DEPARTAMENTO DE BIOLOGÍA CELULAR, FISIOLOGÍA E INMUNOLOGÍA

Role of splicing process alterations in the appearance of endocrine-metabolic diseases and tumor pathologies

Papel de la alteración del proceso de splicing en la aparición de enfermedades endocrino-metabólicas y patologías tumorales

Mercedes del Río Moreno

Córdoba, 2019

TITULO: *Role of splicing process alterations in the appearance of endocrine-metabolic diseases and tumor pathologies*

AUTOR: *Mercedes del Río Moreno*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

DEPARTAMENTO DE BIOLOGÍA CELULAR, FISIOLOGÍA E INMUNOLOGÍA

Role of splicing process alterations in the appearance of endocrine-metabolic diseases and tumor pathologies

Memoria de Tesis Doctoral presentada por **Mercedes del Río Moreno**, Licenciada en Bioquímica, para optar al grado de **Doctora en Biomedicina**

Los Directores,

Dr. Raúl M. Luque Huertas
Profesor Titular de Biología Celular
de la Universidad de Córdoba

Dr. Manuel D. Gahete Ortiz
Profesor Ayudante Doctor de Biología Celular
de la Universidad de Córdoba

En Córdoba, a 8 de abril de 2019

DEPARTAMENTO DE BIOLOGÍA CELULAR, FISIOLOGÍA E INMUNOLOGÍA

Dº Raúl Miguel Luque Huertas y Dº Manuel David Gahete Ortiz, Profesor Titular y Profesor Ayudante Doctor del Departamento de Biología Celular, Fisiología e Inmunología de la Universidad de Córdoba,

INFORMAN

Que Dª Mercedes del Río Moreno, Licenciada en Bioquímica, ha realizado bajo nuestra dirección el trabajo titulado "**Role of splicing process alterations in the appearance of endocrine-metabolic diseases and tumor pathologies"** y que bajo nuestro juicio reúne los méritos suficientes para optar al Grado de Doctora en Biomedicina.

Y para que conste, firmamos la presente en Córdoba, a 8 de abril de 2019.

Fdo.: Dr. Raúl M. Luque Huertas

Fdo.: Dr. Manuel D. Gahete Ortiz

TÍTULO DE LA TESIS: Role of splicing process alterations in the appearance of endocrine-metabolic diseases and tumor pathologies

DOCTORANDO/A: Mercedes del Río Moreno

INFORME RAZONADO DE LOS DIRECTORES DE LA TESIS

Durante el desarrollo de la presente Tesis Doctoral, en el periodo comprendido entre Enero de 2016 y Abril de 2019, la doctoranda Mercedes del Río Moreno no solo ha superado con creces los objetivos planteados al comienzo de la misma, sino que ha desarrollado y validado técnicas experimentales de una gran utilidad para el grupo de investigación, que le han permitido obtener resultados muy relevantes en el campo del estudio del proceso de splicing en diversas patologías como la obesidad, la diabetes mellitus o el cáncer. Concretamente, como fruto de su trabajo durante este periodo, ha publicado tres trabajos directamente relacionados con su Tesis Doctoral, en las revistas "*EBioMedicine*", "*Translational Research*" y "*Journal of Clinical Endocrinology & Metabolism*", revistas de referencia dentro de nuestras áreas de investigación. Además, el trabajo realizado en este periodo ha dado lugar a un cuarto artículo que está actualmente bajo revisión en una revista de prestigio en el área de Endocrinología y Metabolismo. Por último, la doctoranda ha presentado los resultados de su Tesis en diferentes congresos de ámbito nacional e internacional, así como participado en el desarrollo de numerosas patentes.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 8 de abril de 2019

Firma de los directores

Fdo.: Raúl M. Luque Huertas

Fdo.: Manuel D. Gahete Ortiz

Esta Tesis Doctoral ha sido realizada en el Departamento de Biología Celular, Fisiología e Inmunología de la Universidad de Córdoba y en el Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC), bajo la dirección de los Dres. Raúl M. Luque Huertas y Manuel D. Gahete Ortiz. Dicho trabajo ha sido subvencionado mediante los proyectos del Instituto de Salud Carlos III FIS (PIE14/00005, PI16/00264 y PI17/02287). Durante el transcurso de la presente Tesis Doctoral se ha realizado una estancia de tres meses en el Departamento de Medicina, Sección de Endocrinología, Diabetes y Metabolismo en la Universidad de Illinois en Chicago bajo la supervisión de Rhonda D. Kineman, financiada por una ayuda del Instituto Maimónides de Investigación Biomédica (IMIBIC) para la realización de estancias destinadas a la obtención de la Mención Internacional en el Título de Doctor de la Universidad de Córdoba.

List of publications

This Thesis is based on the research articles listed below, which will be referred in the test by their Roman numerals

Article I: Manuel D. Gahete*, **Mercedes del Rio-Moreno***, Antonio Camargo*, Juan F. Alcala-Diaz, Emilia Alors-Perez, Javier Delgado-Lista, Oscar Reyes, Sebastian Ventura, Pablo Perez-Martínez, Justo P. Castaño, José Lopez-Miranda and Raul M. Luque. *Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study*. 2018. EBioMedicine. 37: 356-365. *authors equally contributed for the publication.

[IF: 6.18, 13/133(D1) Medicine Research & Experimental (WOS)]

Article II: Mercedes del Rio-Moreno, Emilia Alors-Perez, Antonio Camargo, Jose Lopez-Miranda, Justo P. Castaño, Raul M. Luque and Manuel D. Gahete. *Dietary intervention modulates the expression of the splicing machinery in patients at high risk of type 2 diabetes development: From the CORDIOPREV study*. Manuscript submitted to "Diabetologia".

Article III: Mercedes del Rio-Moreno, Emilia Alors-Perez, Sandra Gonzalez-Rubio, Gustavo Ferrín, Oscar Reyes, Manuel Rodriguez-Peralvarez, Marina E. Sanchez-Frias, Rafael Sanchez-Sanchez, Sebastian Ventura, Jose Lopez-Miranda, Rhonda D. Kineman, Manuel de la Mata, Justo P. Castaño, Manuel D. Gahete and Raul M. Luque. *Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease*. 2019. Journal of Clinical Endocrinology and Metabolism. [Epub ahead of print; doi: 10.1210/jc.2019-00021].

[IF: 5.78, 20/143(Q1) Endocrinology & Metabolism (WOS)]

Article IV: Mercedes del Rio-Moreno, Emilia Alors-Perez, Patricia Borges de Souza, Maria E. Prados-Gonzalez, Justo P. Castaño, Raul M. Luque and Manuel D. Gahete. *Peptides derived from the extracellular domain of the somatostatin receptor splicing variant SST5TMD4 increase malignancy in multiple cancer cell types.* 2019. Translational Research. [Epub ahead of print; doi: 10.1016/j.trsl.2019.02.013].

[IF: 4.88, 3/30(D1) Medical Laboratory Technology (WOS)]

List of abbreviations

ADA - American Diabetes Association

AHA - American Heart Association

BP - Branch Point

CSCs - Cancer stem cells

CVD - Cardiovascular disease

DASH - Dietary approaches to stop hypertension

ECM - Extracellular matrix

ESE - Exonic splicing enhancer

ESS - Exonic splicing silencer

FADS - Fatty acid desaturase

GGT – Gamma-glutamyl transferase

GPCR - G-protein coupled receptor

HCC - Hepatocellular carcinoma

HIRI – Hepatic insulin resistance index

HOMA-IR - Homeostatic model assessment of insulin resistance

IGF1 - Insulin-like growth factor 1

INSR - Insulin receptor

IR - Insulin resistance

IRS - Insulin receptor substrate

ISE - Intronic splicing enhancer

ISS - Intronic splicing silencer

LF Diet - Low fat diet

MedDiet - Mediterranean diet

MMPs - Matrix metalloproteases

NAFLD - Non-alcoholic fatty liver disease

NASH - Non-alcoholic steatohepatitis

NET - Neuroendocrine tumor

OA - Oleic acid

PA - Palmitic acid

PBMCs - Peripheral blood mononuclear cells

snRNA - small nuclear RNA

snRNP- small nuclear ribonucleoprotein

SR proteins – Serin-arginine proteins

SS - Splice site

SS14 - Somatostatin 14

SSA - Somatostatin analog

SST1-5 - Somatostatin receptors 1-5

SST5TMD4 - Somatostatin receptor 5 variant with 4 transmembrane domains

T2DM - Type 2 Diabetes Mellitus

TCF7L2 - Transcription factor-7-like 2

TMD - Transmembrane domain

TMD - Transmembrane domain

Table of contents

Resumen

La obesidad y la diabetes, así como el cáncer representan graves problemas de salud pública en todo el mundo. Una característica común compartida por estas patologías endocrino-metabólicas y tumorales es la alteración del proceso normal de splicing, lo cual podría conducir a la presencia de variantes aberrantes de ARNm. En concreto, el splicing es un proceso biológico esencial por el cual los intrones del pre-ARNm son escindidos para unir los exones y generar así ARNs maduros capaces de ser traducidos a proteínas funcionales. Concretamente, la mayoría de genes de organismos eucariotas pueden sufrir procesos de splicing alternativo, el cual constituye un proceso fisiológico que permite a estos organismos aumentar exponencialmente el número de transcritos generados a partir de un mismo gen. Durante el proceso de splicing alternativo, las regiones no codificantes y codificantes de un gen se reorganizan para generar diferentes variantes de ARNm. Desafortunadamente, las desregulaciones de este proceso están asociadas a la aparición de diversas isoformas de splicing aberrantes que pueden exhibir un fuerte potencial patológico. En este sentido, debe tenerse en cuenta que el control de la expresión del ARN y la correcta generación de variantes de splicing, así como su homeostasis, se lleva a cabo por el spliceosoma, una maquinaria celular compleja compuesta por ribozimas y proteínas asociadas (factores de splicing) que interaccionan de forma dinámica para catalizar el proceso de splicing. Por tanto, el funcionamiento apropiado de esta maquinaria celular es esencial para mantener la homeostasis celular, tisular y del organismo y, por ello, la desregulación de este proceso ha sido asociada al desarrollo de diferentes enfermedades, incluyendo las patologías endocrino-metabólicas y tumorales.

Por este motivo, las alteraciones en el proceso de splicing podrían representar un nuevo campo para el estudio y la identificación de marcadores diagnósticos, pronósticos y dianas terapéuticas en enfermedades tan prevalentes, como las patologías endocrino-metabólicas (obesidad y diabetes) y el cáncer. Para explorar esta prometedora vía de investigación, se han seguido diferentes enfoques experimentales y metodológicos en la presentes Tesis Doctoral. En primer lugar, para determinar la relación entre la desregulación de los componentes de la maquinaria de splicing y el desarrollo de diabetes mellitus tipo 2 (DMT2), se midieron los niveles de expresión de algunos componentes de esta maquinaria en células mononucleares de sangre periférica de pacientes en alto riesgo de desarrollar DMT2, en condiciones tanto de ayuno como postprandiales. Esta aproximación mostró que los niveles de expresión de algunos componentes clave de esta maquinaria estaban alterados en aquellos pacientes que desarrollarían DMT2 en comparación con aquellos pacientes control que no desarrollan la enfermedad durante los 5 años

que duraba el estudio. Además, la huella molecular compuesta por los niveles de expresión de estos componentes en ayunas y durante el postprandio, era capaz de predecir con elevada precisión el desarrollo futuro de DMT2 en pacientes individuales, mejorando la capacidad predictora de parámetros clásicos como la hemoglobina glicosilada o el test FINDRISK, lo cual sugería que esta huella podría representar una nueva herramienta no invasiva para la detección temprana del riesgo de desarrollar DMT2.

En este contexto, se ha demostrado que el riesgo de desarrollar DMT2 puede reducirse o incluso revertirse como consecuencia de la aplicación de intervenciones dietéticas saludables; sin embargo, los mecanismos moleculares que subyacen a esta asociación están aún por dilucidar. Puesto que los resultados derivados de la presente Tesis han demostrado que el patrón de expresión de la maquinaria de splicing está asociado con el riesgo de desarrollar DMT2, nos propusimos evaluar la influencia de la intervención dietética durante 3 años en el patrón de expresión de la maquinaria de splicing en las células periféricas mononucleares (PBMCs, por sus siglas en inglés) obtenidas de este mismo grupo de pacientes en alto riesgo de desarrollar DMT2. En concreto, el patrón de expresión de los componentes de la maquinaria de splicing se determinó, tras tres años de seguimiento, en las PBMCs de todos los pacientes que desarrollaron DMT2 (n=107) y 108 pacientes que no desarrollaron DMT2, seleccionados aleatoriamente. Estos pacientes fueron asignados aleatoriamente en dos grupos dietéticos saludables, dieta mediterránea o dieta baja en grasa. Los resultados indicaron que la intervención dietética, independientemente del tipo (dieta mediterránea o baja en grasa), modula los niveles de expresión de componentes clave de la maquinaria de splicing (sobreexpresión de *SPFQ*, *RMB45*, *RNU6*, etc. o disminución de *RNU2* o *SRSF6*). También se observó que algunos de estos cambios en los niveles de expresión de componentes de la maquinaria de splicing fueron inducidos diferencialmente entre aquellos pacientes que desarrollaron DMT2 y aquellos pacientes que no la desarrollaron tras cinco años de seguimiento. Por tanto, este estudio sugeriría que una intervención dietética a largo plazo podría modular los niveles de expresión de componentes clave de la maquinaria de splicing en células mononucleares de sangre periférica de pacientes en alto riesgo de desarrollar DMT2, y que dichos cambios serían claramente diferenciales en pacientes que terminan desarrollando DMT2 y aquellos pacientes que no la desarrollan tras cinco años de seguimiento, lo que podría utilizarse como en una herramienta valiosa para monitorizar la progresión de DMT2

Debido a la creciente prevalencia de la obesidad y la DMT2, la enfermedad de hígado graso no alcohólico (NAFLD, por sus siglas en inglés) está emergiendo rápidamente como un problema de salud grave en los países occidentales. Para determinar la posible relación entre la desregulación de la maquinaria reguladora del splicing y el desarrollo de NAFLD, se analizó el patrón de expresión de los componentes del spliceosoma y factores de splicing en biopsias de hígado de pacientes obesas con diferentes grados de esteatosis hepática que fueron sometidas a cirugía bariátrica. Los resultados de esta aproximación mostraron que el hígado de las pacientes obesas con esteatosis presentaba una alteración manifiesta pero diferencial (dependiente del paciente) de la maquinaria celular responsable de la regulación del proceso de splicing. Concretamente, el patrón de expresión de los componentes del spliceosoma y los factores de splicing analizados no se asoció con el grado de esteatosis hepática, pero identificó tres subpoblaciones de pacientes con esteatosis molecularmente definidas y caracterizadas por la expresión alterada de ciertos componentes del spliceosoma y factores de splicing. Cabe destacar, que estas subpoblaciones presentaban características clínicas específicas, así como una respuesta diferencial a la cirugía bariátrica después de un año de seguimiento. El hecho de que estas alteraciones del spliceosoma no estuvieran asociadas con el grado de esteatosis sugiere su papel como desencadenante del inicio de la acumulación de grasa en el hígado. De hecho, una aproximación *in vitro* indicó que las alteraciones en los componentes de la maquinaria de splicing podrían preceder al desarrollo de la esteatosis hepática. Estos resultados arrojan luz sobre los posibles mecanismos moleculares subyacentes al desarrollo de la esteatosis hepática en pacientes obesos y proporcionan información novedosa para explorar el desarrollo de estrategias de diagnóstico, pronóstico o herramientas terapéuticas eficaces para el NAFLD.

Por último, para ampliar el conocimiento sobre la interacción patológica entre los procesos de splicing aberrantes y el cáncer, se realizó una caracterización del papel oncogénico del receptor truncado SST5TMD4 en patologías tumorales endocrinas. En concreto, se evaluaron las consecuencias funcionales del posible procesamiento proteolítico del dominio C-terminal de esta variante de splicing aberrante del receptor SST5 en diferentes tipos celulares representativos de estas patologías tumorales. Específicamente, un análisis *in silico* reveló la existencia de dos sitios de corte para metaloproteasas de matriz (MMP) en el dominio C-terminal extracelular de la variante de splicing SST5TMD4, que podrían ser utilizados por MMP2, 9 y 14 y/o MMP16, respectivamente, para generar tres péptidos con 7, 10 y 17 aminoácidos. El tratamiento con estos tres péptidos fueron capaces de aumentar parámetros funcionales de agresividad en diferentes líneas celulares derivadas de distintas patologías tumorales (tumores neuroendocrinos, cáncer de

mama, próstata e hígado); sin embargo, algunas de estas acciones parecían ejercer diferentes dinámicas, como es el caso de la proliferación celular, o, incluso, ser dependiente de la línea celular, como ocurrió con la migración o la formación de tumorosferas. En cualquier caso, los tres péptidos aumentaron de forma general las características de malignidad de dichas líneas celulares tumorales, probablemente a través de la activación de las rutas PI3K/AKT y/o MEK/ERK y la modulación de genes pro-oncogénicos clave. Este estudio sugiere que los péptidos derivados del receptor truncado SST5TMD4 podrían contribuir al fuerte papel oncogénico de este receptor, observado previamente en múltiples patologías tumorales y, por lo tanto, representan posibles candidatos para identificar nuevas dianas diagnósticas, pronosticas o terapéuticas en el cáncer.

En conjunto, los resultados presentados en esta Tesis Doctoral constituyen una información novedosa y valiosa que respalda la idea de que la desregulación del proceso de splicing, tanto las alteraciones en los componentes del spliceosoma y los factores de splicing, como en las variantes resultantes de dicho proceso, está estrechamente relacionada con la instauración y/o el desarrollo de diversas patologías relevantes como la diabetes, el hígado graso o el cáncer.

Summary

Global human health is currently threatened by highly prevalent diseases such as obesity, diabetes and cancer. A common hallmark shared by these endocrine-metabolic and tumor diseases is the consistent alteration of the normal splicing process, which leads to the presence of aberrant mRNA variants. Splicing represents an essential biological process by which the introns of an immature pre-mRNA are excised and the exons are fused to generate mature mRNAs capable to be translated into functional proteins. However, most eukaryotic genes can also undergo alternative splicing, which is a physiologic process that allows eukaryotic organisms to exponentially increase the number of transcripts generated from a given gene. During the process of alternative splicing, the coding and non-coding regions of a gene are reorganized in order to generate different mRNA variants. Unfortunately, dysregulations of this process are associated to the appearance of diverse protein isoforms that could exhibit strong pathological potential. In this sense, it should be noted that the control of the appropriate RNA expression and the generation of the correct splicing variants and their homeostasis is mainly regulated by the spliceosome, a complex cellular machinery comprised by ribozymes and associated proteins (splicing factors) that dynamically interact to catalyze the splicing process. The appropriate functioning of this cellular machinery is essential to maintain the cellular, tissue and body homeostasis and, therefore, the dysregulation of this process has been associated to different diseases, including endocrine-metabolic and tumor pathologies.

For these reasons, alterations in the splicing process could represent a novel source for the identification of diagnostic, prognostic and therapeutic targets in highly prevalent and health-threatening diseases such as endocrine-metabolic pathologies (obesity and diabetes) and cancer. To explore this promising research avenue, different experimental and methodological approaches have been implemented in the present Doctoral Thesis. First, to determine the relationship between the dysregulation of the splicing machinery components and the development of type 2 diabetes mellitus (T2DM), the expression pattern of selected components of this machinery was determined in the peripheral blood mononuclear cell (PBMCs) of individuals at high risk of T2DM development, under fasting and postprandial conditions. This approach showed that the expression levels of key splicing machinery components was altered in the PBMCs from individuals that developed T2DM compared with non-T2DM controls (patients that did not develop T2DM during the 5 years of follow-up). The molecular fingerprints comprised by the fasting and the postprandial levels of certain of these splicing machinery components were capable to predict the future development of T2DM in individual patients with

high precision, outperforming the capacity of classical predictors of T2DM development, such as HbA1c or FINDRISK, and could become a valuable, non-invasive, new tool for early risk assessment of T2DM.

In this scenario, it has been shown that the risk of T2DM may be reduced or even reversed with healthy dietary interventions; however, the molecular mechanisms underlying this association are still to be elucidated. Since the studies derived from this Thesis have demonstrated that the expression pattern of the splicing machinery is associated with the risk of T2DM development, we aimed to further evaluate the influence of a 3-years dietary intervention in the expression pattern of the splicing machinery components in PBMCs from the same group of individuals at high risk of T2DM development. In particular, the expression pattern of the splicing machinery components was determined, after 3-years of follow-up, in PBMCs from all patients who developed T2DM (n=107) and from 108 randomly selected non-T2DM patients. These patients were randomly enrolled in two healthy dietary interventions (Mediterranean or Low-Fat Diet). The results indicated that a long-term dietary intervention, regardless of the type of diet (Mediterranean or low-fat), influences the expression levels of key components of the splicing machinery (overexpression of *SPFQ*, *RMB45*, *RNU6*, etc. or decrease expression of *RNU2* o *SRSF6*). It was also noted that some of these changes in the expression levels of components of the splicing machinery were induced differentially between those patients who developed T2DM and those patients who did not develop T2DM after five years of follow-up. Therefore, this study suggest that a long-term dietary intervention could modulate the expression levels of key components of the splicing machinery in PBMCs of patients at high risk of developing T2DM, and that such changes would be clearly differential in patients who end up developing T2DM and those patients who do not develop T2DM after five years of follow-up, which could be used as a valuable tool to monitor the progression of T2DM.

Due to the rising prevalence of obesity and T2DM, non-alcoholic fatty liver disease (NAFLD) is rapidly emerging as an important and growing health issue in western countries. In an attempt to determine the possible relationship between the dysregulation of the splicing machinery components and the development of NAFLD, the expression pattern of the components of the splicing machinery was measured in liver biopsies from obese patients with different degrees of liver steatosis that underwent bariatric surgery. This experimental approach shown that the liver of steatotic obese patients exhibited an overt but differential (patient-dependent) alteration of the cellular machinery responsible for the regulation of the splicing

process. The expression pattern of the spliceosome components and splicing factors analyzed herein was not associated with the grade of hepatic steatosis but seemed to identify three discrete molecularly defined subpopulations of steatotic obese patients characterized by the dysregulated expression of certain spliceosome components and splicing factors. Remarkably, these subpopulations presented specific clinical characteristics and also a differential response to bariatric surgery after one year of follow-up. The fact that these spliceosome alterations were not associated with the grade of hepatic steatosis suggests their role as triggers for initiating fat deposition within the liver. Indeed, an *in vitro* approach suggested that alterations in splicing machinery components could precede the development of hepatic steatosis. These findings shed light to the possible underlying molecular mechanisms responsible for the development of hepatic steatosis in obese patients and provide novel information to explore the development of efficient screening strategies, diagnostic, prognostic or therapeutic tools for obesity-related NAFLD.

Finally, to further expand our knowledge on the pathological interplay between aberrant splicing processes and cancer, a more profound characterization of the oncogenic role of the truncated receptor SST5TMD4 on endocrine-related cancers was performed. Specifically, the functional and mechanistic consequences of the SST5TMD4-specific, splicing-derived C-terminal domain were evaluated in different cancer cells types. First, an *in silico* analysis revealed the existence of two putative cleavage sites for matrix metalloproteases (MMPs) in the SST5TMD4 C-terminal extracellular domain, which could be the substrate for MMP2, 9 and 14 and/or MMP16, respectively, to generate three derived peptides with 7, 10 and 17 aminoacids. These three SST5TMD4-derived peptides were capable to enhance the malignant characteristics of multiple cancer cell lines derived from diverse tumor pathologies (neuroendocrine tumor, breast, prostate and liver cancer); however, some of these actions seemed to exhibit different dynamics (i.e. proliferation rate) or, even, be cell-line dependent (i.e. migration or tumorosphere formation). In any case, this study demonstrate that the three derived peptides were able to increase tumor malignancy likely through the activation of PI3K/AKT and/or of MEK/ERK pathways and by the modulation key pro-oncogenic genes. Therefore, this study suggests that peptides derived from the spliced SST5TMD4 receptor could contribute to the strong oncogenic role of SST5TMD4 previously observed in multiple tumor pathologies, and, therefore, represent potential candidates to identify novel diagnostic, prognostic or therapeutic targets in cancer.

Altogether, the results presented in this Doctoral Thesis constitute novel, valuable and germane information supporting the contention that the dysregulation of the splicing process, including the alterations in the spliceosome components, the splicing factors or the surrogate markers of the splicing process, is closely related to the instauration and/or development of several and highly prevalent pathologies such as diabetes, NAFLD or cancer.

1. Introduction

Global human health is currently threatened by different relevant diseases such as obesity, metabolic syndrome, diabetes or tumor pathologies and cancer. The incidence and prevalence of these endocrine-metabolic and tumor diseases is steadily increasing (1-3) and represent a capital health problem for the National and International Health Systems and a scientific challenge for Biomedical Research.

These pathologies curse with diverse and severe alterations at genetic, molecular, cellular and tissue levels but, importantly, **a common hallmark shared by endocrine-metabolic dysregulations and cancers is the atypical and/or ectopic expression and the alteration and/or total or partial loss of relevant molecules (specially splicing variants) of certain key regulatory systems essential in the control of the normal physiology of the cells.** In this sense, a growing body of evidence supports the idea of an association between the aberrant presence of alternative mRNA variants and the development and progression of obesity, metabolic syndrome, diabetes and cancers (4-7). Specifically, during the splicing process, the coding and non-coding regions of a gene are reorganized in order to generate different mRNA variants, a process that is associated to the appearance of diverse protein isoforms that could exhibit strong pathological potential (8-10).

In this sense, it should be noted that the control of the appropriate RNA expression and the generation of the correct splicing variants and their homeostasis is mainly controlled by several intricate and tightly related processes, which are precisely catalyzed and regulated by different cellular machineries. In particular, **the splicing process is regulated by the spliceosome** (9), a complex cellular machinery comprised by ribozymes and associated proteins (splicing factors) that dynamically interact to catalyze the splicing process. The appropriate functioning of this cellular machinery is essential to maintain the cellular, tissue and body endocrine-metabolic homeostasis and, therefore, the dysregulation in the expression of some components of this splicing machinery might be associated to the development and/or progression of different pathologies, including endocrine-metabolic and tumor pathologies (11,12).

Taking into consideration all the information mentioned above, the present **Doctoral Thesis** is based on the **HYPOTHESIS** that the alteration of the normal, physiologic splicing process could be closely implicated in the development and progression of endocrine-metabolic diseases and cancer (Figure 1). Therefore, the main **OBJECTIVE** of this Thesis has been to explore the idea that the altered process of mRNA splicing (including alterations on the spliceosome

components, splicing factors and surrogate splicing variants with pathological potential) could represent a novel source for the identification of diagnostic, prognostic and therapeutic targets in highly prevalent and health-threatening diseases such as endocrine-metabolic pathologies and cancer.

Figure 1. At a glance overview of the Thesis rationale

1.1. Endocrine-metabolic diseases

Obesity and metabolic syndrome are chronic endocrine-metabolic diseases that represent emerging global epidemics and capital health problems for the National and International Health Systems since they increase the risk of developing severe endocrine, cardiovascular and tumor pathologies [World Health Organization (WHO) source]. Obesity and metabolic syndrome promote the development of multiple defects in the neuroendocrine-metabolic system, which cause relevant alterations in the homeostasis of the organism (e.g., dysregulation of hormones and growth factors, glucose/insulin homeostasis, etc.) that often favor the development of the serious pathologies mentioned above (e.g. endocrine disorders, cancer, etc.) (1,13). Obesity affects over a third of world's population today (Figure 2) and it is estimated that, by 2030, 38% of the world's adult population will be overweight and another 20% will be obese (14,15).

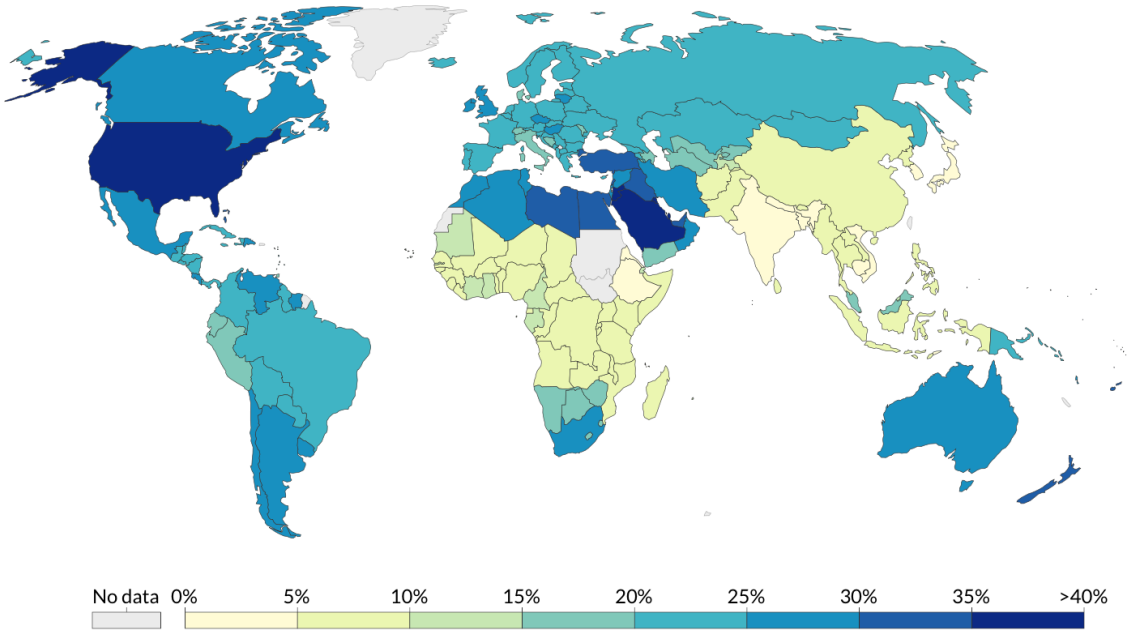


Figure 2. Percentage of obese adults worldwide in 2016. *Percentage of adults aged 18+ years old who are defined as obese based on their body-mass index (BMI). Source: WHO, Global Health Observatory.*

The rise of obesity and metabolic syndrome is closely linked to the concomitant increase of Type 2 diabetes mellitus (T2DM) cases. T2DM is a complex disease, characterized by a combination of impaired insulin action, increased hepatic glucose production and dysfunction in insulin secretion (16). About 90% of T2DM cases are attributable to excess weight, likely due to the predisposition to insulin resistance developed in obese subjects (2,17). The risk of suffering T2DM (425 million of people have T2DM and it is estimated that this number will rise to 629 millions by 2045) results from the combination of genetic factors with environmental influences (18). Due to the fact that T2DM is often linked to overweight and obesity, the prevalence of this disease is dangerously increasing in developed countries (2). One of the main problems associated to T2DM is the high risk of developing different related comorbidities, including cardiovascular disease (CVD). Indeed, patients with myocardial infarction and T2DM have higher risk of developing a new cardiovascular event that those without T2DM (19). Thus, early identification of individuals at high risk for T2DM development, especially among patients with CVD (20,21), is critical for prevention (22,23). Traditionally, this strategy has been based on biomarkers [glycated hemoglobin (HbA1c)] or predictive scores (FINDRISK); however, these approaches have limitations and cannot precisely predict an individual's risk of developing T2DM (24,25). In this sense, the identification of key modifiers of phenotypic plasticity (i.e. the

difficulty to cope with stressors to maintain metabolic homeostasis) that define individual susceptibility to develop T2DM may hold predictive potential (26). In the context of CVD prevention, lifestyle and behavioral intervention have become advantage approaches that have less associated costs and side effects than the current medical treatments (27). In particular, several healthy dietary patterns can be useful tools for the managements of T2DM and reduction of cardiovascular risk, such as the Mediterranean Diet (MedDiet), the Dietary Approaches to Stop Hypertension (DASH), the vegetarian diet, and the low-fat (LF) high carbohydrates diet recommended by the National Cholesterol Education Program and the American Heart Association (AHA) (27-30).

Due to the rising prevalence of obesity and T2DM, non-alcoholic fatty liver disease (NAFLD) is rapidly emerging as an important and growing health issue in western countries (31,32). NAFLD is described as a range of liver disorders characterized by fat accumulation within the liver (steatosis), which is not related to alcohol consumption. These disorders comprise a wide range of diseases, from simple steatosis to hepatic inflammation and fibrosis (non-alcoholic steatohepatitis or NASH), cirrhosis and hepatocellular carcinoma (HCC) (33,34). Although most patients with simple steatosis would remain stable, 10–15% with histologically proven NASH will progress to cirrhosis and HCC. Indeed, due to its high prevalence, NAFLD has emerged as speedily growing cause of end-stage liver disease and HCC, in addition to hepatitis C, hepatitis B and alcohol abuse (35,36). Unfortunately, the molecular mechanisms underlying the heterogeneous outcomes of NAFLD remain unclear, precluding any attempt to anticipate the disease progression to decompensated cirrhosis or HCC (37). Within the natural history of NAFLD, hepatic steatosis is the first stage, wherein an improved understanding of the pathogenesis of liver steatosis would have a critical prognostic impact for preventing disease progression (38,39). In this sense, while hepatic steatosis is closely associated with obesity, there is a meaningful percentage of obese people who have normal intrahepatic triglyceride content and appear to be resistant to developing obesity-related metabolic complications, including NAFLD (40). However, little is known about the mechanisms underlying the apparent resistance of this group of patients.

1.2. Tumor pathologies and cancer

Despite advances in recent years, tumor pathologies and cancer continue to be a critical problem for worldwide health population, which is associated to the lack of specific, sensitive and useful biomarkers for the diagnostic, prognostic and treatment of these pathologies (3).

Indeed, tumor pathologies are associated to leading causes of death in developed countries [World Health Organization (WHO) source] (Figure 3). Cancer encompasses a wide variety of malignancies with a variable etiology and complexity (41). Consequently, the development and progression of cancer is a highly heterogeneous and variable process, that is strongly influenced by genetics, but it is also profoundly conditioned by metabolic, nutritional, ambient and life style factors (42). In this sense, different components (neuropeptides, peptide hormones and/or their receptors) of several endocrine systems (sexual hormones, growth factors, insulin, IGF1, somatostatin, ghrelin, etc.) are commonly dysregulated in tumor pathologies (43-48). This concept is especially relevant in the so-called endocrine-related cancers, a term that classically refers to those hormone-related cancers such as tumor pathologies responsive to sex steroid and pituitary hormones but also to any cancer that exhibit certain "hormone sensitivity", at least at some stages of their development and/or progression (49,50). In this context, diverse examples of endocrine-related cancers types could be pituitary and neuroendocrine tumor (NETs), as well as breast, prostate and liver cancers. Pituitary tumors and NETs constitute a highly heterogeneous group of neoplasms arising from the pituitary gland or from the cells of the neuroendocrine system that are widely distributed throughout the body (51,52). Remarkably, breast and prostate cancers represent the most common malignant tumor types in female and male population respectively, and also some of the major cancer types in terms of mortality (3), with a strong endocrine influence (53), while hepatocellular carcinoma (HCC), the most common type of liver cancer, is the sixth most prevalent cancer type, its incidence is progressively increasing and exhibits only 17% of survival rate after 5-years (3,54,55). Breast and prostate cancers as well as HCC are closely related with the appropriate function of the endocrine-metabolic system in that obesity, and its endocrine, metabolic and inflammatory associated alterations, elevate the risk of developing these tumor pathologies (56,57).

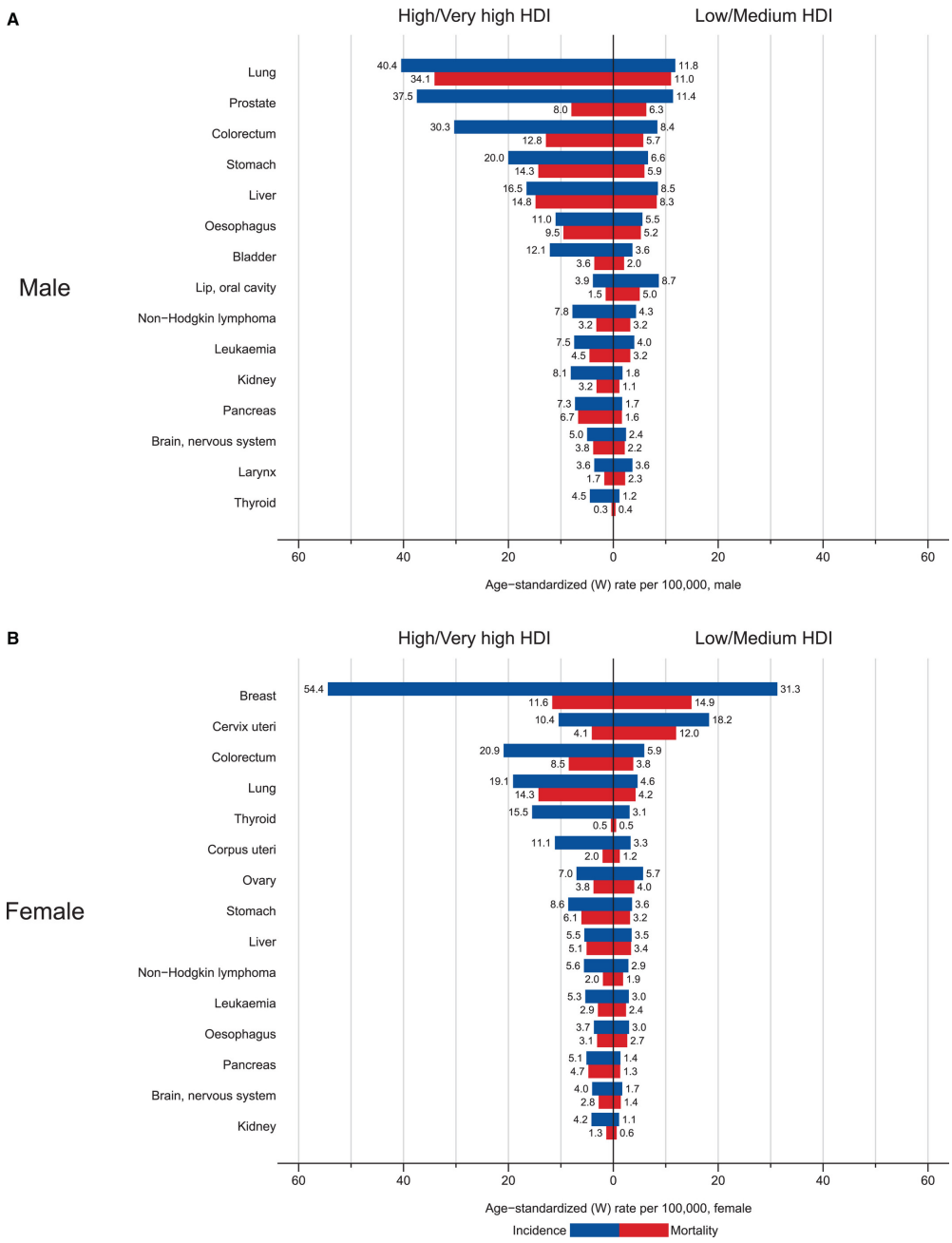


Figure 3. Age-Standardized Incidence and Mortality Rates in High/Very-High Human Development Index (HDI) Regions Versus Low/Medium HDI Regions Among (A) Men and (B) Women in 2018. *Rates are shown in descending order of the world (W) age-standardized rate. Source: GLOBOCAN 2018 (58).*

1.3. Splicing process

As indicated above, a common hallmark shared by endocrine-metabolic dysregulations and cancers is the atypical and/or ectopic expression and the alteration and/or total or partial loss of relevant molecules of certain regulatory systems essential in the control of the normal physiology of the cells. Of particular relevance for this Doctoral Thesis are the dysregulations of the normal, physiologic splicing events, which can be associated to altered expression and function of the cellular machinery responsible for this process (spliceosome and splicing factors) and that can lead to the generation of splicing variants with pathological/oncogenic potential.

In particular, splicing represents an essential biological process by which the introns of an immature pre-mRNA are excised and the exons are combined to generate mature mRNAs capable to be translated to functional proteins. Most genes in eukaryotes contain these intervening sequences, named introns, that are intercalated between the coding sequences or exons and need to be eliminated. From a biochemical perspective, the removal of introns followed by the ligation of the neighboring exons are two sequential trans-esterification reactions initiated by a nucleophilic attack of the 5′ splice site (5′SS) by the branch point (BP) of the intron to be spliced out. This reaction lead to the formation of the intron lariat with a 2′,5′-phosphodiester linkage, which is followed by another trans-esterification reaction between the 5′SS and the 3′ splice site (3′SS) that allow the excision of the intron lariat and the junction between exons (12) (Figure 4).

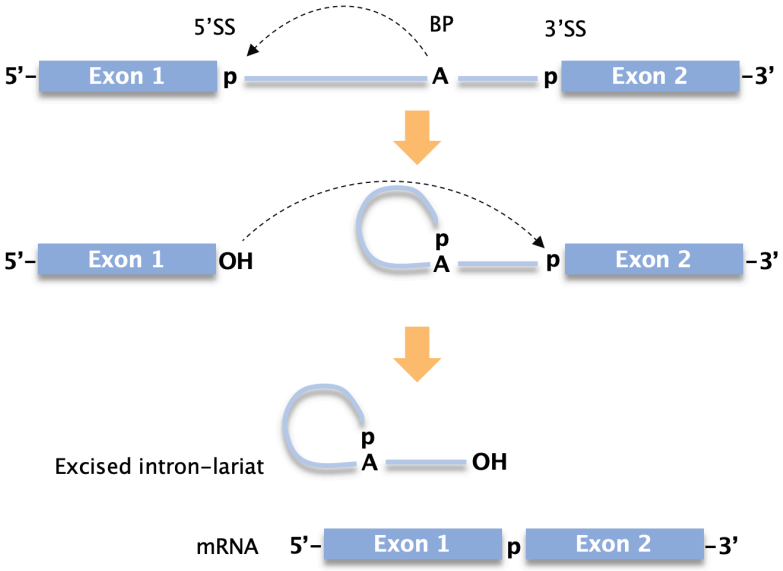


Figure 4. Scheme of intron excision and exon ligation via two consecutive trans-esterification reactions involving the 5′-splice site (SS), a branch point sequence (BP), and the 3′SS.

1.3.1. Regulation of the splicing process

The splicing process is a complex mechanism that requires recognition of RNA sequence elements by regulatory proteins. The cellular machinery responsible for catalyzing the processing and regulation of the splicing process is the **spliceosome**, which is an intricate ribonucleoprotein complex that recognizes specific sequences, both in exons and introns, defining the exact location of an intron. This cellular machinery responsible for mRNA splicing processing is highly dynamic, accurate and flexible (59).

