTTAGCACTC	CGATGGTTAGATGCCACGA	112	Se 9; As 120

	SF3B1	NM_012433.3	CAGTTCCGTCTGTGTGTTCG	GCTGCCTTCTTGCCTTGA	101	Se 65; As 165
	SF3B1 tv1	NM_012433.3	GCAGACCGGGAAGATGAATA	TTTTCCCTCCATCTGCAAAA	88	Se 431; As 518
	SNRNP70	NM_001301069.1	TCTTCGTGGCGAGAGTGAAT	GCTTTCCTGACCGCTTACTG	114	Se 821; As 934
	SNRNP200	NM_014014.4	GGTGCTGTCCCTTGTTGG	CTTTCTTCGCTTGGCTCTTCT	103	Se 249; As 351
	TCERG1	NM_006706.3	GAGGAGCCCAAAGAAGAGGA	CACCAGTCCAAACGACACAC	112	Se 1550; As 1661
	U2AF1	NM_006758.2	GAAGTATGGGGAAGTAGAGGAGATG	TTCAAGTCAATCACAGCCTTTTC	120	Se 424; As 543
	U2AF2	NM_007279.2	CTTTGACCAGAGGCGCTAAA	TACTGCATTGGGGTGATGTG	130	Se 1246; As 1375
Factores de splicing	CELF1	NM_006560.3	AACAGAAGAGAATGGCCCAGC	TGCTGAAGGAGTGCTAAATACTG	121	Se 837; As 957
	CELF4	NM_020180.3	CCCCAGCAGCAGAGAGAA	GAAGCCGAAAGGGAGGAA	108	Se 1627; As 1734
	ESRP1	NM_020180.3	TTTTGGGATCACTGCTGGGG	TGTCCCACCTTCTTGTTGGC	108	Se 216; As 323
	ESRP2	NM_024939.2	AGAGCCCAGCAGTCAATTGTT	GTCTCACTGTCCACCACATCAG	96	Se 833; As 928
	KHDRBS1	NM_006559.2	GAGCGAGTGCTGATACCTGTC	CACCAGTCTCTTCCTGCAGTC	106	Se 774; As 879
	MAGOH	NM_002370.3	GCCAACAACAGCAATTACAAGA	TTATTCTCTTCAGTTCCTCCATCAC	88	Se 265; As 352
	NOVA1	NM_002515.2	TACCCAGGTACTACTGAGCGAG	CTGGTTCTGTCTTGGCCACAT	124	Se 592; As 715
	PTBP1	NM_002819.4	TGGGTCGGTTCCTGCTATT	CAGATCCCCGCTTTGTAC	111	Se 45; As 155
	RAVER1	NM_133452.2	GTAACCGCCGCAAGATACTG	CGAAGGCTGTCCCTTTGTATT	126	Se 298; As 423
	RBM17	NM_032905.4	CAAAGAGCCAAAGGACGAAA	TACATGCGGTGGAGTGTCC	107	Se 345; As 451
	RBM3	NM_006743.4	AAGCTCTTCGTGGGAGGG	TTGACAACGACCACCTCAGA	98	Se 253; As 350
	RBM45	NM_152945.3	CCCATCAAGGTTTTCATTGC	TTCCCGCAGATCTTCTTCTG	123	Se 415; As 537
	SFPQ	NM_005066.2	TGGTAGGGGGTGAAAGTG	TTAAAAACAAGAAATGGGGAAATG	125	Se 2873; As 2997
	SND1	NM_014390.3	ACTACGGCAACAGAGAGGTCC	GAAGGCATACTCCGTGGCT	101	Se 2679; As 2779
	SNW1	NM_001318844.1	ATGCGTGCCCAAGTAGAGAG	TCCCCATCCTCTTTTTCCA	134	Se 937; As 1070
	SRRM1	NM_001303448.1	GTAGCCCAAGAAGACGCAAA	TGGTTCTGTGACGGGGAG	108	Se 733; As 840
	SRRM4	NM_194286.3	CCTTCACCACCTCCTCAC	TTCGGCACATTCCAGACA	113	Se 1386; As 1498
	SRSF1	NM_006924.4	TGTCTCTGGACTGCCTCCA	TGCCATCTCGGTAAACATCA	98	Se 580; As 658
	SRSF10	NM_006625.5	CTACACTCGCCGTCCAAGAG	CCGTCCACAAATCCACTTTC	103	Se 343; As 445
	SRSF2	NM_003016.4	TGTCCAAGAGGGAATCCAAA	GTTTACACTGCTTGCCGATACA	113	Se 835; As 947
	SRSF3	NM_003017.4	TAACCCTAGATCTCGAAATGCATC	CATAGTAGCCAAAAGCCCGTT	117	Se 155; As 271
	SRSF4	NM_005626.4	GGAACTGAAGTCAATGGGAGAA	CTTCGAGAGCGAGACCTTGA	110	Se 857; As 966
	SRSF5	NM_001039465.1	GCAAAAGGCACAGTAGGTCAA	TTTGCGACTACGGGAACG	92	Se 723; As 814
	SRSF6	NM_006275.5	AGACCTCAAAAATGGGTACGG	CTTGCCGTTCAGCTCGTAA	82	Se 263; As 344
	SRSF9	NM_003769.2	CCCTGCGTAAACTGGATGAC	AGCTGGTGCTTCTCTCAGGA	87	Se 628; As 714
	TIA1	NM_022037.2	TAAATCCCGTGCAACAGCAGA	TATGCAGGAACTTGCCAACCA	124	Se 2806; As 2929
	TRA2A	NM_013293.4	TCAAAGGAGGCTATGGAAAGG	TGTGTGCGCTCTCTTGGTTA	90	Se 734; As 823
	TRA2B	NM_004593.2.	GATGATGCCAAGGAAGCTAAAG	AGGTAGGTCTCCCCATGTAAATTC	130	Se 784; As 913
Hk	ACTB	NM_001101	ACTCTTCCAGCCTTCCTTCCT	CAGTGATCTCCTTCTGCATCCT	176	Se 864; As 1039
	GAPDH	NM_002046	AATCCCATCACCATCTTCCA	AAATGAGCCCCAGCCTTC	122	Se 402; As 423
	HPRT	NM_000194.2	CTGAGGATTTGGAAAGGGTGT	TAATCCAGCAGGTCAGCAAAG	157	Se 252; As 409