In mammals, the splicing process is catalyzed by the **major and minor spliceosomes**, which act on different types of introns [U2 type introns (which constitute the 95% of all introns) and U12 type introns] (12,60,61). The functional core of both spliceosomes is comprised by several small ribonucleoproteins (snRNPs) subunits, which dynamically interacts to regulate the splicing process. Specifically, the major spliceosome consists of a nucleus composed by the assembly of 5 snRNP complexes (U1, U2, U4, U5 and U6), which are composed by small nuclear RNAs (snRNAs) that interact via base pairing with pre-mRNA (snRNAs: *RNU1*, *RNU2*, *RNU4*, *RNU5* and *RNU6*) and specific proteins. Similarly, the minor spliceosome is comprised by U11, U12, U4atac, U5 and U6atac (whose principal snRNAs are *RNU11*, *RNU12*, *RNU4ATAC*, *RNU5* and *RNU6ATAC*).

The definition of the different exons is mediated by three major sequence elements: the 5'SS, the 3'SS and the branch point. The spliceosome recognizes these elements and is assembled in a stepwise manner around the nascent pre-mRNA concomitantly with its synthesis by the RNA polymerase II in the nucleus. In addition, the activity of the spliceosome is modulated by more than 300 **splicing factors** that specifically recognize certain sequences in exons and introns (9,62-64). These splicing factors are continuously changing during the splicing process, modifying their position and activity in a coordinated orchestration and under complex regulatory systems that modulate the expression and function of these factors (59,65). These splicing factors regulate the splicing process through binding specific sequences in the pre-mRNA. The recognized sequences are classified as intronic splicing silencers or enhancers (ISS or ISE, respectively) and exonic splicing silencers or enhancer (ESS or ESE, respectively). The recognition of these sequences by the splicing factors has been broadly described (66-69). Finally, the activity and localization of these splicing factors can be modulated by kinases and phosphatases, influencing their functionality and ability to interact with other proteins and snRNPs and, thus, their capacity to bind RNA (61,70,71).

In particular, the identification of the splice sites is carried out by the spliceosome snRNPs in coordination with additional factors. Specifically, spliceosome recognition of the sequence elements at the 5′ and 3′ splice site and branch point is modulated by the ISE, ESE, ESS, which are recognized and bound by the splicing factors (72). Briefly, U1 snRNP form a base-pairing interaction with the 5′SS, while the splicing factor 1 (SF1) interact with the branch point and U2 auxiliary factor (U2AF) with the 3′SS (72,73). This complex, named *E* complex, results in the pre-spliceosomal *A* complex when U2 snRNP substitutes SF1 in the branch point. Then, U4, U5 and U6 form the pre-catalytic *B* complex associating with the formed spliceosome. When U4 is removed, U6 replaces U1 and interacts with U2 bringing the 5′SS and the branch point closer together, forming the catalytic C complex, and allowing the first reaction of trans-esterification (74). Finally, U5 facilitates the second trans-esterification step, approximating the two exons and allowing its junction and the removal of the intron lariat (75) (Figure 5B).

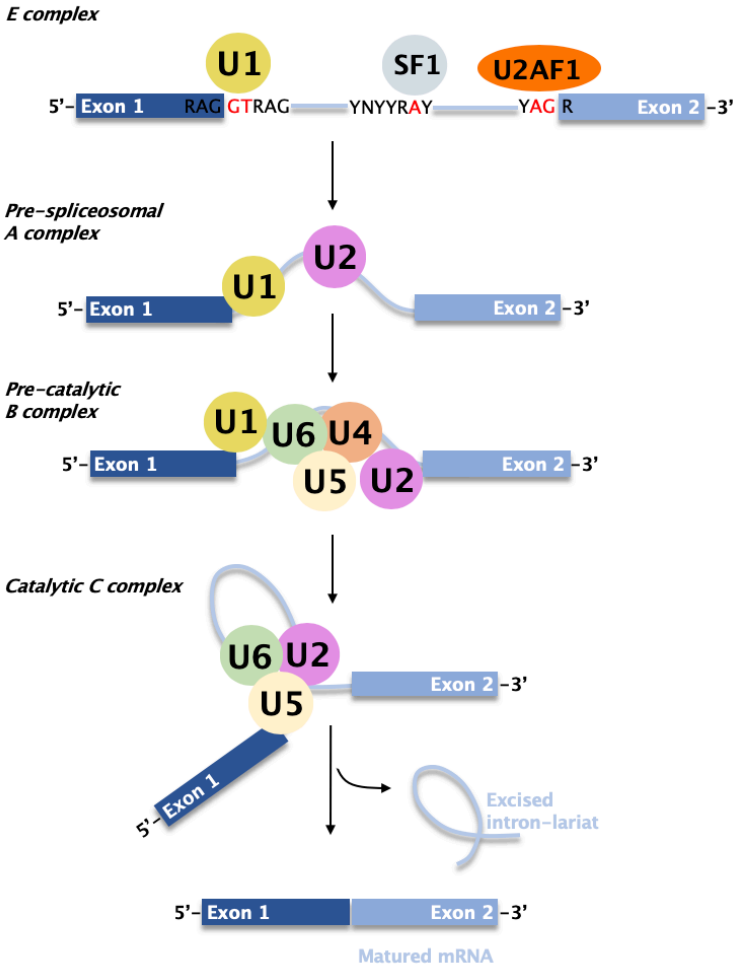


Figure 5. Schematic representation of intron identification by snRNPs and their interaction in order to carry out the splicing reaction.

1.3.2. Alternative splicing

Around 95% of human multi-exonic genes, as well as most genes in other higher eukaryotes, undergo physiological processes of alternative splicing, which largely increases the diversity of mRNAs expressed from the genome, and at the same time, shapes the repertoire of specific proteins present in a given cell (7,9,69). Indeed, alternative splicing of protein-coding genes is an essential regulatory mechanism that increases the complexity of the transcriptome and the diversity and function of the proteome, having a central role forming complex organism (69,76). Alternative splicing occurs by reorganizing the pattern of intron and exon elements that are combined together to generate a mature mRNA, which implies the alteration of the mRNA coding sequence and the generation of alternative protein species. Among the different types of alternative splicing processes that have been reported hitherto, the most frequent events are 1) **exon skipping** (also known as cassette exons), which consists on the elimination of the entire exon in the mature mRNA transcript and constitute the most common type of alternative splicing in vertebrates (10); 2) **intron retention,** which implies the retaining of an intron in the mature mRNA; 3) **mutually exclusive exons**, that occurs when different exon combinations are selected to generate different mRNA splice variants; and 4) **alternative 5′ (donor) or 3′ (acceptor) sites** that leads to changes in the mRNA sequence (10,77) (Figure 6).

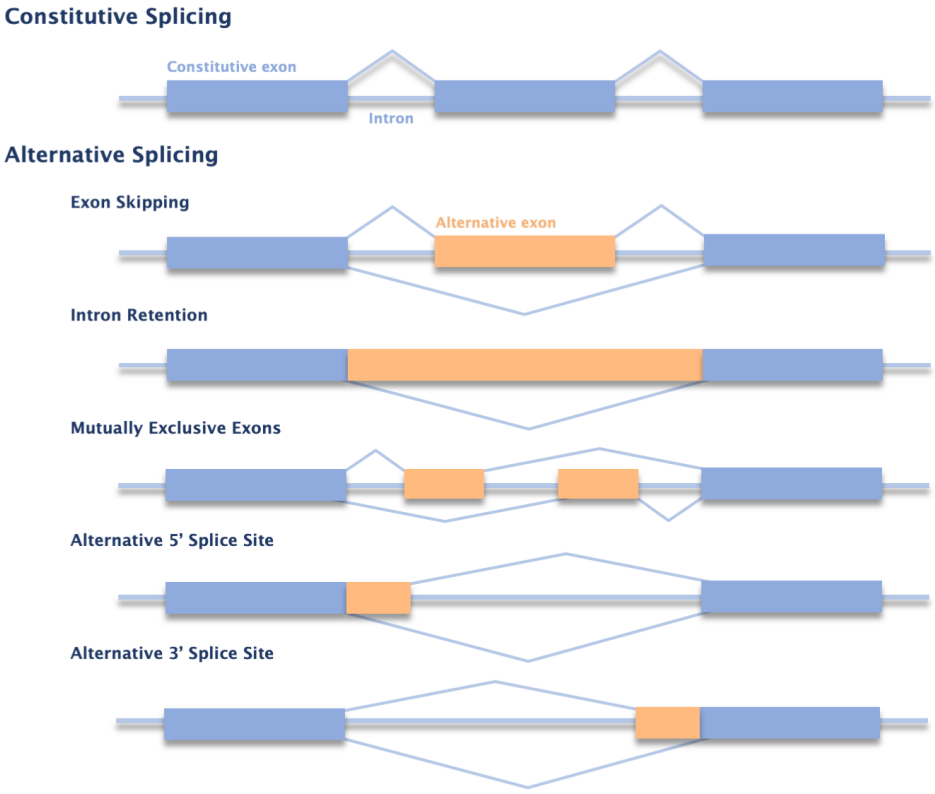


Figure 6. Scheme of constitutive and most common cases of alternative splicing events.

1.4. Alternative splicing in health and disease

Alternative splicing is a central mechanism in gene expression that increases the coding capacity of the human genome and impacts on the protein function. Therefore, the correct function of the splicing machinery (spliceosome components and splicing factors) is an essential mechanism to maintain normal cell physiology and whole-body homeostasis (11,41). Indeed, the splicing process may represent a physiologic mechanism to maintain cellular homeostasis as it has been suggested by different studies demonstrating that: 1) nutrients can modulate gene expression and, particularly, the splicing of pre-mRNAs encoding key regulatory proteins (e.g., insulin receptor, leptin receptor, etc.) (78); 2) alternative splicing processes are essential in the implementation of the adipogenic program (79); 3) a splicing variant of the peroxisome proliferator-activated receptor-γ (PPARγ), named PPARγ2, which is abundantly expressed in adipocytes, is essential in adipogenesis (80); 4) changes in the alternative splicing program during liver development due to alterations in the expression of splicing factors have different impact on the hepatic metabolism and development (6,81,82), and (5) alternative splicing is key for the regulation of pancreatic beta-cells differentiation and function, through the impact on glucose metabolism (11,83,84).

However, **although alternative splicing is a physiological process, a growing body of evidence supports the notion that aberrant alterations of this process can lead to the development and/or progression of human diseases**. Minimal alterations in the tightly regulated alternative splicing process can lead to the generation of deficient proteins resulting in organs failures and contributing to numerous human diseases (9). In this sense, the number of diseases known to be associated with aberrant alternative splicing has increased in the last decades. There are plenty of examples that illustrate how a dysregulation in this process is the underlying cause of many human diseases (10,61). Specifically, the vast majority of the studies are focused on the discovery and description of new aberrant splice variants of proteins that lead to an alteration on their function and, ultimately, to pathological conditions (12). Alterations in the alternative splicing process can cause diseases by different ways (61,85). Mutations in the regulatory sequences, branch point, 3′ and 5′ splice sites, that can impede the recognition of the RNA by splicing factors, have been described in several pathologies conditions like muscular dystrophy (86), cystic fibrosis (87) or Parkinson (88). Remarkably, mutations or aberrant expression of splicing factors are also implicated in numerous diseases including endocrine-related diseases and cancer (89).

1.4.1. Alternative splicing in endocrine-metabolic diseases

There is emerging evidence that under adverse endocrine-metabolic conditions (obesity, metabolic syndrome, insulin resistance or diabetes) the splicing process is dysregulated in a wide variety of tissues, such as pancreas, adipose tissue, liver, muscle or peripheral blood mononuclear cells (PBMCs) (5,6,90). While the precise molecular defects underlying these complications remain unknown, in the last years, studies exploring the impact that alternative splicing may exert on these pathologies are shedding light to this pathological crosstalk.

1.4.1.1. Dysregulations of alternative splicing in type 2 diabetes mellitus

T2DM is a multifactorial metabolic disease, characterized by a combination of progressive insulin resistance and loss of pancreatic β-cell function and/or mass. Insulin resistance arises when insulin receptor (INSR) signaling becomes attenuated in muscle, liver or adipose tissues (91). Defects in insulin signaling contribute to peripheral insulin resistance, including changes in INSR expression or affinity, as well as impaired phosphorylation of insulin receptor substrate (IRS) or glucose transporter translocation (92). The process of alternative splicing is a key modulator of insulin resistance, affecting the activity and sensitivity of INSR. Human INSR gene contains 22 exons and 21 introns (93). Alternative splicing of exon 11, which is responsible for the modulation of the affinity binding of insulin to the receptor (94), generates two isoforms of the INSR: isoform A (INSRA, without exon 11) or isoform B (INSRB, with exon 11) (93,95). The expression profile of these isoforms can be altered in pathological situations including diabetes (96-100). Specifically, alterations in relative abundance of the two isoforms in skeletal muscle, fat, pancreas and liver, might contribute to the development of insulin resistance in T2DM (101-104). Indeed, weight loss induced by bariatric surgery and a very low calorie diet regulates the alternative splicing of INSR in adipose tissue (105). Of interest, it has been found a correlation in the expression levels between INSR and certain splicing factors (91,105). Specifically, it has been described that insulin signaling can up-regulate the expression of the splicing factor SRSF1 in pancreatic beta cells, inducing the alternative splicing of the gene to generate the INSR-B (91). The same study also found a regulation of the protein levels of the splicing factor MBNL1 by high glucose levels.

In addition to INSR, other genetic variants generated by alternative splicing may be related to the development and progression of T2DM. One known example is the association between the alternative splicing of the transcription factor-7-like 2 gene (TCF7L2) and T2DM. Specifically, an association between a single nucleotide polymorphisms (SNPs), linked with a

higher risk of developing T2DM, and TCFL2 splicing was found in diabetic subject (106). Also, a short 3'-end splice variant of TCF7L2 is regulated by weight loss and associated with hyperglycemia and impaired insulin action in adipose tissue (107). Different adiponectin receptor 1 (ADIPOR1) transcripts have been identified in muscle of individuals without or with T2DM. In adult human muscle, the ratio between ADIPOR1T3, a splice variant generated by the insertion of a 67-base pairs exon between exons 1 and 2 in the 5´UTR, and ADIPOR1 is decreased in muscle of T2DM patients compared to those with a normal glucose tolerance (108). However, the role of the components of the cellular machinery responsible to regulate the processes of alternative splicing and their implication in the development and progression of T2DM is still to be fully defined.

1.4.1.2. Dysregulation of alternative splicing in obesity and non-alcoholic fatty liver disease

Obesity represents a global health problem with increasing incidence that has been largely reported to be associated to the dysregulations of the splicing of several genes linked to this pathology (109-112). Specifically, it has been found the existence of a mutation in intron 14 of the low-density lipoprotein (LDL)-receptor (LDLR) mRNA, which led to the discovery of a new cryptic splice site that occurred in patients with familial hypercholesterolemia (FH). In these patients, this new splice site consisting in a 81-bp insert in LDLR gene, encoded an in-frame insertion of new 27 aminoacids in the mutated protein, adding more complexity to the knowledge of LDLR and alternative splicing in the context of FH (113). Other known spliced variants in the context of obesity are the LPIN1 isoforms α and β (based on the skipping or inclusion of exon 7, respectively), which exhibit different roles in adipogenesis (109). Whereas LPIN1-α is necessary for adipocyte differentiation, the predominant effect of variant LPIN1-β is to induce lipogenic genes as well as adipocyte hypertrophy, leading to lipogenesis (109). In this context, it was described that the splicing factor TRA2B (also known as SRSF10) inhibits the inclusion of exon 7 in LPIN1 pre-mRNA promoting the generation of LPIN1α (114,115), and that the expression of TRA2B is reduced in liver and muscle from obese mice and humans, contributing to enhanced lipogenesis (116). In this line, and supporting the importance of the splicing process in the development of obesity, mice lacking the splicing factor KHDRBS1 (also known as SAM68) presented a lean phenotype due to a decreased adipogenesis and an increase in energy expenditure, showing protection against obesity, insulin resistance and glucose intolerance induced with a high fat diet (117,118). Interestingly, genome-wide analysis in adipose tissue of these mice revealed an alteration in the alternative splicing of mTOR, with reduced full-length mTOR expression and an increase of $mTOR_{i5}$ isoform that retains intron 5 and exhibit a shorter

transcript. This reduction implies a decrease in adipocyte cell size as well as a prevention in weight gain (119).

As expected, some splicing factors have also been previously related to liver development and metabolism, as it is the case of the polypyrimidine tract binding protein PTBP1 (HNRNPI), which has been described as a regulator of the splicing of Fatty acid desaturase 2 and 3 (FADS2 and FADS3), implicated in fatty acid elongation and desaturation (120), and multiple genes involved in cholesterol synthesis and uptake, including LDLR, MVK, HMGCS1, and PSCK9 (121). Also, studies with the splicing factor SND1 in rats have indicated that steatogenic conditions promote its action on low density lipid droplets (122). Furthermore, the splicing of the FADS3 has been observed to be modulated in the liver of baboons in response to different diets and in human liver HepG2 cells after treatment with polyunsaturated fatty acids (123).

In this scenario**,** several studies have demonstrated that NAFLD development and progression result from a combination of environmental and genetic factors (38,124-126). Many of these studies have performed transcriptome profiling by microarray in humans with NAFLD and assessed changes in gene expression; however, most studies did not address changes in RNA alternative splicing (37,127-129). Microarray analysis of 19 normal, 10 steatotic and 16 NASH human liver identified RNA splicing as a process that may play a role in NAFLD progression (124). Also, studies in diet-induced obesity and NASH mouse models have shown alterations in the expression of splicing factors. Moreover, NAFLD and NASH are associated with changes in the mRNA expression of certain splicing factors in the liver of obese patients (116,124,130,131). However, the possible contribution of these changes in the expression of RNA splicing factors to the pathophysiology of NAFLD and NASH has not been yet explored.

1.4.2. Alternative splicing in Cancer

Dysregulation of alternative splicing is frequently observed in cancer and has been described as an important factor in several types of tumor pathologies (132). In fact, aberrant splicing process is becoming recognized as one common hallmark for tumor development and progression (41,89,133). It has been postulated that cancer cells progress by developing mechanisms that facilitate the adaptation to their microenvironment and, in this sense, alternative splicing can provide this genetic plasticity that enable cancer cells to grow and progress through the generation of oncogenic isoforms. Numerous abnormal splicing patterns have been reported and associated with tumor biology, especially, an elevated number of alternative splicing

isoforms have been implicated in almost all aspects of cancer development (134,135). There are plenty of examples of proto-oncogenes, tumor suppressor genes, or genes related with apoptosis, cell cycle control, invasion, metastasis or angiogenesis, whose alternative splicing is altered, producing isoforms that favor the transformation process, the acquisition of invasive properties and, in some cases, confer resistance to chemotherapies (7,77,136,137). Of special importance in those cancer types with an endocrine component may be the aberrant splicing processes that may occur in elements of key regulatory endocrine-metabolic systems, such as the somatostatin system.

Somatostatin, a neuropeptide mainly produced in central nervous system and gastrointestinal tract, exerts a broad range of biological functions in endocrine and non-endocrine tissues, including inhibition of hormone secretions or cellular growth (43,138). Somatostatin binds to five G-protein coupled receptors (GPCRs) subtypes with 7-transmembrane-domains (7TMDs), named somatostatin receptors SST1-5, to modulate diverse downstream pathways (139). The five SSTs are encoded by five separated genes and exert distinct effects depending on the particular constellation of SSTs available on cell surface, as they can interact with themselves or with other receptors (140,141). Somatostatin binding to SST1-5 can trigger antitumor effects, inhibiting cell proliferation, angiogenesis and hormone secretion and/or inducing apoptosis (139,142). Consequently, somatostatin synthetic analogues (SSA), like octreotide and lanreotide, were developed and are used as medical therapy for different tumor pathologies including pituitary or NETs (139,142). Nevertheless, although SST1-5 are abundantly expressed in other cancers, such as prostate, breast, or hepatocellular carcinoma (HCC), their potential clinical value in these pathologies is still to be defined (143-145).

In this scenario, several pieces of evidence have demonstrated that the somatostatin system is more complex than originally envisioned. Indeed, recent studies have revealed the existence of diverse mechanisms that increase variability of GPCRs, including the SST1-5, as is the case of alternative splicing processes that could generate non-canonical truncated variants of different regulatory systems with less than 7TMDs (146-149). These truncated forms are functionally active by modulating the physiology of their canonical isoforms or by exerting separate, independent functions (148,149). Moreover, presence of these truncated receptors is commonly associated to development/progression of tumor pathologies, as is the case of the growth hormone-releasing hormone receptor (GHRHR) (146,147), cholecystokinin receptor (150) or adrenergic receptors (151). In this context, our group has also identified a non-canonical

truncated splicing variant of SST5 that only harbors 4TMDs and was therefore named SST5TMD4 (Figure 7). This splicing variant is generated by cryptic splice sites in the coding sequence and the distal, non-coding 3'UTR of the *SST5* gene (140). The SST5TMD4 is barely expressed in normal tissues but is overexpressed in different tumor pathologies as pituitary and NETs, as well as in breast, prostate, and thyroid cancer (48,140,152-156). More importantly, SST5TMD4 acts as an oncogene in these pathologies, wherein its presence is associated to resistance of SSA and with malignancy features, as it correlates with clinical parameters of aggressiveness and promotes cell proliferation, migration, invasion and exacerbated hormone secretion (48,140,152-157). Remarkably, a unique feature of this truncated SST5TMD4 receptor is the presence of 4TMDs, which implies that its C-terminal tail is exposed to the extracellular region (140,158). In this regard, an increasing number of studies suggest that extracellular fragments derived from shedding of plasma membrane receptors can play relevant functional roles in the development/progression of tumor pathologies and might, therefore, provide novel diagnostic/therapeutic tools for these pathologies (137,159-161). Therefore, since SST5TMD4 presents only 4 TMDs and its C-terminal tail is directed towards the extracellular matrix (ECM) instead of the cytoplasm (152,158), the extracellular region of SST5TMD4 may be susceptible to the action of proteases confined in the ECM such as metalloproteinases (MMPs), a group of zync- and calcium-dependent proteolytic enzymes capable to degrade the majority of ECM proteins, such as collagen and elastin, as well as to regulate the activity of other proteinases, growth factors and cell receptors (162,163).

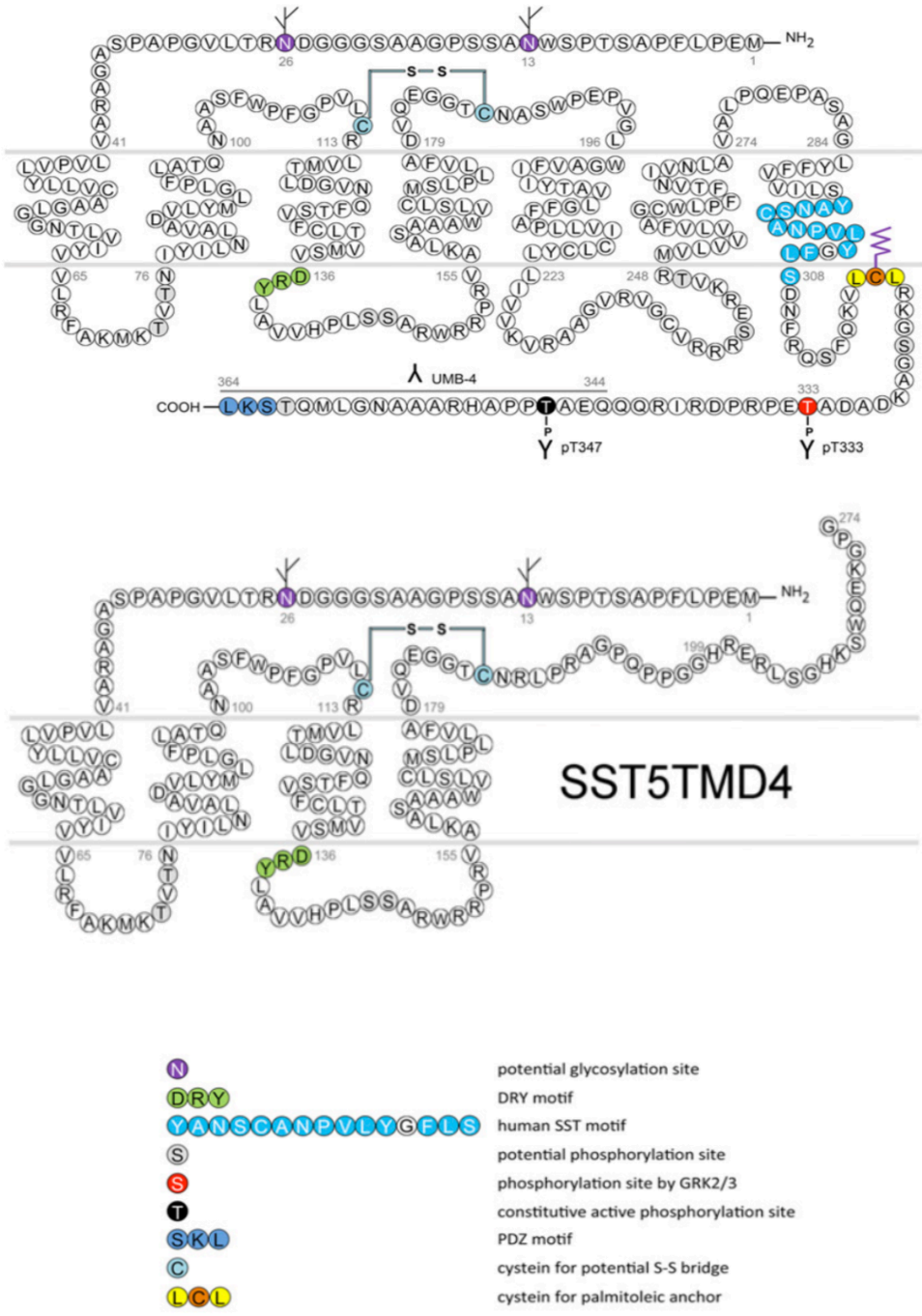


Figure 7. Structure of human SST5 (above) and its truncated variant SST5TMD4 (below). *Adapted from (164).*

2. Aims of the study

Based on the existing evidence mentioned until this point, the present Doctoral Thesis was based on the **HYPOTHESIS** that the alteration of the normal, physiologic splicing process could be closely implicated in the development and progression of endocrine-metabolic diseases (T2DM and NAFLD) and cancer. For this reason, the **GENERAL AIM** of this study was to explore the idea that the altered process of mRNA splicing (including alterations on the spliceosome components, splicing factors and surrogate splicing variants with pathological potential) could represent a novel source for the identification of diagnostic, prognostic and therapeutic targets in highly prevalent and health-threatening diseases such as endocrine-metabolic pathologies and cancer. To achieve this main aim, we have proposed the following **specific objectives**:

OBJECTIVE 1: To determine the relationship between the dysregulation of the splicing machinery components and the development of T2DM. This was pursued through the analysis of the expression pattern of selected spliceosome components and splicing factors in PBMCs, at fasting and post-prandial status, of individuals at high risk of T2DM development that were followed during 5 years and their correlation with the development of T2DM, as well as through the analysis of the *in vitro* modulation of the expression of selected spliceosome components and splicing factors by fasting and prost-prandial serum of these patients.

OBJECTIVE 2: To explore the effects of healthy dietary interventions (Mediterranean and Low-Fat diets) on the expression pattern of the splicing machinery in PBMCs from patients at high-risk of developing T2DM before and after 3 years of dietary intervention. The changes in the expression pattern of selected spliceosome components and splicing factors after the dietary interventions were analyzed considering the type of diet and the development or not of T2DM and were correlated with the medical evolution and clinical parameters of the patients.

OBJECTIVE 3: To determine the relationship between the dysregulation of the splicing machinery components and the development of NAFLD. To do that, a novel and comprehensive approach to identify alterations in the expression pattern of spliceosome components and splicing factors that could be associated to the development of hepatic steatosis was implemented in liver biopsies from obese patients with different degrees of liver steatosis and correlated with clinical parameters. Furthermore, the consequences of the modulation of the expression of

selected spliceosome components and splicing factors on hepatic liver accumulation and *de novo* lipogenesis was explored *in vitro*.

OBJECTIVE 4: To further expand our knowledge on the pathological interplay between aberrant splicing processes and cancer, through a more profound characterization of the oncogenic role of the truncated receptor SST5TMD4 on endocrine-related cancers. Specifically, the functional and mechanistic consequences of the SST5TMD4-specific, splicing derived C-terminal domain peptides were evaluated in different cancer cells types.

3. Results and general discussion

The results generated in the present Doctoral Thesis and their discussion have been summarized in this section and structured in four chapters, corresponding to each objective of the Thesis described above and included in four scientific manuscripts directly derived from this Doctoral Thesis [3 accepted manuscripts in top ranking journals of the "Medicine, Research & Experimental", "Endocrinology & Metabolism" and "Medical Laboratory Technology" fields, and a fourth manuscript that is under review in a specialized journal], which can be found at the end of the Thesis.

3.1. Changes in Splicing Machinery Components Influence, Precede, and Early Predict the development of Type 2 Diabetes: From the CORDIOPREV Study (Article I)

In this study, we aimed to determine the relationship between the dysregulation of the splicing machinery components and the development of T2DM through the analysis of the expression pattern of selected splicing machinery components in the PBMCs of patients at high risk to develop T2DM due to the presence of cardiovascular disease (CVD), using a microfluidic custom-made qPCR-based array. PBMCs were used since gene expression patterns in these cells commonly reflect and accompany disease-characteristic expression patterns (165) and may serve as a general sentinel, biosensor and early indicator of the instauration of metabolic disease. Patients included in this study were followed-up during 5-years [individuals that were included in the CORDIOPREV study (166)]. All the patients that developed T2DM during the 5-years follow-up (Incident-T2DM; n=107) and randomly-selected matched controls (non-T2DM; n=108) were included in the study. This approach revealed that the expression levels of key splicing machinery components (e.g. RNU2, RNU4 or RNU12) were significantly altered in fasting and postprandial PMBCs from Incident-T2DM compared to non-T2DM controls at the inclusion of the study (before the development of T2DM) and associated with T2DM development. Indeed, the molecular fingerprints comprised by the fasting and the postprandial levels of certain of these splicing machinery components were capable to predict the future development of T2DM in individual patients with high precision (AUC=0.881, TPR=0.801, TNR=0.700), after cross-validation analysis, which even outperforms the capacity of classical predictors of T2DM development, such as HbA1c or FINDRISK (AUCs<0.66 in our cohort). Therefore, this splicing machinery-associated molecular fingerprint could become a valuable, non-invasive, new tool for early risk assessment of T2DM in clinical practice in order to prevent disease development. In addition, patients with low PBMCs expression levels of specific splicing machinery components, including *RNU2*, *RNU4* or *RNU12*, which we have shown to be modulated by the serum of

Incident-T2DM patients, were at higher risk (OR>4) of T2DM development as compared with those patients with high PBMC levels of these components.

In addition, this study demonstrates that the expression of some splicing machinery components is altered during the postprandial phase, which is consistent with previous results demonstrating a regulatory response of specific splicing variants to the postprandial environment (167,168). This suggests that the changes observed in the splicing machinery during the postprandial phase may be responsible for the regulation of the expression of particular splicing variants under these conditions (90,100,104), which could be essential to the appropriate response of the organism to metabolic challenges and disturbances. Moreover, our results also show that the response of key splicing machinery components to metabolic insults is altered in individuals who will develop T2DM (Incident-T2DM patients), but especially in those developing T2DM during the first two years of follow-up. Therefore, since postprandial alterations are closely related to the phenotypic flexibility, which is strongly linked to T2DM development (26), these data primarily demonstrate that the alteration in the splicing machinery precedes the instauration of T2DM, thereby suggesting its putative implication as a driving force in the development of this pathology.

Based on all the information mentioned above, it is tempting to propose that the splicing machinery could be acting as a biosensor of the whole body metabolism to adapt cell gene expression to the pathophysiological conditions, and that its dysregulation could lead to an unbalance in the landscape of splicing variants present in a given cell at a given moment (90,100,169), which may be associated to the instauration of T2DM (5,170). This idea is further supported by two pieces of evidence presented herein. First, this study demonstrates that low fasting or postprandial expression levels of certain splicing machinery components drastically increase the relative risk of T2DM development in these patients, suggesting that a dysregulated expression of certain splicing machinery components could augment the risk of developing T2DM. Secondly, *in vitro* studies demonstrate that the expression of relevant spliceosome components, specially RNU2, RNU4 or RNU12, which are key elements responsible for the appropriate function of the spliceosome (171), can be modulated by baseline fasting and/or postprandial serum from Incident-T2DM patients. This observation might suggest the existence of specific factors in the serum of these patients capable to modulate the expression of relevant spliceosome components and, therefore, the function of the splicing machinery. In this sense, previous studies have found a relationship between circulating factors and the modulation of

splicing factors in different tissues. For example, it has been described that insulin signaling can up-regulate the expression of the splicing factor SRSF1 in pancreatic beta cells, inducing the splicing of the insulin receptor to generate the INSR-B isoform (91). The same study also found a regulation of the protein levels of the splicing factor MBNL1 by high glucose levels. Furthermore, the splicing of the Fatty acid desaturase 3 has been observed to be modulated in the liver of baboons in response to different diets and in human liver HepG2 cells after treatment with polyunsaturated fatty acids (123).

Therefore, this study strongly supports the notion that altered expression of spliceosome components and splicing factors may be associated with the development of T2DM, preceding the clinical instauration of this pathology and, therefore, could likely serve as a sensor and early predictor for T2DM development in CVD patients.

3.2. Dietary intervention modulates the expression of the splicing machinery in patients at high risk of type 2 diabetes development: from the CORDIOPREV study (Article II)

This study, also conducted within the CORDIOPREV study framework, had the objective to evaluate the influence of healthy dietary interventions in the expression pattern of the splicing machinery components. Specifically, we evaluated the effect of the consumption of two healthy diets [Mediterranean (MedDiet) or low-fat (LF) diet] in the PBMCs from the same 215 patients included in the previous study (107 patients that developed T2DM during the 5-years follow-up and 108 randomly-selected matched controls), at baseline and after three years of dietary consumption. Specifically, our study provides primary evidence that a dietary intervention can distinctly alter the expression pattern of the splicing machinery, both spliceosome components and splicing factors, in humans at risk of T2DM. In particular, the results demonstrate that the consumption of two healthy diets (MedDiet and LF diet) during three years can modulate the expression pattern of key spliceosome components and splicing factors in PBMCs from the patients enrolled in the CORDIOPREV study, including the overexpression of some molecular components, like *SPFQ*, *RBM45*, *RNU6*, etc. and the downregulation of others, including *RNU2* and *SRSF6*. Interestingly, some of the changes observed in the expression levels of certain splicing machinery components were closely associated with relevant clinical features, as is the case of the increase in the expression levels of the splicing factor *SPFQ*, which was inversely correlated with the decrease in HOMA-IR and HIRI indexes observed in the population. The finding of a diet-related long-term modulation of the expression of the splicing machinery components could

represent a novel valuable piece of information for two reasons. First, because it unveils that the splicing process may represent an adaptive mechanism in response to different nutritional conditions, and that this mechanism could be in place not only in circulating PBMCs but may also operate in cell types from other tissues and organs tightly coupled to nutrient-dependent metabolic homeostasis (e.g. liver, pancreas, adipose tissue), an avenue that is indeed worth exploring. Actually, we and other have already found the delicate and important role that the regulation of the splicing machinery can play in those organs (6,105,172-174). Secondly, inasmuch as PBMCs can be an accessible and suitable sentinel to detect relevant changes related to nutrient- and diet-dependent metabolic homeostasis, our current results support the idea that changes in the expression of key splicing machinery components could provide a fine screening marker for the development or progression of T2DM and their diet-related dynamics. Indeed, within the CORDIOPREV study, the long-term intake of a MedDiet, rich in olive oil, or a LF diet improved insulin sensitivity and beta-cell function (175), and therefore the increase in the expression of specific splicing factors found herein under both diets, and their inverse correlation with insulin resistance indexes, strongly suggest that the molecular changes might be related to the beneficial consequence of the healthy diet consumption and, therefore, that they could represent a novel mechanism linking healthy dietary intervention and the improvement the metabolic status of the patients and the protection from cardiovascular complications. Given the very scarce information available on the functional roles and implications of many of the molecules identified in the present study to be altered in PBMCs (e.g. *SPFQ*, *RBM45*, *RNU6*, etc.) the present findings open novel, unexplored avenues in this field of research.

One of the findings from this study that we consider most noteworthy is that the diet-induced alterations in the splicing machinery of PBMCs was independent of the type of healthy diet in which CORDIOPREV participants were enrolled (MedDiet or LF Diet), except for three splicing factors (*SNW1, SPFQ* and *NOVA1*) that showed a more pronounced modulation in patients under the LF Diet. To date, and to the best of our knowledge, nothing has been reported regarding the influence of diet intervention in the modulation of the expression of *SNW1*, *SPFQ* and *NOVA1*. However, some of these factors have been described to contribute to the alternative splicing of key genes whose splicing processing changes in response to a fatty diet (174). Previously, several studies have shown that PBMCs gene expression pattern is influenced by the diet (176-179) and that this might reflect changes related to both metabolic and immune responses (180,181). In addition, it has been demonstrated that the splicing process of key regulatory proteins for metabolic homeostasis, like the receptors for insulin or leptin, can be markedly

influenced by nutrient metabolism, directly or indirectly (78,123). Thus, it seems reasonable to think that those splicing-related changes would rely on upstream changes in the function of the machinery responsible for generating the splice variants. However, little or nothing is known in this regard in PBMCs, for there are no reports on how diet can influence the expression of the components of the spliceosome and the splicing factors, which altogether are responsible of the modulation of the splicing process. Nevertheless, in this context, some studies have highlighted that nutritional status can induce changes in the activity of serin-arginine (SR) proteins, an important family of splicing factors, further supporting the contention that different nutrients may be able to modulate the expression of metabolic genes at the level of its splicing processing. Specifically, it has been described that insulin signaling can up-regulate the expression of the splicing factor SRSF1 in pancreatic beta cells, inducing the splicing of the insulin receptor to generate the INSR-B isoform (91). The same study also found a regulation of the protein levels of the splicing factor MBNL1 by high glucose levels. In addition, other splicing factors belonging to the serin-arginine (SR) proteins family, SRSF2, is decreased under Vitamin E-deficient diet in the liver (182). Thus, although still limited, growing evidence, included the results from this study, points to a link between diet and nutrient and regulation of the splicing process, including its underlying operating machinery.

Another intriguing implication of our present results relates to the predictive capacity of studying changes in the splicing machinery in at-risk patients. To be more specific, nutrient-induced changes in specific splicing machinery components may provide hints on the predictive potential and possible functional correlation of key molecules, which had not been explored hitherto in this regard. Thus, within this study, regardless of the type of dietary intervention, the expression of some of the splicing factors studied was differentially altered in patients that develop T2DM after the 5 years of the study compared to non-T2DM subjects. For example, *RNU12*, a component of the minor spliceosome, showed a significant increase after the 3 years of dietary intervention in non-T2DM patients. Interestingly, we have previously described that the expression of this small nuclear RNA (snRNA), which is essential to form U12 snRNP and carry out the appropriate splicing of type 12 introns (183), was lower, at baseline (inclusion of the study), in Incident-T2DM compared to non-T2DM patients and that was associated with the risk of T2DM development (172). Therefore, since lower expression levels of RNU12 were associated with higher risk of T2DM, the dietary-induced increase in the expression of this component in non-T2DM may be associated to the protective effects of the healthy dietary consumption. Furthermore, in the same study, 4 hours incubation with baseline postprandial serum from

Incident-T2DM patients induced a significant reduction of *RNU12* expression compared to non-T2DM treated PBMCs from healthy patients (172). Therefore, the modulation in the expression of these spliceosomal components may represent a link between the dietary intervention and the beneficial effects on the metabolic status of the patient. Remarkably, the difference in *RNU12* expression, at year 3, between incident and non-T2DM patients, was more pronounced under LF Diet. Thus, although the possible mechanisms linking nutrient-induced changes in the splicing machinery, their functional consequences and the regulatory implications thereof are still to be fully elucidated, our present study provide suggestive evidence that it is worth exploring both the mechanistic/functional and the predictive components of this plausible link, for it may provide original, valuable biological knowledge as well as practical information for the patients.

Altogether, this study reveals that expression of the splicing machinery components in PBMCs from patients at risk of T2DM can be notably and selectively influenced by long-term dietary intervention; also, that the two dietary interventions tested herein, MedDiet and LF Diet, induced remarkably similar changes on the expression of spliceosome components; and finally, that there are distinct, diet type-induced changes in PBMCs from both non-T2DM and incident-T2DM patients, that may have an as yet unknown functional significance. Therefore, we propose that the machinery that controls and performs the alternative splicing process and is consequently responsible for changes in the pattern of functionally and pathologically relevant splice variants in the regulation of metabolic homeostasis is a plausible target to be operated by dietary intervention. As such, our results pave the way to explore in experimental models the possible mechanistic role and relevance of the splicing machinery and its components in diet-related metabolic regulation, and the investigate the value of screening changes in specific splicing machinery components to monitor and early predict relevant diet-related changes in patients at risk of T2DM.

3.3. Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease (Article III)

The present study aimed to determine the relationship between the dysregulation of the splicing machinery components and the development of NAFLD, which represents a novel and comprehensive approximation for understanding the molecular dysregulations underlying the development of NAFLD in obese patients as compared to non-steatotic obese patients. Considering that hepatic steatosis is the first and reversible stage of NAFLD, improving the

understanding of the pathogenesis of liver steatosis would have a critical impact for preventing NAFLD progression (38,39). In this study, 41 liver biopsies from non-alcoholic obese individuals with or without hepatic steatosis that underwent bariatric surgery were collected and the expression pattern of splicing machinery components was determined using a microfluidic custom-made qPCR-based array. This experimental approach has shown that the liver of steatotic obese patients exhibit an overt but differential (patient-dependent) alteration of the cellular machinery responsible for the regulation of the splicing process (spliceosome components and splicing factors). Specifically, an alteration in a representative set of splicing factors (*PTBP1*, *SRRM1*, *SND1* or *SRSF2*) have been shown in the liver of obese steatotic patients. Some of these altered splicing factors have been previously related to liver development and metabolism, as it is the case of the polypyrimidine tract binding protein *PTBP1* (*HNRNPI*), which has been shown herein to be increased in the liver of steatotic patients, and has been reported to regulate the splicing of *FADS2* and *FADS*3, which are implicated in fatty acid elongation and desaturation (120), and multiple genes involved in cholesterol synthesis and uptake including *LDLR*, *MVK*, *HMGCS1*, and *PSCK9* (121). Also, studies with the splicing factor *SND1* in rats have indicated that steatogenic conditions promote its action on low density lipid droplets (122). However, it is also described herein a dysregulation in some elements of the spliceosome core (*RNU2*, *RNU6*, *SF3B1, RNU6ATAC* or *RNU4ATAC*), which are essential for the appropriate recognition of intron sequences and the assembly of spliceosome to initiate the splicing process (184). These results show that a spliceosome-associated molecular fingerprint is able to discriminate between steatotic and non-steatotic livers in obese patients (with AUC>0.96).

Remarkably, the expression pattern of the spliceosome components and splicing factors analyzed herein was not associated with the grade of hepatic steatosis but seemed to identify three discrete molecularly defined subpopulations of steatotic obese patients characterized by the dysregulated expression of certain spliceosome components and splicing factors (Clusters A, B and C). Interestingly, these clusters were associated to specific clinical characteristics and also with a differential response to bariatric surgery after one year. In particular, patients included in the Cluster A, which were mainly characterized by altered expression of *SRSF4*, *RBM22* and *TRA2B* compared to non-steatotic and Clusters B and C steatotic patients, presented an increase in glucose plasma levels, indicating a possible relationship between this splicing factors and glucose homeostasis. Indeed, one of the splicing factors dysregulated in this cluster of patients (*TRA2B*, also named *SRSF10*) has been previously described in relation to obesity, wherein *TRA2B* seems to be downregulated in liver samples from insulin resistant humans with obesity

(130). Similarly, patients in Cluster B were characterized by altered hepatic expression of *RBM45* and *TRA2A* and presented higher triglycerides and GGT levels at the diagnosis. This represents a very interesting finding inasmuch as hepatic *TRA2A* has been found to be modulated in response to estrogens in order to control the alternative splicing of class B scavenger receptors BI (SR-BI) and BII (SR-BII) (185), which are crucial player in the hepatic uptake of triglycerides (186). Finally, obese patients included in Cluster C presented elevated levels of insulin and aminotransferases and a poor evolution of biochemical parameters and of the recovery in the grade of steatosis after bariatric surgery. Interestingly, *SND1*, one of the splicing factors whose altered expression characterized Cluster C, is a splicing factor associated to the physiological function and pathological transformation of the liver (122,187). Indeed, as it has been mentioned in the previous articles, hepatic *SND1* has been shown to be associated to different aspects of lipid metabolism in the liver and with the development of hepatic steatosis (122).

Furthermore, the fact that these spliceosome alterations were not associated with the grade of hepatic steatosis, suggest their role as triggers for initiating fat deposition within the liver. Indeed, the *in vitro* results suggest that alterations in splicing machinery components could precede the development of hepatic steatosis in obese patients since the modulation of the expression of *PTBP1*, *SRSF4*, *RBM22*, *RBM45*, *SND1* and *RAVER1* lowered lipid accumulation in HepG2 cells, while an exogenous overload of oleic acid (OA) was not able to alter the expression of the mentioned spliceosome components and/or splicing factors. Remarkably, these results also indicate that the dysregulations in the expression of certain splicing machinery elements could have an impact on other key hepatic processes, inasmuch as silencing of *RBM22* and *SND1* decreased the expression of *FASN* and *SCD* enzymes, indicating a possible connection with *de novo* lipogenesis in the liver. In fact, it has been described that the inhibition of *SRPK2*, a regulator of the splicing factors RNA-binding SR proteins, results in intron retention and mRNA instability of lipogenic genes (188). It is also worth noticing that, although OA treatment did not alter the expression of these splicing machinery components, other metabolic factors, such as glucose, leptin, IGF1 or PA, modulated the expression of some of these splicing factors. In this sense, several studies have previously shown a regulation of specific splicing factors by metabolic factors as it has been discussed in the above-described articles.

Therefore, this study demonstrates a novel relationship between the dysregulation of splicing machinery and the development of NAFLD and its associated metabolic co-morbidities, as well as with the biochemical improvement after bariatric surgery. These findings shed light to

the possible underlying molecular mechanisms responsible for the development of hepatic steatosis in obese patients and, thus, provide novel information to explore the development of efficient screening strategies, diagnostic, prognostic or therapeutic tools for obesity-related NAFLD, which is becoming one of the main source of morbidity and mortality in developed countries.

3.4. Peptides derived from the extracellular domain of the somatostatin receptor splicing variant SST5TMD4 increase malignancy in multiple cancer cell types (Article IV)

This study aimed to further expand our knowledge on the pathological interplay between aberrant splicing processes and cancer, through a more profound characterization of the oncogenic role of the truncated receptor SST5TMD4 on endocrine-related cancers. SST5TMD4 constitutes an example of the oncogenic consequences of aberrant splicing processes since this spliced truncated receptor is markedly overexpressed in several tumors, wherein its presence is associated to higher aggressiveness and/or resistance to medical treatment with SSAs (48,140,152-155,157). Like other truncated receptors (146-148), it exhibits a preferential intracellular localization and disrupts the function of other SSTs, mainly SST2 and SST5 (140), and inhibits the ability of MCF-7 and SST2-transfected CHO-K1 cells to respond to somatostatin/SSAs (152). However, a substantial proportion of SST5TMD4 (140) also resides at the cell membrane, wherein it acts as a functional receptor capable to mediate ligand-induced intracellular responses to somatostatin and cortistatin (140,155,157). This unique location at the cell membrane, together with its distinctive feature of harboring only 4 TMDs, confers SST5TMD4 one of its most relevant particularities: its C-terminal tail is exposed towards the ECM, rendering this extracellular region as a potential substrate for MMPs, which are tightly associated to cancer angiogenesis, invasiveness and metastasis (163).

In this study, we performed a detailed *in silico* analysis that revealed the existence of two putative cleavage sites in the unique SST5TMD4 C-terminal extracellular domain, which could be used by MMP2, 9 and 14 and/or MMP16, respectively, to generate three derived peptides with 7, 10 and 17 aminoacids. MMPs are able to degrade the majority of ECM proteins, including cell receptors (162,163), and have been functionally linked to the aggressiveness of different tumors types (163,189-192). These tumor-promoting features of MMPs may be, at least in part, mediated by the processing of key receptors present at the cellular surface, which can generate receptor-derived peptides with oncogenic capacity (137,160,161).

We next demonstrated that SST5TMD4-derived peptides are capable to enhance the malignant characteristics of multiple cancer cells derived from diverse tumor pathologies as NETs, breast, prostate and liver cancer; however, some of these actions seem to exhibit different dynamics (i.e. proliferation rate) or, even, be cell-line dependent (i.e. migration or tumorosphere formation). Certainly, SST5TMD4-derived peptides were able to drastically increase the capacity of the BON-1 cells to proliferate, migrate, and induce tumorsphere formation; while in the QGP-1 cells, SST5TMD4-derived peptides clearly increased tumorsphere formation and hampered the cellular response to somatostatin, but exerted lesser effects on proliferation and had no effect on migration rate. In this sense, SST5TMD4 has been previously found overexpressed in gasteroenteropancreatic-NETs (GEP-NETs) and associated to aggressiveness, metastasis and worse prognosis. Consistent with the present results, SST5TMD4 transfected BON-1 cells showed an increase in proliferation rate (156), whereas SST5TMD4-transfected QGP-1 did not (156). These differences could relate to the distinct nature of the two NET cell models (156,193), since QGP-1 cells have constitutive high expression of somatostatin, which imparts a constant inhibition pattern that might hinder the stimulatory action of the SST5TMD4-derived peptides. Similar divergences were found in breast and prostate cancer cell models. Thus, whereas MDA-MB-231 cells exhibited an increase in cell proliferation, migration and tumorspheres formation in response to the SST5TMD4-derived peptides, in MCF-7, a stimulatory effect of the three peptides was only found in cell migration capacity and tumorspheres formation, these differences being probably attributable to the distinct origin and aggressiveness of each cell line (194). Consistent with these results, it was previously observed that forced overexpression of SST5TMD4 in MCF-7 and MDA-MB-231 cells increased proliferation, cell invasion and/or tumorsphere formation capacity (152,155). In the case of prostate cancer cells, the bone metastasis-derived PC-3 cell line showed increased cell proliferation and migration in response to the treatment with the SST5TMD4-derived peptides. In contrast, in 22Rv1 cells, a cell line more representative of the early disease, exhibited a clear stimulatory effect of the three peptides only on cell proliferation (195,196). It should be noted that SST5TMD4 is overexpressed in prostate cancer, especially in metastatic disease, and has been shown to hamper the response to SSAs (48). Indeed, SST5TMD4 has been shown to play a functional role in prostate cancer cells, wherein its overexpression enhanced cell proliferation, migration and promoted tumor growth *in vivo* (48). Of note, this study describes, for the first time, an increase in cell proliferation and migration in the HCC cell line SNU-387 in response to SST5TMD4-derived peptides. These results open a new research avenue in the study of this pathology in that, although the presence and putative functional role of the SST5TMD4 has not yet been investigated in detail, it has been reported that HCC could

present resistance to SSAs (143,145), an important therapeutic limitation that has been shown to be associated, in several instances, with the presence of the SST5TMD4 (48,153,157). Moreover, these data on HCC cells demonstrate that the malignancy-prone effects of the SST5TMD4-derived peptides could have a wide range of action across different cancer types.

SST5TMD4-derived peptides could exert their function through the activation of different cancer-relevant signalling pathways (197-199), such as PI3K/AKT, MEK/ERK, but not JNK, in different cancer cell lines. In line with this, it has been also previously reported that SST5TMD4 exerts its functions via modulation of several pathways, including activation of PI3K/AKT (152) and MEK/ERK (48,152), two signalling pathways associated to malignancy promotion in several tumor pathologies (197-199). Interestingly, the three SST5TMD4-derived peptides induced a clear upregulation of the proliferation markers *CCND3* and *MKI67*, also used to assess the grade and differentiation of NETs (200), the *MMP2*, involve in ECM degradation (163) and likely in the own production of SST5TMD4-derived peptides, and *ACTR2/3* complex, closely related to cell migration and invasion (201). Furthermore, two well-accepted stem cells surface markers *CD24* and *CD44* were also increased in BON-1 cells in response to the three SST5TMD4-derived peptides.

Altogether, this study suggests that peptides derived from the spliced SST5TMD4 receptor could contribute to the strong oncogenic role of SST5TMD4 observed in multiple tumor pathologies, and, therefore, represent potential candidates to identify novel diagnostic, prognostic or therapeutic targets in cancer.

4. General conclusion

The results presented in this Doctoral Thesis constitute novel, valuable and germane information supporting the contention that the dysregulation of the splicing process, including the alterations in the spliceosome components, the splicing factors or the surrogate markers of the splicing process, is closely related to the instauration and/or development of several and diverse pathologies such as diabetes, NAFLD or cancer. The main conclusions associated to each objective/article included in the Thesis are described below:

Article I

1) The expression pattern of selected spliceosome components and splicing factors is altered in the PBMCs of high-risk patients with CVD that develop T2DM compared to non-T2DM control patients. This alteration may be associated with the development of T2DM, preceding the clinical instauration of this pathology.
2) The expression pattern of these splicing machinery components can be dynamically modulated during the post-prandial phase.
3) The existence of a spliceosome-associated molecular fingerprint is capable to early predict the future development of T2DM in individual patients with CVD, with high precision, outperforming the capacity of classical predictors of T2DM development. This molecular fingerprint could become a valuable, non-invasive, new tool for early risk assessment of T2DM in clinical practice to prevent disease development.

Article II

4) The expression pattern of the splicing machinery components in PBMCs from patients at risk of T2DM can be notably and selectively influenced by long-term dietary intervention.
5) The two dietary interventions tested herein, Mediterranean Diet and Low Fat Diet, induced remarkably similar changes on the expression of spliceosome components.
6) There are distinct, diet type-induced changes in PBMCs from both non-T2DM and incident-T2DM patients, that may have an as yet unknown functional significance.

Article III

7) The expression pattern of selected spliceosome components and splicing factors is dysregulated in the liver of steatotic obese patients compared to non-steatotic patients.
8) This alteration is not associated with the grade of hepatic steatosis but seems to identify three discrete molecularly defined subpopulations of steatotic obese patients

characterized by the dysregulated expression of certain spliceosome components and splicing factors and associated with specific metabolic co-morbidities, as well as with the biochemical improvement after bariatric surgery.

9) The dysregulation in the expression of the splicing machinery components could play a role in the initiation of fat deposition within the liver, wherein these alterations in splicing machinery components could precede the development of hepatic steatosis, shedding light to the possible underlying molecular mechanisms responsible for the development of hepatic steatosis in obese patients.

Article IV

10) The extracellular SST5TMD4 C-terminal domain presents two MMPs cleavage sites that could lead to the release of three derived peptides.
11) The three SST5TMD4-derived peptides are capable to enhance the malignant characteristics (proliferation, migration and tumorspheres formation) of cancer cells derived from diverse tumor pathologies (neuroendocrine tumors, breast, prostate and liver cancer), and blunted the anti-proliferative response to somatostatin in QGP-1 cells.
12) The three SST5TMD4-derived peptides could increase malignancy likely through the activation of PI3K/AKT and/or of MEK/ERK pathways and by the modulation key pro-oncogenic genes.
13) The peptides derived from the SST5TMD4 extracellular domain could have important biological activities and pathological implications since they could contribute to the strong oncogenic role of SST5TMD4 previously reported in multiple tumor pathologies.

Altogether, these results demonstrate that the alteration of the splicing process could contribute to the development and pathogenesis of different endocrine-metabolic and tumor pathologies and also provides novel information to explore the development of efficient screening strategies, diagnostic, prognostic and potential therapeutic targets that could be of potential utility in clinical practice.

5. References

1. Smith KB, Smith MS. Obesity Statistics. *Prim Care*. 2016;43(1):121-135, ix.
2. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract*. 2018;138:271-281.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34.
4. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-724.
5. Dlamini Z, Mokoena F, Hull R. Abnormalities in alternative splicing in diabetes: therapeutic targets. *J Mol Endocrinol*. 2017;59(2):R93-R107.
6. Webster NJG. Alternative RNA Splicing in the Pathogenesis of Liver Disease. *Front Endocrinol (Lausanne)*. 2017;8:133.
7. El Marabti E, Younis I. The Cancer Spliceome: Reprograming of Alternative Splicing in Cancer. *Front Mol Biosci*. 2018;5:80.
8. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*. 2013;497(7447):127-131.
9. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*. 2014;15(2):108-121.
10. Kim HK, Pham MHC, Ko KS, Rhee BD, Han J. Alternative splicing isoforms in health and disease. *Pflugers Arch*. 2018;470(7):995-1016.
11. Juan-Mateu J, Villate O, Eizirik DL. Mechanisms in Endocrinology: Alternative splicing. The new frontier in diabetes research. *Eur J Endocrinol*. 2016;174(5):R225-R238.
12. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19-32.
13. Apovian CM. Obesity: definition, comorbidities, causes, and burden. *Am J Manag Care*. 2016;22(7 Suppl):s176-185.
14. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes (Lond)*. 2008;32(9):1431-1437.
15. Collaborators GBDO, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, Marczak L, Mokdad AH, Moradi-Lakeh M, Naghavi M, Salama JS, Vos T, Abate KH, Abbafati C, Ahmed MB, Al-Aly Z, Alkerwi A, Al-Raddadi R, Amare AT, Amberbir A, Amegah AK, Amini E, Amrock SM, Anjana RM, Arnlov J, Asayesh H, Banerjee A, Barac A, Baye E, Bennett DA, Beyene AS, Biadgilign S, Biryukov S, Bjertness E, Boneya DJ, Campos-Nonato I, Carrero JJ, Cecilio P, Cercy K, Ciobanu LG, Cornaby L, Damtew SA, Dandona L, Dandona R, Dharmaratne SD, Duncan BB, Eshrati B, Esteghamati A, Feigin VL, Fernandes JC, Furst T, Gebrehiwot TT, Gold A, Gona PN, Goto A, Habtewold TD,

Hadush KT, Hafezi-Nejad N, Hay SI, Horino M, Islami F, Kamal R, Kasaeian A, Katikireddi SV, Kengne AP, Kesavachandran CN, Khader YS, Khang YH, Khubchandani J, Kim D, Kim YJ, Kinfu Y, Kosen S, Ku T, Defo BK, Kumar GA, Larson HJ, Leinsalu M, Liang X, Lim SS, Liu P, Lopez AD, Lozano R, Majeed A, Malekzadeh R, Malta DC, Mazidi M, McAlinden C, McGarvey ST, Mengistu DT, Mensah GA, Mensink GBM, Mezgebe HB, Mirrakhimov EM, Mueller UO, Noubiap JJ, Obermeyer CM, Ogbo FA, Owolabi MO, Patton GC, Pourmalek F, Qorbani M, Rafay A, Rai RK, Ranabhat CL, Reinig N, Safiri S, Salomon JA, Sanabria JR, Santos IS, Sartorius B, Sawhney M, Schmidhuber J, Schutte AE, Schmidt MI, Sepanlou SG, Shamsizadeh M, Sheikhbahaei S, Shin MJ, Shiri R, Shiue I, Roba HS, Silva DAS, Silverberg JI, Singh JA, Stranges S, Swaminathan S, Tabares-Seisdedos R, Tadese F, Tedla BA, Tegegne BS, Terkawi AS, Thakur JS, Tonelli M, Topor-Madry R, Tyrovolas S, Ukwaja KN, Uthman OA, Vaezghasemi M, Vasankari T, Vlassov VV, Vollset SE, Weiderpass E, Werdecker A, Wesana J, Westerman R, Yano Y, Yonemoto N, Yonga G, Zaidi Z, Zenebe ZM, Zipkin B, Murray CJL. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med*. 2017;377(1):13-27.

16. American Diabetes A. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2014;37 Suppl 1:S81-90.

17. Verma S, Hussain ME. Obesity and diabetes: An update. *Diabetes Metab Syndr*. 2017;11(1):73-79.

18. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881-885.

19. Martin-Timon I, Sevillano-Collantes C, Segura-Galindo A, Del Canizo-Gomez FJ. Type 2 diabetes and cardiovascular disease: Have all risk factors the same strength? *World J Diabetes*. 2014;5(4):444-470.

20. Blanco-Rojo R, Perez-Martinez P, Lopez-Moreno J, Martinez-Botas J, Delgado-Lista J, van-Ommen B, Yubero-Serrano E, Camargo A, Ordovas JM, Perez-Jimenez F, Gomez-Coronado D, Lopez-Miranda J. HDL cholesterol efflux normalised to apoA-I is associated with future development of type 2 diabetes: from the CORDIOPREV trial. *Sci Rep*. 2017;7(1):12499.

21. Camargo A, Jimenez-Lucena R, Alcala-Diaz JF, Rangel-Zuniga OA, Garcia-Carpintero S, Lopez-Moreno J, Blanco-Rojo R, Delgado-Lista J, Perez-Martinez P, van Ommen B,

Malagon MM, Ordovas JM, Perez-Jimenez F, Lopez-Miranda J. Postprandial endotoxemia may influence the development of type 2 diabetes mellitus: From the CORDIOPREV study. *Clin Nutr*. 2018;[Epub ahead of print].

22. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103.
23. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, van der AD, Moons KG, Navis G, Bakker SJ, Beulens JW. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ*. 2012;345:e5900.
24. Cohen RM, Haggerty S, Herman WH. HbA1c for the diagnosis of diabetes and prediabetes: is it time for a mid-course correction? *J Clin Endocrinol Metab*. 2010;95(12):5203-5206.
25. Brodovicz KG, Dekker JM, Rijkelijkhuizen JM, Rhodes T, Mari A, Alssema M, Nijpels G, Williams-Herman DE, Girman CJ. The Finnish Diabetes Risk Score is associated with insulin resistance but not reduced beta-cell function, by classical and model-based estimates. *Diabet Med*. 2011;28(9):1078-1081.
26. Stroeve JHM, van Wietmarschen H, Kremer BHA, van Ommen B, Wopereis S. Phenotypic flexibility as a measure of health: the optimal nutritional stress response test. *Genes Nutr*. 2015;10(3):13.
27. Garcia-Rios A, Ordovas JM, Lopez-Miranda J, Perez-Martinez P. New diet trials and cardiovascular risk. *Curr Opin Cardiol*. 2018;33(4):423-428.
28. Perez-Martinez P, Ordovas JM, Garcia-Rios A, Delgado-Lista J, Delgado-Casado N, Cruz-Teno C, Camargo A, Yubero-Serrano EM, Rodriguez F, Perez-Jimenez F, Lopez-Miranda J. Consumption of diets with different type of fat influences triacylglycerols-rich lipoproteins particle number and size during the postprandial state. *Nutr Metab Cardiovasc Dis*. 2011;21(1):39-45.
29. Salas-Salvado J, Martinez-Gonzalez MA, Bullo M, Ros E. The role of diet in the prevention of type 2 diabetes. *Nutr Metab Cardiovasc Dis*. 2011;21 Suppl 2:B32-48.
30. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. *Diabetes Care*. 2019;42(Suppl 1):S13-S28.
31. Younossi Z, Tacke F, Arrese M, Sharma BC, Mostafa I, Bugianesi E, Wong VW, Yilmaz Y, George J, Fan J, Vos MB. Global Perspectives on Non-alcoholic Fatty Liver Disease and Non-alcoholic Steatohepatitis. *Hepatology*. 2018.

32. Iqbal U, Perumpail BJ, Akhtar D, Kim D, Ahmed A. The Epidemiology, Risk Profiling and Diagnostic Challenges of Nonalcoholic Fatty Liver Disease. *Medicines (Basel)*. 2019;6(1).

33. Masuoka HC, Chalasani N. Nonalcoholic fatty liver disease: an emerging threat to obese and diabetic individuals. *Annals of the New York Academy of Sciences*. 2013;1281:106-122.

34. Asrih M, Jornayvaz FR. Metabolic syndrome and nonalcoholic fatty liver disease: Is insulin resistance the link? *Molecular and cellular endocrinology*. 2015;418 Pt 1:55-65.

35. Baffy G, Brunt EM, Caldwell SH. Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace. *J Hepatol*. 2012;56(6):1384-1391.

36. Masuzaki R, Karp SJ, Omata M. NAFLD as a risk factor for HCC: new rules of engagement? *Hepatol Int*. 2016;10(4):533-534.

37. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, Ashley-Koch AE, Choi SS, Michelotti GA, Hampton DD, Chen Y, Tillmann HL, Hauser MA, Abdelmalek MF, Diehl AM. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology*. 2014;59(2):471-482.

38. Koutsari C, Lazaridis KN. Emerging genes associated with the progression of nonalcoholic fatty liver disease. *Hepatology*. 2010;52(3):807-810.

39. Fabbrini E, Magkos F. Hepatic Steatosis as a Marker of Metabolic Dysfunction. *Nutrients*. 2015;7(6):4995-5019.

40. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology*. 2010;51(2):679-689.

41. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674.

42. Caiado F, Silva-Santos B, Norell H. Intra-tumour heterogeneity - going beyond genetics. *FEBS J*. 2016;283(12):2245-2258.

43. Gahete MD, Cordoba-Chacon J, Duran-Prado M, Malagon MM, Martinez-Fuentes AJ, Gracia-Navarro F, Luque RM, Castano JP. Somatostatin and its receptors from fish to mammals. *Ann N Y Acad Sci*. 2010;1200:43-52.

44. Chopin LK, Seim I, Walpole CM, Herington AC. The ghrelin axis--does it have an appetite for cancer progression? *Endocrine reviews*. 2012;33(6):849-891.

45. Gahete MD, Rincon-Fernandez D, Villa-Osaba A, Hormaechea-Agulla D, Ibanez-Costa A, Martinez-Fuentes AJ, Gracia-Navarro F, Castano JP, Luque RM. Ghrelin gene

products, receptors, and GOAT enzyme: biological and pathophysiological insight. *J Endocrinol*. 2014;220(1):R1-R24.

46. Hormaechea-Agulla D, Gomez-Gomez E, Ibanez-Costa A, Carrasco-Valiente J, Rivero-Cortes E, F LL, Pedraza-Arevalo S, Valero-Rosa J, Sanchez-Sanchez R, Ortega-Salas R, Moreno MM, Gahete MD, Lopez-Miranda J, Requena MJ, Castano JP, Luque RM. Ghrelin O-acyltransferase (GOAT) enzyme is overexpressed in prostate cancer, and its levels are associated with patient's metabolic status: Potential value as a non-invasive biomarker. *Cancer letters*. 2016;383(1):125-134.

47. Hormaechea-Agulla D, Gahete MD, Jimenez-Vacas JM, Gomez-Gomez E, Ibanez-Costa A, F LL, Rivero-Cortes E, Sarmento-Cabral A, Valero-Rosa J, Carrasco-Valiente J, Sanchez-Sanchez R, Ortega-Salas R, Moreno MM, Tsomaia N, Swanson SM, Culler MD, Requena MJ, Castano JP, Luque RM. The oncogenic role of the In1-ghrelin splicing variant in prostate cancer aggressiveness. *Mol Cancer*. 2017;16(1):146.

48. Hormaechea-Agulla D, Jimenez-Vacas JM, Gomez-Gomez E, F LL, Carrasco-Valiente J, Valero-Rosa J, Moreno MM, Sanchez-Sanchez R, Ortega-Salas R, Gracia-Navarro F, Culler MD, Ibanez-Costa A, Gahete MD, Requena MJ, Castano JP, Luque RM. The oncogenic role of the spliced somatostatin receptor sst5TMD4 variant in prostate cancer. *FASEB J*. 2017;31(11):4682-4696.

49. Henderson BE, Ross RK, Pike MC, Casagrande JT. Endogenous hormones as a major factor in human cancer. *Cancer Res*. 1982;42(8):3232-3239.

50. Belfiore A, Perks CM. Grand challenges in cancer endocrinology: endocrine related cancers, an expanding concept. *Front Endocrinol (Lausanne)*. 2013;4:141.

51. Williams RA, Whitehead R. Non-carcinoid epithelial tumours of the appendix--a proposed classification. *Pathology*. 1986;18(1):50-53.

52. Pedraza-Arevalo S, Gahete MD, Alors-Perez E, Luque RM, Castano JP. Multilayered heterogeneity as an intrinsic hallmark of neuroendocrine tumors. *Rev Endocr Metab Disord*. 2018;19(2):179-192.

53. Packer JR, Maitland NJ. The molecular and cellular origin of human prostate cancer. *Biochim Biophys Acta*. 2016;1863(6 Pt A):1238-1260.

54. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, Gores G. Hepatocellular carcinoma. *Nat Rev Dis Primers*. 2016;2:16018.

55. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-386.

56. Margini C, Dufour JF. The story of HCC in NAFLD: from epidemiology, across pathogenesis, to prevention and treatment. *Liver Int*. 2016;36(3):317-324.
57. Cholankeril G, Patel R, Khurana S, Satapathy SK. Hepatocellular carcinoma in non-alcoholic steatohepatitis: Current knowledge and implications for management. *World J Hepatol*. 2017;9(11):533-543.
58. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
59. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*. 2013;14(3):153-165.
60. Patel AA, Steitz JA. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*. 2003;4(12):960-970.
61. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. *Biochim Biophys Acta*. 2009;1792(1):14-26.
62. Jurica MS, Moore MJ. Pre-mRNA Splicing: awash in a sea of proteins. *Mol Cell*. 2003;12(1):5-14.
63. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev*. 2009;10(11):741-754.
64. Hoskins AA, Moore MJ. The spliceosome: a flexible, reversible macromolecular machine. *Trend Biochem Sci*. 2012;37(5):179-188.
65. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136(4):701-718.
66. Smith CW, Valcarcel J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*. 2000;25(8):381-388.
67. House AE, Lynch KW. Regulation of alternative splicing: more than just the ABCs. *J Biol Chem*. 2008;283(3):1217-1221.
68. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008;14(5):802-813.
69. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. Function of alternative splicing. *Gene*. 2013;514(1):1-30.
70. Huang Y, Yario TA, Steitz JA. A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A*. 2004;101(26):9666-9670.

71. Stamm S. Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem*. 2008;283(3):1223-1227.

72. Berget SM. Exon recognition in vertebrate splicing. *J Biol Chem*. 1995;270(6):2411-2414.

73. Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Luhrmann R. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell*. 2010;38(2):223-235.

74. Sterne-Weiler T, Sanford JR. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol*. 2014;15(1):201.

75. Valadkhan S, Manley JL. Splicing-related catalysis by protein-free snRNAs. *Nature*. 2001;413(6857):701-707.

76. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010;463(7280):457-463.

77. Wang BD, Lee NH. Aberrant RNA Splicing in Cancer and Drug Resistance. *Cancers (Basel)*. 2018;10(11).

78. Ravi S, Schilder RJ, Kimball SR. Role of precursor mRNA splicing in nutrient-induced alterations in gene expression and metabolism. *J Nutr*. 2015;145(5):841-846.

79. Noon LA, Bakmanidis A, Clark AJL, O'Shaughnessy PJ, King PJ. Identification of a novel melanocortin 2 receptor splice variant in murine adipocytes: implications for post-transcriptional control of expression during adipogenesis. *J Mol Endocrinol*. 2006;37(3):415-420.

80. Jumaa H, Wei G, PJ N. Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Current Biology*. 1999;9(16):899-902.

81. Elizalde M, Urtasun R, Azkona M, Latasa MU, Goni S, Garcia-Irigoyen O, Uriarte I, Segura V, Collantes M, Di Scala M, Lujambio A, Prieto J, Avila MA, Berasain C. Splicing regulator SLU7 is essential for maintaining liver homeostasis. *J Clin Invest*. 2014;124(7):2909-2920.

82. Cheng Y, Luo C, Wu W, Xie Z, Fu X, Feng Y. Liver-Specific Deletion of SRSF2 Caused Acute Liver Failure and Early Death in Mice. *Mol Cell Biol*. 2016;36(11):1628-1638.

83. Lin J-C, Yan Y-T, Hsieh W-K, Peng P-J, Su C-H, Tarn W-Y. RBM4 Promotes Pancreas Cell Differentiation and Insulin Expression. *Mol Cell Biol*. 2013;33(2):319-327.

84. Villate O, Turatsinze J-V, Mascali LG, Grieco FA, Nogueira TC, Cunha DA, Nardelli TR, Sammeth M, Salunkhe Vishal A, Esguerra Jonathan L S, Eliasson L, Marselli L, Marchetti P, Eizirik DL. Nova1 is a master regulator of alternative splicing in pancreatic beta cells. *Nucleic Acids Research*. 2014;42(18):11818-11830.

85. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell*. 2009;136(4):777-793.

86. Fletcher S, Meloni PL, Johnsen RD, Wong BL, Muntoni F, Wilton SD. Antisense suppression of donor splice site mutations in the dystrophin gene transcript. *Mol Genet Genomic Med*. 2013;1(3):162-173.

87. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harb Perspect Med*. 2013;3(2):a009472.

88. Niblock M, Gallo JM. Tau alternative splicing in familial and sporadic tauopathies. *Biochem Soc Trans*. 2012;40(4):677-680.

89. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*. 2016;35(19):2413-2427.

90. Mercader JM, Liao RG, Bell AD, Dymek Z, Estrada K, Tukiainen T, Huerta-Chagoya A, Moreno-Macias H, Jablonski KA, Hanson RL, Walford GA, Moran I, Chen L, Agarwala V, Ordonez-Sanchez ML, Rodriguez-Guillen R, Rodriguez-Torres M, Segura-Kato Y, Garcia-Ortiz H, Centeno-Cruz F, Barajas-Olmos F, Caulkins L, Puppala S, Fontanillas P, Williams AL, Bonas-Guarch S, Hartl C, Ripke S, Diabetes Prevention Program Research G, Tooley K, Lane J, Zerrweck C, Martinez-Hernandez A, Cordova EJ, Mendoza-Caamal E, Contreras-Cubas C, Gonzalez-Villalpando ME, Cruz-Bautista I, Munoz-Hernandez L, Gomez-Velasco D, Alvirde U, Henderson BE, Wilkens LR, Le Marchand L, Arellano-Campos O, Riba L, Harden M, Broad Genomics P, Gabriel S, Consortium TDG, Abboud HE, Cortes ML, Revilla-Monsalve C, Islas-Andrade S, Soberon X, Curran JE, Jenkinson CP, DeFronzo RA, Lehman DM, Hanis CL, Bell GI, Boehnke M, Blangero J, Duggirala R, Saxena R, MacArthur D, Ferrer J, McCarroll SA, Torrents D, Knowler WC, Baier LJ, Burtt N, Gonzalez-Villalpando C, Haiman CA, Aguilar-Salinas CA, Tusie-Luna T, Flannick J, Jacobs SBR, Orozco L, Altshuler D, Florez JC, Consortium STDG. A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for Type 2 Diabetes. *Diabetes*. 2017;66(11):2903-2914.

91. Malakar P, Chartarifsky L, Hija A, Leibowitz G, Glaser B, Dor Y, Karni R. Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. *Sci Rep*. 2016;6:31222.

92. Frojdo S, Vidal H, Pirola L. Alterations of insulin signaling in type 2 diabetes: a review of the current evidence from humans. *Biochim Biophys Acta*. 2009;1792(2):83-92.

93. Ullrich A, Bell JR, Chen EY, Herrera R, Petruzzelli LM, Dull TJ, Gray A, Coussens L, Liao YC, Tsubokawa M, et al. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. *Nature*. 1985;313(6005):756-761.

94. Belfiore A, Frasca F, Pandini G, Sciacca L, Vigneri R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocr Rev*. 2009;30(6):586-623.

95. Seino S, Bell GI. Alternative splicing of human insulin receptor messenger RNA. *Biochem Biophys Res Commun*. 1989;159(1):312-316.

96. Mosthaf L, Vogt B, Häring HU, Ullrich A. Altered expression of insulin receptor types A and B in the skeletal muscle of non-insulin-dependent diabetes mellitus patients. *Proc Natl Acad Sci USA*. 1991;88(11):4728-4730.

97. Norgren S, Zierath J, Galuska D, Wallberg-Henriksson H, Luthman H. Differences in the ratio of RNA encoding two isoforms of the insulin receptor between control and NIDDM patients: the RNA variant without exon 11 predominates in both groups. *Diabetes*. 1993;42(5):675-681.

98. Benecke H, Flier JS, Moller DE. Alternatively spliced variants of the insulin receptor protein. Expression in normal and diabetic human tissues. *J Clin Invest*. 1992;89(6):2066-2070.

99. Anderson CM, Henry RR, Knudson PE, Olefsky JM, Webster NJ. Relative expression of insulin receptor isoforms does not differ in lean, obese, and noninsulin-dependent diabetes mellitus subjects. *J Clin Endocrinol Metab*. 1993;76(5):1380-1382.

100. Escribano O, Beneit N, Rubio-Longas C, Lopez-Pastor AR, Gomez-Hernandez A. The Role of Insulin Receptor Isoforms in Diabetes and Its Metabolic and Vascular Complications. *J Diabetes Res*. 2017;2017:1403206.

101. Sesti G. Insulin receptor variant forms and type 2 diabetes mellitus. *Pharmacogenomics*. 2000;1(1):49-61.

102. Nevado C, Valverde AM, Benito M. Role of insulin receptor in the regulation of glucose uptake in neonatal hepatocytes. *Endocrinology*. 2006;147(8):3709-3718.

103. Besic V, Shi H, Stubbs RS, Hayes MT. Aberrant liver insulin receptor isoform A expression normalises with remission of Type 2 Diabetes after gastric bypass surgery. *PLoS ONE*. 2015;10(3):e0119270.

104. Belfiore A, Malaguarnera R, Vella V, Lawrence MC, Sciacca L, Frasca F, Morrione A, Vigneri R. Insulin Receptor Isoforms in Physiology and Disease: An Updated View. *Endocr Rev*. 2017;38(5):379-431.

105. Kaminska D, Hämäläinen M, Cederberg H, Käkelä P, Venesmaa S, Miettinen P, Ilves I, Herzig K-H, Kolehmainen M, Karhunen L, Kuusisto J, Gylling H, Laakso M, Pihlajamäki J. Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. *Diabetologia*. 2014;57(2):347-351.

106. Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320-323.

107. Kaminska D, Kuulasmaa T, Venesmaa S, Käkelä P, Vaittinen M, Pulkkinen L, Pääkkönen M, Gylling H, Laakso M, Pihlajamäki J. Adipose Tissue TCF7L2 Splicing Is Regulated by Weight Loss and Associates With Glucose and Fatty Acid Metabolism. *Diabetes*. 2012;61(11):2807-2813.

108. Ashwal R, Hemi R, Tirosh A, Gordin R, Yissachar E, Cohen-Dayag A, Rosenberg A, Karasik A, Blüher M, Kanety H. Differential Expression of Novel Adiponectin Receptor-1 Transcripts in Skeletal Muscle of Subjects With Normal Glucose Tolerance and Type 2 Diabetes. *Diabetes*. 2011;60(3):936-946.

109. Péterfy M, Phan J, Reue K. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *J Biol Chem*. 2005;280(38):32883-32889.

110. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320-323.

111. Kulseth MA, Berge KE, Bogsrud MP, Leren TP. Analysis of LDLR mRNA in patients with familial hypercholesterolemia revealed a novel mutation in intron 14, which activates a cryptic splice site. *J Hum Genet*. 2010;55(10):676-680.

112. Kaminska D, Hamalainen M, Cederberg H, Kakela P, Venesmaa S, Miettinen P, Ilves I, Herzig KH, Kolehmainen M, Karhunen L, Kuusisto J, Gylling H, Laakso M, Pihlajamaki J. Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. *Diabetologia*. 2014;57(2):347-351.

113. Kulseth MA, Berge KE, Bogsrud MP, Leren TP. Analysis of LDLR mRNA in patients with familial hypercholesterolemia revealed a novel mutation in intron 14, which activates a cryptic splice site. *J Hum Genet*. 2010;55(10):676-680.

114. Péterfy M, Phan J, Xu P, Reue K. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat Genet*. 2001;27(1):121-124.

115. Li H, Cheng Y, Wu W, Liu Y, Wei N, Feng X, Xie Z, Feng Y. SRSF10 regulates alternative splicing and is required for adipocyte differentiation. *Mol Cell Biol*. 2014;34(12):2198-2207.

116. Brosch M, von Schonfels W, Ahrens M, Nothnagel M, Krawczak M, Laudes M, Sipos B, Becker T, Schreiber S, Rocken C, Schafmayer C, Hampe J. SFRS10--a splicing factor gene reduced in human obesity? *Cell Metab*. 2012;15(3):265-266; author reply 267-269.

117. Huot M-É, Richard S. Stay lean without dieting. *Adipocyte*. 2012;1(4):246-249.

118. Huot M-É, Vogel G, Zabarauskas A, Ngo Chau T-A, Coulombe-Huntington J, Majewski J, Richard S. The Sam68 STAR RNA-Binding Protein Regulates mTOR Alternative Splicing during Adipogenesis. *Molecular Cell*. 2012;46(2):187-199.

119. Rovira J, Marcelo Arellano E, Burke JT, Brault Y, Moya-Rull D, Bañón-Maneus E, Ramírez-Bajo MJ, Gutiérrez-Dalmau A, Revuelta I, Quintana LF, Campistol JM, Diekmann F. Effect of mTOR inhibitor on body weight: from an experimental rat model to human transplant patients. *Transplant Int*. 2008;21(10):992-998.

120. Reardon HT, Park WJ, Zhang J, Lawrence P, Kothapalli KS, Brenna JT. The polypyrimidine tract binding protein regulates desaturase alternative splicing and PUFA composition. *J Lipid Res*. 2011;52(12):2279-2286.

121. Medina MW, Krauss RM. Alternative splicing in the regulation of cholesterol homeostasis. *Curr Opin Lipidol*. 2013;24(2):147-152.

122. Garcia-Arcos I, Rueda Y, Gonzalez-Kother P, Palacios L, Ochoa B, Fresnedo O. Association of SND1 protein to low density lipid droplets in liver steatosis. *J Physiol Biochem*. 2010;66(1):73-83.

123. Reardon HT, Hsieh AT, Park WJ, Kothapalli KS, Anthony JC, Nathanielsz PW, Brenna JT. Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. *Prostaglandins Leukot Essent Fatty Acids*. 2013;88(1):15-19.

124. Ye H, Liu W. Transcriptional networks implicated in human nonalcoholic fatty liver disease. *Molecular genetics and genomics : MGG*. 2015;290(5):1793-1804.

125. Yu J, Marsh S, Hu J, Feng W, Wu C. The Pathogenesis of Nonalcoholic Fatty Liver Disease: Interplay between Diet, Gut Microbiota, and Genetic Background. *Gastroenterology research and practice*. 2016:2862173.

126. Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2017;23(1):1-12.

127. Lake AD, Novak P, Fisher CD, Jackson JP, Hardwick RN, Billheimer DD, Klimecki WT, Cherrington NJ. Analysis of global and absorption, distribution, metabolism, and elimination gene expression in the progressive stages of human nonalcoholic fatty liver disease. *Drug Metab Dispos*. 2011;39(10):1954-1960.

128. Murphy SK, Yang H, Moylan CA, Pang H, Dellinger A, Abdelmalek MF, Garrett ME, Ashley-Koch A, Suzuki A, Tillmann HL, Hauser MA, Diehl AM. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2013;145(5):1076-1087.

129. Zhu R, Baker SS, Moylan CA, Abdelmalek MF, Guy CD, Zamboni F, Wu D, Lin W, Liu W, Baker RD, Govindarajan S, Cao Z, Farci P, Diehl AM, Zhu L. Systematic transcriptome analysis reveals elevated expression of alcohol-metabolizing genes in NAFLD livers. *J Pathol*. 2016;238(4):531-542.

130. Pihlajamaki J, Lerin C, Itkonen P, Boes T, Floss T, Schroeder J, Dearie F, Crunkhorn S, Burak F, Jimenez-Chillaron JC, Kuulasmaa T, Miettinen P, Park PJ, Nasser I, Zhao Z, Zhang Z, Xu Y, Wurst W, Ren H, Morris AJ, Stamm S, Goldfine AB, Laakso M, Patti ME. Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. *Cell metabolism*. 2011;14(2):208-218.