Tabla 1. Primers específicos para los transcritos humanos medidos, incluyendo componentes de los spliceosomas mayor y menor, factores de splicing asociados y tres genes housekeeping (hk) que fueron específicamente diseñados y utilizados en un array de qPCR basado en tecnología microfluídica. Los números de acceso de NCBI, secuencias de los primers, tamaño de producto esperado y posición nucleotídica de los genes estudiados han sido incluidos.

Los cebadores se seleccionaron utilizando el software Primer3 con parámetros de selección establecidos para identificar pares de cebadores que: (1) abarcan un intrón (cuando es posible), (2) difieren en no más de 1°C en la temperatura de recocido, (3) tienen una longitud de al menos 20 pb, (4) tienen un contenido de GC entre 45 y 55%, pero (5) excluyen los cebadores que pueden formar dímeros de cebador. Las secuencias de los cebadores seleccionados se utilizaron en las búsquedas de BLAST (NCBI) para verificar la posible homología con secuencias distintas al objetivo designado. La selección inicial de la eficacia del cebador utilizando la detección en

tiempo real se realizó amplificando diluciones de 2 veces de productos de RT, donde la eficiencia óptima se demostró mediante una diferencia de un umbral de ciclo entre diluciones y un pico de fusión claro seguido de una disociación graduada dependiente de la temperatura a Verifique que solo un producto fue amplificado. El perfil de termociclado consistió en: 1) 95ºC durante 1 minuto; 2) 35 ciclos de desnaturalización (95ºC durante 5 segundos) y recocido / extensión (60ºC durante 20 segundos); y 3) un último ciclo en el que los productos de la PCR final se sometieron a una disociación graduada dependiente de la temperatura (60ºC a 95ºC, aumentando 1ºC / 3 segundos). Los productos de la PCR se purificaron en la columna (FAVORGEN Biotech, Viena, Austria) y se secuenciaron para confirmar la especificidad del objetivo. Después de la confirmación de la eficacia y especificidad del cebador, se determinó la concentración de productos purificados y los productos de PCR se diluyeron en serie para obtener estándares que contenían 1, 10^1,10^2, 10^3, 10^4, 10^5 y 10^6 copias de la plantilla sintética. Los estándares luego se amplificaron mediante qPCR, y las curvas estándar se generaron utilizando el software Stratagene Mx3000p. La pendiente de una curva estándar para cada plantilla examinada fue de aproximadamente −3.33 (R2≈1), lo que indica que la eficiencia de la amplificación de nuestros cebadores fue del 100%, lo que significa que se copiaron todas las plantillas de cada ciclo.

Las preamplificaciones, el tratamiento con exonucleasas y la matriz dinámica de qPCR basada en tecnología microfluídica se implementaron siguiendo las instrucciones del fabricante utilizando el sistema Biomark y el software de análisis de PCR en tiempo real (Fluidigm). Finalmente, en el caso de las muestras de PBMC de pacientes no T2DM e *Incident-T2DM*, el nivel de expresión de cada transcripción se ajustó mediante un factor de normalización (NF) obtenido a partir de los niveles de expresión de dos genes de mantenimiento diferentes [Beta-actina (ACTB) y gliceraldehído-3-fosfato deshidrogenasa (GAPDH)] utilizando Genorm 3.3 mientras que ACTB fue el gen de mantenimiento utilizado para normalizar el nivel de expresión de cada transcripción en el experimento *in vitro*.

RESULTADOS

La expresión basal de los componentes de la maquinaria de splicing es diferente entre pacientes *incident-T2DM* y *non-T2DM*.