131. Lopez-Vicario C, Gonzalez-Periz A, Rius B, Moran-Salvador E, Garcia-Alonso V, Lozano JJ, Bataller R, Cofan M, Kang JX, Arroyo V, Claria J, Titos E. Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. *Gut*. 2014;63(2):344-355.

132. Kahles A, Lehmann KV, Toussaint NC, Huser M, Stark SG, Sachsenberg T, Stegle O, Kohlbacher O, Sander C, Cancer Genome Atlas Research N, Ratsch G. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell*. 2018;34(2):211-224 e216.

133. Climente-Gonzalez H, Porta-Pardo E, Godzik A, Eyras E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep*. 2017;20(9):2215-2226.

134. Xu Q, Lee C. Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucleic Acids Res*. 2003;31(19):5635-5643.

135. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, Lucier JF, Thibault P, Rancourt C, Tremblay K, Prinos P, Chabot

B, Elela SA. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol*. 2009;16(6):670-676.

136. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene*. 2014;33(46):5311-5318.

137. Carriero MV, Bifulco K, Minopoli M, Lista L, Maglio O, Mele L, Di Carluccio G, De Rosa M, Pavone V. UPARANT: a urokinase receptor-derived peptide inhibitor of VEGF-driven angiogenesis with enhanced stability and in vitro and in vivo potency. *Mol Cancer Ther*. 2014;13(5):1092-1104.

138. Ruscica M, Arvigo M, Steffani L, Ferone D, Magni P. Somatostatin, somatostatin analogs and somatostatin receptor dynamics in the biology of cancer progression. *Current molecular medicine*. 2013;13(4):555-571.

139. Theodoropoulou M, Stalla GK. Somatostatin receptors: from signaling to clinical practice. *Front Neuroendocrinol*. 2013;34(3):228-252.

140. Duran-Prado M, Gahete MD, Martinez-Fuentes AJ, Luque RM, Quintero A, Webb SM, Benito-Lopez P, Leal A, Schulz S, Gracia-Navarro F, Malagon MM, Castano JP. Identification and characterization of two novel truncated but functional isoforms of the somatostatin receptor subtype 5 differentially present in pituitary tumors. *J Clin Endocrinol Metab*. 2009;94(7):2634-2643.

141. Gün̈ther T TG, Dournaud P, Bousquet C, Csaba Z, Kreienkamp HJ, Lupp A, Korbonits M, Castaño JP, Wester HJ, Culler MD, Melmed S, Schulz S. International Union of Basic and Clinical Pharmacology. Somatostatin Receptors: Structure, Function, Ligands and New Nomenclature. . *Pharmacol Rev*. 2018;In Press.

142. Chalabi M, Duluc C, Caron P, Vezzosi D, Guillermet-Guibert J, Pyronnet S, Bousquet C. Somatostatin analogs: does pharmacology impact antitumor efficacy? *Trends in endocrinology and metabolism: TEM*. 2014;25(3):115-127.

143. Samonakis DN, Notas G, Christodoulakis N, Kouroumalis EA. Mechanisms of action and resistance of somatostatin analogues for the treatment of hepatocellular carcinoma: a message not well taken. *Dig Dis Sci*. 2008;53(9):2359-2365.

144. Watt HL, Kharmate G, Kumar U. Biology of somatostatin in breast cancer. *Molecular and cellular endocrinology*. 2008;286(1-2):251-261.

145. Ji XQ, Ruan XJ, Chen H, Chen G, Li SY, Yu B. Somatostatin analogues in advanced hepatocellular carcinoma: an updated systematic review and meta-analysis of randomized controlled trials. *Med Sci Monit*. 2011;17(8):RA169-176.

146. Rekasi Z, Czompoly T, Schally AV, Halmos G. Isolation and sequencing of cDNAs for splice variants of growth hormone-releasing hormone receptors from human cancers. *Proc Natl Acad Sci U S A*. 2000;97(19):10561-10566.

147. Havt A, Schally AV, Halmos G, Varga JL, Toller GL, Horvath JE, Szepeshazi K, Koster F, Kovitz K, Groot K, Zarandi M, Kanashiro CA. The expression of the pituitary growth hormone-releasing hormone receptor and its splice variants in normal and neoplastic human tissues. *Proc Natl Acad Sci U S A*. 2005;102(48):17424-17429.

148. Markovic D, Challiss RAJ. Alternative splicing of G protein-coupled receptors: physiology and pathophysiology. *Cellular and molecular life sciences : CMLS*. 2009;66(20):3337-3352.

149. Cordoba-Chacon J, Gahete MD, Duran-Prado M, Luque RM, Castano JP. Truncated somatostatin receptors as new players in somatostatin-cortistatin pathophysiology. *Ann N Y Acad Sci*. 2011;1220:6-15.

150. McWilliams DF, Watson SA, Crosbee DM, Michaeli D, Seth R. Coexpression of gastrin and gastrin receptors (CCK-B and delta CCK-B) in gastrointestinal tumour cell lines. *Gut*. 1998;42(6):795-798.

151. Hawrylyshyn KA, Michelotti GA, Coge F, Guenin SP, Schwinn DA. Update on human alpha1-adrenoceptor subtype signaling and genomic organization. *Trends Pharmacol Sci*. 2004;25(9):449-455.

152. Duran-Prado M, Gahete MD, Hergueta-Redondo M, Martinez-Fuentes AJ, Cordoba-Chacon J, Palacios J, Gracia-Navarro F, Moreno-Bueno G, Malagon MM, Luque RM, Castano JP. The new truncated somatostatin receptor variant sst5TMD4 is associated to poor prognosis in breast cancer and increases malignancy in MCF-7 cells. *Oncogene*. 2012;31(16):2049-2061.

153. Puig-Domingo M, Luque RM, Reverter JL, Lopez-Sanchez LM, Gahete MD, Culler MD, Diaz-Soto G, Lomena F, Squarcia M, Mate JL, Mora M, Fernandez-Cruz L, Vidal O, Alastrue A, Balibrea J, Halperin I, Mauricio D, Castano JP. The Truncated Isoform of Somatostatin Receptor5 (sst5TMD4) Is Associated with Poorly Differentiated Thyroid Cancer. *PloS one*. 2014;9(1):e85527.

154. Luque RM, Ibanez-Costa A, Neto LV, Taboada GF, Hormaechea-Agulla D, Kasuki L, Venegas-Moreno E, Moreno-Carazo A, Galvez MA, Soto-Moreno A, Kineman RD, Culler MD, Gahete MD, Gadelha MR, Castano JP. Truncated somatostatin receptor variant sst5TMD4 confers aggressive features (proliferation, invasion and reduced octreotide response) to somatotropinomas. *Cancer letters*. 2015;359(2):299-306.

155. Gahete MD, Rincon-Fernandez D, Duran-Prado M, Hergueta-Redondo M, Ibanez-Costa A, Rojo-Sebastian A, Gracia-Navarro F, Culler MD, Casanovas O, Moreno-Bueno G, Luque RM, Castano JP. The truncated somatostatin receptor sst5TMD4 stimulates the angiogenic process and is associated to lymphatic metastasis and disease-free survival in breast cancer patients. *Oncotarget*. 2016;7(37):60110-60122.

156. Sampedro-Nunez M, Luque RM, Ramos-Levi AM, Gahete MD, Serrano-Somavilla A, Villa-Osaba A, Adrados M, Ibanez-Costa A, Martin-Perez E, Culler MD, Marazuela M, Castano JP. Presence of sst5TMD4, a truncated splice variant of the somatostatin receptor subtype 5, is associated to features of increased aggressiveness in pancreatic neuroendocrine tumors. *Oncotarget*. 2016;7(6):6593-6608.

157. Duran-Prado M, Saveanu A, Luque RM, Gahete MD, Gracia-Navarro F, Jaquet P, Dufour H, Malagon MM, Culler MD, Barlier A, Castano JP. A potential inhibitory role for the new truncated variant of somatostatin receptor 5, sst5TMD4, in pituitary adenomas poorly responsive to somatostatin analogs. *J Clin Endocrinol Metab*. 2010;95(5):2497-2502.

158. Domingo B, Gasset M, Duran-Prado M, Castano JP, Serrano A, Fischer T, Llopis J. Discrimination between alternate membrane protein topologies in living cells using GFP/YFP tagging and pH exchange. *Cell Mol Life Sci*. 2010;67(19):3345-3354.

159. Craig SE, Brady-Kalnay SM. Tumor-derived extracellular fragments of receptor protein tyrosine phosphatases (RPTPs) as cancer molecular diagnostic tools. *Anticancer Agents Med Chem*. 2011;11(1):133-140.

160. Bifulco K, Longanesi-Cattani I, Liguori E, Arra C, Rea D, Masucci MT, De Rosa M, Pavone V, Stoppelli MP, Carriero MV. A urokinase receptor-derived peptide inhibiting VEGF-dependent directional migration and vascular sprouting. *Mol Cancer Ther*. 2013;12(10):1981-1993.

161. Ebben JD, Lubet RA, Gad E, Disis ML, You M. Epidermal growth factor receptor derived peptide vaccination to prevent lung adenocarcinoma formation: An in vivo study in a murine model of EGFR mutant lung cancer. *Mol Carcinog*. 2016;55(11):1517-1525.

162. Gurevich LE. Role of matrix metalloproteinases 2 and 9 in determination of invasive potential of pancreatic tumors. *Bull Exp Biol Med*. 2003;136(5):494-498.

163. Hadler-Olsen E, Winberg JO, Uhlin-Hansen L. Matrix metalloproteinases in cancer: their value as diagnostic and prognostic markers and therapeutic targets. *Tumour Biol*. 2013;34(4):2041-2051.

164. Gunther T, Tulipano G, Dournaud P, Bousquet C, Csaba Z, Kreienkamp HJ, Lupp A, Korbonits M, Castano JP, Wester HJ, Culler M, Melmed S, Schulz S. International Union

of Basic and Clinical Pharmacology. CV. Somatostatin Receptors: Structure, Function, Ligands, and New Nomenclature. *Pharmacol Rev*. 2018;70(4):763-835.

165. Burczynski ME, Dorner AJ. Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. *Pharmacogenomics*. 2006;7(2):187-202.

166. Delgado-Lista J, Perez-Martinez P, Garcia-Rios A, Alcala-Diaz JF, Perez-Caballero AI, Gomez-Delgado F, Fuentes F, Quintana-Navarro G, Lopez-Segura F, Ortiz-Morales AM, Delgado-Casado N, Yubero-Serrano EM, Camargo A, Marin C, Rodriguez-Cantalejo F, Gomez-Luna P, Ordovas JM, Lopez-Miranda J, Perez-Jimenez F. CORonary Diet Intervention with Olive oil and cardiovascular PREVention 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;177:42-50.

167. Pfaffenbach KT, Nivala AM, Reese L, Ellis F, Wang D, Wei Y, Pagliassotti MJ. Rapamycin inhibits postprandial-mediated X-box-binding protein-1 splicing in rat liver. *J Nutr*. 2010;140(5):879-884.

168. Fruhbeck G, Gomez-Ambrosi J, Martinez JA. Pre- and postprandial expression of the leptin receptor splice variants OB-Ra and OB-Rb in murine peripheral tissues. *Physiol Res*. 1999;48(3):189-195.

169. Gortan Cappellari G, Barazzoni R, Cattin L, Muro AF, Zanetti M. Lack of Fibronectin Extra Domain A Alternative Splicing Exacerbates Endothelial Dysfunction in Diabetes. *Sci Rep*. 2016;6:37965.

170. Newman JRB, Conesa A, Mika M, New FN, Onengut-Gumuscu S, Atkinson MA, Rich SS, McIntyre LM, Concannon P. Disease-specific biases in alternative splicing and tissue-specific dysregulation revealed by multitissue profiling of lymphocyte gene expression in type 1 diabetes. *Genome Res*. 2017;27(11):1807-1815.

171. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol*. 2017;18(11):655-670.

172. Gahete MD, Del Rio-Moreno M, Camargo A, Alcala-Diaz JF, Alors-Perez E, Delgado-Lista J, Reyes O, Ventura S, Perez-Martinez P, Castano JP, Lopez-Miranda J, Luque RM. Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study. *EBioMedicine*. 2018;37:356-365.

173. Del Rio-Moreno M, Alors-Perez E, Gonzalez-Rubio S, Ferrin G, Reyes O, Rodriguez-Peralvarez M, Sanchez-Frias ME, Sanchez-Sanchez R, Ventura S, Lopez-Miranda J,

Kineman RD, de la Mata M, Castano JP, Gahete MD, Luque RM. Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease. *J Clin Endocrinol Metab*. 2019.

174. Vernia S, Edwards YJ, Han MS, Cavanagh-Kyros J, Barrett T, Kim JK, Davis RJ. An alternative splicing program promotes adipose tissue thermogenesis. *Elife*. 2016;5.

175. Blanco-Rojo R, Alcala-Diaz JF, Wopereis S, Perez-Martinez P, Quintana-Navarro GM, Marin C, Ordovas JM, van Ommen B, Perez-Jimenez F, Delgado-Lista J, Lopez-Miranda J. The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomised clinical trial. *Diabetologia*. 2015;59(1):67–76.

176. Yubero-Serrano EM, Delgado-Casado N, Delgado-Lista J, Perez-Martinez P, Tasset-Cuevas I, Santos-Gonzalez M, Caballero J, Garcia-Rios A, Marin C, Gutierrez-Mariscal FM, Fuentes F, Villalba JM, Tunez I, Perez-Jimenez F, Lopez-Miranda J. Postprandial antioxidant effect of the Mediterranean diet supplemented with coenzyme Q10 in elderly men and women. *Age (Dordr)*. 2011;33(4):579-590.

177. Yubero-Serrano EM, Gonzalez-Guardia L, Rangel-Zuniga O, Delgado-Casado N, Delgado-Lista J, Perez-Martinez P, Garcia-Rios A, Caballero J, Marin C, Gutierrez-Mariscal FM, Tinahones FJ, Villalba JM, Tunez I, Perez-Jimenez F, Lopez-Miranda J. Postprandial antioxidant gene expression is modified by Mediterranean diet supplemented with coenzyme Q(10) in elderly men and women. *Age (Dordr)*. 2013;35(1):159-170.

178. de Mello VD, Kolehmanien M, Schwab U, Pulkkinen L, Uusitupa M. Gene expression of peripheral blood mononuclear cells as a tool in dietary intervention studies: What do we know so far? *Molecular nutrition & food research*. 2012;56(7):1160-1172.

179. Gahete MD, Luque RM, Yubero-Serrano EM, Cruz-Teno C, Ibanez-Costa A, Delgado-Lista J, Gracia-Navarro F, Perez-Jimenez F, Castano JP, Lopez-Miranda J. Dietary fat alters the expression of cortistatin and ghrelin systems in the PBMCs of elderly subjects: putative implications in the postprandial inflammatory response. *Mol Nutr Food Res*. 2014;58(9):1897-1906.

180. Fuchs D, Piller R, Linseisen J, Daniel H, Wenzel U. The human peripheral blood mononuclear cell proteome responds to a dietary flaxseed-intervention and proteins identified suggest a protective effect in atherosclerosis. *Proteomics*. 2007;7(18):3278-3288.

181. Rendo-Urteaga T, Garcia-Calzon S, Gonzalez-Muniesa P, Milagro FI, Chueca M, Oyarzabal M, Azcona-Sanjulian MC, Martinez JA, Marti A. Peripheral blood

mononuclear cell gene expression profile in obese boys who followed a moderate energy-restricted diet: differences between high and low responders at baseline and after the intervention. *Br J Nutr*. 2015;113(2):331-342.

182. Malatesta M, Bertoni-Freddari C, Fattoretti P, Baldelli B, Fakan S, Gazzanelli G. Aging and vitamin E deficiency are responsible for altered RNA pathways. *Annals of the New York Academy of Sciences*. 2004;1019:379-382.

183. Verma B, Akinyi MV, Norppa AJ, Frilander MJ. Minor spliceosome and disease. *Semin Cell Dev Biol*. 2018;79:103-112.

184. Matera G, Wang Z. A day in the life of the spliceosome. *Nature reviews molecular cell biology*. 2014;15:108-121.

185. Zhang X, Moor AN, Merkler KA, Liu Q, McLean MP. Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors. *Endocrinology*. 2007;148(11):5295-5304.

186. Thuahnai ST, Lund-Katz S, Williams DL, Phillips MC. Scavenger receptor class B, type I-mediated uptake of various lipids into cells. Influence of the nature of the donor particle interaction with the receptor. *J Biol Chem*. 2001;276(47):43801-43808.

187. Jariwala N, Rajasekaran D, Mendoza RG, Shen XN, Siddiq A, Akiel MA, Robertson CL, Subler MA, Windle JJ, Fisher PB, Sanyal AJ, Sarkar D. Oncogenic Role of SND1 in Development and Progression of Hepatocellular Carcinoma. *Cancer Res*. 2017;77(12):3306-3316.

188. Lee G, Zheng Y, Cho S, Jang C, England C, Dempsey JM, Yu Y, Liu X, He L, Cavaliere PM, Chavez A, Zhang E, Isik M, Couvillon A, Dephoure NE, Blackwell TK, Yu JJ, Rabinowitz JD, Cantley LC, Blenis J. Post-transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. *Cell*. 2017;171(7):1545-1558 e1518.

189. Shuman Moss LA, Jensen-Taubman S, Stetler-Stevenson WG. Matrix metalloproteinases: changing roles in tumor progression and metastasis. *Am J Pathol*. 2012;181(6):1895-1899.

190. Salem N, Kamal I, Al-Maghrabi J, Abuzenadah A, Peer-Zada AA, Qari Y, Al-Ahwal M, Al-Qahtani M, Buhmeida A. High expression of matrix metalloproteinases: MMP-2 and MMP-9 predicts poor survival outcome in colorectal carcinoma. *Future Oncol*. 2016;12(3):323-331.

191. Turunen SP, Tatti-Bugaeva O, Lehti K. Membrane-type matrix metalloproteases as diverse effectors of cancer progression. *Biochim Biophys Acta*. 2017;1864(11 Pt A):1974-1988.

192. Itoh Y. Membrane-type matrix metalloproteinases: Their functions and regulations. *Matrix Biol*. 2015;44-46:207-223.

193. Vandamme T, Peeters M, Dogan F, Pauwels P, Van Assche E, Beyens M, Mortier G, Vandeweyer G, de Herder W, Van Camp G, Hofland LJ, Op de Beeck K. Whole-exome characterization of pancreatic neuroendocrine tumor cell lines BON-1 and QGP-1. *J Mol Endocrinol*. 2015;54(2):137-147.

194. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. *Breast Cancer Res*. 2011;13(4):215.

195. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest Urol*. 1979;17(1):16-23.

196. Sramkoski RM, Pretlow TG, 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS, Marengo SR, Rhim JS, Zhang D, Jacobberger JW. A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell Dev Biol Anim*. 1999;35(7):403-409.

197. Lopez-Bergami P, Huang C, Goydos JS, Yip D, Bar-Eli M, Herlyn M, Smalley KS, Mahale A, Eroshkin A, Aaronson S, Ronai Z. Rewired ERK-JNK signaling pathways in melanoma. *Cancer Cell*. 2007;11(5):447-460.

198. McCubrey JA, Steelman LS, Chappell WH, Abrams SL, Wong EW, Chang F, Lehmann B, Terrian DM, Milella M, Tafuri A, Stivala F, Libra M, Basecke J, Evangelisti C, Martelli AM, Franklin RA. Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochim Biophys Acta*. 2007;1773(8):1263-1284.

199. De Luca A, Maiello MR, D'Alessio A, Pergameno M, Normanno N. The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. *Expert Opin Ther Targets*. 2012;16 Suppl 2:S17-27.

200. Jernman J, Valimaki MJ, Louhimo J, Haglund C, Arola J. The novel WHO 2010 classification for gastrointestinal neuroendocrine tumours correlates well with the metastatic potential of rectal neuroendocrine tumours. *Neuroendocrinology*. 2012;95(4):317-324.

201. Molinie N, Gautreau A. The Arp2/3 Regulatory System and Its Deregulation in Cancer. *Physiol Rev*. 2018;98(1):215-238.

Articles

Article I: Manuel D. Gahete*, **Mercedes del Rio-Moreno***, Antonio Camargo*, Juan F. Alcala-Diaz, Emilia Alors-Perez, Javier Delgado-Lista, Oscar Reyes, Sebastian Ventura, Pablo Perez-Martínez, Justo P. Castaño, José Lopez-Miranda and Raul M. Luque. *Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study*. 2018. EBioMedicine. 37: 356-365. *authors equally contributed for the publication.

Article II: **Mercedes del Rio-Moreno,** Emilia Alors-Perez, Antonio Camargo, Jose Lopez-Miranda, Justo P. Castaño, Raul M. Luque and Manuel D. Gahete. *Dietary intervention modulates the expression of the splicing machinery in patients at high risk of type 2 diabetes development: From the CORDIOPREV study*. Manuscript submitted to "Diabetologia".

Article III: **Mercedes del Rio-Moreno,** Emilia Alors-Perez, Sandra Gonzalez-Rubio, Gustavo Ferrín, Oscar Reyes, Manuel Rodriguez-Peralvarez, Marina E. Sanchez-Frias, Rafael Sanchez-Sanchez, Sebastian Ventura, Jose Lopez-Miranda, Rhonda D. Kineman, Manuel de la Mata, Justo P. Castaño, Manuel D. Gahete and Raul M. Luque. *Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease*. 2019. Journal of Clinical Endocrinology and Metabolism. [Epub ahead of print; doi: 10.1210/jc.2019-00021].

Article IV: **Mercedes del Rio-Moreno**, Emilia Alors-Perez, Patricia Borges de Souza, Maria E. Prados-Gonzalez, Justo P. Castaño, Raul M. Luque and Manuel D. Gahete. *Peptides derived from the extracellular domain of the somatostatin receptor splicing variant SST5TMD4 increase malignancy in multiple cancer cell types.* 2019. Translational Research. [Epub ahead of print; doi: 10.1016/j.trsl.2019.02.013].

Article I

EBioMedicine 37 (2018) 356–365

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study

Manuel D. Gahete [a,b,c,d,1,2], Mercedes del Rio-Moreno [a,b,c,d,1], Antonio Camargo [a,b,d,e,1], Juan F. Alcala-Diaz [a,b,d,e], Emilia Alors-Perez [a,b,c,d], Javier Delgado-Lista [a,b,d,e], Oscar Reyes [a,b,c,d,f], Sebastian Ventura [a,b,c,d,f], Pablo Perez-Martínez [a,b,d,e], Justo P. Castaño [a,b,c,d,2], José Lopez-Miranda [a,b,d,e,2], Raul M. Luque [a,b,c,d,*,2]

[a] Maimonides Institute for Biomedical Research of Cordoba (IMIBIC), Córdoba, Spain.
[b] Universidad de Córdoba, Córdoba, Spain.
[c] Reina Sofia University Hospital, Córdoba, Spain.
[d] CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn), Córdoba, Spain.
[e] Lipid and Atherosclerosis Unit, Department of Medicine, Reina Sofia University Hospital, University of Córdoba, Córdoba, Spain.
[f] Department of Computer Sciences, University of Cordoba, Córdoba, Spain

ARTICLE INFO

Article history:
Received 4 September 2018
Received in revised form 18 October 2018
Accepted 24 October 2018
Available online 13 November 2018

Keywords:
Peripheral blood mononuclear cells (PBMCs)
Postprandial state
Spliceosome
OGTT
RNUs

ABSTRACT

Background: Type-2 diabetes mellitus (T2DM) is a major health problem with increasing incidence, which severely impacts cardiovascular disease. Because T2DM is associated with altered gene expression and aberrant splicing, we hypothesized that dysregulations in splicing machinery could precede, contribute to, and predict T2DM development.

Methods: A cohort of patients with cardiovascular disease (CORDIOPREV study) and without T2DM at baseline (at the inclusion of the study) was used ($n = 215$). We determined the expression of selected splicing machinery components in fasting and 4 h-postprandial peripheral blood mononuclear cells (PBMCs, obtained at baseline) from all the patients who developed T2DM during 5-years of follow-up ($n = 107$ incident-T2DM cases) and 108 randomly selected non-T2DM patients (controls). Serum from incident-T2DM and control patients was used to analyze *in vitro* the modulation of splicing machinery expression in control PBMCs from an independent cohort of healthy subjects.

Findings: Expression of key splicing machinery components (e.g. *RNU2*, *RNU4* or *RNU12*) from fasting and 4 h-postprandial PBMCs of incident-T2DM patients was markedly altered compared to non-T2DM controls. Moreover, *in vitro* treatment of healthy individuals PBMCs with serum from incident-T2DM patients (compared to non-T2DM controls) reduced the expression of splicing machinery elements found down-regulated in incident-T2DM patients PBMCs. Finally, fasting/postprandial levels of several splicing machinery components in the PBMCs of CORDIOPREV patients were associated to higher risk of T2DM (Odds Ratio > 4) and could accurately predict (AUC > 0.85) T2DM development.

Interpretation: Our results reveal the existence of splicing machinery alterations that precede and predict T2DM development in patients with cardiovascular disease.

Fund: ISCIII, MINECO, CIBERObn.

1. Introduction

Type 2 diabetes mellitus (T2DM) is a major health problem [1] closely associated to cardiovascular disease (CVD), wherein patients with myocardial infarction and T2DM have a higher risk of developing a new cardiovascular event than those without T2DM [2]. Thus, early identification of individuals at high risk for T2DM development, especially among patients with CVD [3,4], is critical for prevention [5,6]. Traditionally, this strategy has been based on biomarkers [glycated

* Corresponding author at: Lipid and Atherosclerosis Unit, Department of Medicine, Reina Sofia University Hospital, University of Córdoba, Córdoba 14004, Spain.
E-mail address: raul.luque@uco.es (R.M. Luque).
[1] These authors should be considered as first authors
[2] These authors should be considered as corresponding authors and contributed equally to the study.

https://doi.org/10.1016/j.ebiom.2018.10.056

Research in context

Evidence before this study

Early identification of individuals at high risk of type 2 diabetes mellitus (T2DM) development, especially among patients with cardiovascular disease (CVD), is critical for prevention. Traditionally, the identification strategy has been based on biomarkers such as glycated hemoglobin (HbA1c) or predictive scores like FINDRISK. However, these approaches have limitations and cannot precisely predict an individual's risk of developing T2DM. In this context, the identification of key modifiers of phenotypic plasticity that define individual susceptibility to develop T2DM may hold predictive potential. Particularly, there is emerging evidence that under adverse metabolic conditions, including obesity and T2DM, the splicing process and, consequently, the generation of splicing variants is markedly dysregulated in most tissues. However, until the present study, the association between potential changes in the regulatory machinery governing the splicing process and alterations in phenotypic plasticity —which is essential to maintain whole body homeostasis, and therefore, to prevent the development of metabolic pathologies— was still unknown. In this context, we took advantage of the emerging evidence showing that gene expression pattern of peripheral blood mononuclear cells (PBMCs) commonly reflects and accompany disease-characteristic expression patterns, and may thus serve as a general sentinel, biosensor and early indicator of the instauration of metabolic disease. Until now, it was not known whether the expression of the splicing machinery components is dysregulated in PBMCs under conditions of loss of phenotypic plasticity and, therefore, if it could be associated with the subsequent development of T2DM.

Added value of this study

We took advantage of the CORDIOPREV study (CORonary Diet Intervention with Olive oil and cardiovascular PREVention), an ongoing, prospective, randomized trial that includes the 5-year follow-up of patients with CVD, but without T2DM at the inclusion of the study (baseline), to show, for the first time, that striking changes in the expression of key splicing machinery components of PBMCs precede and can early predict the development of T2DM. Moreover, our study demonstrates that expression pattern of the splicing machinery components can be dynamically modulated during the post-prandial phase. These rapid changes likely reflect the phenotypic flexibility of these patients in response to a metabolic challenge, and, most importantly, were clearly altered at the inclusion of the study in patients that developed T2DM during the 5-year follow-up. In particular, the fasting/postprandial expression levels of a set of bioinformatically-selected splicing machinery components in the PBMCs of CORDIOPREV patients were associated to higher risk of T2DM (Odds Ratio > 4) and could accurately predict (AUC > 0.85) diabetes development in this cohort, which is significantly higher than the current standard procedures (HbA1c and FINDRISK exhibited AUCs < 0.66 in our cohort). Finally, incubation of PBMCs with serum from patients that developed, or not, T2DM revealed a potential causal relationship in that the latter was able to reproduce some of the changes observed in the expression of key splicing machinery components.

Implications of all the available evidence

Altered expression of splicing machinery components may be associated with the development of T2DM, preceding the clinical instauration of this pathology, and could serve as a sensor and early predictor for T2DM development in CVD patients. Our data demonstrate the existence of a splicing machinery-associated molecular fingerprint capable to predict the future development of T2DM in individual patients with high precision, which even outperforms the capacity of classical predictors of T2DM development. Therefore, this splicing machinery-associated molecular fingerprint could become a valuable, non-invasive, new tool for early risk assessment of T2DM in clinical practice to prevent disease development.

hemoglobin (HbA1c)] or predictive scores (FINDRISK); however, these approaches have limitations and cannot precisely predict an individual's risk of developing T2DM [7,8].

T2DM development is critically affected by the loss of phenotypic flexibility (i.e. the difficulty to cope with stressors to maintain metabolic homeostasis), wherein the identification of key modifiers of phenotypic plasticity that define individual susceptibility to develop T2DM may hold predictive potential [9]. Particularly, there is emerging evidence that under adverse metabolic conditions (obesity, insulin resistance, etc.) the splicing machinery is markedly dysregulated in most tissues [10–12], including peripheral blood mononuclear cells (PBMCs); and that dysregulations in splicing processes are associated with development of several pathologies [10,13,14]. Specifically, splicing process is catalyzed by the minor and major spliceosomes, which act on different types of introns [15]. The spliceosome is an intricate macromolecular complex, whose functional core is comprised by several small nuclear ribonucleoproteins (snRNPs) subunits, which dynamically interact to regulate the splicing process. In addition, the activity of the spliceosome is modulated by >300 splicing factors (SFs) that specifically recognize certain sequences in exons and introns. Consequently, the dysregulation of the expression and/or function of certain spliceosomal components may drive the aberrant alteration of the normal splicing process [16]. Indeed, correct function of the splicing machinery, i.e. spliceosome components and SFs, is an essential mechanism to maintain whole body homeostasis [11,12,17,18]. Therefore, perturbed splicing may play a major role as a pathogenic factor for and may serve as a predictive marker of the development of T2DM [12,19].

In this sense, since gene expression patterns in PBMCs commonly reflect and accompany disease-characteristic expression patterns [20], the hypothesis of this study is that the pattern of expression of certain splicing machinery elements in the PBMCs, especially during the post-prandial response, when changes in phenotypic flexibility are more evident, could be associated with the risk and could early predict the development of T2DM. To test this idea, we took advantage of the CORDIOPREV study (CORonary Diet Intervention with Olive oil and cardiovascular PREVention), an ongoing, prospective and randomized trial that includes the follow-up of patients with CVD and, therefore, with high risk to develop T2DM.

2. Material and methods

2.1. Study cohort

This work has been conducted in the context of the CORDIOPREV study (Clinical Trials Registry NCT0092493741) [21]. The study protocol was approved by the Human Investigation Review Committee of the

Reina Sofia University Hospital (HURS, Cordoba, Spain), according to institutional and Good Clinical Practice guidelines. The CORDIOPREV study is a prospective, randomized, controlled trial that includes 1002 coronary heart disease (CHD) patients, who had their last coronary event over six months before joining the study. Details of the study, including inclusion and exclusion criteria and cardiovascular risk factors of the patients have been published elsewhere [21]. Briefly, patients aged 20–75 years, with established CHD but without clinical events in the last six months with no other serious illnesses and a life expectancy of at least five years were eligible, with the intention of following a long-term monitoring study. In addition to conventional treatment for CHD, subjects were randomized in two different dietary models (Mediterranean and low-fat diets). The intervention phase is still in progress, and will have a follow up of seven years. Specifically, 462 of those patients were non-T2DM at the inclusion of the study (baseline) [22] and, from those, 107 patients developed T2DM (Incident-T2DM cases) after a mean follow-up of 60 months, according to all the American Diabetes Association (ADA) diagnosis criteria, evaluated on the basis of glucose tolerance tests performed each year (Supplementary Fig. 1). T2DM was diagnosed if one or more of following criteria were present in the study subjects: fasting plasma glucose (FPG) concentration ≥ 126 mg/dL, impaired fasting glucose (IFG); FPG ≥200 mg/dL after 2 h of oral glucose test (OGTT), impaired glucose tolerance (IGT); glycated hemoglobin (HbA1c) ≥6.5%. Specifically, among the 107 incident-T2DM patients, 43 subjects were diagnosed during the first year of follow-up, 24 during the second year, 11 during the third year, 19 during the fourth year and 10 during the fifth year (Supplementary Fig. 1). The remaining 355 subjects did not develop T2DM during the study period. In the present study, all the incident-T2DM cases ($n = 107$) were included together with 108 matched, randomly selected controls, who did not develop T2DM during the follow-up (non-T2DM subjects) (Supplementary Table 1). The random selection of the non-T2DM subjects was performed using computational stratified sampling from the 355 non-T2DM subjects of the CORDIOPREV study according to the following clinical, anthropometric and biochemical variables: diet, age, gender, fasting plasma glucose, body mass index, LDL-cholesterol and HDL-cholesterol. To implement this type of sampling, the target population was first divided into separate strata and then, samples were randomly selected within each stratum through simple automatic sampling by using the R software.

2.2. Metabolic study design

Oral glucose tolerance test (OGTT) and fat-rich meal tests were implemented in all patients to dynamically determine the metabolic status of the patient, as previously reported [21]. Briefly, OGTT (75 g dextrose monohydrate in 250 ml water, NUTER. TEC GLUCOSA, Subra, Toulouse, France) was started at 8:00 am, and plasma samples were collected at 0, 30, 60, 90 and 120 min to determine plasma glucose and insulin levels. Fat-rich meal test (0.7 g fat and 5 mg cholesterol per kg body weight with 12% saturated fatty acids, 10% polyunsaturated fatty acids, 43% monounsaturated fatty acids, 10% protein, and 25% carbohydrates) was performed at 8:00 am. Blood samples were collected before the meal and after 4 h. Biochemical determination of metabolic parameters and calculation of insulin resistance and sensitivity indexes were performed as previously reported [21]. Further details are provided in Supplementary Material and Methods.

2.3. Blood sampling and processing to isolate PBMCs

Venous blood from the participants (12 h overnight fast) was collected in tubes containing EDTA during the fat-rich meal test, at 0 and 4 h. PBMCs were isolated as previously described [21,23].

2.4. RNA extraction, quantification and reverse transcribed

Total RNA from PBMCs was isolated using Direct-zol RNA kit (Zymo Research, Irvine, CA, USA) following manufacturer's instructions. The amount of RNA recovered was determined and its quality assessed by the NanoDrop2000 spectrophotometer (Thermo Fisher). Specifically, all the RNA samples passed the quality controls, being the 260/280 and 230/260 absorbance ratios among 1.8–2.0. One μg of RNA was reverse transcribed (RT) to cDNA using random hexamer primers with the First Strand Synthesis Kit (Thermo Fisher).

2.5. In vitro culturing and treatment of PBMCs

PBMCs from $n = 7$ healthy subjects (Supplementary Table 2) were extracted as described above, and then cultured in serum-free RPMI medium (Lonza). 500,000 cells/well from each subject were seeded on ultra-low attachment multi 12-well plates (Corning Costar, Sigma) and treated per duplicate with 10% baseline fasting and postprandial serum derived from control and incident-T2DM patients (specifically, we used serum form individuals that developed T2DM during the first two years of follow-up). After the incubation periods (4 and 24 h), PBMCs were centrifuged and RNA was extracted and isolated using TRI-reagent (Sigma) [23].

2.6. Analysis of splicing machinery components by qPCR dynamic array based on microfluidic technology

A 48.48 Dynamic Array based on microfluidic technology (Fluidigm, San Francisco, CA, USA) was implemented to determine the expression of 48 transcripts in 48 samples, simultaneously. Specific primers for human transcripts including components of the major ($n = 13$) and minor spliceosome ($n = 4$), associated SFs ($n = 28$) and three housekeeping genes were specifically designed (Supplementary Table 3). The panel of splicing machinery components was selected on the basis of two main criteria: 1) the relevance of the given spliceosome components in the splicing process (such as the components of the spliceosome core) and 2) the demonstrated implication in the regulation of splicing variants implicated in the pathophysiology of T2DM (as is the case of the 28 splicing factors selected in this study).

Primers were selected using Primer3 software with selection parameters set to identify primer pairs that: 1) span an intron (when possible), 2) differ by no more that 1 °C in annealing temperature, 3) are at least 20 bp in length, 4) have a GC content between 45 and 55%, but 5) exclude primers that may form primer-dimers. Sequences of selected primers were used in BLAST (NCBI) searches to check for potential homology to sequences other than the designated target. Initial screening of primer efficiency using real-time detection was performed by amplifying 2-fold dilutions of RT products, where optimal efficiency was demonstrated by a difference of one cycle threshold between dilutions and a clear melting peak followed by a graded temperature-dependent dissociation to verify that only one product was amplified. The thermocycling profile consisted of: 1) 95 °C for 1 min; 2) 35 cycles of denaturing (95 °C for 5 s) and annealing/extension (60 °C for 20s); and 3) a last cycle where final PCR products were subjected to graded temperature-dependent dissociation (60 °C to 95 °C, increasing 1 °C/3 s). PCR products were then column-purified (FAVORGEN Biotech, Vienna, Austria) and sequenced to confirm target specificity. After confirmation of primer efficiency and specificity, the concentration of purified products was determined, and PCR products were serial diluted to obtain standards containing 1, 10^1,10^2, 10^3, 10^4, 10^5, and 10^6 copies of the synthetic template. Standards were then amplified by qPCR, and standard curves were generated using Stratagene Mx3000p software. The slope of a standard curve for each template examined was approximately −3.33 ($R2 \approx 1$), indicating that the efficiency of amplification of our primers was 100%, meaning that all templates in each cycle were copied.

Preamplification, exonuclease treatment and qPCR dynamic array based on microfluidic technology were implemented following manufacturer's instructions using the Biomark System and the Real-Time PCR Analysis Software (Fluidigm). Additional details are provided in Supplementary Material and Methods. Finally, in the case of PBMC samples from non-T2DM and Incident-T2DM patients, the expression level of each transcript was adjusted by a normalization factor (NF) obtained from the expression levels of two different housekeeping genes [Beta-actin (*ACTB*) and Glyceraldehyde-3-Phosphate Dehydrogenase (*GAPDH*)] using Genorm 3.3 [24], while *ACTB* was the housekeeping gene used to normalize the expression level of each transcript in the *in vitro* experiment. This selection was based on the stability of these housekeeping genes among the experimental groups to be compared, wherein the expression of these housekeeping genes was not significantly different among groups.

2.7. Statistical and bioinformatical analysis

Data were evaluated for heterogeneity of variance using the Kolmogorov–Smirnov test and are expressed as mean ± SEM. Statistical analysis was carried out using unpaired *t*-test, Mann Whitney *U* test or one-way ANOVA followed by Dunnett's test. Significant correlations were studied using bivariate Spearman correlation methods. The odds ratio (OR) was calculated by logistic regression analysis by comparing T1 vs. T3. The subjects were categorized into tertiles of expression levels as low (T1), intermediate (T2) and high (T3). Predictive models were constructed by logistic regression (first with the SPSS software and later validated with the R language), Random Forest and C4.5 (both with R language) algorithms and followed by cross-validation studies by using the full-dataset of variables or a selection of them (obtained by feature ranking processes) as described in Supplementary Material and Methods. AUCs from ROC curves were compared by DeLong test [25]. *P*-values smaller than 0.05 were considered statistically significant. When appropriate, analyses were adjusted for age, diet, gender, BMI, glycated hemoglobin, HDL and triglycerides. Statistical analyses were carried out with GraphPad Prism 6 (La Jolla, CA, USA) and SPSS 17.0 (IBM).

3. Results

3.1. Description of the cohort

The group of patients who developed T2DM during the study (Incident-T2DM; $n = 107$) presented higher weight, BMI, HbA1c levels, fasting insulin and TGs at baseline compared to non-T2DM controls ($n = 108$) (Table 1). HOMA-IR, HIRI, DI and ISI were also significantly different at baseline between both groups (Table 1). Remarkably, the cohort of n = 108 randomly selected non-T2DM patients showed comparable levels of all the parameters determined except for age and DI compared to the total population ($n = 355$) of non-T2DM individuals (Supplementary Table 1).

Table 1
Baseline characteristics of subjects who did not develop T2DM (Non-T2DM) vs subjects who developed T2DM (Incident-T2DM) after a median follow-up of 5 years.

Variables	Non-T2DM	Incident-T2DM	*p*-value
n	108	107	
Sex (male; female)	93; 15	89; 18	0.550
Diabetes family history n (%)	35 (32.4)	41 (38.3)	0.220
Age (years)	60.30 ±0.806	58.75 ±0.873	0.191
Waist circumference (cm)	102.48 ±0.958	105.28 ±1.08	0.053
Weight (kg)	81.92 ±1.194	85.69 ±1.47	0.047
BMI (kg/m2)	30.16 ±0.362	31.39 ±0.466	0.038
Glucose (mg/dl)	94.4 ±0.952	96.18 ±1.403	0.208
HbA1c (%)	5.88 ±0.032	6.03 ±0.033	0.001
Insulin (mU/l)	8.07 ±0.514	10.5 ±0.656	0.004
TG (mg/dl)	109.24 ±4.699	132.60 ±6.608	0.004
Total cholesterol (mg/dl)	159.55 ±3.027	164.97 ±3.409	0.235
c-LDL (mg/dl)	91.20 ±2.38	93.4 ±2.657	0.538
c-HDL (mg/dl)	44.58 ±0.899	43.52 ±1.039	0.440
NEFA (mmol/L)	0.286 ±0.015	0.317 ±0.016	0.174
Apo A1 (mg/dl)	133.5 ±2.093	135.15 ±2.312	0.596
Apo B (mg/dl)	71.57 ±1.934	76.22 ±1.835	0.083
hs-CRP (mg/dl)	2.428 ±0.32	2.878 ±0.292	0.300
HOMA-IR	2.5424 ±0.126	3.3734 ±0.302	0.012
HIRI	1024.55 ±50.85	1370.214 ±120.93	0.009
MISI ($x10^2$)	0.021 ±0.002	0.019 ±0.002	0.402
DI	0.8948 ±0.041	0.7685 ±0.041	0.030
ISI	4.0815 ±0.256	3.2758 ±0.186	0.012
IGI	1.0646 ±0.103	0.6633 ±0.294	0.200

Values expressed as mean ± SEM. *BMI:* Body mass index; *HbA1c:* Glycated hemoglobin; *TG:* Triglycerides; *c- LDL:* Low density lipoprotein; *c-HDL:* High density lipoprotein; *NEFA:* Non-esterified fatty acids; *Apo A1:* Apolipoprotein A1; *Apo B:* Apolipoprotein B; *hs-CRP:* High sensitivity C-reactive protein; *HOMA-IR:* Homeostasis model assessment- insulin resistance; *HIRI:* Hepatic insulin resistance index; *MISI:* Muscle insulin sensitivity index; *DI:* Disposition index; *ISI:* Insulin sensitivity index; *IGI:* Insulinogenic index.

3.2. Expression of splicing machinery components is different between incident-T2DM and non-T2DM patients

The expression pattern of several spliceosome components and SFs was dysregulated in the PBMCs of incident-T2DM compared to non-T2DM patients at baseline (Fig. 1a). Specifically, PBMCs of incident-T2DM patients exhibited significantly lower levels of *RNU2*, *RNU4*, *RNU6ATAC*, *SNRNP200*, *ESRP1*, *SRSF1* and *SRSF5* (Fig. 1b).

Most notably, expression of many of these spliceosome components and SFs was dynamically, and differentially, regulated in the PBMCs of these individuals during a fat-rich meal test (a stress condition that challenges phenotypic flexibility, enabling its analysis) (Supplementary Fig. 2). Specifically, some spliceosome components and SFs were selectively altered during the post-prandial phase compared to the baseline state in a specific group of patients (e.g. *RNU12* was increased in non-T2DM patients, whereas *SNRNP70* was increased in incident-T2DM patients), leading to the appearance of significant differences between both groups of patients (Supplementary Fig. 2a). Interestingly, several of the changes observed at baseline (*RNU2*, *RNU4*, *RNU6ATAC*, *ESRP1* and *SRSF1*) were also evident in the postprandial state, while others did not (*SNRNP200* and *SRSF5*) (Supplementary Fig. 2B and C, respectively).

Comparison between the expression profiles of PBMCs from non-T2DM and incident-T2DM patients at post-prandial state revealed a drastic dysregulation of many spliceosome components and SFs after this meal challenge (Fig. 2a). Specifically, expression levels of *RNU2*, *RNU4*, *RNU6*, *RNU4ATAC*, *RNU6ATAC*, *RNU12*, *NOVA1*, *ESRP1* and *SRSF1* were lower, whereas those of *SNRNP70* were higher in PBMCs from incident-T2DM patients vs. non-T2DM during the post-prandial phase (Fig. 2b). Remarkably, these fasting and postprandial changes observed between non-T2DM and incident-T2DM patients were also observed when only considering the male population, which were age and BMI matched (data not shown).

3.3. Expression of splicing machinery components is different according to the year of T2DM diagnosis

Subjects diagnosed of T2DM in the first 2-years of follow-up exhibited more pronounced changes in the expression pattern of spliceosome components and SFs as compared to non-T2DM controls and subjects diagnosed in subsequent years, at both baseline fasting and post-prandial states (Supplementary Figs. 3 and 4). Specifically, fasting *RNU4*, and post-prandial *RNU4*, *RNU4ATAC* and *RNU6ATAC* expression levels were altered in patients who developed T2DM during the first and second years of follow-up compared to non-T2DM patients (Supplementary Figs. 3 and 4). In contrast, fasting *RNU2*, *SNRNP200*, *RNU6ATAC* and *SRSF5* and post-prandial *RNU2*, *RNU6*, *NOVA1*, *ESRP1* and *SRSF1* expression levels were lower only in patients

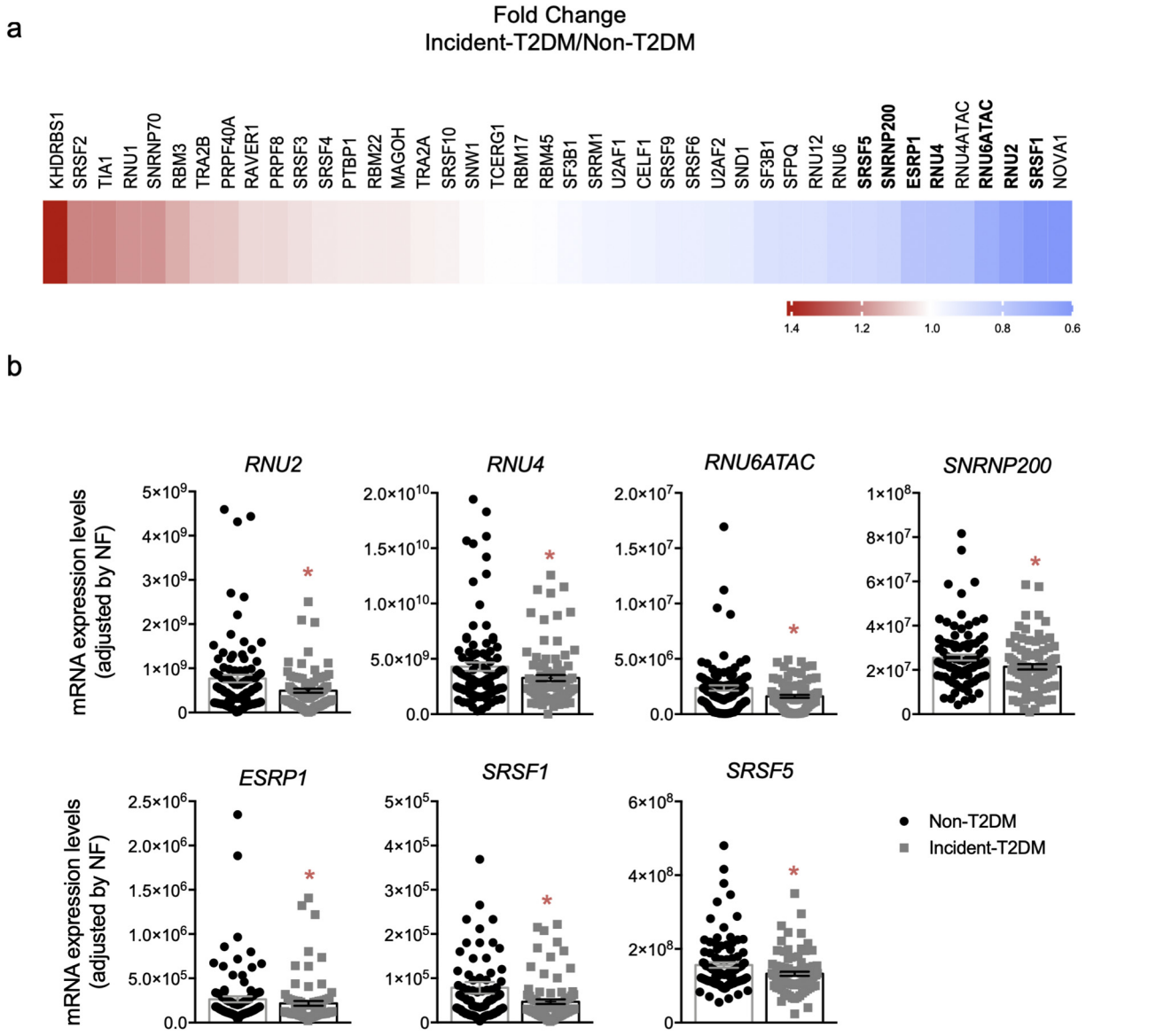


Fig. 1. Baseline expression pattern of the selected spliceosome components and splicing factors in the PBMCs of Incident-T2DM and control non-T2DM patients. a) Fold-change levels between Incident-T2DM and non-T2DM subjects, represented in red (increase) or blue (decrease). Specific spliceosome components or splicing factors significantly altered are highlighted in bold. b) mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *GAPDH* and *ACTB*] of specific spliceosome components (first row) and splicing factors (second row) in the PBMCs from non-T2DM and Incident-T2DM subjects. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (*t*-test: *, $p < .05$).

that developed T2DM during the first year of follow-up compared to non-T2DM patients (Supplementary Figs. 3 and 4). Also, post-prandial expression levels of *RNU1* and *RNU12* were significantly altered during the second year of follow-up in patients that developed T2DM compared to non-T2DM patients (Supplementary Fig. 4).

3.4. Baseline expression of splicing machinery components was correlated with key clinical parameters

Baseline expression of several spliceosome components and SFs was correlated with clinically relevant parameters, thus suggesting a potential pathophysiological role. Although the baseline fasting or postprandial expression levels of the splicing components were not associated with BMI, baseline fasting levels of *ESRP1*, *SRSF1* and *RNU4* was directly correlated with fasting c-HDL and NEFA levels. Furthermore, fasting levels of *SRSF1* and *RNU4* were directly correlated with postprandial NEFA levels and fasting levels of *SNRNP200* were directly correlated with postprandial glucose. Fasting *SRSF5* expression levels were directly correlated with fasting and postprandial c-HDL, postprandial Apo A1 and ISI values, and inversely correlated with fasting HbA1c and postprandial triglycerides and C-reactive protein. Additionally, baseline postprandial expression levels of *ESRP1*, *RNU4*, *RNU6* and *NOVA1* were inversely correlated with HbA1c levels, while postprandial levels of *RNU2* were directly correlated with DI and postprandial *RNU12* levels inversely correlated with ISI. Finally, postprandial *SRSF1* levels were directly correlated with postprandial NEFA levels (Supplementary Table 4).

3.5. Expression of some splicing machinery components was associated to the risk of developing T2DM

The possible association between the expression of the components of the splicing machinery and the risk of developing T2DM was assessed by logistic regression analysis (odds ratio, OR). This analysis revealed

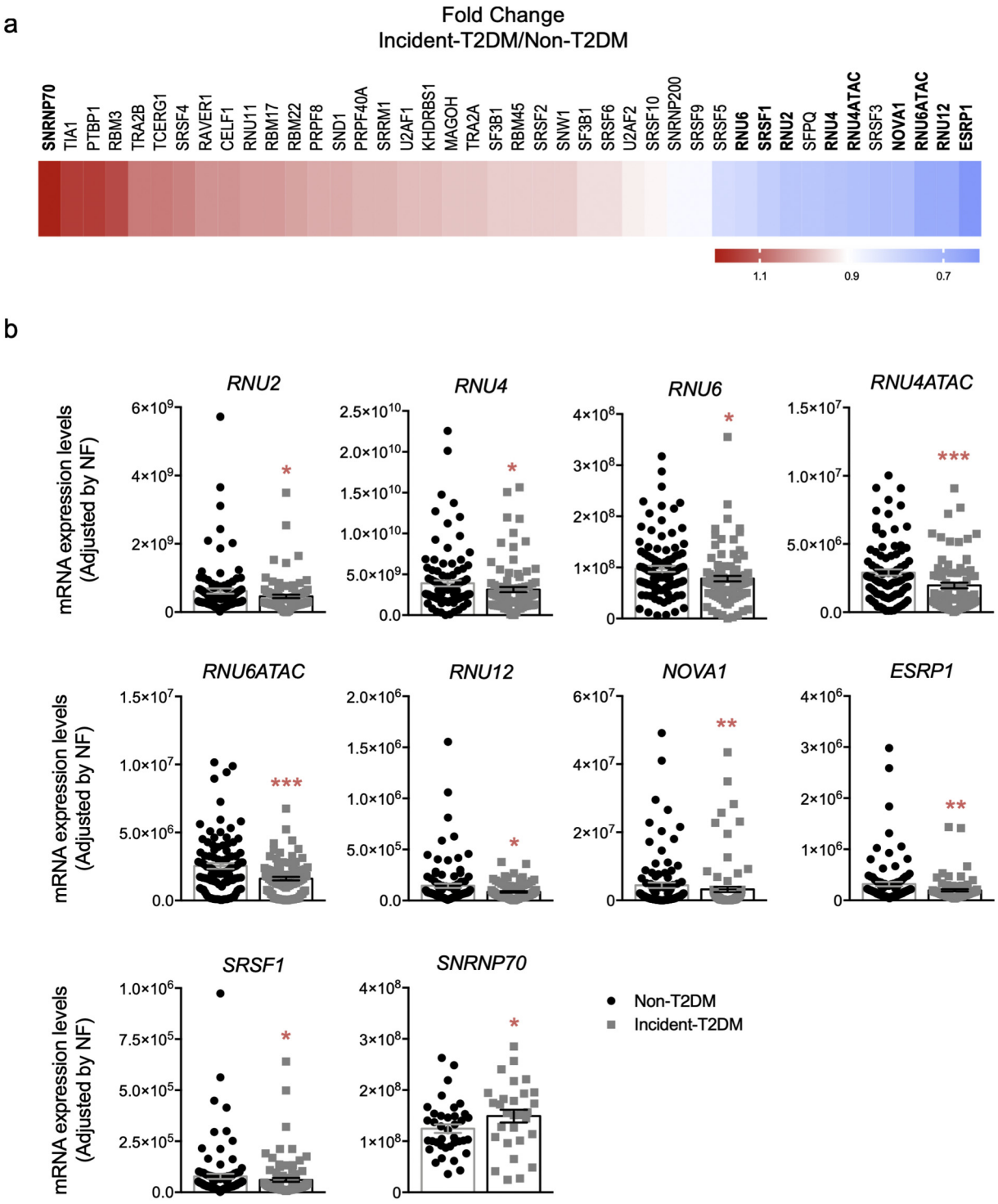


Fig. 2. Baseline expression pattern of the selected spliceosome components and splicing factors in the PBMCs of incident-T2DM and control non-T2DM patients during the postprandial state. a) Fold-change levels between Incident-T2DM and non-T2DM subjects, represented in red (increase) or blue (decrease). Specific spliceosome components or splicing factors significantly altered are highlighted in bold. b) mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *GAPDH* and *ACTB*] of specific spliceosome components and SFs in the PBMCs of non-T2DM and Incident-T2DM subjects. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, $p < .05$; **, $p < .01$; ***, $p < .001$).

that low baseline fasting expression levels of *RNU4* and *RNU2*, as well as low baseline postprandial levels of *RNU4ATAC*, *NOVA1*, *RNU6ATAC*, *RNU6*, *RNU12*, *SRSF1* and *RNU4* were strongly associated to the relative risk of T2DM development when adjusting by age, gender, diet, BMI, HbA1c, c-HDL and TG levels (Table 2).

3.6. Alterations in the expression of the splicing machinery components predict the development of T2DM

In our cohort, incident-T2DM and non-T2DM patients exhibited different baseline characteristics (Table 1); however, HbA1c and/or

Table 2
Association between the PBMC expression of the components of the splicing machinery and the relative risk of developing T2DM by logistic regression analysis of relative risk (odds ratio, OR).

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

Subjects were categorized in tertiles according to the expression level of each spliceosome component or splicing factor as follows: low expression levels (T1), medium expression levels (T2) and high expression levels (T3), and the OR estimated between T1 and T3 for each element of interest. *OR*: Odds Ratio; *C.I.*: Confidence intervals.

FINDRISK, the classic predictors of T2DM development, exhibited low predictive potential (HbA1c: AUC = 0.643, $p < .001$, TPR = 0.669, TNR = 0.565; FINDRISK: AUC = 0.548, $p = .231$, TPR = 0.622, TNR = 0.393; HbA1c and FINDRISK: AUC = 0.643, $p < .001$, TPR = 0.651, TNR = 0.606) (Table 3). In this sense, although the expression levels of spliceosome-associated elements altered in PBMCs of incident-T2DM compared to non-T2DM patients generated significant but low (<0.65) AUCs in ROC curves (Supplementary Table 5), the combination of all these elements led to more significant and accurate ROC curves. Specifically, the fingerprints comprised by the baseline expression of spliceosome components and SFs during fasting, post-prandial or their combination generated significant ROC curves with AUC > 0.85 using different algorithms (Table 3 and Supplementary Table 6), which was validated (AUCs ranging 0.65–0.81) by cross-validation analysis (Supplementary Table 6). In particular, the fingerprint comprised by the fasting and postprandial expression of spliceosome components and SFs presented an AUC = 0.813 (TPR = 0.802, TNR = 0.689) in the cross-validation analysis (Fig. 3a – green line), which was significantly higher than the capacity of the clinically-relevant HbA1c and FINDRISK to predict T2DM development (Fig. 3a), as demonstrated by DeLong test comparing the AUCs ($p < .05$ vs. HbA1c or FINDRISK).

Table 3
Capacity of the molecular fingerprint comprised by baseline fasting and/or postprandial levels of spliceosome components and splicing factors as T2DM predictive models by logistic regression and ROC curve analysis.

Model	AUC	*p*-value
Splicing machinery components baseline expression during fasting	0.894	0.000
Splicing machinery components baseline expression during postprandial phase	0.853	0.000
Splicing machinery components baseline expression during fasting and postprandial phase	1	0.000
HbA1c	0.643	0.000
FINDRISK	0.548	0.231
HbA1c + FINDRISK	0.643	0.000
Splicing machinery components baseline expression during fasting + HbA1c	0.898	0.000
Splicing machinery components baseline expression during postprandial phase + HbA1c	0.867	0.000
Splicing machinery components baseline expression during fasting and postprandial phase + HbA1c	1	0.000
Splicing machinery components baseline expression during fasting + FINDRISK	0.895	0.000
Splicing machinery components baseline expression during postprandial phase + FINDRISK	0.856	0.000
Splicing machinery components baseline expression during fasting and postprandial phase + FINDRISK	1	0.000

Logistic regression models considering the fasting and/or postprandial baseline levels of the measured elements, alone or in commination with the classic predictors of T2DM (HbA1c and FINDRISK). *AUC:* Area under curve; *HbA1c:* Glycated hemoglobin.

More remarkably, a subset of splicing machinery elements selected by computational approaches (feature ranking analysis) exhibited even higher predictive capacity compared to the whole dataset. Indeed, a predictive model comprised by fasting expression levels of *RNU4ATAC*, *SRSF3*, *SRSF6*, *SRSF10*, *TRA2B*, *PTBP1*, *SF3B1* and post-prandial levels of *RNU4*, *RNU6*, *RNU4ATAC*, *RAVER1*, *NOVA1*, and *PRPF8* exhibited AUC = 0.881 (TPR = 0.801, TNR = 0.700) in cross-validation analysis (Fig. 3b – green line), which was validated by different modeling methods, clearly outperformed the capacity of the classic predictors of T2DM development ($p < .001$ by DeLong test vs. HbA1c or FINDRISK AUCs) (Fig. 3 b).

Finally, the capacity of splicing machinery alterations to predict T2DM development was even superior when considering only the patients that developed T2DM during the first 2 years of follow-up. Indeed, the fasting and/or postprandial baseline fingerprints comprised by spliceosome components and SFs generated ROC curves with AUC ranging 0.630–0.851 by cross-validation analysis (Fig. 3c and Supplementary Table 6), which were again clearly higher than those generated by HbA1c or FINDRISK ($p < .05$ by DeLong test in all cases). Similarly, models generated after selection of the most relevant elements exhibited AUCs>0.8 when predicting T2DM development in the first two years of follow-up (Fig. 3d), which were clearly higher than those generated by HbA1c or FINDRISK ($p < .001$ by DeLong test in all cases).

3.7. Incident-T2DM patient serum altered spliceosome components expression in PBMCs of healthy patients

To test whether changes in the expression of spliceosome components and SFs in PBMCs could be caused by the metabolic dysregulation occurring in incident-T2DM patients, PBMCs from healthy subjects (Supplementary Table 1) were incubated with serum from non-T2DM and incident-T2DM patients. Remarkably, 24 h incubation with baseline fasting serum from Incident-T2DM patients markedly reduced *RNU4* expression compared to the serum from non-T2DM patients (Fig. 4a); whereas expression of other spliceosome components and SFs altered at fasting baseline was not significantly modulated (Supplementary Fig. 5). More strikingly, 4 h incubation with baseline postprandial serum from Incident-T2DM patients induced a significant reduction of *RNU4* and *RNU12* expression and a non-significant reduction trend of *RNU2* ($p = .09$) compared to non-T2DM treated samples (Fig. 4b). It should be noted that PBMCs survival was evaluated to assess the possible effect of human sera and that after 24 h of culture with baseline fasting and postprandial serum derived from control and incident-T2DM patients, the survival of the PBMCs was minimally affected (>95% of cell survival in all cases) by the different sera used and that there were not significant changes among the different types of serum (data not shown).

4. Discussion

In this study, we analyzed the expression pattern of selected splicing machinery elements in PBMCs from non-T2DM patients with CVD and, therefore, with high risk to develop T2DM, who were followed-up during 5-years (individuals that suffered a cardiovascular event and were included in the CORDIOPREV study [21]). This revealed, for the first time, that PMBCs expression levels of certain splicing machinery components were significantly altered in Incident-T2DM at the inclusion of the study (before the development of T2DM) and associated with T2DM development. Indeed, the molecular fingerprints comprised by the fasting and, especially, by the postprandial levels of certain of these splicing machinery components were able to differentiate between patients who subsequently developed T2DM (Incident-T2DM) from those patients who did not develop the disease with an AUC > 0.8 after cross-validation analysis, which is significantly higher than

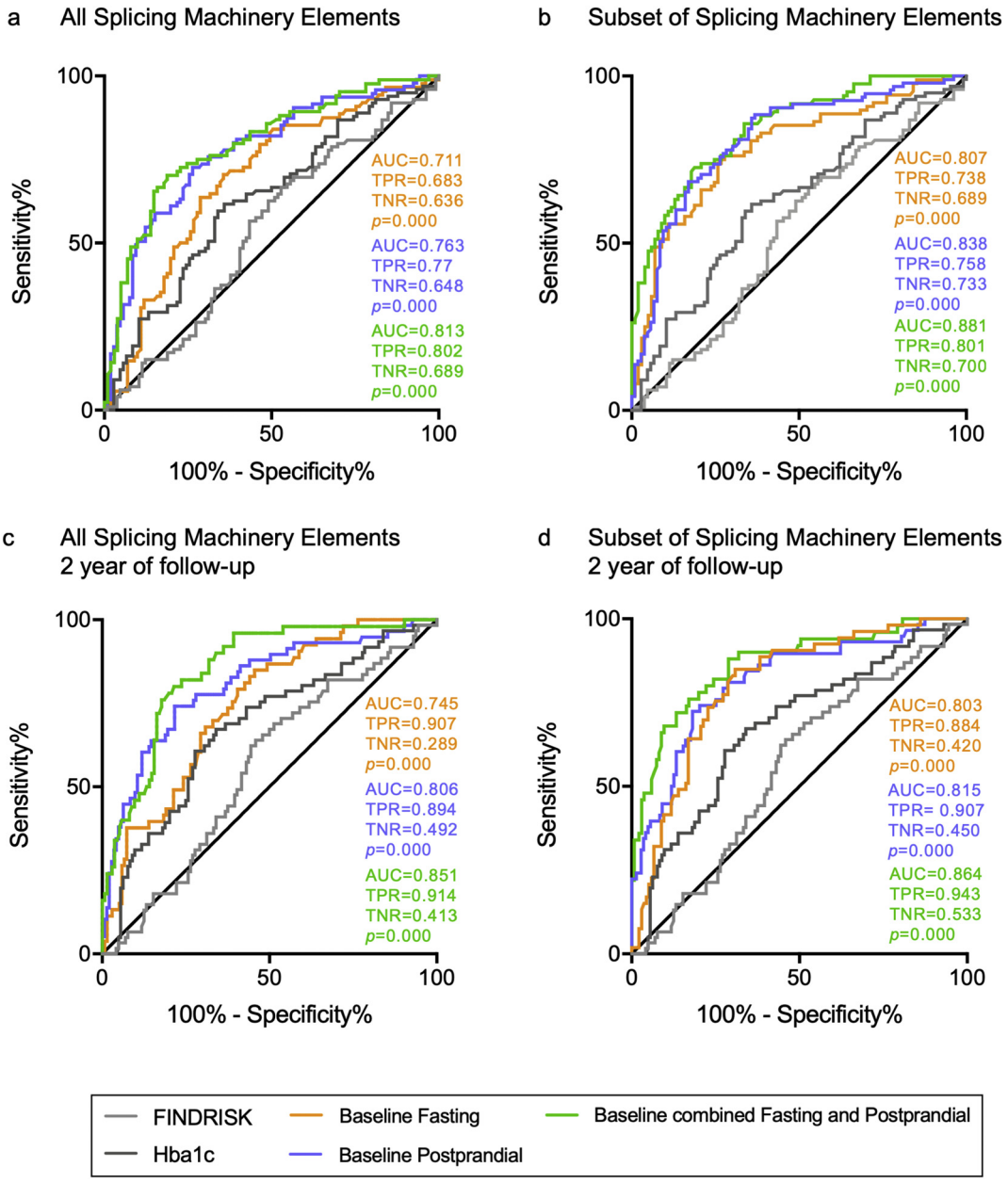


Fig. 3. Spliceosome components and splicing factors-based predictive models generated by Random Forest computational algorithm and ROC curve analysis. ROC curves parameters were calculated for the predictive models generated by Random Forest algorithm considering the expression of all the splicing machinery elements determined at fasting (orange), postprandial (blue) or their combination (green) in non-T2DM and Incident-T2DM patients (a) or using a selection of the most relevant and discriminatory splicing machinery components (b). Specifically, the subset of specific splicing machinery components were *RNU4ATAC, TIA1, KHDRBS1, SRSF10, PTBP1, RAVER1, RNU2, RNU5, SRSF9, U2AF2, RBM45, SRSF4, RBM3* for baseline fasting, *RNU4ATAC, RNU6ATAC, RAVER1, SF3B1, SRSF3, NOVA1, SRM160, SRSF6, ESRP1, U2AF1* for baseline postprandial and fasting *SRSF3, SRSF10, SRSF6, TRA2B, PTBP1, SF3B1* and postprandial *RNU4ATAC, RAVER1, RNU4ATAC, NOVA1, RNU4, RNU6, PRPF8* for combined analysis. The same ROC curves were calculated considering the patients that developed T2DM during the first two years of follow-up (c and d). In this case, the subset of specific splicing machinery components were *RNUU4ATAC, PTBP1, TRA2A, RM17, RNU12, TIA1, SRSF5, RNU2* for baseline fasting, *RNU4ATA, PTBP1, MAGOH, SRSF9, RAVER1, PRP8, PRPF40A, SRRM1, SRSF6, SNRNP200, TIA1, RNU2* for baseline postprandial, and baseline fasting *RNU4ATAC, TRA2A, SRSF5, RBM17, SRSF10, SRSF3, RNU2* and postprandial *RNU4ATAC, RNU4, RNU6ATAC, MAGOH, RAVER1, PRPF40A, RBM4, U2AF2, SRSF10, RNU11 TRA2B, SND1* for combined analysis. ROC curves of HbA1c and FINDRISK were also estimated. *HbA1c:* glycosylated hemoglobin; *AUC*: Area under the curve; *TPR*: True positive ratio; *TNR*: True negative ratio; *p*: *p* value.

the current standard procedures (HbA1c and FINDRISK exhibited AUCs<0.66 in our cohort). In addition, patients with low PBMCs expression levels of specific splicing machinery components, including *RNU2*, *RNU4* or *RNU12*, which we have shown to be modulated by the serum of Incident-T2DM patients, were at higher risk (OR > 4) of T2DM development as compared with those patients with high PBMC levels of these components.

In line with our findings, previous reports have indicated that under adverse metabolic conditions (e.g. obesity, insulin resistance, etc.) the splicing machinery is markedly altered in most tissues [10–12] and associated with the development of several pathologies [10,13,14]. Actually, alternative splicing seems to reside at the crossroad between hyperinsulinemia, insulin resistance, obesity and T2DM [11,12,17], and, consequently, the correct function of the splicing machinery (spliceosome components and SFs) is essential to maintain whole body homeostasis [18]. However, we provide herein novel, primary evidence demonstrating that the alteration in the expression of certain spliceosome components and SFs precedes the development of T2DM in CVD patients and, most notably, our results provide the first indication that the expression of certain spliceosome components and SFs in

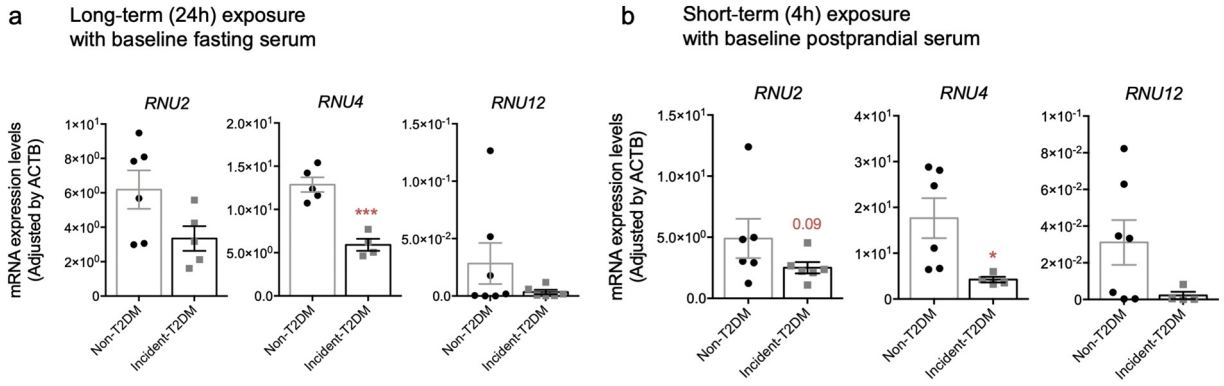


Fig. 4. Expression levels of key spliceosome components in PBMCs from healthy subjects treated with non-T2DM and Incident-T2DM derived baseline fasting during 24 h (a) or postprandial serum during 4 h (b). mRNA expression levels, adjusted by *ACTB* expression levels, of specific spliceosome components. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from PBMCs treated with non-T2DM derived serum (*t*-test: *, $p < .05$; ***, $p < .001$).

PBMCs can be dynamically modulated, according to the metabolic status of the individual, and specially under stress conditions as those observed during the post-prandial response, thus suggesting an adaptive capacity of the splicing machinery in order to respond to metabolic disturbances.

In particular, this study demonstrated that the expression of some splicing machinery components is altered in control individuals during the postprandial phase, which is consistent with previous results demonstrating a regulatory response of specific splicing variants to the postprandial environment [26,27]. Therefore, these data suggest that the changes observed in the splicing machinery during the postprandial phase may be responsible for the regulation of the expression of particular splicing variants under these conditions [12,19,28], which could be essential to the appropriate response of the organism to metabolic challenges and disturbances. Moreover, our results also demonstrate for the first time that the response of key splicing machinery components to metabolic insults is altered in individuals who will develop T2DM (Incident-T2DM patients), but especially in those developing T2DM during the first two years of follow-up. Therefore, since postprandial alterations are closely related to the phenotypic flexibility, which is strongly linked to T2DM development [9], our data primarily demonstrate that the alteration in the splicing machinery precedes the instauration of T2DM, thereby suggesting its putative implication as a driving force in the development of this pathology. Based on all the information mentioned above, it is tempting to propose that the splicing machinery could be acting as a biosensor of the whole body metabolism to adapt cell gene expression to the pathophysiological conditions, and that its dysregulation could lead to an unbalance in the landscape of splicing variants present in a given cell at a given moment [12,28,29], which may be associated to the instauration of T2DM [10,30]. This idea is further supported by two pieces of evidence presented herein. First, we have demonstrated that low fasting or postprandial expression levels of certain splicing machinery components drastically increase the relative risk of T2DM development in these patients, suggesting that a dysregulated expression of certain splicing machinery components could augment the risk of developing T2DM. Secondly, our *in vitro* studies demonstrate that the expression of relevant spliceosome components, specially *RNU2*, *RNU4* or *RNU12*, which are key elements responsible for the appropriate function of the spliceosome [31], can be modulated by baseline fasting and/or postprandial serum from Incident-T2DM patients. This observation might suggest the existence of specific factors in the serum of these patients capable to modulate the expression of relevant spliceosome components and, therefore, the function of the splicing machinery. In this sense, previous studies have found a relationship between circulating factors and the modulation of SFs in different tissues. For example, it has been described that insulin signaling can up-regulate the expression of the splicing factor *SRSF1* in pancreatic beta cells, inducing the splicing of the insulin receptor to generate the INSR-B isoform [32]. The same study also found a regulation of the protein levels of the splicing factor MBNL1 by high glucose levels. Furthermore, the splicing of the Fatty acid desaturase 3 has been observed to be modulated in the liver of baboons in response to different diets and in human liver HepG2 cells after treatment with polyunsaturated fatty acids [33]. Obviously, further studies will be required to attain a more comprehensive understanding of the changes in the splicing process and the contribution of the dysregulation in the splicing machinery to the generation of alternative spliced isoforms bearing pathological implications; as well, additional, and more detailed and refined experiments will be needed to assess the specific contribution of each blood cell subset to the effects observed herein. Ultimately, all of this information would help to elucidate the nature of the elements causing the changes in the splicing machinery.

In conclusion, although this study was implemented in CVD patients from the CORDIOPREV study [21], which limits our findings to people with these characteristics, and precludes generalization to healthy people, the data presented herein strongly support the notion that altered expression of spliceosome components and SFs variants may be associated with the development of T2DM, preceding the clinical instauration of this pathology and, therefore, could likely serve as a sensor and early predictor for T2DM development in CVD patients. Certainly, our data demonstrate the existence of a spliceosome-associated molecular fingerprint capable to predict the future development of T2DM in individual patients with high precision (AUC = 0.881, TPR = 0.801, TNR = 0.700), which even outperforms the capacity of classical predictors of T2DM development, such as HbA1c or FINDRISK. Therefore, this splicing machinery-associated molecular fingerprint could become a valuable, non-invasive, new tool for early risk assessment of T2DM in clinical practice to prevent disease development.

Role of the funding sources

This work was funded by Instituto de Salud Carlos III, co-funded by European Union (ERDF/ESF, "Investing in your future") [PIE14/00005, PIE 14/00031, PI16/00264, CP15/00156, PI17/002287], MINECO (BFU2016-80360-R, TIN2017–83445-P, PI13/00023, AGL2012/39615, AGL2015-67896-P), Junta de Andalucía (BIO-0139, CTS-1406, CTS-525, PI-0541-2013, CVI-7450) and CIBERobn. CIBER is an initiative of Instituto de Salud Carlos III, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain. The CORDIOPREV study is supported by the Fundación Patrimonio Comunal Olivarero, Junta de Andalucía (Consejería de Salud, Consejería de Agricultura y Pesca, Consejería de Innovación, Ciencia y Empresa), Diputaciones de Jaén y Córdoba, Centro de Excelencia en Investigación sobre Aceite de Oliva y Salud and Ministerio de Medio Ambiente, Medio Rural y Marino, Gobierno de España.

The following funding sources were essential for data collection, analysis, interpretation, and patient recruitment: PIE14/00005, PIE 14/

00031, TIN2017-83445-P, PI13/00023, AGL2012/39615, AGL2015–67896-P, CTS-525, CVI-7450.

The following funding sources were essential for data collection, analysis, and interpretation: CP15/00156, PI16/00264, BFU2016–80360-R, BIO-0139, CTS-1406, PI17/002287, PI-0541-2013.

The corresponding authors declare that they had full access to all the data in the study and had the final responsibility for the decision to submit for publication.

Conflicts of interest

The authors declare that they have no conflict of interest.

Authors contributions

Manuel D. Gahete contributed to the conception and design of the work, literature search, acquisition, analysis, and interpretation of the data, drafting and revising the work and final approval of the manuscript.

Mercedes del Rio-Moreno contributed to the design of the work, literature search, acquisition, analysis, and interpretation of the data, drafting and revising the work and final approval of the manuscript.

Antonio Camargo contributed to the conception and design of the work, analysis and interpretation of the data, drafting and revising the work and final approval of the manuscript.

Juan F Alcala contributed to the conception and design of the work, acquisition and analysis of the data and final approval of the manuscript.

Emilia Alors-Perez contributed to the acquisition and analysis of the data, drafting the work, and final approval of the manuscript.

Javier Delgado-Lista contributed to the conception and design of the work, acquisition of the data and final approval of the manuscript.

Oscar Reyes contributed to the analysis and interpretation of the data, drafting the work and final approval of the manuscript.

Sebastian Ventura contributed to the analysis and interpretation of the data, revising the work, and final approval of the manuscript.

Pablo Perez-Martinez contributed to the conception and design of the work, literature search and interpretation of the data; revising the work and final approval of the manuscript.

Justo P. Castaño contributed to the conception and design of the work, literature search and interpretation of the data, revising the work and final approval of the manuscript.

Jose Lopez-Miranda contributed to the conception and design of the work, literature search and interpretation of the data, revising the work and final approval of the manuscript.

Raúl M. Luque contributed to the conception and design of the work, literature search and interpretation of the data, drafting and revising the work and final approval of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.10.056.

References

[1] World Health Organization. World Health Organization: Global Status Report on Noncommunicable Diseases. Geneva: World Health Organization; 2014.
[2] Martin-Timon I, Sevillano-Collantes C, Segura-Galindo A, Del Canizo-Gomez FJ. Type 2 diabetes and cardiovascular disease: have all risk factors the same strength? World J Diabetes 2014;5(4):444–70.
[3] Blanco-Rojo R, Perez-Martinez P, Lopez-Moreno J, Martinez-Botas J, Delgado-Lista J, Van-Ommen B, 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;7(1):12499.
[4] Camargo A, Jimenez-Lucena R, Alcala-Diaz JF, Rangel-Zuniga OA, Garcia-Carpintero S, Lopez-Moreno J, et al. Postprandial endotoxemia may influence the development of type 2 diabetes mellitus: from the CORDIOPREV study. Clin Nutr 2018 [in press].
[5] Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. BMJ 2012;345:e5900.
[6] Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011;9:103.
[7] Cohen RM, Haggerty S, Herman WH. HbA1c for the diagnosis of diabetes and prediabetes: is it time for a mid-course correction? J Clin Endocrinol Metab 2010;95(12):5203–6.
[8] Brodovicz KG, Dekker JM, Rijkelijkhuizen JM, Rhodes T, Mari A, Alssema M, et al. The Finnish Diabetes Risk score is associated with insulin resistance but not reduced beta-cell function, by classical and model-based estimates. Diabet Med 2011;28(9):1078–81.
[9] Stroeve JHM, van Wietmarschen H, Kremer BHA, van Ommen B, Wopereis S. Phenotypic flexibility as a measure of health: the optimal nutritional stress response test. Genes Nutr 2015;10(3):13.
[10] Dlamini Z, Mokoena F, Hull R. Abnormalities in alternative splicing in diabetes: therapeutic targets. J Mol Endocrinol 2017;59(2):R93-107.
[11] Webster NJG. Alternative RNA Splicing in the Pathogenesis of Liver Disease. Front Endocrinol (Lausanne) 2017;8:133.
[12] Mercader JM, Liao RG, Bell AD, Dymek Z, Estrada K, Tukiainen T, et al. A Loss-of-Function Splice Acceptor Variant in IGF2 is protective for Type 2 Diabetes. Diabetes 2017;66(11):2903–14.
[13] Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat Med 2016;22(9):976–86.
[14] Gallego-Paez LM, Bordone MC, Leote AC, Saraiva-Agostinho N, Ascensao-Ferreira M, Barbosa-Morais NL. Alternative splicing: the pledge, the turn, and the prestige : the key role of alternative splicing in human biological systems. Hum Genet 2017;136(9):1015–42.
[15] Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet 2016;17(1):19–32.
[16] Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol 2014;15(2):108–21.
[17] Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: pathogenesis and treatment. Lancet 2008;371(9631):2153–6.
[18] Juan-Mateu J, Villate O, Eizirik DL. MECHANISMS IN ENDOCRINOLOGY: Alternative splicing: the new frontier in diabetes research. Eur J Endocrinol 2016;174(5):R225–38.
[19] Belfiore A, Malaguarnera R, Vella V, Lawrence MC, Sciacca L, Frasca F, et al. Insulin Receptor Isoforms in Physiology and Disease: an Updated View. Endocr Rev 2017;38(5):379–431.
[20] Burczynski ME, Dorner AJ. Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. Pharmacogenomics 2006;7(2):187–202.
[21] Delgado-Lista J, Perez-Martinez P, Garcia-Rios A, Alcala-Diaz JF, Perez-Caballero AI, Gomez-Delgado F, et al. CORonary Diet intervention with Olive oil and cardiovascular PREVention 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;177:42–50.
[22] Blanco-Rojo R, Alcala-Diaz JF, Wopereis S, Perez-Martinez P, Quintana-Navarro GM, Marin C, et al. The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomised clinical trial. Diabetologia 2015;59(1):67–76.
[23] Gahete MD, Luque RM, Yubero-Serrano EM, Cruz-Teno C, Ibanez-Costa A, Delgado-Lista J, et al. Dietary fat alters the expression of cortistatin and ghrelin systems in the PBMCs of elderly subjects: putative implications in the postprandial inflammatory response. Mol Nutr Food Res 2014;58(9):1897–906.
[24] Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3(7).
[25] Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–45.
[26] Pfaffenbach KT, Nivala AM, Reese L, Ellis F, Wang D, Wei Y, et al. Rapamycin inhibits postprandial-mediated X-box-binding protein-1 splicing in rat liver. J Nutr 2010;140(5):879–84.
[27] Fruhbeck G, Gomez-Ambrosi J, Martinez JA. Pre- and postprandial expression of the leptin receptor splice variants OB-Ra and OB-Rb in murine peripheral tissues. Physiol Res 1999;48(3):189–95.
[28] Escribano O, Beneit N, Rubio-Longas C, Lopez-Pastor AR, Gomez-Hernandez A. The Role of Insulin Receptor Isoforms in Diabetes and its Metabolic and Vascular Complications. J Diabetes Res 2017;2017:1403206.
[29] Gortan Cappellari G, Barazzoni R, Cattin L, Muro AF, Zanetti M. Lack of Fibronectin Extra Domain a Alternative Splicing Exacerbates Endothelial Dysfunction in Diabetes. Sci Rep 2016;6:37965.
[30] Newman JRB, Conesa A, Mika M, New FN, Onengut-Gumuscu S, Atkinson MA, et al. Disease-specific biases in alternative splicing and tissue-specific dysregulation revealed by multitissue profiling of lymphocyte gene expression in type 1 diabetes. Genome Res 2017;27(11):1807–15.
[31] Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol 2017;18(11):655–70.
[32] Malakar P, Chartarifsky L, Hija A, Leibowitz G, Glaser B, Dor Y, et al. Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. Sci Rep 2016;6:31222.
[33] Reardon HT, Hsieh AT, Park WJ, Kothapalli KS, Anthony JC, Nathanielsz PW, et al. Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. Prostaglandins Leukot Essent Fatty Acids 2013;88(1):15–9.