El patrón de expresión de diferentes componentes del spliceosoma y SFs se encontró alterada en PBMCs de pacientes *incident-T2DM* comparado con pacientes *non-T2DM* (Figura 1A). Así, las PBMCs de pacientes que desarrollaron DMT2 (*incident-T2DM*) mostraron niveles de expresión significativamente menores de *RNU2, RNU4, RNU6ATAC, SNRNP200, ESRP1, SRSF1* y *SRSF5* (Figura 1B).

Además, la expresión de algunos componentes del spliceosoma y SFs se encontró dinámicamente y diferencialmente regulada en PBMCs de estos individuos durante un test de sobrecarga de lípidos (una situación de estrés que estima la flexibilidad metabólica (Figura suplementaria 1). Así, algunos componentes del spliceosoma y SFs se alteraron selectivamente durante el post-prandio en un grupo específico de pacientes (p.e. RNU12 se incrementó en pacientes non-T2DM, mientras que SNRNP70 se incrementó en incident-T2DM), lo que dio lugar a la aparición de cambios significativos (Figura suplementaria 1A). Además, muchos de los cambios observados en el ayuno se mantuvieron en el post-prandio (*RNU2, RNU4, RNU6ATAC, ESRP1* y *SRSF1*) aunque otros no se mantuvieron (*SNRNP200* y *SRSF5*) (Figura suplementaria 1B y C).

Así, la comparación entre los perfiles de expresión de las PBMCs de los pacientes *incident-T2DM* y *non-T2DM* en el estado post-prandial reveló una drástica desregulación de varios componentes del spliceosoma y SFs (Figura 2A). Específicamente, los niveles de expresión de *RNU2, RNU4, RNU6, RNU4ATAC, RNU6ATAC, RNU12, NOVA1, ESRP1 y SRSF1* fueron menores, mientras que los de *SNRNP70* fueron mayores en PBMCs de pacientes *incident-T2DM* vs. *non-T2DM* durante la fase post-prandial (Figura 2B).

La expresión de algunos componentes del spliceosoma y SFs está asociado con el riesgo de desarrollar DMT2.

La posible asociación entre la expresión de los componentes del spliceosoma y SFs con el riesgo de desarrollar DMT2 se analizó mediante estudios de regresión logística (odds ratio, OR). Este análisis reveló que bajos niveles basales de expresión en ayuno de *RNU4 y RNU2*, así como bajos niveles basales de expresión en post-prandio de *RNU4ATAC, NOVA1, RNU6ATAC, RNU6, RNU12, SRSF1 y RNU4* están fuertemente asociados con el riesgo relativo de desarrollar DMT2, una vez ajustado por edad, genero, dieta, IMC, HbA1c, c-HDL y TG (Tabla 2).

Las alteraciones en la expresión de los componentes de la maquinaria de *splicing* pueden predecir el desarrollo de DMT2.

Las huellas moleculares determinadas por los niveles de expresión de los componentes del spliceosoma y SFs durante el ayuno, el post-prandio o su combinación generaron curvas ROC para la predicción del desarrollo de DMT2 con AUC>0.85 usando diferentes algoritmos (Table 3 y tabla suplementaria 1), las cuales fueron validadas por métodos de validación cruzada (AUCs entre 0.65-0.81). En concreto, la huella molecular formada por los niveles de expresión en ayuno y post-prandio de los componentes del spliceosoma y SFs presentó una AUC=0.813

(TPR=0.802, TNR=0.689) tras la validación cruzada (Figura 3A – línea verde), la cual fue significativamente mejor que la capacidad predictiva de los predictores de DMT2 clínicamente utilizados en la actualidad HbA1c y FINDRISK (Figura 3A), tal y como demuestra el test de DeLong de comparación de las AUCs ($p<0.05$ vs. HbA1c o FINDRISC).

Además, un subgrupo de estos componentes de la maquinaria y SFs seleccionados computacionalmente (mediante *feature ranking analysis*) muestra incluso una mayor capacidad de predicción comparada con la del conjunto completo de biomarcadores. Así, un modelo predictivo comprendido por los niveles en ayunas de *RNU4ATAC, SRSF3, SRSF6, SRSF10, TRA2B, PTBP1, SF3B1* y los niveles post-prandiales de *RNU4, RNU6, RNU4ATAC, RAVER1, NOVA1 y PRPF8* presenta un AUC=0.881 (TPR=0.801, TNR=0.700) tras la validación cruzada (Figura 3B – línea verde), lo cual fue también confirmado por diferentes modelos y mejoró claramente la capacidad predictiva de los predictores de DMT2 clínicamente utilizados en la actualidad HbA1c y FINDRISK ($p<0.001$ en el test de DeLong vs. HbA1c o FINDRISC) (Figura 3B).

Finalmente, la capacidad de las alteraciones de la maquinaria de splicing de predecir el desarrollo de DMT2 es incluso superior cuando se trata de predecir el desarrollo de DMT2 durante los dos primeros años de seguimiento (Figura 3C y tabla suplementaria 1).