Supplementary Material and Methods

Calculation of Insulin resistance and sensitivity indexes

The Matsuda insulin sensitivity index (ISI) was calculated from the OGTT using the following formula: ISI = 10.000 ÷ √([fasting plasma glucose X fasting plasma insulin] X [mean glucose in OGTT X mean insulin in OGTT]) [1]. HOMA-IR was calculated as previously described Song, Y., et al [2]. Insulin secretion was measured by the insulinogenic index (IGI): IGI= [30 min insulin−fasting insulin (pmol/l)]/[30 min glucose−fasting glucose (mmol/l)] [3]. Beta-cell function was estimated by calculating the disposition index (DI) as follows: DI=ISI×[AUC30 min insulin / AUC30 min glucose], where AUC30 min is the area under the curve between baseline and 30 min of the OGTT for insulin (pmol/l) and glucose (mmol/l) measurements, respectively, calculated by the trapezoidal method [4]. The indices used to determine tissue-specific IR were: the hepatic insulin resistance index (HIRI) and the muscle insulin sensitivity index (MISI), which were calculated as described in previous work by our group [5], following the methods described by Matsuda and DeFronzo for HIRI and Abdul-Ghani et al for MISI [1,6]. The FINDRISC score was calculated as previously defined [7].

Biochemical determinations of metabolic parameters

Glucose levels were determined by the hexokinase method. The hs-C-Reactive Protein (hs-CRP) was determined by ELISA (BioCheck, Inc., Foster City, CA, USA). Insulin concentrations were measured by microparticle enzyme immunoassay (Abbott Diagnostics, Matsudo-shi, Japan). Lipid variables were assessed with the modular auto analyzer DDPPII Hitachi (Roche, Basel, Switzerland) using specific reagents (Boehringer-Mannheim, Mannheim, Germany). Measurements of total cholesterol (TC) and triglycerides (TG) levels were made by colorimetric enzymatic methods [8,9], high density lipoprotein-cholesterol (HDL-c) by colorimetric assay [10] and low density lipoprotein (LDL-C) concentration was calculated by the Friedewald equation, using the following formula: LDLc = CT- (HDL + TG / 5). Non-esterified fatty acid concentrations were measured by enzymatic colorimetric assay (Roche Diagnostics, Penzberg, Germany). ApoA-1 and ApoB concentrations were determined by immunoturbidimetry.

Analysis of splicing machinery components by qPCR dynamic array based on microfluidic technology

A 48.48 Dynamic Array based on microfluidic technology (Fluidigm, San Francisco, CA, USA) was implemented to determine the expression of 48 transcripts in 48 samples, simultaneously. Specific primers for human transcripts including components of the major (n=13) and minor spliceosome (n=4), associated SFs (n=28) and three housekeeping genes were specifically designed with the Primer3 software and StepOne™ Real-Time PCR System software v2.3 (Applied Biosystems, Foster City, CA, USA) (Supplementary Table 3). Following manufacturer's instructions, 12.5ng of cDNA of each sample were pre-amplified using 1µL of PreAmp Master Mix (Fluidigm) and 0.5µL of all primers mix (500nM) in a T100 Thermal-cycler (BioRad, Hercules, CA, USA), using the following program: 1) 2 min at 95ºC; 2) 15 sec at 94ºC and 4 min at 60ºC (14 cycles). Then, samples were treated with 2µL of 4U/µL Exonuclease I solution (New England BioLabs, Ipswich, MA, USA) following manufacturer's instructions. Then, samples were diluted with 18µL of TE Buffer (Thermo Scientific), and 2.7µL were mixed with 3µL of EvaGreen Supermix (Bio-Rad) and 0.3µL of DNA Binding Dye Sample Loading Reagent (Fluidigm). Primers were diluted to 5µM with 2X Assay Loading Reagent (Fluidigm). Control line fluid was charged in the chip and Prime script program was run into the IFC controller MX (Fluidigm). Finally, 5µL of each primer and 5µL of each sample were pipetted into their respective inlets on the chip and the Load Mix script in the IFC controller software was run. After this program, the qPCR was run in the Biomark System (Fluidigm) following the thermal cycling program: 1) 95ºC for 1min; 2) 35 cycles of denaturing (95ºC for 5sec) and annealing/extension (60ºC for 20sec); and 3) a last cycle where final PCR products were subjected to graded temperature-dependent dissociation (60ºC to 95ºC, increasing 1ºC/3 sec). Data were processed with Real-Time PCR Analysis Software 3.0 (Fluidigm).

Bioinformatical analysis

1. Data preprocessing

Before conducting the computational study, the dataset was preprocessed as follows: 1) An univariate analysis of outliers was conducted and those values that lie outside 1.5 times the inter-quartile range were removed; 2) All genes with more than 70% of missing values were eliminated; 3) All missing values and outliers were replaced by the median of the variable in each group of patients; 4) All variables with zero variance were removed; 5) All variables were centered by means of subtracting the original values by the mean, and then, they were scaled by means of dividing by the standard deviation. By this way, all variables had the same impact and, therefore, the posterior estimation of the variables' relevance was not biased by those variables with extreme values.

2. Construction of predictive models with all variables

In this work, we aimed to analyze the effectiveness of the considered variables at predicting T2DM development in the population of non-T2DM patients. A logistic regression model was constructed with the SPSS software, and later validated with the R language; the aim was to confirm the high AUC values obtained by SPSS. Also, Random Forest and C4.5

models were constructed with R, serving as a point of comparison to determine the effectiveness of the aforementioned logistic regression model.

Two scenarios were considered in this work. First, the capacity of the model to memorize the data distributions and properties was studied by means of assessing the models with the same data used to construct them (SPSS implements this type of evaluation by default). Secondly, the capacity of generalization of the models was studied by means of assessing the models in a data partition that has not been seen previously (Cross-validation studies). This last scenario was performed in R language by repeating three times a 10-fold cross validation to construct the models, using the same seed to partition the database, so allowing the replication of the computational study. A single 10-fold cross-validation procedure was carried out as follows: (I) the database was partitioned in 10 parts; (II) in each fold execution the model was trained over the union of nine data partitions, and then it was tested on the remaining partition that was not used at training the model; (III) the predictive performance of the model was averaged across all fold executions.

3. Estimation of the variables' relevance

So far, the predictive models were constructed by considering all the variables of the database. However, it is well-known that a better predictive performance can be attained if redundant, noisy and interacting variables are removed at the time of constructing models [11,12]. In this work, several well-known feature-weighting algorithms [13] were used to compute the relevance of variables. The variables were ordered from higher to lower relevance according to a weight that represents the ability to distinguish between Incident-T2DM and non-T2DM patients. To avoid possible biases in the process of estimating the variables' relevance, five well-known feature estimation methods were used, namely Correlation-Feature-Selection (CFS), Correlation-Attribute-Evaluation (CAE), Gain-Ratio (GR), Information-Gain (IG), and Relief-F (RF) [14,15]. These algorithms are filter methods that evaluate the usefulness of a variable (or a set of variables) through measures of distance, dependency, information or correlation on data [11,12], so they are not influenced by classification algorithms in the feature estimation process. The use of these supervised feature weighting methods can lead to a superior estimation of the variables' relevance, having as main advantages: (I) consideration of the expert knowledge unlike of several traditional approaches, such as Principal Component Analysis (PCA), that do not exploit the a priori classification of patients; (II) detection of redundant information; (III) detection of interacting features.

In this work, the five feature-weighting methods were assessed by repeating three times a 10-fold cross-validation, and the process was implemented in the R language. In the case of RF method, it was executed with a set of number of nearest neighbors equal to [11]. The five estimation methods returned ranking of variables and, therefore, an average ranking can be computed as presented in Figure 1.

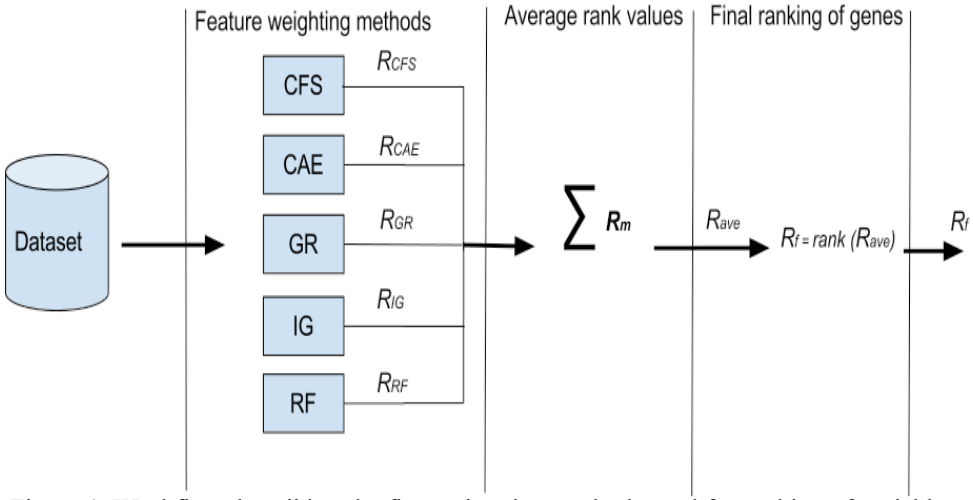


Figure 1. Workflow describing the five estimation methods used for ranking of variables and their integration to obtain the average ranking

3.1. Construction of predictive models with subsets of relevant variables

Subsequently, it was also determined the subset of variables that can best predict the development of T2DM. To determine the best subset of variables by means of a final ranking of variables, we implemented a method previously published by our group [12], which provides a heuristic for searching on feature rankings. The number of possible subsets of variables is of exponential size but through a series of steps that are performed for each sub-ranking (Figure 2), the best subset of variables is selected as this one that produces the best classifier at predicting whether a patient will develop T2DM or not.

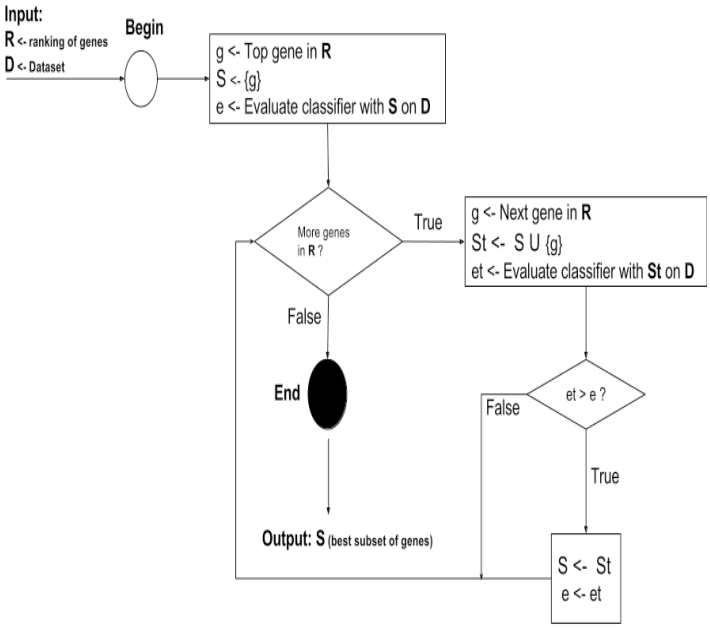


Figure 2. Workflow describing the series of steps that are performed for each sub-ranking

References

1. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes care* 1999; **22**(9): 1462-70.
2. Song Y, Manson JE, Tinker L, et al. Insulin sensitivity and insulin secretion determined by homeostasis model assessment and risk of diabetes in a multiethnic cohort of women: the Women's Health Initiative Observational Study. *Diabetes care* 2007; **30**(7): 1747-52.
3. Hanson RL, Pratley RE, Bogardus C, et al. Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *American journal of epidemiology* 2000; **151**(2): 190-8.
4. Tang W, Fu Q, Zhang Q, et al. The association between serum uric acid and residual beta -cell function in type 2 diabetes. *J Diabetes Res* 2014; **2014**: 709691.
5. Blanco-Rojo R, Alcala-Diaz JF, Wopereis S, et al. The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomised clinical trial. *Diabetologia* 2015.
6. Abdul-Ghani MA, Matsuda M, Balas B, DeFronzo RA. Muscle and liver insulin resistance indexes derived from the oral glucose tolerance test. *Diabetes care* 2007; **30**(1): 89-94.
7. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003; **26**(3): 725-31.
8. Allain CC, Poon LS, Chan CS, Richmond W, Fu PC. Enzymatic determination of total serum cholesterol. *Clinical chemistry* 1974; **20**(4): 470-5.
9. Bucolo G, David H. Quantitative determination of serum triglycerides by the use of enzymes. *Clinical chemistry* 1973; **19**(5): 476-82.
10. Briggs CJ, Anderson D, Johnson P, Deegan T. Evaluation of the polyethylene glycol precipitation method for the estimation of high-density lipoprotein cholesterol. *Ann Clin Biochem* 1981; **18**(Pt 3): 177-81.
11. Reyes O, Morell C, Ventura S. Evolutionary feature weighting to improve the performance of multi-label lazy algorithms. *Integr Comput-Aid E* 2014; **21**(4): 339-54.
12. Reyes O, Morell C, Ventura S. Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. *Neurocomputing* 2015; **161**: 168-82.
13. Wettschereck D, Aha DW, Mohri T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artif Intell Rev* 1997; **11**(1-5): 273-314.
14. Garcia S, Luengo J, Herrera F. Data Preprocessing in Data Mining. *Intel Syst Ref Libr* 2015; **72**: 1-320.
15. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* 2003; **53**(1-2): 23-69.

Supplementary Tables

Supplementary Table 1

Variables	Total Non-T2DM	Randomly selected Non-T2DM	*p- value*
n	355	108	
Sex (male; female)	304; 51	93; 15	
Age (years)	57.33 ± 0.504	60.30 ± 0.806	0.002
Waist circumference (cm)	101.73 ± 0.575	102.48 ± 0.958	0.525
Weight (kg)	82.45 ± 0.717	81.92 ± 1.194	0.577
BMI (kg/m2)	29.87 ± 0.221	30.16 ± 0.362	0.522
Glucose (mg/dl)	92.59 ± 0.531	94.4 ± 0.952	0.099
HbA1c (%)	5.86 ± 0.018	5.88 ± 0.032	0.551
Insulin (mU/l)	8.34 ± 0.305	8.07 ± 0.514	0.66
TG (mg/dl)	119.45 ±3.239	109.24 ± 4.699	0.112
Total cholesterol (mg/dl)	160.65 ± 1.612	159.55 ± 3.027	0.744
c-LDL (mg/dl)	91.1 ± 1.332	91.20 ± 2.38	0.971
c-HDL (mg/dl)	44.58 ± 0.535	44.58 ± 0.899	0.999
NEFA (mmol/L)	0.278 ± 0.009	0.286 ± 0.015	0.894
Apo A1 (mg/dl)	133.36 ± 1.144	133.5 ± 2.093	0.952
Apo B (mg/dl)	72.15 ± 0.939	71.57 ± 1.934	0.773
hs-CRP (mg/dl)	2.51 ± 0.186	2.428 ± 0.32	0.826
HOMA-IR	2.58 ± 0.091	2.5424 ± 0.126	0.825
HIRI	1052.27 ± 36.783	1024.55 ± 50.85	0.702
MISI (x10^2)	0.02 ± 0.001	0.021 ± 0.002	0.855
DI	1.01 ± 0.07	0.8948 ± 0.041	0.02
ISI	4.3 ± 0.142	4.0815 ± 0.256	0.449
IGI	1.12 ± 0.072	1.0646 ± 0.103	0.675

Supplementary table 1. Baseline characteristics of total non-T2DM patients *vs* randomly selected non-T2DM patients cohort included in the study. Values expressed as mean ± SEM. ***BMI:*** Body mass index; ***HbA1c:*** Glycated hemoglobin; ***TG:*** Triglycerides; ***c- LDL:*** Low density lipoprotein; ***c-HDL:*** High density lipoprotein; ***NEFA:*** Non esterified fatty acids; ***Apo A1:*** Apolipoprotein A1; ***Apo B:*** Apolipoprotein B; ***hs-CRP:*** High sensitivity C-reactive protein; ***HOMA-IR:*** Homeostasis model assessment of insulin resistance; ***HIRI:*** Hepatic insulin resistance index; ***MISI:*** Muscle insulin sensitivity index; ***DI:*** Disposition index; ***ISI:*** Insulin sensitivity index; ***IGI:*** Insulinogenic index.

Supplementary Table 2

Variables	Healthy Subjects
Age (years)	24.7 ± 1.32
Waist circumference (cm)	0.87 ± 0.01
Weight (kg)	81.57 ± 3.77
BMI (kg/m2)	23.53 ± 1.059
Glucose (mg/dl)	93.2 ± 2.97
HbA1c (%)	5.18 ± 0.13
Insulin (mU/l)	8.14 ± 0.62
TG (mg/dl)	74 ± 6.46
Total cholesterol (mg/dl)	169 ± 5.97
c-LDL (mg/dl)	93.2 ± 7.32
c-HDL (mg/dl)	60 ± 7.82
Apo A1 (mg/dl)	128 ±6.22
Apo B (mg/dl)	62.4 ±3.2
hs-CRP (mg/dl)	0.68 ± 0.42

Supplementary Table 2. Clinical characteristics of the healthy subjects (n=7) that donated PBMCs to test the modulation of the expression of the spliceosome components and splicing factors by metabolic dysregulations associated to T2DM patients. Values are expressed as mean ± standard error. ***BMI:*** Body mass index; ***HbA1c:*** Glycated hemoglobin; ***TG:*** Triglycerides; ***c-LDL:*** Low density lipoprotein; ***c-HDL:*** High density lipoprotein; ***NEFA:*** Non esterified fatty acids; ***Apo A1:*** Apolipoprotein A1; ***Apo B:*** Apolipoprotein B; ***hs-CRP:*** High sensitivity C-reactive protein.

Supplementary Table 3

	Gene	Accession Number	Primer Sequence (Sense. Se)	Primer Sequence (Antisense. As)	Product Size	Nucleotide Position
Splicosome Components	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
Splicing Factors	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 genes	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

Supplementary Table 3. Specific primers for human transcripts used in this study, including components of the major and minor spliceosomes, associated splicing factors and three housekeeping genes (Hk) that were specifically designed and used in qPCR-based microfluidic assays. NCBI accession number, primers sequences, product sizes and nucleotide positions for the genes studied are included.

Supplementary Table 4

		Fasting					Postprandial						
		ESRP1	***SRSF1***	***SRSF5***	***RNU4***	***SNRNP200***	***ESRP1***	***SRSF1***	***RNU4***	***RNU6***	***NOVA1***	***RNU2***	***RNU12***
Fasting c-HDL (mg/dl)	r^2	0.246	0.155	0.179	0.236								
	p	**0.001****	**0.038***	**0.016***	**0.001****								
Fasting NEFA (mmol/L)	r^2	0.159	0.157		0.186								
	p	**0.033***	**0.036***		**0.012***								
Fasting HbA1c (%)	r^2			-0.147			-0.188		-0.189	-0.194	-0.194		
	p			**0.048***			**0.009****		**0.009****	**0.007****	**0.007****		
Postprandial c-HDL (mg/dl)	r^2			0.236									
	p			**0.002****									
Postprandial NEFA (mmol/L)	r^2		0.158		0.153			0.180					
	p		**0.039***		**0.045***			**0.015***					
Postprandial Glucose (mg/dl)	r^2					0.194							
	p					**0.01****							
Postprandial TG (mg/dl)	r^2			-0.170									
	p			**0.023***									
Postprandial hs-CRP (mg/dl)	r^2			-0.197									
	p			0.008**									
Postprandial Apo A1 (mg/dl)	r^2			0.182									
	p			**0,015***									
ISI	r^2			0.160									-0.273
	p			**0.034***									**0.000****
DI	r^2											0.180	
	p											**0.014***	

Supplementary table 4. Spearman correlation between the PBMC expression levels of baseline fasting and postprandial spliceosome components and splicing factors, and relevant T2DM related parameters. ***c-HDL****: cholesterol high density lipoprotein;* ***NEFA****: non-esterified fatty acids;* ***HbA1c****:* glycated hemoglobin; ***TG****: Triglycerides;* ***hs-CRP****: High sensitivity C-reactive protein;* ***Apo A1****: Apolipoprotein A1;* ***ISI****:* insulin sensitivity index; ***DI****:* disposition index. Non significative correlations are not depicted.

Supplementary Table 5

ROC Curve Non-T2DM / Incident-T2DM		
Fasting	**AUC**	**p-value**
RNU2	0.60	0.010
SRSF5	0.60	0.013
RNU6ATAC	0.60	0.014
RNU4	0.59	0.019
SNRNP200	0.59	0.027
ESRP1	0.58	0.046
SRSF1	0.58	0.050
Postprandial	**AUC**	**p-value**
RNU4ATAC	0.65	0.0007
RNU6ATAC	0.64	0.0005
SNRNP70	0.64	0.04
NOVA1	0.62	0.004
ESRP1	0.62	0.005
RNU6	0.61	0.006
SRSF1	0.59	0.02
RNU12	0.59	0.03
RNU4	0.59	0.02
RNU2	0.58	0.06

Supplementary table 5. Capacity of baseline fasting and postprandial levels of selected spliceosome components and splicing factors to predict T2DM development by ROC curve analysis. ***AUC:*** Area under curve.

Supplementary Table 6

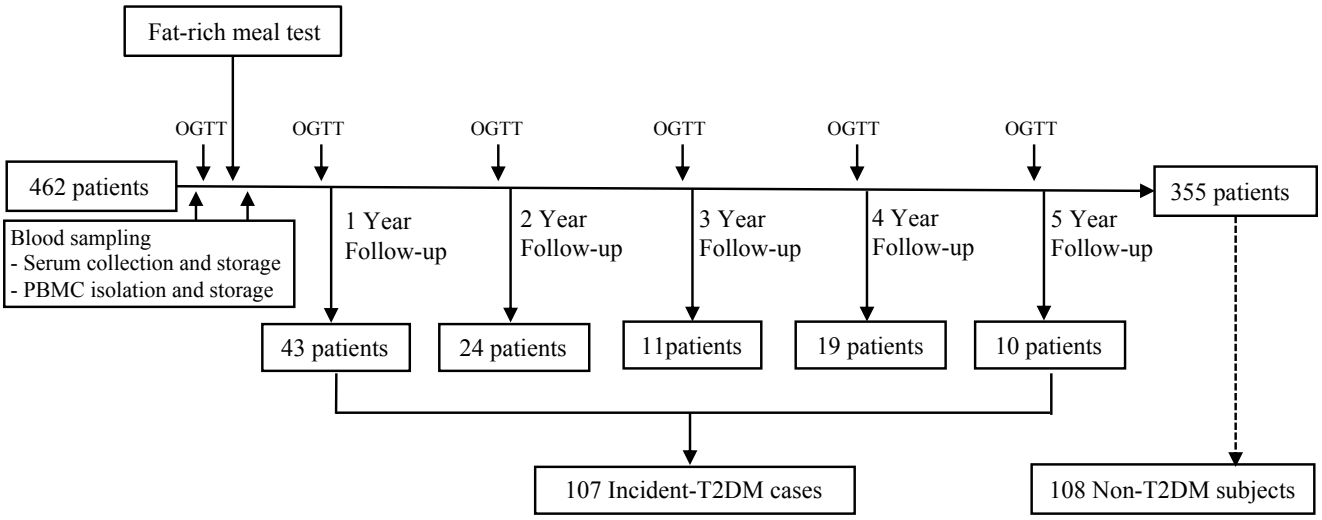
All patients during 5 years of follow-up	Logistic Regression		Random Forest		C4.5	
	Training	CV	Training	CV	Training	CV
HbA1c	0.643	0.634	0.676	0.567	0.618	0.557
FINDRISK	0.548	0.537	0.661	0.544	0.500	0.449
HbA1c and FINDRISK	0.643	0.624	0.929	0.509	0.618	0.562
Baseline fasting SFs levels	0.894	0.686	1	0.711	0.942	0.704
Baseline postprandial SFs levels	0.853	0.656	1	0.763	0.963	0.703
Baseline fasting and postprandial SFs levels	1	0.654	1	0.813	0.953	0.771
Baseline fasting and postprandial SFs levels + HbA1c	1	0.675	1	0.795	0.947	0.743
Baseline fasting and postprandial SFs levels + FINDRISK	1	0.704	1	0.808	0.972	0.737

Patients that developed T2DM during the twofirst years of follow-up	Logistic Regression		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 and FINDRISK	0.691	0.667	0.950	0.611	0.706	0.647
Baseline fasting SFs levels	0.924	0.72	1	0.745	0.988	0.630
Baseline postprandial SFs levels	0.906	0.678	1	0.806	0.998	0.739
Baseline fasting and postprandial SFs levels	1	0.699	1	0.851	0.872	0.705
Baseline fasting and postprandial SFs levels + HbA1c	1	0.681	1	0.871	0.989	0.756
Baseline fasting and postprandial SFs levels + FINDRISK	1	0.682	1	0.861	0.971	0.745

Supplementary table 6. Capacity of the fingerprints comprised by baseline fasting and/or postprandial levels of spliceosome components and splicing factors to predict T2DM development using different computational (machine-learning) algorithms. Logistic Regression**,** Random Forest and C4.5 algorithms were applied using R language to determine the capacity of the fingerprints comprised by baseline fasting and/or postprandial levels of spliceosome components and splicing factors to predict T2DM development using the full cohort the patients or only those patients that developed T2DM during the two first years of follow-up. These analysis (training) were confirmed by cross-validation studies as described in Supplementary Material and Methods. AUC from training cohort and cross-validation (CV) analysis are shown. ***HbA1c:*** Glycosylated hemoglobin; ***SFs:*** Splicing factors.

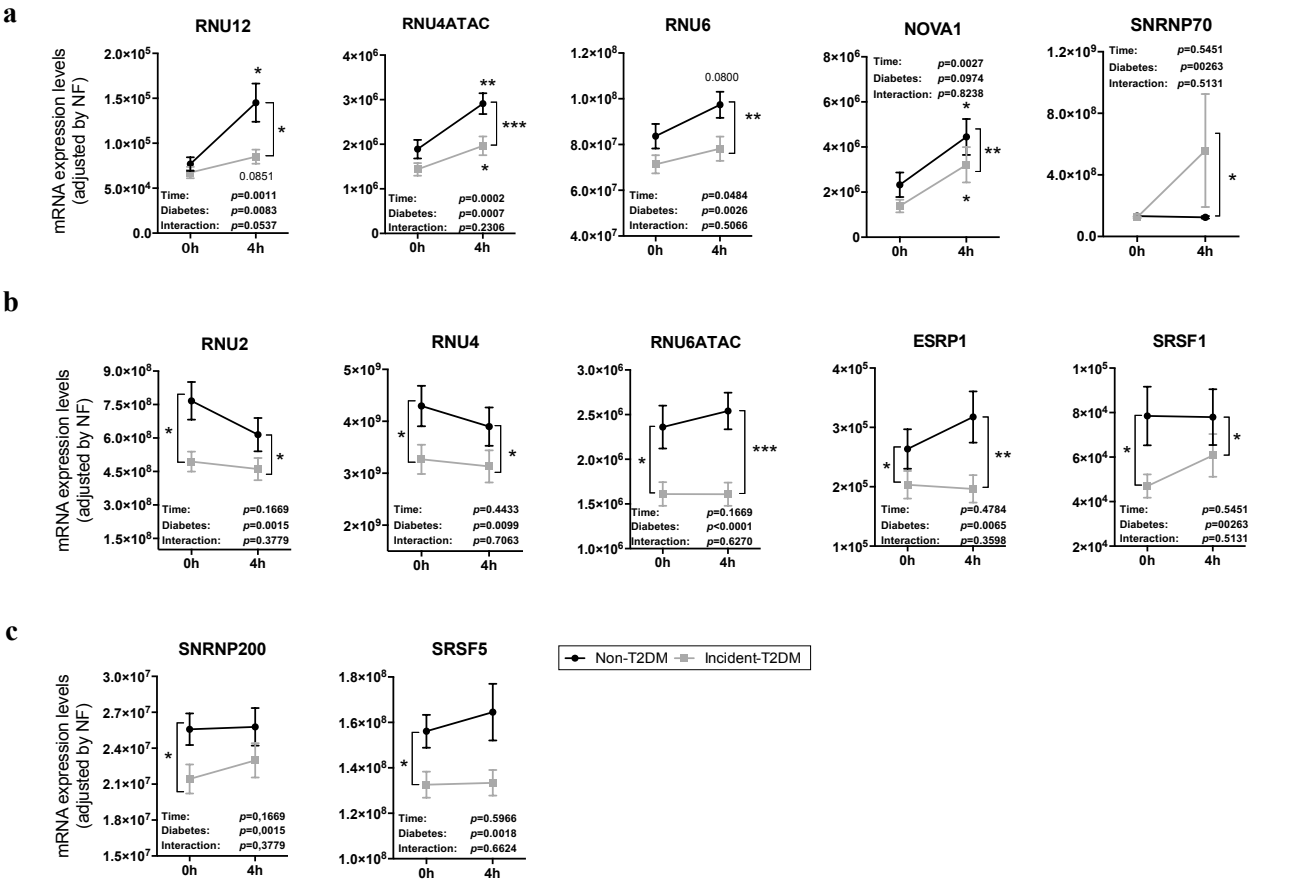
Supplementary Figures

Supplementary Figure 1



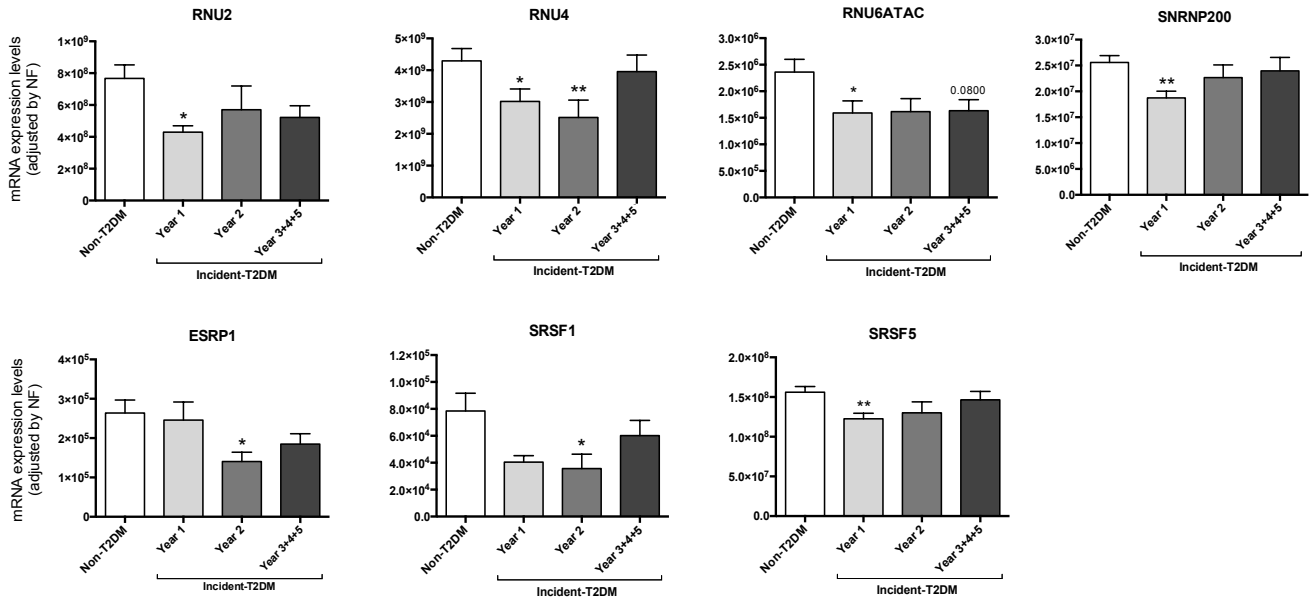
Supplementary figure 1. Graphical scheme of the patients study timeline. Specifically, from the initial 462 non-T2DM patients included in the CORDIOPREV study, 107 patients developed T2DM (Incident-T2DM cases) after a mean follow-up of 5-years according to all the American Diabetes Association (ADA) diagnosis criteria evaluated on the basis of glucose tolerance tests (OGTT) performed each year. In the present study, all these 107 incident-T2DM cases and 108 matched controls (non-T2DM subjects, randomly selected from the remaining 355 subjects that did not develop T2DM during the study period) were included. A fat-rich meal test was performed at baseline under fasting conditions, and blood was taken (at 0h and 4h-postprandial) to isolate the PBMCs and determine the expression profile of splicing machinery components.

Supplementary Figure 2



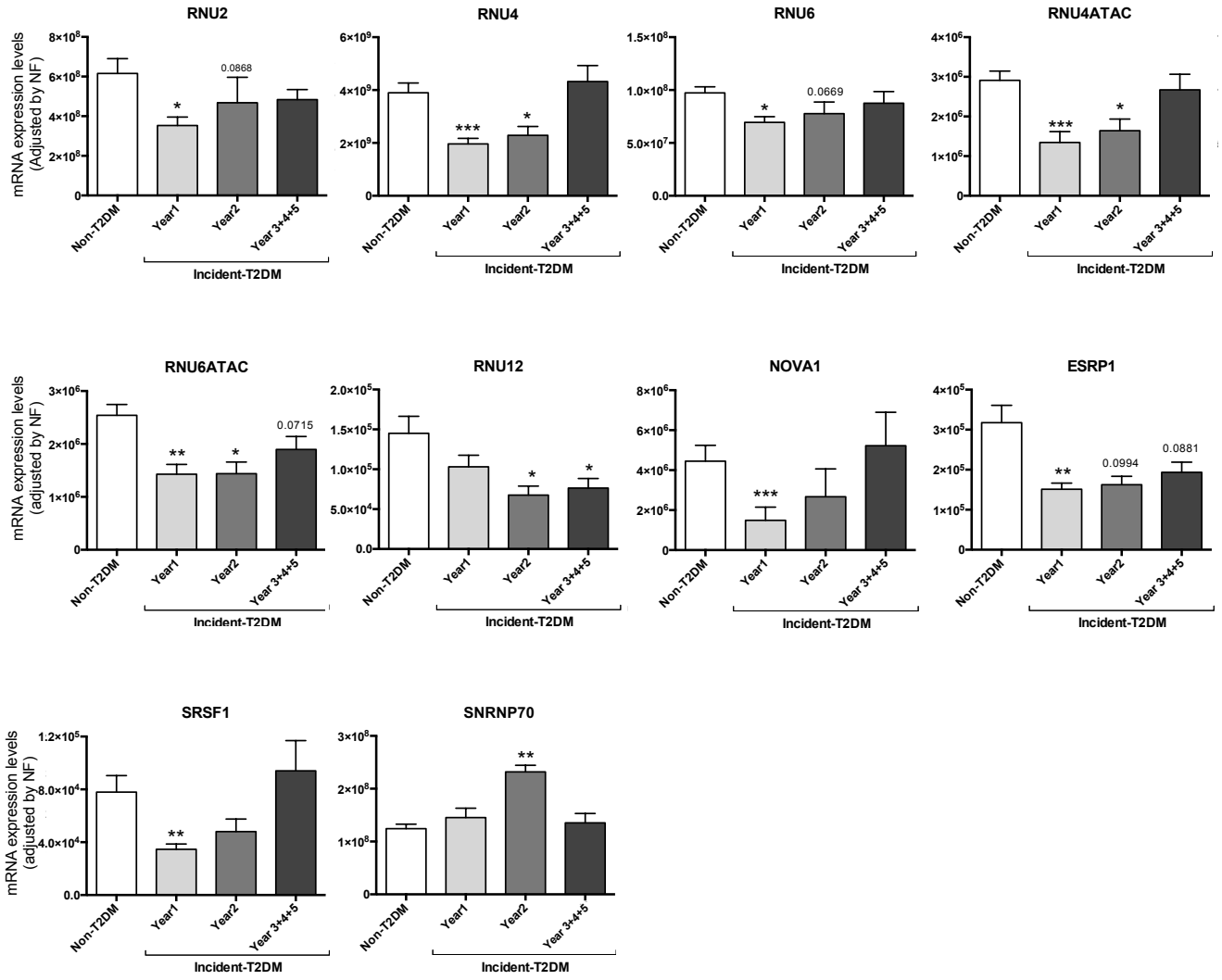
Supplementary figure 2. Modulation of the expression levels of specific spliceosome components and splicing factors during a fat-rich meal test (fasting vs. 4h of postprandial). a) Spliceosome components and splicing factors that were selectively altered during the postprandial phase (in comparison with baseline levels) in a specific group of patients (Incident-T2DM or non-T2DM patients). b) Altered factors at baseline that remained altered in postprandial phase. c) Factors that were only altered at baseline. Data represent mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *GAPDH* and *ACTB*] of the different spliceosome components and SFs in the PBMCs from non-T2DM and Incident-T2DM subjects during fat-rich meal test. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$).

Supplementary Figure 3



Supplementary figure 3. Expression pattern of altered spliceosome components and splicing factors at baseline in the PBMCs of non-T2DM and Incident-T2DM subjects sub-classified according with the year of T2DM diagnosis. mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components (first row) and SFs (second row) in the PBMCs from non-T2DM and Incident-T2DM subjects. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, $p<0.05$; **, $p<0.01$).

Supplementary Figure 4



Supplementary figure 4. Expression pattern of altered spliceosome components and splicing factors during the postprandial state in the PBMCs of non-T2DM and Incident-T2DM subjects sub-classified according with the year of T2DM diagnosis. mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *GAPDH* and *ACTB*] of specific spliceosome components and SFs in the PBMCs from non-T2DM and Incident-T2DM subjects. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$).

Article II

Title: Dietary intervention modulates the expression of the splicing machinery in patients at high risk of type 2 diabetes development: From the CORDIOPREV study.

Running title: Splicing machinery modulation by dietary intervention

Authors: Mercedes del Rio-Moreno, MS[1,2,3,4], Emilia Alors-Pérez, MS[1,2,3,4], Antonio Camargo, PhD[1,2,4,5], José López-Miranda, MD[1,2,4,5*], Justo P. Castaño, PhD [1,2,3,4*], Raúl M. Luque, PhD [1,2,3,4*], Manuel D. Gahete, PhD[1,2,3,4*],

Affiliations: [1]Maimonides Institute for Biomedical Research of Cordoba (IMIBIC), Córdoba, Spain. [2]Universidad de Córdoba, Córdoba, Spain. [3]Reina Sofia University Hospital, Córdoba, Spain. [4]CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn), Córdoba, Spain. [5]Lipid and Atherosclerosis Unit, Department of Medicine, Reina Sofia University Hospital, University of Córdoba, Córdoba, Spain.

[*]These authors contributed equally to the study and should be considered as corresponding authors.

Address correspondence to: José López-Miranda, Justo P. Castaño, Raúl M. Luque or Manuel D. Gahete. Edificio IMIBIC. Av Menéndez Pidal s/n. 14004-Córdoba, Spain. [$]Unidad de Lípidos y Arteriosclerosis, Departamento de Medicina, Hospital Universitario Reina Sofía. Universidad de Córdoba. 14004 Córdoba, Spain. E-mail: jlopezmir@gmail.com, justo@uco.es, raul.luque@uco.es and bc2gaorm@uco.es.

Abstract

Type-2 diabetes mellitus (T2DM) has become a major health problem worldwide. T2DM risk can be reduced with healthy dietary interventions, but the precise molecular mechanisms underlying this association are still incompletely understood. We recently discovered that the expression pattern of the splicing machinery is associated with the risk of T2DM development. Thus, the aim of this work was to evaluate the influence of 3-year dietary intervention in the expression pattern of the splicing machinery components in PBMCs from patients within the CORDIOPREV-study. Expression of splicing machinery components were determined in PBMCs, at baseline and after 3-years of follow-up, from all patients who developed T2DM (n=107) and 108 randomly selected non-T2DM patients, who were randomly enrolled in two healthy dietary patterns (Mediterranean or Low-Fat Diet). Dietary intervention modulated the expression of key splicing machinery components (i.e. up-regulation of *SPFQ*/*RMB45*/*RNU6*, etc., down-regulation of *RNU2*/*SRSF6*) after three years, independently of the type of healthy diet. Some of the changes observed were associated with key clinical features and were differentially induced by dietary intervention in incident-T2DM and non-T2DM patients. This study reveals that splicing machinery can be modulated by long-term dietary intervention, and could become a valuable tool to screen the progression of T2DM.

Introduction

Type 2 diabetes mellitus (T2DM) has become a major global health problem in recent decades due to its rising incidence, prevalence, and its tight causal association with diverse comorbidities, including cardiovascular disease (CVD) (1; 2). In the context of CVD prevention, lifestyle and behavioral interventions have been shown as advantageous approaches with less associated costs and side effects compared to available medical treatments (3). In particular, several healthy dietary patterns have been shown to be useful tools for the management of T2DM and the concomitant reduction of cardiovascular risk. This is the case of Mediterranean Diet (MedDiet), Dietary Approaches to Stop Hypertension (DASH), vegetarian diet, and the low-fat (LF) high carbohydrates diet recommended by the National Cholesterol Education Program and the American Diabetes Association (ADA) (2-5). In line with this, our group has shown that within the CORonary Diet Intervention with Olive oil and cardiovascular PREVention study (CORDIOPREV), a prospective, randomized, controlled trial that includes CVD patients at high-risk for T2DM development (6), long-term consumption of MedDiet rich in olive oil and LF diet have beneficial effects on the patients, improving insulin sensitivity and beta-cell function (7). In this context, early identification of patients at higher risk for T2DM development is critical for prevention of a new cardiovascular events (8; 9). Interestingly, we recently discovered that the expression pattern of certain splicing machinery elements in the peripheral blood mononuclear cells (PBMCs) of patients is tightly associated with the risk of T2DM and could accurately predict T2DM development in the individuals from the CORDIOPREV study, outperforming the capacity of classical predictors of T2DM development, such as glycated hemoglobin (HbA1c) or predictive scores (FINDRISK) (10), two stablished strategies that have limitations and cannot precisely predict an individual's risk of developing T2DM (11; 12).