El suero de pacientes *incident-T2DM* alteró la expresión de componentes del spliceosoma en PBMCs de pacientes sanos

Incubación durante 24h con suero basal en ayuno de pacientes *incident-T2DM* redujo la expresión de *RNU4* comparado con el suero de los pacientes control en PBMCs de individuos sanos (Figura 4a); mientras que la expresión de otros componentes del spliceosoma y SFs no fue alterada significativamente (Figura suplementaria 2). Además, la incubación durante 4h con suero post-prandial de pacientes *incident-T2DM* redujo la expresión de *RNU4* y *RNU12* comparado con el suero de los pacientes control en PBMCs de individuos sanos (Figura 4b)

	OR	95% C.I.		p-value
		Inferior	Superior	
RNU4 en ayuno	2.521	1.117	5.688	0.026
RNU2 en ayuno	2.283	1.012	5.153	0.047
RNU4ATAC en postprandio	4.456	1.821	10.903	0.001
NOVA1 en postprandio	4.099	1.836	9.154	0.001
RNU6ATAC en postprandio	3.762	1.706	8.298	0.001
RNU6 en postprandio	3.762	1.706	8.298	0.001
RNU12 en postprandio	2.274	1.041	4.967	0.039
SRSF1 en postprandio	2.204	1.007	4.825	0.048
RNU4 en postprandio	2.109	0.963	4.619	0.062

Tabla 2. Asociación entre la expresión de los componentes de la maquinaria de splicing en PBMCs durante el ayuno o postprandio y el riesgo relativo de desarrollo de DMT2 mediante análisis de regresión logística del riesgo relativo (odds ratio, OR). Los sujetos fueron categorizados en terciles de acuerdo con el nivel de expresión de cada componente del spliceosoma o factor de splicing del siguiente modo: niveles de expresión bajos (T1), niveles de expresión medios (T2) y niveles de expresión altos (T3), y el riesgo relativo (OR) se estimó entre los tercies T1 y T3 para cada uno de los elementos de interés. ***CI***: intervalo de confianza; ***p-value***: p valor

Modelo	AUC	p-value
Expresión basal de los componentes de la maquinaria de splicing en el ayuno	0.894	0.000
Expresión basal de los componentes de la maquinaria de splicing en el postprandio	0.853	0.000
Expresión basal de los componentes de la maquinaria de splicing en el ayuno y el postprandio	1	0.000
HbA1c	0.643	0.000
FINDRISK	0.548	0.231
HbA1c + FINDRISK	0.643	0.000
Expresión basal de los componentes de la maquinaria de splicing en el ayuno + HbA1c	0.898	0.000
Expresión basal de los componentes de la maquinaria de splicing en el postprandio + HbA1c	0.867	0.000
Expresión basal de los componentes de la maquinaria de splicing en el ayuno y el postprandio + HbA1c	1	0.000
Expresión basal de los componentes de la maquinaria de splicing en el ayuno + FINDRISK	0.895	0.000
Expresión basal de los componentes de la maquinaria de splicing en el postprandio + FINDRISK	0.856	0.000
Expresión basal de los componentes de la maquinaria de splicing en el ayuno y el postprandio + FINDRISK	1	0.000

Tabla 3. Capacidad de la huella molecular compuesta por los niveles de expresión de los componentes del spliceosoma y factores de splicing en ayunas y/o postprandio como modelos predictivos de DMT2 mediante regresión logística y análisis de curva ROC. Se construyeron modelos de regresión logística considerando los niveles basales de expresión en el ayuno y/o postprandio de los elementos medidos, solos o en combinación con los predictores clásicos de DMT2 [HbA1c (hemoglobina glicosilada) y FINDRISK]. ***AUC*:** Área bajo la curva. ***p-value***: p valor.

Pacientes que desarrollaron DMT2 durante los dos primeros años de seguimiento	Regresión logística		Random Forest		C4.5	
	Training	CV	Training	CV	Training	CV
HbA1c	0.691	0.687	0.716	0.606	0.681	0.633
FINDRISK	0.568	0.553	0.671	0.530	0.500	0.486
HbA1c y FINDRISK	0.691	0.667	0.950	0.611	0.706	0.647
Niveles basales en ayuno de SFs	0.924	0.72	1	0.745	0.988	0.630
Niveles basales en postprandio de SFs	0.906	0.678	1	0.806	0.998	0.739
Niveles basales en ayuno y postprandio de SFs	1	0.699	1	0.851	0.872	0.705
Niveles basales en ayuno y postprandio de SFs + HbA1c	1	0.681	1	0.871	0.989	0.756
Niveles basales en ayuno y postprandio de SFs + FINDRISK	1	0.682	1	0.861	0.971	0.745
Pacientes que desarrollaron DMT2 durante los dos primeros años de seguimiento	**Regresión logística**		**Random Forest**		**C4.5**	
	Training	CV	Training	CV	Training	CV
HbA1c	0.691	0.687	0.716	0.606	0.681	0.633
FINDRISK	0.568	0.553	0.671	0.530	0.500	0.486
HbA1c y FINDRISK	0.691	0.667	0.950	0.611	0.706	0.647
Niveles basales en ayuno de SFs	0.924	0.72	1	0.745	0.988	0.630
Niveles basales en postprandio de SFs	0.906	0.678	1	0.806	0.998	0.739
Niveles basales en ayuno y postprandio de SFs	1	0.699	1	0.851	0.872	0.705
Niveles basales en ayuno y postprandio de SFs + HbA1c	1	0.681	1	0.871	0.989	0.756
Niveles basales en ayuno y postprandio de SFs + FINDRISK	1	0.682	1	0.861	0.971	0.745

Tabla suplementaria 1. Capacidad de la huella molecular compuesta por los niveles de expresión de los componentes del spliceosoma y factores de splicing (SFs) en ayunas y/o postprandio como modelos predictivos de DMT2 utilizando diferentes algoritmos computacionales. Se aplicaron los algoritmos Regresión logística, Random Forest y C4.5 utilizando lenguaje R para determinar la capacidad de la huella molecular compuesta por los niveles de expresión de la maquinaria de splicing en ayuno y/o postprandio para predecir el desarrollo de DMT2 utilizando la cohorte completa de pacientes o únicamente aquellos que desarrollaron diabetes durante los dos primeros años de seguimiento. Estos análisis (training) fueron confirmados mediante análisis de validación cruzada (CV). Las áreas bajo la curva de training y CV están indicadas. ***HbA1c:*** hemoglobina glicosilada; ***SFs:*** factores de splicing.

REIVINDICACIONES

1.- Uso del producto de expresión de los genes *RNU2*, *RNU4*, *RNU4ATAC, RNU6, RNU6ATAC*, *RNU12*, *SNRNP200*, *ESRP1*, *SRSF1 SRSF5*, *SNRNP70* y/o *NOVA1* para predecir o pronosticar el desarrollo de DMT2 en un individuo.

2.- Un método *in vitro* de obtención de obtención de datos útiles para predecir o pronosticar el desarrollo de Diabetes Mellitus tipo 2 (DMT2) en un individuo, que comprende:

a) medir en una muestra biológica aislada de dicho individuo los niveles del producto de expresión de los genes *RNU2*, *RNU4*, *RNU4ATAC, RNU6, RNU6ATAC*, *RNU12*, *SNRNP200*, *ESRP1*, *SRSF1 SRSF5*, *SNRNP70* y/o *NOVA1, o cualquiera de sus combinaciones,* según se define en las reivindicaciones 1-2, y

b) comparar los niveles obtenidos en el paso anterior con los estándares,

3.- Un método *in vitro* para predecir o pronosticar el desarrollo de Diabetes Mellitus tipo 2 (DMT2) en un individuo, que comprende los pasos (a) y (b) según la reivindicación 3, y además comprende incluir en el grupo de individuos con mayor riesgo de desarrollo de DMT2 aquellos pacientes con bajos niveles de los factores de splicing *RNU2*, *RNU4*, *RNU4ATAC, RNU6, RNU6ATAC*, *RNU12*, *SNRNP200*, *ESRP1*, *SRSF1 SRSF5* y/o *NOVA1* y/o aquellos pacientes con altos niveles de *SNRNP70*, en comparación con una muestra de referencia.

4.- El método según cualquiera de las reivindicaciones 2-3 donde la muestra biológica es sangre periférica.

5.- El método según cualquiera de las reivindicaciones 2-4, donde los niveles de los componentes del spliceosoma o factores de splicing concretos se obtienen mediante cualquier técnica de expresión génica, como métodos analíticos para examinar niveles de expresión del ARNm.

6.- El método según la reivindicación anterior donde la detección del producto de expresión de los genes se realiza mediante RT-PCR y/o PCR cuantitativa (qPCR).

7.- El método según la reivindicación anterior donde el producto de expresión son proteínas, preferiblemente *SNRNP200, SRSF1, SRSF5, NOVA1* y/o *SNRNP70,* 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 o tener DMT2, cuando estos componentes de maquinaria de splicing específicos, incluyendo *RNU2*, *RNU4*, *RNU4ATAC, RNU6, RNU6ATAC*, *RNU12*, *SNRNP200*, *ESRP1*, *SRSF1 SRSF5* y/o *NOVA1,* o cualquiera de sus combinaciones, presentan bajos niveles de expresión por debajo del tercil inferior de expresión, con un AUC de 0.902 (sensibilidad del 85% y especificidad del 82%).