The splicing machinery is comprised by the spliceosome, an intricate macromolecular complex whose functional core is composed of several small nuclear ribonucleoprotein (snRNP) subunits, which dynamically interact to regulate the splicing process. The activity of the

spliceosome is precisely modulated by more than 300 auxiliary proteins, the so-called splicing factors, that specifically recognize certain sequences in exons and introns (13; 14). An emerging body of evidence indicates that, under adverse health conditions, there is a profound dysregulation of certain spliceosomal components and splicing factors that results in altered and even aberrant splicing processes, which, in turn, substantially contribute to the development of severe pathologies, including cancer, neurodegeneration, liver disease and diabetes (13; 15-19). Indeed, the correct function of the splicing machinery is essential to maintain cell homeostasis (17; 18; 20; 21). In this scenario, it has been proposed that nutrients can modulate processes required for cell homeostasis through the alteration of gene expression and, particularly, the splicing of pre-mRNAs encoding key regulatory proteins (e.g., insulin receptor, leptin receptor) (22). Moreover, several studies have shown that the gene expression pattern of PBMCs is severely influenced by the diet (23-26) and might reflect metabolic and immune responses of adipocytes or hepatocytes (27; 28), thus providing valuable information to advance in the study of diseases such as T2DM and CVDs, using less invasive sampling methods (10; 25).

Based on all the above, the aim of this work was to evaluate the influence of the dietary intervention in the expression pattern of the components of the splicing machinery in PBMCs from patients included in the CORDIOPREV study. Specifically, we sought to ascertain if the consumption of two healthy diets (MedDiet and LF diet) during three years modified the expression pattern of the splicing machinery in PBMCs from patients at high risk of T2DM development.

Research Design and Methods

Study population

The present study was conducted within the framework of the CORDIOPREV study (Clinical Trials Registry NCT0092493741), a prospective, randomized, controlled trial that includes 1002 CHD patients, who had their last coronary event over six months before joining the study (6). Patients gave written informed consent to participate in the trial and the study protocol was approved by the Human Investigation Review Committee of the Reina Sofia University Hospital (HURS, Cordoba, Spain), according to institutional and Good Clinical Practice guidelines. Details of the study, including inclusion and exclusion criteria and cardiovascular risk factors of the patients have been previously reported. In brief, patients aged 20-75 years, with established CHD but without clinical events in the last six months with no other serious diseases and a life expectancy of at least five years were eligible. Participants were randomly enrolled in 2 study dietary models: the Mediterranean diet (MedDiet) and the low-fat (LF), high–complex carbohydrate diet recommended by the National Cholesterol Education Program and the ADA, both providing a wide variety of foods. The intervention phase is still in progress, and will have a follow up of seven years.

At baseline, 462 participants were non-T2DM (7) and, after a mean follow-up of 60 months, 107 from those patients developed T2DM (Incident-T2DM cases) (**Supplementary Figure 1**). Each year, T2DM was diagnosed according to the ADA diagnosis criteria. Specifically, if one or more of following criteria were present in the study subjects: fasting plasma glucose (FPG) concentration ≥126 mg/dL, impaired fasting glucose (IFG); FPG ≥200 mg/dL after 2h of oral glucose test (OGTT), impaired glucose tolerance (IGT); glycated hemoglobin (HbA1c) ≥6.5 % (≥48 mmol/mol). Among the 107 incident-T2DM cases, 43 were diagnosed during the first year of follow-up, 24 during the second year, 11 during the third year, 19 during the fourth year and 10 during the fifth year. In the present study, we included the incident-T2DM cases after the 60

months of follow-up (n=107) together with 108 matched, randomly selected controls, who did not develop T2DM from the remaining 355 subjects that did not develop T2DM (non-T2DM subjects) during the study period, as previously reported (10) (**Supplementary Figure 1**). The random selection of the non-T2DM subjects was performed using computational stratified sampling from the 462 non-T2DM subjects of the CORDIOPREV study according to the following clinical, anthropometric and biochemical variables: diet, age, gender, fasting plasma glucose, body mass index, LDL-cholesterol and HDL-cholesterol. To implement this type of sampling, the target population was first divided into separate strata and then, samples were randomly selected within each stratum through simple automatic sampling by using the R software (10). We collected PBMCs samples at baseline and at the third year of follow-up, as established in the study protocol guidelines (6).

Study diets

The MedDiet was composed by a minimum of 35% calories from fat [22% Monounsaturated Fatty Acid (MUFA), 6% Polyunsaturated Fatty Acid (PUFA) and 10% saturated fat], 15% proteins, and a maximum of 50% carbohydrates; and the LF diet by 30% total fat (12%-14% MUFA, 6-8% PUFA 10% and 10% saturated fat), 15% protein, and a minimum of 55% carbohydrates. In both diets, the cholesterol content was adjusted to 300 mg/d. Participants receive the same intensive dietary counseling and were monitored by nutritionists, dietitians, internists and cardiologists. Details about diets and randomization has been previously reported and summarized (6). In the present study, 93 patients were assigned to the LF Diet group and 122 to the MedDiet group (**Supplementary Figure 1**).

Metabolic study design

Oral glucose tolerance test (OGTT) and fat-rich meal tests were implemented in all patients to dynamically determine the metabolic status of the patient. Blood samples were

collected before the meal and after 4h, and biochemical determination of metabolic parameters and calculation of insulin resistance and sensitivity indexes were performed as previously reported (6; 10).

Blood sampling and processing to isolate PBMCs

Venous blood from the participants at the inclusion of the study and after three years of follow-up (12h overnight fast) was collected in tubes containing EDTA and PBMCs were isolated as previously described (6; 26)

RNA extraction and quantification

Total RNA from PBMCs was isolated using Direct-zol RNA kit (Zymo Research, Irvine, CA, USA) following manufacturer's instructions. The amount of RNA recovered was determined and its quality assessed by the NanoDrop2000 spectrophotometer (Thermo Fisher). One μg of RNA was reverse transcribed to cDNA using random hexamer primers with the First Strand Synthesis Kit (Thermo Fisher).

Analysis of splicing machinery components by microfluidic-based dynamic qPCR array

A 48.48 Dynamic Array based on microfluidic technology (Fluidigm, San Francisco, CA, USA) was implemented to determine the expression of 48 transcripts in 48 samples, simultaneously. Specific primers for human transcripts including components of the major (n=13) and minor spliceosome (n=4), associated splicing factors (n=28) and three housekeeping genes were specifically designed, as previously reported (10). Preamplification, exonuclease treatment and qPCR dynamic array were implemented following manufacturer's instructions using the Biomark System and the Real-Time PCR Analysis Software (Fluidigm). Details are provided in (10).

Statistical and bioinformatical analysis

Data were assessed for normality of distribution using the Kolmogorov–Smirnov test and are expressed as mean ± SEM. Statistical analysis was carried out using paired Student's t-test, for the alteration of the expression pattern in each patient over the three years of the study, unpaired t-test (Mann Whitney U test), when comparing values at baseline or third year, or ANOVA depending on the existence of ≥2 groups in each comparison. Significant correlations were studied using bivariate Spearman correlation methods, for these analyses, the fold change between the third year and baseline gene expression or clinical parameters were calculated, in order to determine dynamic correlations. P-values smaller than 0.05 were considered statistically significant. Statistical analyses were carried out with GraphPad Prism 6 (La Jolla, CA, USA) and SPSS 17.0 (IBM).

Results

Dietary intervention modulated the expression of several splicing machinery components

The effect of dietary intervention on the expression pattern of splicing machinery components was evaluated in PBMCs from n=215 patients at high risk of T2DM-development included in the CORDIO-PREV study (**Supplementary Figure 1**). Demographic, anthropometric and clinical parameters at basal (inclusion of the study) and after 3 years of follow-up are depicted in **Table 1**. Dietary intervention during 3 years improved HbA1c levels and HOMA-IR and HIRI indexes in this study cohort. However, glucose levels were significantly increased, and HDL levels decreased (**Table 1**). These changes were observed under both LF Diet and MedDiet (**Supplementary Table 1 and 2**). Microfluidic-based qPCR analysis revealed that the expression pattern of several splicing machinery components was altered in PBMCs from these patients after three years of dietary intervention. Specifically, results unveiled an increase in the expression of the spliceosome components *RNU6, RNU4ATAC, U2AF1, PRPF40A,* and *RNU12* and in the splicing factors *NOVA1, SRSF3, RBM45, SPFQ, ESRP1*, and *SNW1*, as well as a decrease of the spliceosome component *RNU2* and the splicing factor *SRSF6* (**Figure 1**). Further analysis indicated that the increase in the expression of *SPFQ* observed herein was inversely correlated with the evolution of HOMA-IR and HIRI indexes, which decreased during the follow-up in the full population (**Table 2**).

The modulation of the expression of several splicing machinery components was not diet dependent

At baseline, all CORDIO-PREV participants were randomly enrolled in one of the two dietary model groups, LF Diet and MedDiet showing comparable levels of all parameters determined except for C-reactive protein, which was lower for MedDiet (**Table 3**). Interestingly, while most of the splicing machinery components altered during the dietary intervention showed

similar trends and changes when the population was separated by diets (**Supplementary Figure 2),** a distinct, differential response was observed in the case of three splicing factors *SNW1*, *SPFQ* and *NOVA1*. In particular, the observed increase in the expression of these factors was more pronounced in PBMCs from patients under LF Diet than in those from MedDiet patients (**Figure 2**).

The expression pattern of specific splicing machinery components was differentially modulated by dietary intervention in Incident-T2DM and non-T2DM patients

After 5 years of dietary intervention, 107 participants had developed T2DM (29). As previously reported, at baseline, the group of patients who developed T2DM during the five year of the study (Incident-T2DM; n=107) presented higher weight, BMI, HbA1c levels, fasting insulin and TGs compared to randomly selected non-T2DM controls (n=108) (10). The cohort of n=108 randomly selected non-T2DM patients showed comparable levels of all the parameters determined except for age and DI compared to the total population (n=355) of non-T2DM individuals (10). The expression of some of the previously mentioned spliceosome components and splicing factors was differentially altered in the PBMCs of patients that develop T2DM after the 5 years of the study compared to non-T2DM subjects. Specifically, *RNU12* was clearly increased in non-T2DM patients as compared to incident-T2DM patients (**Figure 3A**). On the other hand, *ESRP1* and *RNU6* were markedly increased, and *SRSF6* reduced, in Incident-T2DM patients while no such changes were observed in non-T2DM subjects (**Figure 3B**). Of note, *SRSF6* decrease was directly correlated with the decrease in HbA1c levels observed in Incident-T2DM patients (Table 5). In addition, we observed that *NOVA1* and *RNU4ATAC* were significantly altered both in Incident-T2DM and non-T2DM; however, the increase of *NOVA1* was clearly more pronounced in Incident-T2DM, wherein it inversely correlated with HOMA-IR, while the increase of *RNU4ATAC* was more pronounced in non-T2DM (**Figure 3C**).

From non-T2DM patients, 51 were initially assigned to LF Diet and 57 to MedDiet, while from Incident-T2DM patients, 42 were initially assigned to LF Diet and 65 to MedDiet (**Supplementary figure 1**). During the follow-up, participants that did not develop T2DM presented a significant decrease in HbA1c, HOMA-IR, HIRI and c-Hdl, although glucose levels significantly increased (**Table 4**). Similar changes to those found for the overall population were observed under LF Diet, although changes in HbA1c and glucose levels appeared to be attributable to MedDiet group (**Table B A and C**). On the other hand, Incident-T2DM patients presented a decrease in HbA1c levels and an increase in glucose levels, with a significant decrease in HOMA-IR and HIRI and increase in cholesterol in patients submitted to LF Diet and a decrease in DI and c-HDL in patients from MedDiet group (**Table 5**). When analyzing the expression pattern according to the diabetic status and the dietary group, several differences emerged between groups. In particular, *RNU12,* which was increased in non-T2DM subject after the three years of follow-up (**Figure 3A**), showed a trend to a more pronounced increase under LF Diet compared to Incident-T2DM cases, although it did not reach statistical significance (**Figure 4A**). On the other hand, the differences observed in the splicing factor *ESRP1* and *RNU6* between non-T2DM and Incident-T2DM patients were stronger under MedDiet and LF Diet, respectively (**Figure 4B**); while changes in SRSF6 were similar in patients under both dietary interventions. Finally, the increase in *NOVA1* was strikingly more pronounced under LF Diet than under MedDiet in Incident-T2DM patients, whereas the increase in *RNU4ATAC* was more obvious in MedDiet group in non-T2DM patients (**Figure 4C**).

Discussion

This study represents, to the best of our knowledge, the first comprehensive analysis of the regulatory role of a healthy dietary pattern intervention on the expression of the components of the splicing machinery, including spliceosome elements and splicing factors. This study was implemented using PBMCs from patients at high risk of T2DM development included in the CORDIOPREV trial, in that we have previously demonstrated that the expression pattern of certain splicing machinery components is associated with the risk of T2DM development and could accurately predict this development in individuals from the CORDIOPREV study [10].

Importantly, our present study provides primary evidence that a dietary intervention can distinctly alter the expression pattern of the splicing machinery, both spliceosome components and splicing factors, in humans at risk of T2DM. In particular, the results demonstrate that the consumption of two healthy diets (MedDiet and LF diet) during three years can modulate the expression pattern of key spliceosome components and splicing factors in PBMCs from the patients enrolled in the CORDIOPREV study, including the overexpression of some molecular components, like *SPFQ*, *RBM45*, *RNU6*, etc. and the downregulation of others, including *RNU2* and *SRSF6*. Interestingly, some of the changes observed in the expression levels of certain splicing machinery components were closely associated with relevant clinical features, as is the case of the increase in the expression levels of the splicing factor *SPFQ*, which was inversely correlated with the decrease in HOMA-IR and HIRI indexes observed in the population. The finding of a diet-related long-term modulation of the expression of the splicing machinery components could represent a novel valuable piece of information for two reasons. First, because it unveils that the splicing process may represent an adaptive mechanism in response to different nutritional conditions, and that this mechanism could be in place not only in circulating PBMCs but may also operate in cell types from other tissues and organs tightly coupled to nutrient-dependent metabolic homeostasis (e.g. liver, pancreas, adipose tissue), an avenue that is indeed worth exploring. Actually, we and other have already found the delicate and important role that the

regulation of the splicing machinery can play in those organs (10; 18; 30-32). Secondly, inasmuch as PBMCs can be an accessible and suitable sentinel to detect relevant changes related to nutrient- and diet-dependent metabolic homeostasis, our current results support the idea that changes in the expression of key splicing machinery components could provide a fine screening marker for the development or progression of T2DM and their diet-related dynamics. Indeed, within the CORDIOPREV study, the long-term intake of a MedDiet, rich in olive oil, or a LF diet improved insulin sensitivity and beta-cell function (7), and therefore the increase in the expression of specific splicing factors found herein under both diets, and their inverse correlation with insulin resistance indexes, strongly suggest that the molecular changes might be related to the beneficial consequence of the healthy diet consumption and, therefore, that they could represent a novel mechanism linking healthy dietary intervention and the improvement the metabolic status of the patients and the protection from cardiovascular complications. Given the very scarce information available on the functional roles and implications of many of the molecules identified in the present study to be altered in PBMCs (e.g. *SPFQ*, *RBM45*, *RNU6*, etc.) the present findings open novel, unexplored avenues in this field of research.

One of the findings from this study that we consider most noteworthy is that the diet-induced alterations in the splicing machinery of PBMCs was independent of the type of healthy diet in which CORDIOPREV participants were enrolled (MedDiet or LF Diet), except for three splicing factors (*SNW1, SPFQ* and *NOVA1*) that showed a more pronounced modulation in patients under the LF Diet. To date, and to the best of our knowledge, nothing has been reported regarding the influence of diet intervention in the modulation of the expression of *SNW1*, *SPFQ* and *NOVA1*. However, some of these factors have been described to contribute to the alternative splicing of key genes whose splicing processing changes in response to a fatty diet (31). Previously, several studies have shown that PBMCs gene expression pattern is influenced by the diet (23-26) and that this might reflect changes related to both metabolic and immune responses (27; 28). In addition, it has been demonstrated that the splicing process of key regulatory proteins

for metabolic homeostasis, like the receptors for insulin or leptin, can be markedly influenced by nutrient metabolism, directly or indirectly (22; 33). Thus, it seems reasonable to think that those splicing-related changes would rely on upstream changes in the function of the machinery responsible for generating the splice variants. However, little or nothing is known in this regard in PBMCs, for there are no reports on how diet can influence the expression of the components of the spliceosome and the splicing factors, which altogether are responsible of the modulation of the splicing process. Nevertheless, in this context, some studies have highlighted that nutritional status can induce changes in the activity of serin-arginine (SR) proteins, an important family of splicing factors, further supporting the contention that different nutrients may be able to modulate the expression of metabolic genes at the level of its splicing processing. Specifically, it has been described that insulin signaling can up-regulate the expression of the splicing factor SRSF1 in pancreatic beta cells, inducing the splicing of the insulin receptor to generate the INSR-B isoform (34). The same study also found a regulation of the protein levels of the splicing factor MBNL1 by high glucose levels. In addition, other splicing factors belonging to the SR proteins family, SRSF2, is decreased under Vitamin E-deficient diet in the liver (35). Thus, although still limited, growing evidence, included the results from this study, points to a link between diet and nutrient and regulation of the splicing process, including its underlying operating machinery.

Another intriguing implication of our present results relates to the predictive capacity of studying changes in the splicing machinery in at-risk patients. To be more specific, nutrient-induced changes in specific splicing machinery components may provide hints on the predictive potential and possible functional correlation of key molecules, which had not been explored hitherto in this regard. Thus, within this study, regardless of the type of dietary intervention, the expression of some of the splicing factors studied was differentially altered in patients that develop T2DM after the 5 years of the study compared to non-T2DM subjects. For example, *RNU12*, a component of the minor spliceosome, showed a significant increase after the 3 years of dietary intervention in non-T2DM patients. Interestingly, we have previously described that the

expression of this small nuclear RNA (snRNA), which is essential to form U12 snRNP and carry out the appropriate splicing of type 12 introns (36), was lower, at baseline (inclusion of the study), in Incident-T2DM compared to non-T2DM patients and that was associated with the risk of T2DM development (10). Therefore, since lower expression levels of RNU12 were associated with higher risk of T2DM, the dietary-induced increase in the expression of this component in non-T2DM may be associated to the protective effects of the healthy dietary consumption. Furthermore, in the same study, 4 hours incubation with baseline postprandial serum from Incident-T2DM patients induced a significant reduction of *RNU12* expression compared to non-T2DM treated PBMCs from healthy patients (10). Therefore, the modulation in the expression of these spliceosomal components may represent a link between the dietary intervention and the beneficial effects on the metabolic status of the patient. Remarkably, the difference in *RNU12* expression, at year 3, between incident and non-T2DM patients, was more pronounced under LF Diet. Thus, although the possible mechanisms linking nutrient-induced changes in the splicing machinery, their functional consequences and the regulatory implications thereof are still to be fully elucidated, our present study provide suggestive evidence that it is worth exploring both the mechanistic/functional and the predictive components of this plausible link, for it may provide original, valuable biological knowledge as well as practical information for the patients.

In conclusion, this study reveals that expression of the splicing machinery components in PBMCs from patients at risk of T2DM can be notably and selectively influenced by long-term dietary intervention; also, that the two dietary interventions tested herein, MEdDiet and LF Diet, induced remarkably similar changes on the expression of spliceosome components; and finally, that there are distinct, diet type-induced changes in PBMCs from both non-T2DM and incident-T2DM patients, that may have an as yet unknown functional significance. Therefore, we propose that the machinery that controls and performs the alternative splicing process and is consequently responsible for changes in the pattern of functionally and pathologically relevant splice variants in the regulation of metabolic homeostasis is a plausible target to be operated by

dietary intervention. As such, our results pave the way to explore in experimental models the possible mechanistic role and relevance of the splicing machinery and its components in diet-related metabolic regulation, and the investigate the value of screening changes in specific splicing machinery components to monitor and early predict relevant diet-related changes in patients at risk of T2DM.

References

1. Wahl MC, Will CL, Lührmann R: The spliceosome: design principles of a dynamic RNP machine. Cell 2009;136:701-718

2. American Diabetes A: 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019;42:S13-S28

3. Garcia-Rios A, Ordovas JM, Lopez-Miranda J, Perez-Martinez P: New diet trials and cardiovascular risk. Curr Opin Cardiol 2018;33:423-428

4. Perez-Martinez P, Ordovas JM, Garcia-Rios A, Delgado-Lista J, Delgado-Casado N, Cruz-Teno C, Camargo A, Yubero-Serrano EM, Rodriguez F, Perez-Jimenez F, Lopez-Miranda J: Consumption of diets with different type of fat influences triacylglycerols-rich lipoproteins particle number and size during the postprandial state. Nutr Metab Cardiovasc Dis 2011;21:39-45

5. Salas-Salvado J, Martinez-Gonzalez MA, Bullo M, Ros E: The role of diet in the prevention of type 2 diabetes. Nutr Metab Cardiovasc Dis 2011;21 Suppl 2:B32-48

6. Delgado-Lista J, Perez-Martinez P, Garcia-Rios A, Alcala-Diaz JF, Perez-Caballero AI, Gomez-Delgado F, Fuentes F, Quintana-Navarro G, Lopez-Segura F, Ortiz-Morales AM, Delgado-Casado N, Yubero-Serrano EM, Camargo A, Marin C, Rodriguez-Cantalejo F, Gomez-Luna P, Ordovas JM, Lopez-Miranda J, Perez-Jimenez F: CORonary Diet Intervention with Olive oil and cardiovascular PREVention 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;177:42-50

7. Blanco-Rojo R, Alcala-Diaz JF, Wopereis S, Perez-Martinez P, Quintana-Navarro GM, Marin C, Ordovas JM, van Ommen B, Perez-Jimenez F, Delgado-Lista J, Lopez-Miranda J: The insulin resistance phenotype (muscle or liver) interacts with the type of diet to determine changes in disposition index after 2 years of intervention: the CORDIOPREV-DIAB randomised clinical trial. Diabetologia 2015;59:67–76

8. Camargo A, Jimenez-Lucena R, Alcala-Diaz JF, Rangel-Zuniga OA, Garcia-Carpintero S, Lopez-Moreno J, Blanco-Rojo R, Delgado-Lista J, Perez-Martinez P, van Ommen B, Malagon MM, Ordovas JM, Perez-Jimenez F, Lopez-Miranda J: Postprandial endotoxemia may influence the development of type 2 diabetes mellitus: From the CORDIOPREV study. Clin Nutr 2018;

9. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, van der AD, Moons KG, Navis G, Bakker SJ, Beulens JW: Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. BMJ 2012;345:e5900

10. Gahete MD, Del Rio-Moreno M, Camargo A, Alcala-Diaz JF, Alors-Perez E, Delgado-Lista J, Reyes O, Ventura S, Perez-Martinez P, Castano JP, Lopez-Miranda J, Luque RM: Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study. EBioMedicine 2018;37:356-365

11. Burczynski ME, Dorner AJ: Transcriptional profiling of peripheral blood cells in clinical pharmacogenomic studies. Pharmacogenomics 2006;7:187-202

12. Cohen RM, Haggerty S, Herman WH: HbA1c for the diagnosis of diabetes and prediabetes: is it time for a mid-course correction? J Clin Endocrinol Metab 2010;95:5203-5206

13. Matera AG, Wang Z: A day in the life of the spliceosome. Nat Rev Mol Cell Biol 2014;15:108-121

14. Scotti MM, Swanson MS: RNA mis-splicing in disease. Nat Rev Genet 2016;17:19-32

15. Dlamini Z, Mokoena F, Hull R: Abnormalities in alternative splicing in diabetes: therapeutic targets. J Mol Endocrinol 2017;59:R93-R107

16. Lee SC, Abdel-Wahab O: Therapeutic targeting of splicing in cancer. Nat Med 2016;22:976-986

17. Mercader JM, Liao RG, Bell AD, Dymek Z, Estrada K, Tukiainen T, Huerta-Chagoya A, Moreno-Macias H, Jablonski KA, Hanson RL, Walford GA, Moran I, Chen L, Agarwala V, Ordonez-Sanchez ML, Rodriguez-Guillen R, Rodriguez-Torres M, Segura-Kato Y, Garcia-Ortiz H, Centeno-Cruz F, Barajas-Olmos F, Caulkins L, Puppala S, Fontanillas P, Williams AL, Bonas-

Guarch S, Hartl C, Ripke S, Diabetes Prevention Program Research G, Tooley K, Lane J, Zerrweck C, Martinez-Hernandez A, Cordova EJ, Mendoza-Caamal E, Contreras-Cubas C, Gonzalez-Villalpando ME, Cruz-Bautista I, Munoz-Hernandez L, Gomez-Velasco D, Alvirde U, Henderson BE, Wilkens LR, Le Marchand L, Arellano-Campos O, Riba L, Harden M, Broad Genomics P, Gabriel S, Consortium TDG, Abboud HE, Cortes ML, Revilla-Monsalve C, Islas-Andrade S, Soberon X, Curran JE, Jenkinson CP, DeFronzo RA, Lehman DM, Hanis CL, Bell GI, Boehnke M, Blangero J, Duggirala R, Saxena R, MacArthur D, Ferrer J, McCarroll SA, Torrents D, Knowler WC, Baier LJ, Burtt N, Gonzalez-Villalpando C, Haiman CA, Aguilar-Salinas CA, Tusie-Luna T, Flannick J, Jacobs SBR, Orozco L, Altshuler D, Florez JC, Consortium STDG: A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for Type 2 Diabetes. Diabetes 2017;66:2903-2914

18. Webster NJG: Alternative RNA Splicing in the Pathogenesis of Liver Disease. Front Endocrinol (Lausanne) 2017;8:133

19. Gallego-Paez LM, Bordone MC, Leote AC, Saraiva-Agostinho N, Ascensao-Ferreira M, Barbosa-Morais NL: Alternative splicing: the pledge, the turn, and the prestige : The key role of alternative splicing in human biological systems. Hum Genet 2017;136:1015-1042

20. Stumvoll M, Goldstein BJ, van Haeften TW: Type 2 diabetes: pathogenesis and treatment. Lancet 2008;371:2153-2156

21. Juan-Mateu J, Villate O, Eizirik DL: MECHANISMS IN ENDOCRINOLOGY: Alternative splicing: the new frontier in diabetes research. Eur J Endocrinol 2016;174:R225-238

22. Ravi S, Schilder RJ, Kimball SR: Role of precursor mRNA splicing in nutrient-induced alterations in gene expression and metabolism. J Nutr 2015;145:841-846

23. Yubero-Serrano EM, Delgado-Casado N, Delgado-Lista J, Perez-Martinez P, Tasset-Cuevas I, Santos-Gonzalez M, Caballero J, Garcia-Rios A, Marin C, Gutierrez-Mariscal FM, Fuentes F, Villalba JM, Tunez I, Perez-Jimenez F, Lopez-Miranda J: Postprandial antioxidant effect of the Mediterranean diet supplemented with coenzyme Q10 in elderly men and women. Age (Dordr) 2011;33:579-590

24. Yubero-Serrano EM, Gonzalez-Guardia L, Rangel-Zuniga O, Delgado-Casado N, Delgado-Lista J, Perez-Martinez P, Garcia-Rios A, Caballero J, Marin C, Gutierrez-Mariscal FM, Tinahones FJ, Villalba JM, Tunez I, Perez-Jimenez F, Lopez-Miranda J: Postprandial antioxidant gene expression is modified by Mediterranean diet supplemented with coenzyme Q(10) in elderly men and women. Age (Dordr) 2013;35:159-170

25. de Mello VD, Kolehmanien M, Schwab U, Pulkkinen L, Uusitupa M: Gene expression of peripheral blood mononuclear cells as a tool in dietary intervention studies: What do we know so far? Mol Nutr Food Res 2012;56:1160-1172

26. Gahete MD, Luque RM, Yubero-Serrano EM, Cruz-Teno C, Ibanez-Costa A, Delgado-Lista J, Gracia-Navarro F, Perez-Jimenez F, Castano JP, Lopez-Miranda J: Dietary fat alters the expression of cortistatin and ghrelin systems in the PBMCs of elderly subjects: putative implications in the postprandial inflammatory response. Molecular nutrition & food research 2014;58:1897-1906

27. Fuchs D, Piller R, Linseisen J, Daniel H, Wenzel U: The human peripheral blood mononuclear cell proteome responds to a dietary flaxseed-intervention and proteins identified suggest a protective effect in atherosclerosis. Proteomics 2007;7:3278-3288

28. Rendo-Urteaga T, Garcia-Calzon S, Gonzalez-Muniesa P, Milagro FI, Chueca M, Oyarzabal M, Azcona-Sanjulian MC, Martinez JA, Marti A: Peripheral blood mononuclear cell gene expression profile in obese boys who followed a moderate energy-restricted diet: differences between high and low responders at baseline and after the intervention. Br J Nutr 2015;113:331-342

29. Roncero-Ramos I, Jimenez-Lucena R, Alcala-Diaz JF, Vals-Delgado C, Arenas-Larriva AP, Rangel-Zuniga OA, Leon-Acuna A, Malagon MM, Delgado-Lista J, Perez-Martinez P, Ordovas JM, Camargo A, Lopez-Miranda J: Alpha cell function interacts with diet to modulate prediabetes and Type 2 diabetes. J Nutr Biochem 2018;62:247-256

30. Del Rio-Moreno M, Alors-Perez E, Gonzalez-Rubio S, Ferrin G, Reyes O, Rodriguez-Peralvarez M, Sanchez-Frias ME, Sanchez-Sanchez R, Ventura S, Lopez-Miranda J, Kineman RD,

de la Mata M, Castano JP, Gahete MD, Luque RM: Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease. J Clin Endocrinol Metab 2019;

31. Vernia S, Edwards YJ, Han MS, Cavanagh-Kyros J, Barrett T, Kim JK, Davis RJ: An alternative splicing program promotes adipose tissue thermogenesis. Elife 2016;5

32. Kaminska D, Hämäläinen M, Cederberg H, Käkelä P, Venesmaa S, Miettinen P, Ilves I, Herzig K-H, Kolehmainen M, Karhunen L, Kuusisto J, Gylling H, Laakso M, Pihlajamäki J: Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. Diabetologia 2014;57:347-351

33. Reardon HT, Hsieh AT, Park WJ, Kothapalli KS, Anthony JC, Nathanielsz PW, Brenna JT: Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. Prostaglandins Leukot Essent Fatty Acids 2013;88:15-19

34. Malakar P, Chartarifsky L, Hija A, Leibowitz G, Glaser B, Dor Y, Karni R: Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. Sci Rep 2016;6:31222

35. Malatesta M, Bertoni-Freddari C, Fattoretti P, Baldelli B, Fakan S, Gazzanelli G: Aging and vitamin E deficiency are responsible for altered RNA pathways. Annals of the New York Academy of Sciences 2004;1019:379-382

36. Verma B, Akinyi MV, Norppa AJ, Frilander MJ: Minor spliceosome and disease. Semin Cell Dev Biol 2018;79:103-112

Figures

Figure 1

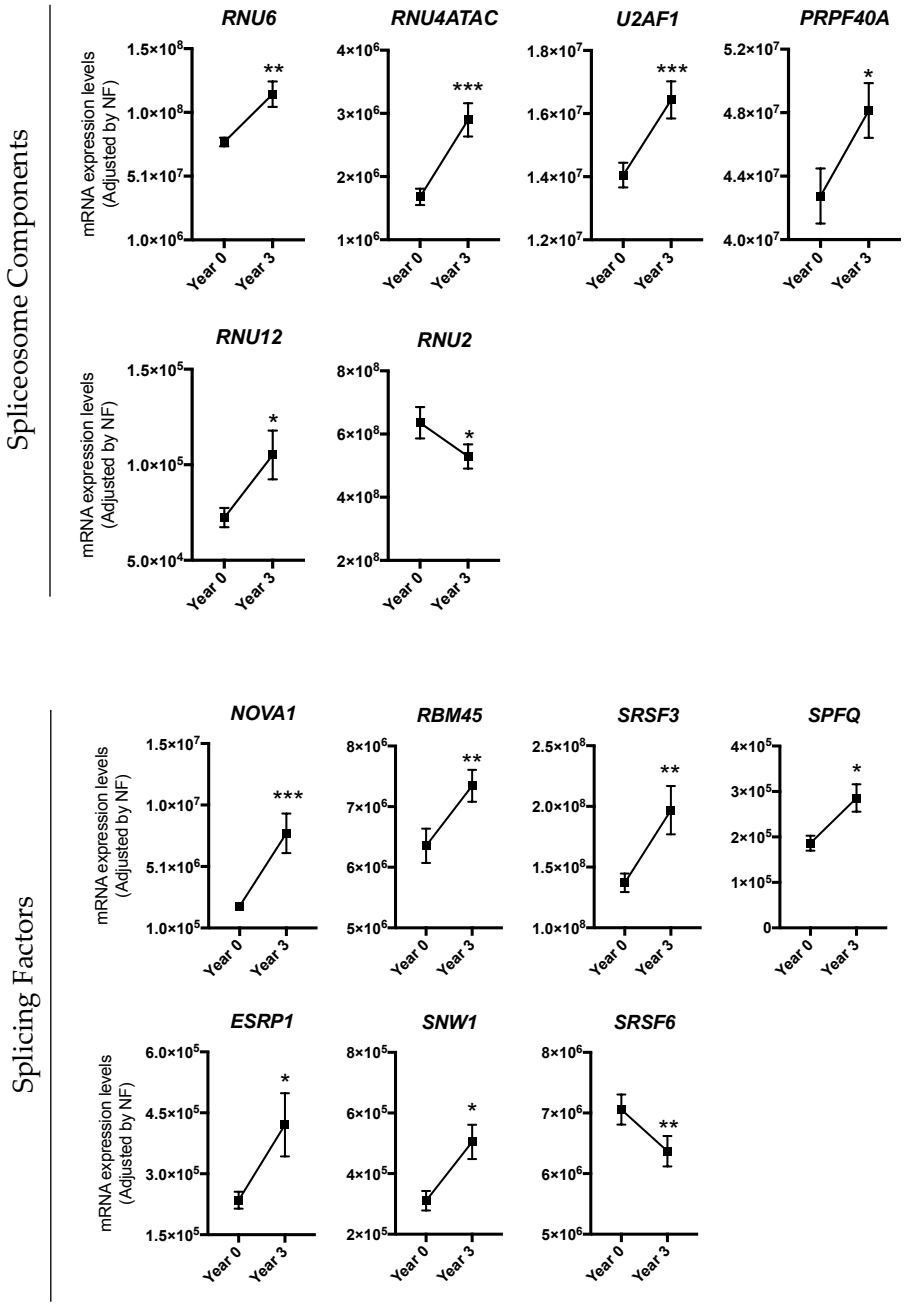


Figure 1. PBMCs expression pattern of specific splicing machinery components after three years of follow-up. *mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components and splicing factors in the PBMCs from all the patients included in the study. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, p<0.05; **, p<0.01; ***, p<0.001).*

Figure 2

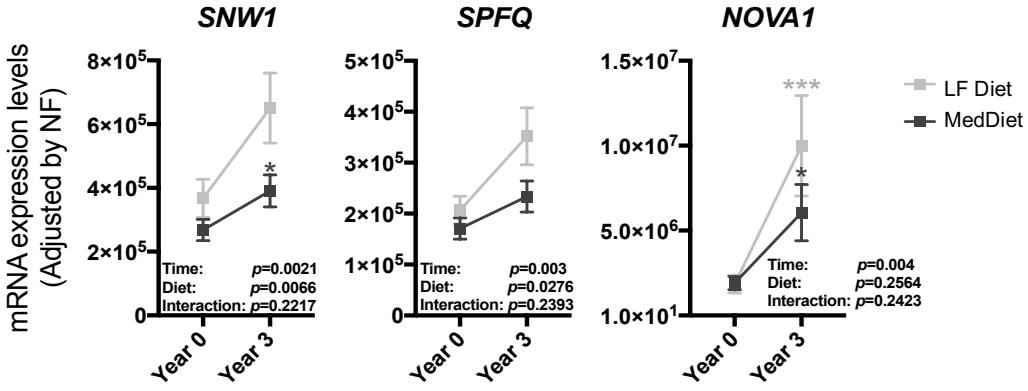


Figure 2. PBMCs expression pattern of the specific splicing machinery components after three years of follow-up under two healthy dietary pattern (LF Diet and MedDiet). *mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components and splicing factors in the PBMCs from all the patients included in the study. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, p<0.05; **, p<0.01).* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Figure 3

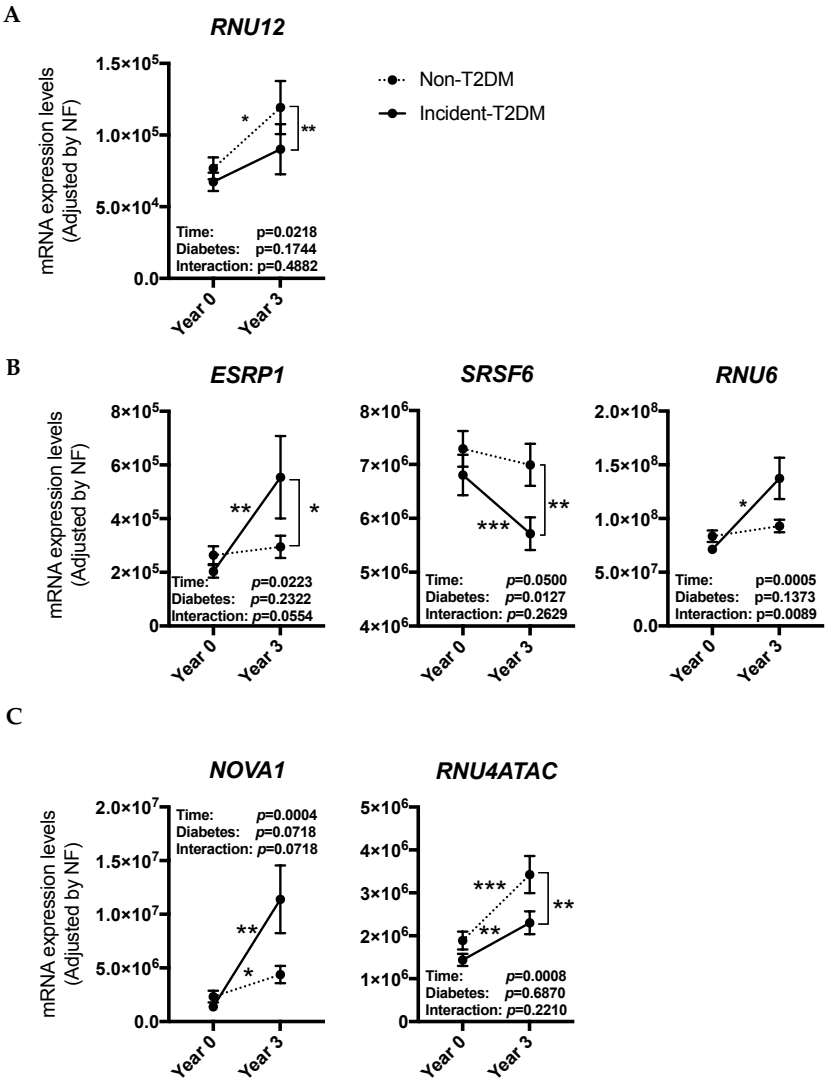


Figure 3. PBMCs expression pattern of specific splicing machinery components after three years of follow-up in non-T2DM and incident-T2DM cases. *mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components and splicing factors in the PBMCs from all the patients included in the study. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, p<0.05; **, p<0.01; ***<0.001).*

Figure 4

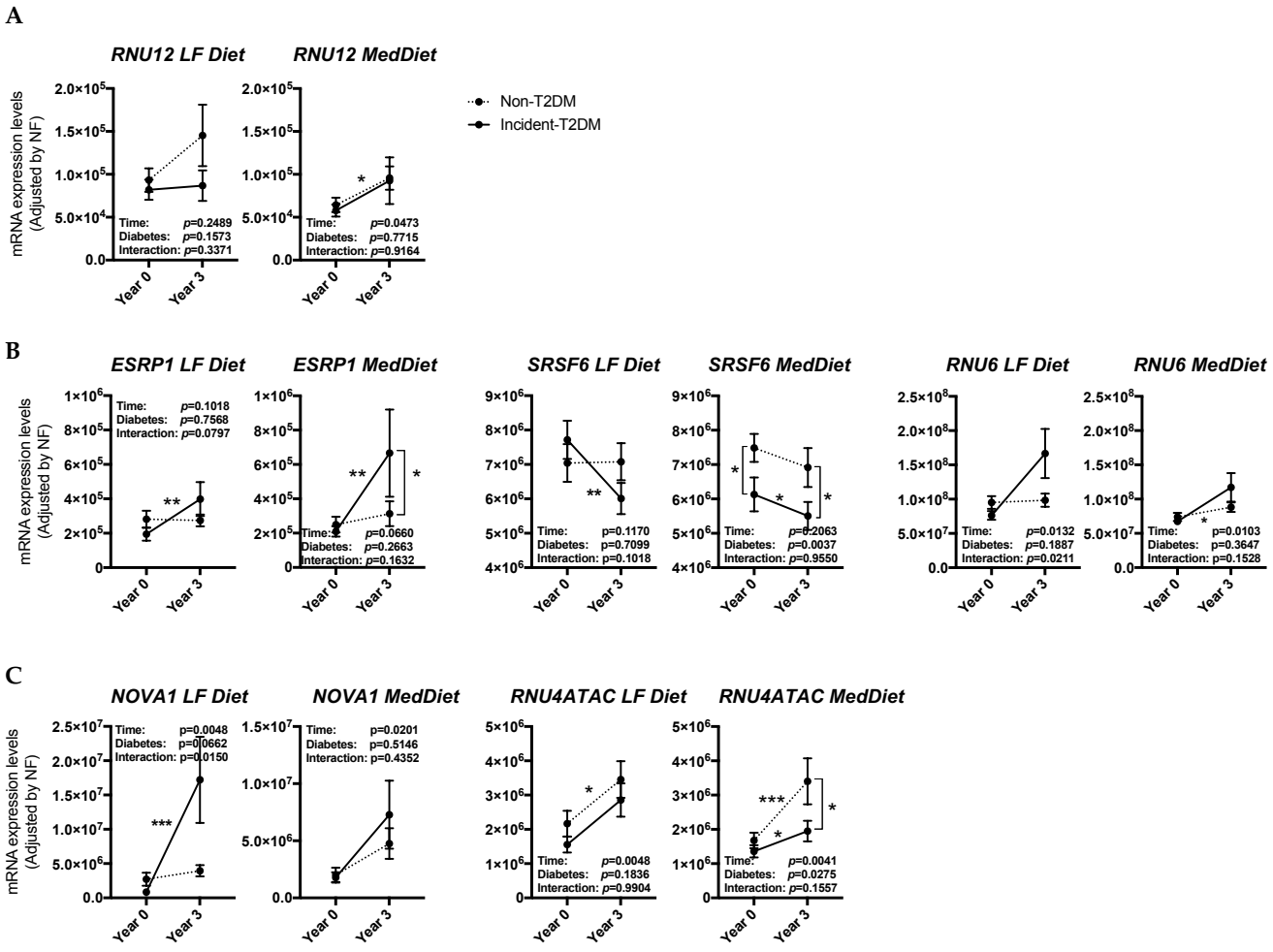


Figure 4. PBMCs expression pattern of specific splicing machinery components after three years of follow-up in non-T2DM and incident-T2DM cases under two healthy dietary pattern (LF Diet and MedDiet). *mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components and splicing factors in the PBMCs from all the patients included in the study. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: *, p<0.05; **, p<0.01; ***<0.001).* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Tables