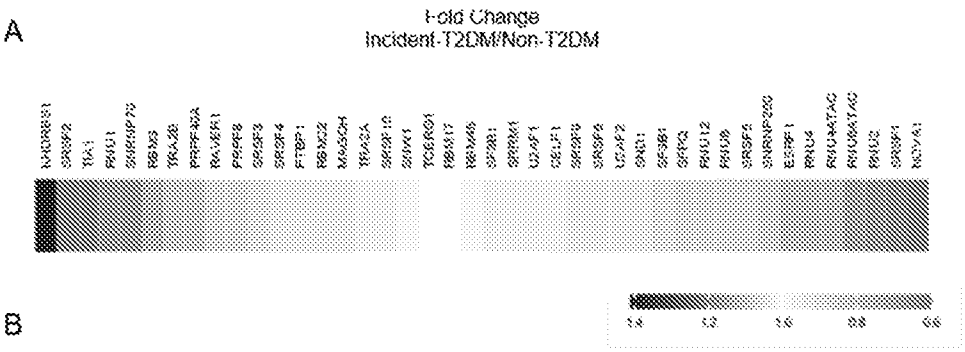
10.-Un kit o dispositivo que comprende los elementos necesarios para cuantificar los niveles de expresión de los genes, y donde:

- 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 Nº 1 (*RNU2*), SEQ ID Nº: 2 (*RNU4*), SEQ ID Nº 3 (*RNU4ATAC),* SEQ ID Nº 4 (*RNU6*), SEQ ID Nº 5 (*RNU6ATAC),* SEQ ID Nº 6 (*RNU12*), SEQ ID Nº 7 (*SNRNP200*), SEQ ID Nº 9 (*ESRP1*), SEQ ID Nº 10 (*SRSF1*), SEQ ID Nº 12 (*SRSF5*), SEQ ID Nº 14 (*NOVA1*), y/o SEQ ID Nº 16 (*SNRNF70*),
- 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 Nº 1 (*RNU2*), SEQ ID Nº: 2 (*RNU4*), SEQ ID Nº 3 (*RNU4ATAC),* SEQ ID Nº 4 (*RNU6*), SEQ ID Nº 5 (*RNU6ATAC),* SEQ ID Nº 6 (*RNU12*), SEQ ID Nº 7 (*SNRNP200*), SEQ ID Nº 9 (*ESRP1*), SEQ ID Nº 10 (*SRSF1*), SEQ ID Nº 12 (*SRSF5*), SEQ ID Nº 14 (*NOVA1*), y/o SEQ ID Nº 16 (*SNRNF70*),

- los anticuerpos son capaces de unirse específicamente a una región formada por cualquiera de las secuencias aminoacídicas SEQ ID Nº 8 (*SNRNP200*), SEQ ID Nº 11 (*SRSF1*), SEQ ID Nº 13 (*SRSF5*), SEQ ID Nº 15 (*NOVA1*), y/o SEQ ID Nº 17 (*SNRNF70*).

11.- El uso del kit o dispositivo según la reivindicación anterior para predecir o pronosticar el desarrollo de Diabetes Mellitus tipo 2 (DMT2) en un individuo.

Figura 1.


A
Fold Change
Incident-T2DM/Non-T2DM
B

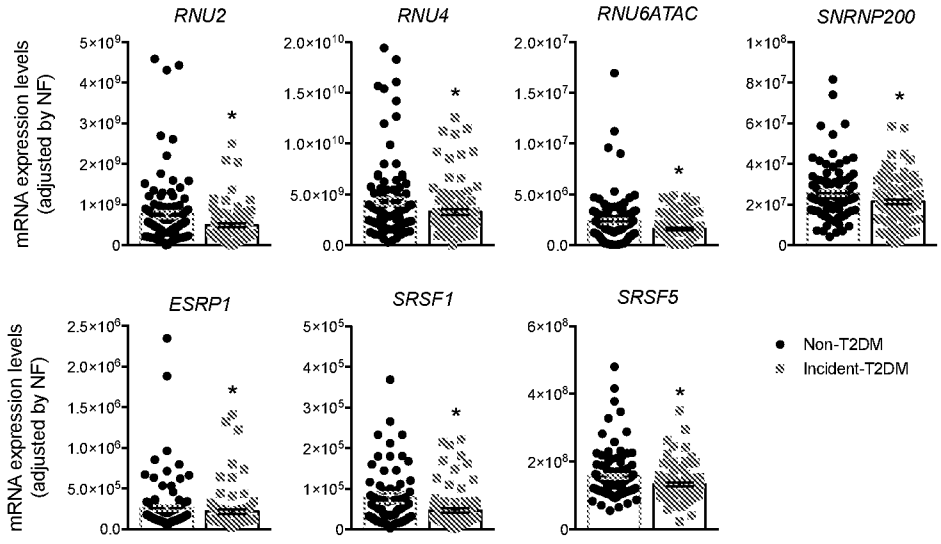

RNU2
RNU4
RNU6ATAC
SNRNP200
ESRP1
SRSF1
SRSF5
mRNA expression levels (adjusted by NF)
Non-T2DM
Incident-T2DM

Figura 2

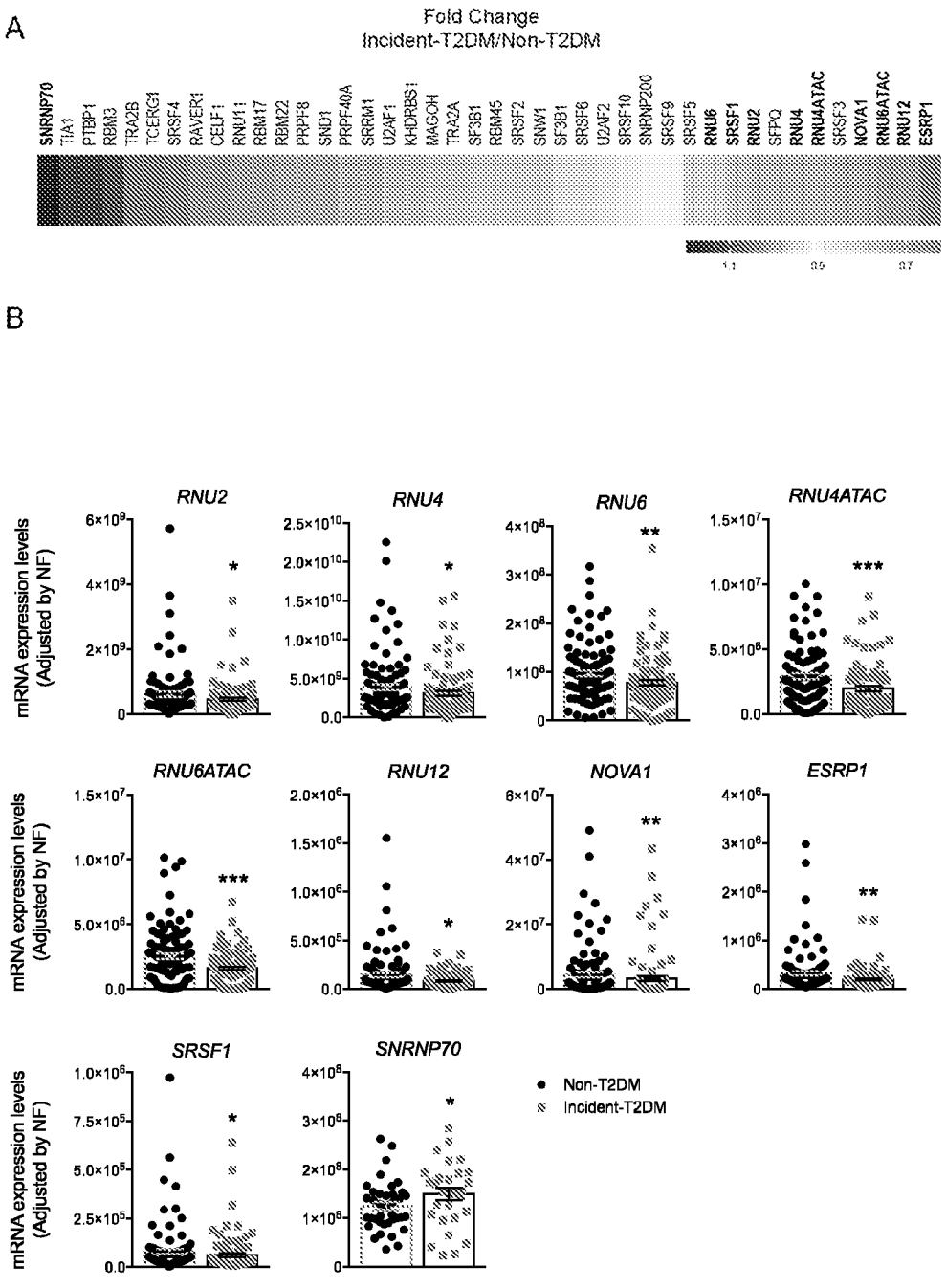

A
Fold Change
Incident-T2DM/Non-T2DM
SNRNP70
TIA1
PTBP1
RBM3
TRA2B
TCERG1
SRSF4
RAVER1
CELF1
RNU11
RBM17
RBM22
PRPF8
SND1
PRPF40A
SRRM1
U2AF1
KHDRBS1
MAGOH
TRA2A
SF3B1
RBM45
SRSF2
SNW1
SF3B1
SRSF6
U2AF2
SRSF10
SNRNP200
SRSF9
SRSF5
RNU6
SRSF1
RNU2
SFPQ
RNU4
RNU4ATAC
SRSF3
NOVA1
RNU6ATAC
RNU12
ESRP1
1.1
0.9
0.7
B
RNU2
RNU4
RNU6
RNU4ATAC
RNU6ATAC
RNU12
NOVA1
ESRP1
SRSF1
SNRNP70
mRNA expression levels
(Adjusted by NF)
Non-T2DM
Incident-T2DM

Figura 3

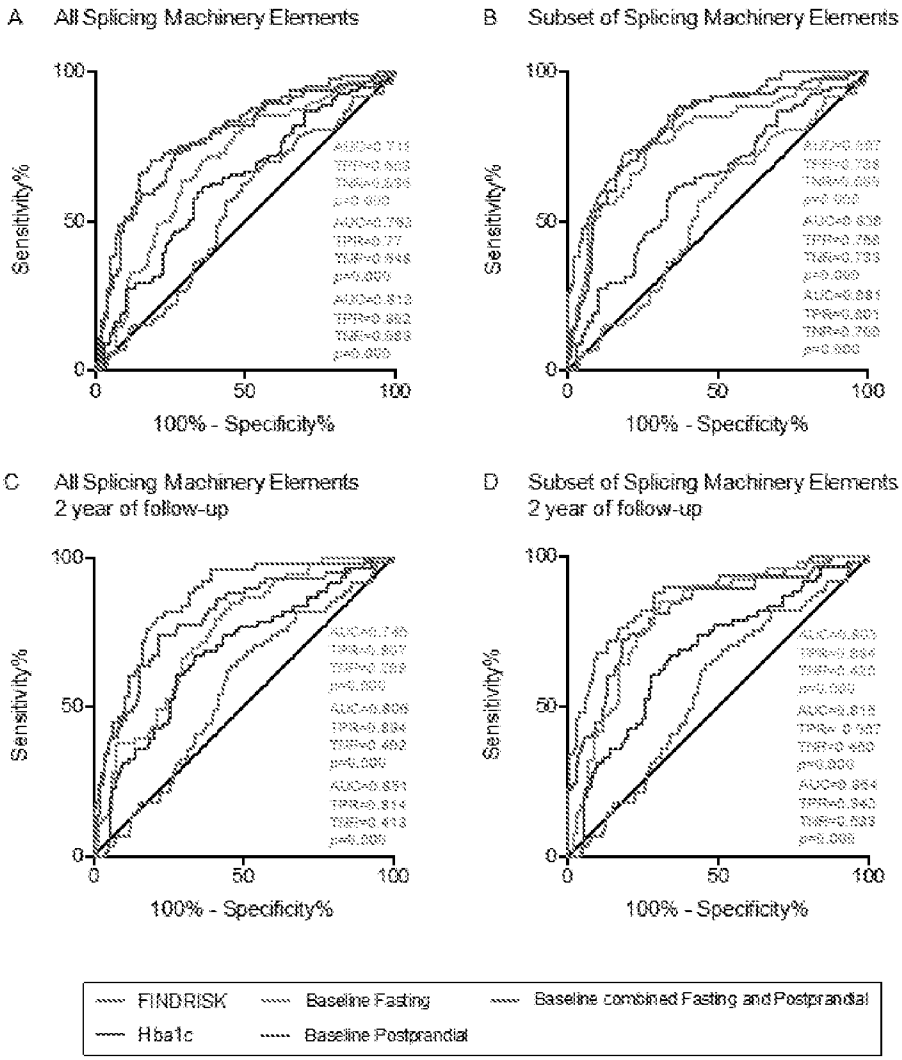
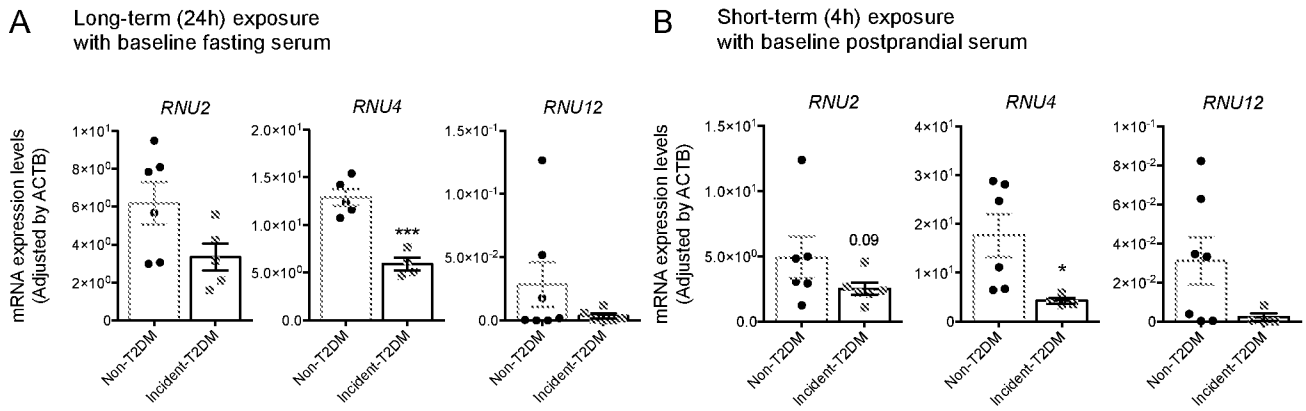

A All Splicing Machinery Elements
B Subset of Splicing Machinery Elements
C All Splicing Machinery Elements
2 year of follow-up
D Subset of Splicing Machinery Elements
2 year of follow-up
Sensitivity%
100% - Specificity%
0
50
100
FINDRISK
Baseline Fasting
Baseline combined Fasting and Postprandial
Hba1c
Baseline Postprandial

Figura. 4


A
Long-term (24h) exposure
with baseline fasting serum
B
Short-term (4h) exposure
with baseline postprandial serum
RNU2
RNU4
RNU12
mRNA expression levels
(Adjusted by ACTB)
Non-T2DM
Incident-T2DM
***
0.09
*