Table 1

	Basal mean ± SEM	3rd year mean ± SEM	*p*-value
HbA1c (%)	**5.945 ± 0.025**	**5.599 ± 0.028**	**<0.001**
Glucose (mg/dl)	**95.24 ± 0.71**	**99.22 ± 0.775**	**<0.001**
Insulin (mU/l)	9.351 ± 0.434	9.467 ± 0.547	0.836
ISI	3.797 ± 0.181	3.916 ± 0.202	0.517
HOMA-IR	**2.936 ± 0.170**	**2.375 ± 0.152**	**0.001**
HIRI	**1188.537 ±68.348**	**970.754 ± 63.257**	**0.002**
IGI	0.804 ± **0.175**	1.047 ± 0.127	0.227
DI	0.841 ± 0.030	0.832 ± 0.036	0.825
MISI ($x10^2$)	0.021 ± 0.030	0.019 ± 0.001	0.455
Total cholesterol (mg/dl)	162.701 ± 2.350	165.132 ± 2.446	0.241
c-HDL (mg/dl)	**44.306 ± 0.699**	**42.065 ± 0.686**	**<0.001**
TG (mg/dl)	120.552 ± 4.245	124.478 ± 4.729	0.33
hs-CRP (mg/dl)	2.569 ± 0.22	2.352 ± 0.171	0.342

Table 1. Demographic and metabolic characteristics after three years of follow-up. *Values expressed as mean ± SEM.* ***BMI:*** *Body mass index;* ***HbA1c:*** *Glycated hemoglobin;* ***TG:*** *Triglycerides;* ***c- LDL:*** *Low density lipoprotein;* ***c-HDL:*** *High density lipoprotein;* ***NEFA:*** *Non-esterified fatty acids;* ***Apo A1:*** *Apolipoprotein A1;* ***Apo B:*** *Apolipoprotein B;* ***hs-CRP:*** *High sensitivity C-reactive protein;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index.*

Table 2

			Fold change during follow-up	
			Study population	**Incident-T2DM**
			SPFQ	***SRSF6***
Fold change during follow-up	***HOMA-IR***	r^2	-0.157	
		p	**0.03***	
	HIRI	r^2	-0.176	
		p	**0.018***	
	HbA1c (%)	r^2		0.227
		p		**0.047***

Table 2. Spearman correlations between the fold change in the expression levels of splicing machinery components (3rd year/baseline levels) and the fold change in relevant T2DM related parameters. HbA1c*: glycated hemoglobin;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index; Non-significant correlations are not depicted.*

Table 3

	LF Diet mean ± SEM	MedDiet mean ± SEM	*p*-value
Age (years)	59.667 ± 0.825	59.426 ± 0.841	0.842
Waist circumference (cm)	104.457 ± 1.172	103.428 ± 0.919	0.484
Weight (kg)	84.276 ± 1.596	83.439 ± 1.160	0.664
BMI (kg/m2)	30.895 ± 0.489	30.676 ± 0.369	0.716
Glucose (mg/dl)	94.55 ± 1.098	95.84 ± 0.921	0.365
HbA1c (%)	5.985 ± 0.032	5.9256 ± 0.033	0.21
Insulin (mU/l)	9.032 ± 0.612	9.467 ± 0.585	0.612
TG (mg/dl)	119.3804 ± 5.979	121.823 ± 5.645	0.769
Total cholesterol (mg/dl)	161.301 ± 3.544	162.966 ± 2.980	0.718
c-LDL (mg/dl)	92.374 ± 2.748	92.183 ± 2.333	0.958
c-HDL (mg/dl)	43.25 ± 0.910	44.721 ± 0.992	0.276
NEFA (mmol/L)	0.306 ± 0.016	0.298 ± 0.015	0.709
Apo A1 (mg/dl)	133.132 ± 2.217	135.227 ± 2.163	0.506
Apo B (mg/dl)	74.467 ± 2.163	73.433 ± 1.703	0.704
hs-CRP (mg/dl)	**3.509 ± 0.421**	**1.989 ± 0.187**	**0.001**
HOMA-IR	2.896 ± 0.189	3.001 ± 0.252	0.755
HIRI	1145.944 ± 76.126	1235.176 ± 101.531	0.506
MISI ($x10^2$)	0.022 ± 0.003	0.018 ± 0.001	0.214
DI	0.813 ± 0.045	0.849 ± 0.038	0.546
ISI	3.841 ± 0.291	3.574 ± 0.181	0.417
IGI	0.888 ± 0.072	0.85 ± 0.264	0.903

Table 3. Baseline demographic and metabolic characteristics according to dietary pattern. *Values expressed as mean ± SEM.* ***BMI:*** *Body mass index;* ***HbA1c:*** *Glycated hemoglobin;* ***TG:*** *Triglycerides;* ***c-LDL:*** *Low density lipoprotein;* ***c-HDL:*** *High density lipoprotein;* ***NEFA:*** *Non-esterified fatty acids;* ***Apo A1:*** *Apolipoprotein A1;* ***Apo B:*** *Apolipoprotein B;* ***hs-CRP:*** *High sensitivity C-reactive protein;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index.* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Table 4

A

Non-T2DM	**Basal mean ± SEM**	**3rd year mean ± SEM**	***p*-value**
HbA1c (%)	**5.873 ± 0.033**	**5.538 ± 0.033**	**<0.001**
Glucose (mg/dl)	**94.58 ± 0.931**	**96.58 ± 0.923**	**0.033**
Insulin (mU/l)	7.960 ± 0.516	7.674 ± 0.607	0.711
ISI	4.113 ± 0.286	4.316 ± 0.295	0.432
HOMA-IR	**2.538 ± 0.129**	**1.79 ± 0.125**	**<0.001**
HIRI	**1022.711 ± 51.75**	**754.442 ± 61.552**	**<0.001**
IGI	0.999 ± 0.081	1.075 ± 0.135	0.537
DI	0.904 ± 0.046	0.97 ± 0.051	0.254
MISI (x10^2)	0.022 ± 0.002	0.018 ± 0.001	0.211
Total cholesterol (mg/dl)	159.962 ± 3.052	158.457 ± 2.936	0.554
c-HDL (mg/dl)	**44.856 ± 0.893**	**42.135 ± 0.844**	**0.002**
TG (mg/dl)	109.228 ± 4.789	110.057 ± 5.709	0.859
hs-CRP (mg/dl)	2.422 ± 0.329	2.038 ± 0.206	0.268

B

Non-T2DM (LF Diet)	**Basal mean ± SEM**	**3rd year mean ± SEM**	***p*-value**
HbA1c (%)	**5.906 ± 0.042**	**5.492 ± 0.038**	**<0.001**
Glucose (mg/dl)	94.44 ± 1.208	95.24 ± 1.312	0.574
Insulin (mU/l)	8.016 ± 0.753	8.13 ± 1.019	0.926
ISI	4.25 ± 0.545	4.581 ± 0.492	0.467
HOMA-IR	**2.613 ± 0.209**	**1.811 ± 0.177**	**<0.001**
HIRI	**1033.922 ± 83.559**	**795.222 ± 103.476**	**0.057**
IGI	0.938 ± 0.107	1.07 ± 0.16	0.418
DI	**0.847 ± 0.059**	**1.04 ± 0.069**	**0.008**
MISI (x10^2)	0.024 ± 0.005	0.02 ± 0.003	0.413
Total cholesterol (mg/dl)	157.66 ± 4.446	156.42 ± 4.39	0.697
c-HDL (mg/dl)	**43.735 ± 1.244**	**40.449 ± 0.95**	**0.013**
TG (mg/dl)	113.88 ± 7.876	109.46 ± 7.32	0.466
hs-CRP (mg/dl)	3.246 ± 0.623	2.177 ± 0.285	0.105

C

Non-T2DM (MedDiet)	**Basal mean ± SEM**	**3rd year mean ± SEM**	***p*-value**
HbA1c (%)	**5.844 ± 0.049**	**5.579 ± 0.053**	**<0.001**
Glucose (mg/dl)	**94.7 ± 1.403**	**97.77 ± 1.288**	**0.014**
Insulin (mU/l)	7.911 ± 0.715	7.268 ± 0.705	0.508
ISI	4.015 ± 0.303	4.128 ± 0.365	0.714
HOMA-IR	**2.474 ± 0.159**	**1.773 ± 0.177**	**<0.001**
HIRI	**1012.701 ± 64.185**	**718.032 ± 71.573**	**<0.001**
IGI	1.045 ± 0.116	1.079 ± 0.204	0.852
DI	0.945 ± 0.067	0.919 ± 0.073	0.761
MISI (x10^2)	0.019 ± 0.002	0.017 ± 0.002	0.238
Total cholesterol (mg/dl)	162.054 ± 4.215	160.302 ± 3.957	0.658
c-HDL (mg/dl)	**45.854 ± 1.27**	**43.636 ± 1.329**	**0.059**
TG (mg/dl)	105 ± 5.698	110.6 ± 8.697	0.426
hs-CRP (mg/dl)	1.689 ± 0.248	1.915 ± 0.297	0.431

Table 4. Demographic and metabolic characteristics after three years of follow-up of Non-T2DM cases (A), under LF Diet (B) and under Mediterranean Diet (C). *Values expressed as mean ± SEM.* ***HbA1c:*** *Glycated hemoglobin;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index;* ***c-HDL:*** *High density lipoprotein;* ***TG:*** *Triglycerides;* ***hs-CRP:*** *High sensitivity C-reactive protein.* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Table 5

A

Incident-T2DM	Basal mean ± SEM	3rd year mean ± SEM	*p*-value
HbA1c (%)	**6.016 ± 0.036**	**5.667 ± 0.046**	**<0.001**
Glucose (mg/dl)	**95.94 ± 1.081**	**102.03 ± 1.205**	**<0.001**
Insulin (mU/l)	10.81 ± 0.677	11.349 ±0.889	0.515
ISI	3.459 ±0.213	3.487 ± 0.269	0.914
HOMA-IR	3.354 ±0.318	2.989 ± 0.272	0.25
HIRI	1362.571 ± 127.201	1197.774 ± 108.225	0.192
IGI	0.596 ± 0.35	1.017 ± 0.22	0.287
DI	**0.771 ± 0.037**	**0.68 ± 0.044**	**0.077**
MISI (x10^2)	0.02 ± 0.002	0.021 ± 0.002	0.834
Total cholesterol (mg/dl)	165.606 ±3.596	172.212 ±3.852	0.046
c-HDL (mg/dl)	**43.705 ± 1.093**	**41.989 ±1.104**	**0.011**
TG (mg/dl)	132.684 ± 6.96	139.929 ± 7.364	0.28
hs-CRP (mg/dl)	2.725 ±0.292	2.686 ± 0.273	0.895

B

Incident-T2DM (LF Diet)	Basal mean ± SEM	3rd year mean ± SEM	*p*-value
HbA1c (%)	**6.081 ± 0.051**	**5.689 ± 0.072**	**<0.001**
Glucose (mg/dl)	**95.15 ± 1.895**	**101 ± 1.896**	**0.002**
Insulin (mU/l)	10.276 ± 0.975	9.112 ± 0.744	0.17
ISI	3.564 ± 0.388	3.811 ± 0.424	0.516
HOMA-IR	**3.184 ± 0.338**	**2.319 ± 0.201**	**0.018**
HIRI	**1260.763 ± 136.871**	**925.57 ± 80.505**	**0.021**
IGI	0.787 ± 0.088	0.634 ± 0.097	0.105
DI	0.739 ± 0.061	0.719 ± 0.092	0.84
MISI (x10^2)	0.023 ± 0.004	0.021 ± 0.003	0.708
Total cholesterol (mg/dl)	**167.8 ± 5.874**	**178.025 ± 7.355**	**0.064**
c-HDL (mg/dl)	43.5 ± 1.346	43.2 ± 1.361	0.765
TG (mg/dl)	126.103 ± 9.817	136.667 ± 11.105	0.23
hs-CRP (mg/dl)	3.754 ± 0.585	136.667 ± 11.105	0.282

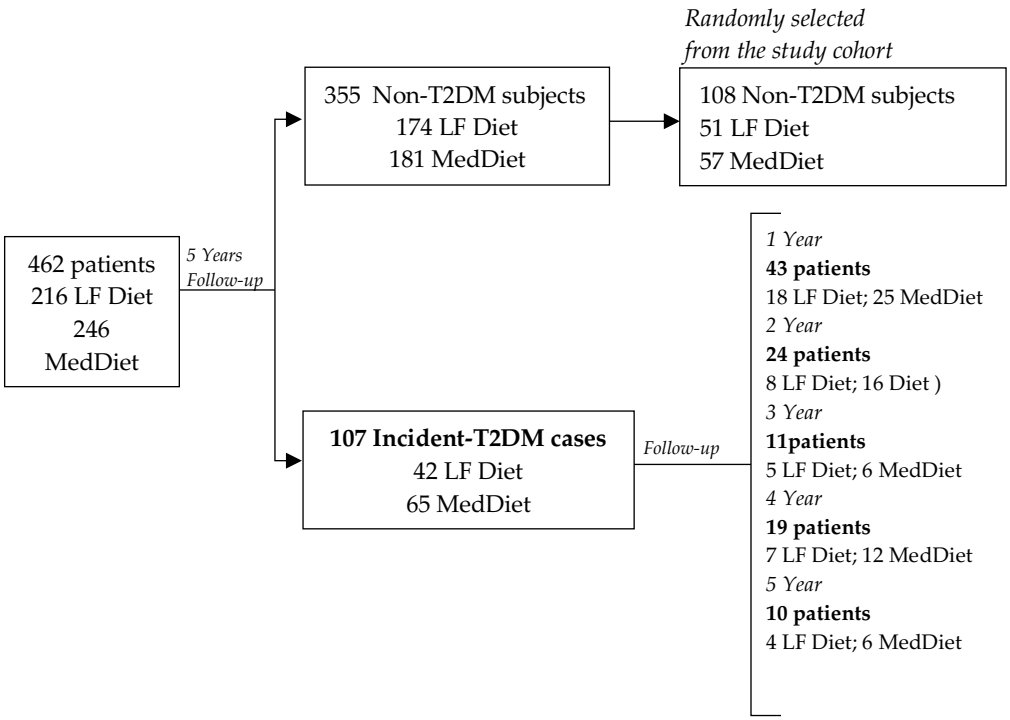
C

Incident-T2DM (MedDiet)	Basal mean ± SEM	3rd year mean ± SEM	*p*-value
HbA1c (%)	**5.975 ± 0.048**	**5.652 ± 0.06**	**<0.001**
Glucose (mg/dl)	**96.49 ± 1.284**	**102.75 ± 1.567**	**<0.001**
Insulin (mU/l)	11.175 ± 0.929	12.878 ± 1.379	0.179
ISI	3.393 ± 0.25	3.283 ± 0.348	0.759
HOMA-IR	3.469 ± 0.483	3.442 ± 0.4260.426	0.955
HIRI	1432.14 ± 193.086	1383.778 ± 170.206	0.798
IGI	0.472 ± 0.577	1.267 ± 0.356	0.222
DI	**0.791 ± 0.047**	**0.656 ± 0.043**	**0.013**
MISI (x10^2)	0.018 ± 0.002	0.021 ± 0.004	0.574
Total cholesterol (mg/dl)	164.119 ± 4.564	168.271 ± 4.091	0.315
c-HDL (mg/dl)	**43.854 ± 1.624**	**41.109 ± 1.632**	**0.003**
TG (mg/dl)	137.034 ± 9.592	142.085 ± 9.851	0.599
hs-CRP (mg/dl)	**2.021 ± 0.250**	**2.426 ± 0.341**	**0.079**

Table 5. Demographic and metabolic characteristics after three years of follow-up of Incident-T2DM patients (A), under LF Diet (B) and under Mediterranean Diet (C). *Values expressed as mean ± SEM.* ***HbA1c:*** *Glycated hemoglobin;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index;* ***c-HDL:*** *High density lipoprotein;* ***TG:*** *Triglycerides;* ***hs-CRP:*** *High sensitivity C-reactive protein.* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

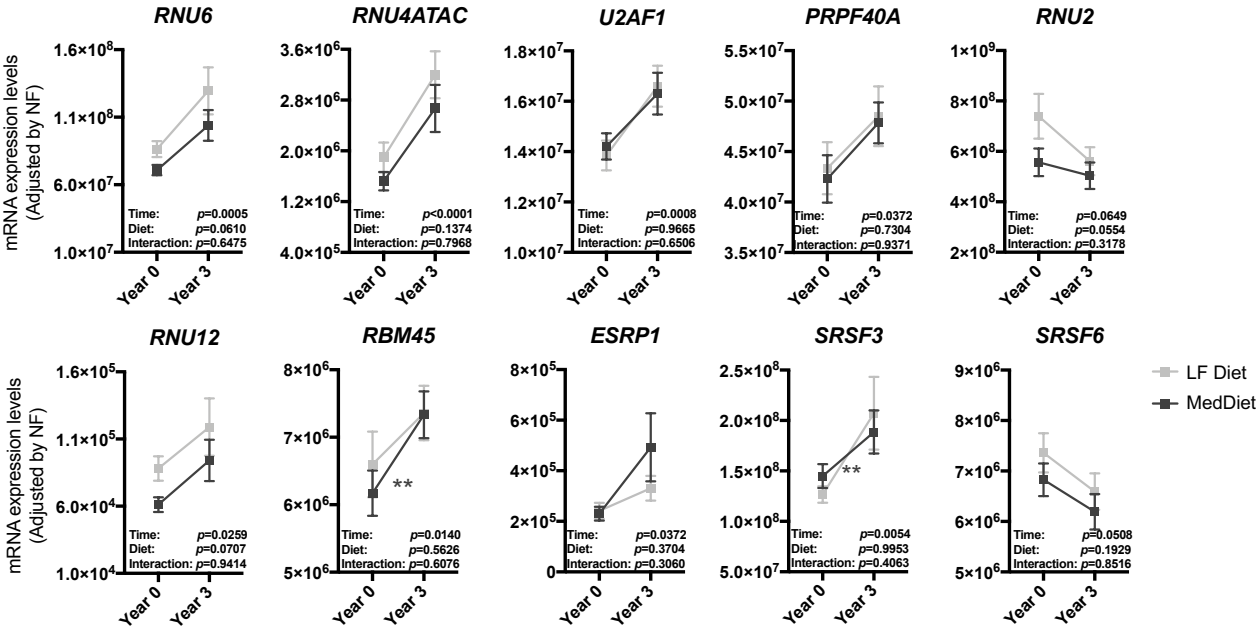
Supplementary Figures

Supplementary Figure 1



Supplementary figure 1. *Graphical scheme of the patients study timeline. Specifically, from the initial 462 non-T2DM patients included in the CORDIOPREV study, 107 patients developed T2DM (Incident-T2DM cases) after a mean follow-up of 5-years according to all the American Diabetes Association (ADA) diagnosis criteria evaluated on the basis of glucose tolerance tests (OGTT) performed each year. In the present study. all these 107 incident-T2DM cases and 108 matched controls (non-T2DM subjects. randomly selected from the remaining 355 subjects that did not develop T2DM during the study period) were included. The number of patients enrolled in each dietary pattern is included.* ***LF Diet****: Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Supplementary Figure 2



Supplementary Figure 2. PBMCs expression pattern of specific splicing machinery components after three years of follow-up under two healthy dietary pattern (LF Diet and MedDiet). *mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of GAPDH and ACTB] of specific spliceosome components and splicing factors in the PBMCs from all the patients included in the study. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-T2DM patients (t-test: **, p<0.01).* ***LF Diet:*** *Low-fat Diet;* ***MedDiet:*** *Mediterranean Diet.*

Supplementary Tables

Supplementary Table 1

	Basal mean ± SEM	3rd year mean ± SEM	*p*-value
HbA1c (%)	**5.981 ± 0.034**	**5.576 ± 0.039**	**<0.001**
Glucose (mg/dl)	**94.76 ± 1.076**	**97.84 ± 1.152**	**0.008**
Insulin (mU/l)	9.034 ± 0.611	8.572 ± 0.651	0.551
ISI	3.932 ± 0.343	4.224 ± 0.33	0.327
HOMA-IR	**2.872 ± 0.192**	**2.042 ± 0.135**	**<0.001**
HIRI	**1136.125 ± 77.340**	**853.952 ± 67.436**	**0.003**
IGI	0.868 ± **0.071**	0.867 ± 0.099	0.996
DI	0.798 ± 0.043	0.894 ± 0.059	0.116
MISI ($x10^2$)	0.024 ± 0.003	0.02 ± 0.002	0.369
Total cholesterol (mg/dl)	162.167 ± 3.613	166.022 ± 4.21	0.203
c-HDL (mg/dl)	**43.629 ± 0.909**	**41.685 ± 0.813**	**0.024**
TG (mg/dl)	119.236 ± 6.169	121.382 ± 6.493	0.676
hs-CRP (mg/dl)	**3.473 ± 0.431**	**2.5759 ± 0.258**	**0.051**

Supplementary Table 1. Demographic and metabolic characteristics after three years of follow-up of patients under Low-Fat Diet. *Values expressed as mean ± SEM.* ***BMI:*** *Body mass index;* ***HbA1c:*** *Glycated hemoglobin;* ***TG:*** *Triglycerides;* ***c- LDL:*** *Low density lipoprotein;* ***c-HDL:*** *High density lipoprotein;* ***NEFA:*** *Non-esterified fatty acids;* ***Apo A1:*** *Apolipoprotein A1;* ***Apo B:*** *Apolipoprotein B;* ***hs-CRP:*** *High sensitivity C-reactive protein;* ***HOMA-IR:*** *Homeostasis model assessment-insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index.*

Supplementary Table 2

	Basal mean ± SEM	**3rd year mean ± SEM**	***p*-value**
HbA1c (%)	**5.912 ± 0.035**	**5.617 ± 0.04**	**<0.001**
Glucose (mg/dl)	**95.62 ± 0.949**	**100.32 ± 1.041**	**<0.001**
Insulin (mU/l)	9.599 ± 0.608	10.169 ± 0.83	0.478
ISI	3.707 ± 0.198	3.71 ± 0.255	0.991
HOMA-IR	**2.984 ± 0.263**	**2.629 ± 0.247**	**0.173**
HIRI	**1229.653 ± 105.956**	**1062.383 ± 99.161**	**0.111**
IGI	0.759 ± 0.294	1.173 ± 0.204	0.218
DI	0.87 ± 0.042	0.790 ± 0.045	0.117
MISI ($x10^2$)	0.019 ± 0.001	0.019 ± 0.002	0.944
Total cholesterol (mg/dl)	163.123 ± 3.105	164.43 ± 2.863	0.646
c-HDL (mg/dl)	**44.854 ± 1.03**	**42.373 ± 1.054**	**0.001**
TG (mg/dl)	121.579 ± 5.848	126.895 ± 6.738	0.374
hs-CRP (mg/dl)	1.859 ± 0.176	2.177 ± 0.227	0.081

Supplementary Table 2. Demographic and metabolic characteristics after three years of follow-up of patients under Mediterranean Diet. *Values expressed as mean ± SEM.* ***BMI:*** *Body mass index;* ***HbA1c:*** *Glycated hemoglobin;* ***TG:*** *Triglycerides;* ***c- LDL:*** *Low density lipoprotein;* ***c-HDL:*** *High density lipoprotein;* ***NEFA:*** *Non-esterified fatty acids;* ***Apo A1:*** *Apolipoprotein A1;* ***Apo B:*** *Apolipoprotein B;* ***hs-CRP:*** *High sensitivity C-reactive protein;* ***HOMA-IR:*** *Homeostasis model assessment- insulin resistance;* ***HIRI:*** *Hepatic insulin resistance index;* ***MISI:*** *Muscle insulin sensitivity index;* ***DI:*** *Disposition index;* ***ISI:*** *Insulin sensitivity index;* ***IGI:*** *Insulinogenic index.*

Article III

Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease

Mercedes del Río-Moreno, MS, Emilia Alors-Pérez, MS, Sandra González-Rubio, PhD, Gustavo Ferrín, PhD, Oscar Reyes, PhD, Manuel Rodríguez-Perálvarez, MD PhD, Marina E. Sánchez-Frías, MD PhD, Rafael Sánchez-Sánchez, MD, Sebastián Ventura, PhD, José López-Miranda, MD, Rhonda D. Kineman, PhD, Manuel de la Mata, MD, Justo P. Castaño, PhD, Manuel D. Gahete, PhD, Raúl M. Luque, PhD

The Journal of Clinical Endocrinology & Metabolism
Endocrine Society

Submitted: January 04, 2019
Accepted: March 18, 2019
First Online: March 22, 2019

The Journal of Clinical Endocrinology & Metabolism; Copyright 2019 DOI: 10.1210/jc.2019-00021

Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease

Mercedes del Río-Moreno, MS[1,2,3,4], Emilia Alors-Pérez, MS[1,2,3,4], Sandra González-Rubio, PhD[1,3,5], Gustavo Ferrín, PhD [1,3,5], Oscar Reyes, PhD [1,6], Manuel Rodríguez-Perálvarez, MD PhD [1,3,5], Marina E. Sánchez-Frías, MD PhD[1,3,7], Rafael Sánchez-Sánchez, MD[1,3,7], Sebastián Ventura, PhD [1,6,8], José López-Miranda, MD[1,3,4,9], Rhonda D. Kineman, PhD [10], Manuel de la Mata, MD[1,3,5], Justo P. Castaño, PhD [1,2,3,4], Manuel D. Gahete, PhD [1,2,3,4*], Raúl M. Luque, PhD [1,2,3,4*].

[1]Maimonides Institute for Biomedical Research of Cordoba (IMIBIC), Córdoba, Spain. [2]Department of Cell Biology, Physiology and Immunology, University of Córdoba, Córdoba, Spain. [3]Reina Sofia University Hospital, Córdoba, Spain. [4]CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn), Córdoba, Spain. [5]CIBER Digestive and Liver Diseases (CIBERehd), Córdoba, Spain. [6]Department of Computer Sciences, University of Cordoba, Cordoba, Spain. [7]Anatomical Pathology Service, Reina Sofia University Hospital (HURS), Cordoba, Spain. [8]Department of Information Systems, King Abdulaziz University, Saudi Arabia Kingdom. [9]Lipids and Atherosclerosis Unit, Reina Sofia University Hospital, Córdoba, Spain. [10]Research and Development Division, Jesse Brown Veterans Affairs Medical Center; Section of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Illinois at Chicago, Chicago, Illinois, USA.

ORCiD numbers:

0000-0002-7585-1913

Luque

Raul M

Received 04 January 2019. Accepted 18 March 2019.

*These authors should be considered as corresponding authors.

Context: Non-alcoholic fatty liver disease (NAFLD) is a common obesity-associated pathology characterized by hepatic fat accumulation, which can progress to fibrosis, cirrhosis and hepatocellular carcinoma. Obesity is associated with profound changes in gene expression patterns of the liver, which could contribute to the onset of comorbidities.
Objective: Since these alterations might be linked to a dysregulation of the splicing process, we aimed to determine whether the dysregulation in the expression of splicing machinery components could be associated with NAFLD.
Methods: We collected n=41 liver biopsies from non-alcoholic obese individuals with or without hepatic steatosis that underwent bariatric surgery. The expression pattern of splicing machinery components was determined using a microfluidic qPCR-based array. An *in vitro* approximation to determine lipid accumulation using HepG2 cells was also implemented.
Results: The liver of obese steatotic patients exhibited a severe dysregulation of certain splicing machinery components compared to non-steatotic obese patients. Non-supervised clustering analysis allowed the identification of three molecular phenotypes of NAFLD with a unique fingerprint of alterations in splicing machinery components, which also presented distinctive hepatic and clinical-metabolic alterations, and a differential response to bariatric surgery after one year. In addition, *in vitro* silencing of certain splicing machinery components (i.e. PTBP1, RBM45, SND1) reduced fat accumulation and modulated the expression of key *de novo* lipogenesis enzymes, whereas, conversely, fat accumulation did not alter spliceosome components expression.

Conclusion: There is a close relationship between splicing machinery dysregulation and NAFLD development, which should be further investigated to identify novel therapeutic targets.

Dysregulations of the splicing machinery precede the development of steatosis and define different molecular phenotypes of NAFLD. This study provides novel information to explore strategies for NAFLD.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is rapidly emerging as an important and growing health issue in western countries, due to the rising prevalence of obesity, type-2 diabetes (T2D) and metabolic syndrome (1). NAFLD is described as a range of liver disorders characterized by fat accumulation within the liver (steatosis), which is not related to alcohol consumption. These disorders comprise a wide range of diseases, from simple steatosis to hepatic inflammation and fibrosis (non-alcoholic steatohepatitis or NASH), cirrhosis and hepatocellular carcinoma (HCC) (2,3). Although most patients with simple steatosis would remain stable, 10–15% with histologically proven NASH will progress to cirrhosis and HCC. Indeed, due to its high prevalence, NAFLD has emerged as speedily growing cause of end-stage liver disease and HCC, in addition to hepatitis C, hepatitis B and alcohol abuse (4,5). Unfortunately, the molecular mechanisms underlying the heterogeneous outcomes of NAFLD remain unclear, precluding any attempt to anticipate the disease progression to decompensated cirrhosis or HCC (6). Within the natural history of NAFLD, hepatic steatosis is the first stage, wherein an improved understanding of the pathogenesis of liver steatosis would have a critical prognostic impact for preventing disease progression (7,8). In this sense, while hepatic steatosis is closely associated with obesity, there is a meaningful percentage of obese people who have normal intrahepatic triglyceride content and appear to be resistant to developing obesity-related metabolic complications, including NAFLD (9). However, little is known about the mechanisms underlying the apparent resistance of this group of patients.

Several studies have demonstrated that NAFLD development and progression result from a combination of environmental and genetic factors (8,10-12). Many of these studies have performed transcriptome profiling by microarray in humans with NAFLD and assessed changes in gene expression; however, most studies do not address changes in RNA alternative splicing. This process is of relevance because numerous diseases course with a dysregulation in the process of alternative RNA splicing and because alterations in alternative RNA splicing are associated with inflammation, metabolic disorders and cancer, which are critical hallmarks in the natural history of NAFLD (13-15). Splicing process is catalyzed by the minor and major spliceosomes, which act on different types of introns (13). The spliceosome is an intricate macromolecular complex, whose functional core is comprised by several small nuclear ribonucleoproteins (snRNPs) subunits, which dynamically interact to regulate the splicing process. In addition, the activity of the spliceosome is modulated by many splicing factors that specifically recognize certain sequences in exons and introns (16). Consequently, the dysregulation of the expression and/or function of certain spliceosomal components may result into an aberrant splicing process (16).

In recent years, a growing body of evidence pinpoints an obvious association between several metabolism-related elements and processes of alternative splicing, which could represent potential tools for the translational research on NAFLD. Specially, genes linked to obesity are regulated by alternative splicing (17-20) and some studies have underlined the importance of splicing variants in NAFLD and NASH development (21,22). The appearance of these aberrant alternative splicing variants might be linked to a dysregulation of the cellular machinery responsible for this process, including the spliceosome and splicing

factors (13,14,23). Actually, some studies have found correlations in the expression levels between specific splicing isoforms and certain splicing factors in the context of obesity and diabetes (24,25). Moreover, NAFLD and NASH are associated with changes in the mRNA expression of certain splicing factors in the liver of obese patients (10,26-28).

Taken as a whole, the existing evidence suggests that the cellular machinery responsible for the catalysis and regulation of the splicing process (spliceosome components and splicing factors) could be dysregulated and associated with obesity-derived NAFLD, wherein we hypothesize that these alterations could be causally related to the development of NAFLD. Therefore, the aim of the present study was to implement a novel and comprehensive approach to identify alterations in the expression of spliceosome components and splicing factors that could be associated to the development of hepatic steatosis in obese patients.

Material and Methods

Patients and samples

All experimental protocols were carried out in accordance with approved guidelines of the IMIBIC/University Hospital Ethics Committees. Informed consent was obtained from all participants. Excessive alcohol consumption, as defined by a daily alcohol intake >20g, was the main exclusion criterion. Liver biopsies were obtained during bariatric surgery from obese women (BMI>30) with (n=32) and without (n=9) hepatic steatosis. The presence and grade of steatosis was assessed by the same operator using liver ultrasound. Patients were classified in: absence of steatosis (normal liver pattern), mild steatosis (increase in echoes of the liver parenchyma as compared with the kidney cortex, but with adequate visualization of vessels and diaphragm), moderate steatosis (increased fine echoes within the liver parenchyma with impaired visualization of vascular structures, but not diaphragm) and severe steatosis (marked increase in echoes of the liver parenchyma with poor or no visualization of vascular structures, posterior right lobe and diaphragm). The presence of fibrosis and inflammation was evaluated by liver histology by two experienced pathologists. Lipid content was also evaluated by liver histology; however, the result of these analyses did not completely match with the classification obtained according to liver ultrasound. In this sense, biopsy histology, considered the gold standard, has important limitations (29-32), including that is not capable to evaluate the liver completely and it is not suitable to be used to determine the follow-up of the patients, one of the main aims of this study. For these reasons, as one of the major aims of this study was to correlate the changes observed with the follow-up of the patients, and liver echography is also a well-recognized established tool for the screening of fatty liver (32-34), we used the classification established by liver echography. Since NALFD is a tremendously gender dimorphic disease (35) and the gene expression patterns in the liver are clearly gender-dependent (36), in this study only women were included, as they represented the vast majority of subject undergoing bariatric surgery in our hospital during the study period.

RNA extraction and reverse transcription

Details regarding RNA extraction, quantification and reverse transcription have been previously reported elsewhere by our group (37-39). Specifically, total RNA from liver samples was isolated using AllPrep DNA/RNA/Protein Mini Kit following the manufacturer's protocol (Qiagen, Madrid, Spain) and RNA from the HepG2 cell line using TRI Reagent (Sigma-Aldrich, Madrid, Spain), followed by DNase treatment. The amount and purity of RNA recovered was determined using the NanoDrop2000 spectrophotometer (Thermo Scientific, Madrid, Spain). One μg of RNA was reverse transcribed to cDNA using random hexamer primers and RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Madrid, Spain).

Analysis of splicing machinery components by qPCR dynamic array based on microfluidic technology

A 48.48 Dynamic Array (Fluidigm, San Francisco, CA, USA) using a microfluidic-based technique for gene expression analysis was implemented to determine the expression of 48 transcripts in 48 samples, simultaneously, as previously reported (40). Specific primers for human transcripts including the components of the major (n=13) and minor spliceosome (n=4), as well as associated splicing factors (n=28) and three housekeeping genes [Supplementary Table 1 (41)] were specifically designed with the Primer3 software and StepOne™ Real-Time PCR System software v2.3 (Applied Biosystems, Foster City, CA, USA). Preamplification, exonuclease treatment and qPCR dynamic array based on microfluidic technology were implemented following manufacturer's instructions using the Biomark System and the Real-Time PCR Analysis Software (Fluidigm). Additional details are provided in Supplementary Material and Methods (41). To control for variations in the amount of RNA used and the efficiency of the reverse-transcription reaction, the expression level of each transcript was adjusted by *ACTB* (silencing experiments in HepG2 cells) or by a Normalization Factor (NF) obtained from the expression levels of different housekeeping genes (*ACTB* and *HPRT* in human samples and *GAPDH* and *HPRT* in HepG2 cells) using Genorm 3.3 (42).

Quantitative real time PCR (qPCR)

qPCR reactions were carried out using the Stratagene Mx3000p system with the Brilliant III SYBR Green Master Mix (Stratagene, La Jolla, CA, USA) as previously reported (37-39). The primers for human transcripts were designed and qPCR results generated adjusted by a NF as explained above. Specific primers for human transcripts for *de novo* lipogenesis enzymes genes are included in Supplementary Table 2 (41).

Cell culture and treatments

Human HepG2 cell line was purchased from the American Type Culture Collection (HB-8065) and maintained following manufacturer´s instructions at 37°C and 5% CO_2. This cell line was validated by the analysis of short tandem repeat analysis (STR; GenePrint 10 System, Promega, Madrid, Spain) and routinely tested for mycoplasma contamination as previously reported (43). To explore the modulation of the expression of key spliceosome components and splicing factors, 24 hours treatments with glucose (25mM, Sigma Aldrich), insulin (10nM, Sigma Aldrich), IGF1 (10nM, I3769, Sigma Aldrich), leptin (10ng/ml, 003-12, Phoenix Pharmaceuticals), oleic acid (OA, 500µM, O1383, Sigma-Aldrich) and palmitic acid (PA, 500µM, P9767, Sigma-Aldrich) were performed. Doses were chosen based on previous studies (44,45).

Transfection with specific siRNAs

To perform silencing experiments, specific siRNAs for the factors of interest were used (Thermo Scientific, Madrid, Spain; and Origene, Herford, Germany). A commercial negative control (scramble; Thermo Scientific, Madrid, Spain) was used in all the experiments. For transfection assays, 200,000 cells were seeded in FBS-containing medium during 48h. Then, medium was replaced by antibiotic-antimycotic free medium and cells were transfected with 100nM of each specific siRNA using Lipofectamine RNAiMAX reagent (Thermo Scientific, Madrid, Spain). After 48h, cells were detached and seeded to analyze *in vitro* lipid accumulation experiments or to extract RNA to validate the transfection efficiency and to determine the expression of key *de novo* lipogenesis enzymes (measured by qPCR).

Lipid accumulation experiments and Oil Red O staining in HepG2

HepG2 cells were treated with OA for two different purposes: 1) to assess the effect of the silencing of selected splicing machinery components on lipid accumulation; or, 2) to evaluate the effect of lipid accumulation on the expression of the selected splicing machinery

components. Briefly, cells were seeded and, once 80% of cell confluence was reached, cells were cultured in FBS-free medium and 0,5% free fatty acids bovine serum albumin (BSA) for 1h. Then, HepG2 cells were treated with 500µM OA in FBS-free medium and 2% BSA for 24h (for the silencing experiment in HepG2 cells) or 10, 24, 48h (naive HepG2 cells). After the incubation time, medium was removed and silenced cells were washed with PBS and fixed with 10% paraformaldehyde for 15 minutes, while naive HepG2 cells were lysed with Trizol reagent to isolate RNA as described above. Control cells were treated with OA-free medium. In silenced HepG2 cells, after removing paraformaldehyde, cells were washed and incubated with 60% isopropanol. Then, Oil Red O working solution was added for 10 minutes in agitation and washed. At this point, cells were examined under a light inverse microscope to acquire images. OD was measured at a wavelength of 520 nm in the FlexStation III system (Molecular Devices, Sunnyvale, CA, USA).

Alamar Blue assay

Alamar Blue assay (Bio-Rad Laboratories, Madrid, Spain) was used to determine cell viability, as previously reported (43). Briefly, cell viability was analyzed at 0h (before OA treatment) and 24h (after OA treatment and before Oil Red O staining) by measurement of fluorescent signal exciting at 560nm and reading at 590nm (Flex Station 3). Specifically, the day of the measurement, Alamar blue reduction was measured after cells were incubated for 3h with 10% Alamar blue/serum free media.

Statistical and bioinformatical analysis

Data are expressed as mean ± standard error of the mean (SEM) or compared with the corresponding controls (set at 100%). Data were evaluated for heterogeneity of variance using the Kolmogorov–Smirnov test and, consequently, parametric (Student t) or nonparametric (Mann-Whitney U) tests were implemented to analyze the statistical differences. Heatmaps were generated using MetaboAnalyst 3.0 (46). Predictive models were constructed by Random Forest algorithm (with R language) as classifier, followed by a leave one-out cross-validation using the full-dataset of variables or a selection of them (obtained by feature ranking processes) as described in Supplementary Material and Methods (41). P-values smaller than 0.05 were considered statistically significant. All statistics analyses were performed using the GraphPad Prism 6.0 software (La Jolla, CA, USA) (46).

RESULTS

Patients and clinical correlations

Demographic and clinical features of the 41 patients included in the study (9 without and 32 with hepatic steatosis) are summarized in Table 1, while same parameters in the steatotic patients according to the grade of hepatic steatosis are shown in Supplementary Table 3 (41). In general, steatotic patients presented higher levels of plasma alanine aminotransferase (ALT; p=0.041), which validates the ultrasound diagnosis, and higher Homeostatic Model Assessment for insulin resistance (HOMA-IR; p=0.049) and fasting insulin levels (p=0.049) compared to non-steatotic patients. Within the steatotic patients, HOMA-IR and total protein plasma levels were associated with the grade of steatosis (p=0.03 and 0.007, respectively). Histology analysis revealed no signal of fibrosis or NASH. It should be noted that these patients presented an early stage of NAFLD, only with the presence of steatosis and not fibrosis, inflammation or liver damage.

Hepatic steatosis was associated with alterations in splicing machinery components

Dynamic qPCR array showed that livers from steatotic patients showed a marked dysregulation of the expression levels of several (16 out of 45) spliceosomal components and splicing factors compared to livers from non-steatotic patients (Fig. 1A), such as the snRNPs *RNU6ATAC*, *RNU6* or *RNU2* and the splicing factors *PTPBP1* or *SRRM1* (Fig. 1B). Indeed,

individual ROC curves analyses of each of the altered elements revealed that the expression of these spliceosome components and splicing factors was clearly different between patients with or without steatosis [Fig. 1B; specificity and sensitivity values for each ROC curve are shown in Supplementary Table 4 (41)]. Random Forest analysis demonstrated the existence of splicing machinery-associated molecular fingerprints that were significantly different between steatotic and non-steatotic subjects. Particularly, the combination of the expression levels of the factors *PTBP1*, *SRRM1*, *RBM45*, *RNU4* and *RNU6ATAC* generated a ROC curve that was able to classify our cohort in steatotic and non-steatotic patients with an AUC of 0.960 (Fig. 1C). The validity and consistency of this and other splicing machinery-associated molecular fingerprints were demonstrated by cross-validation analysis (Fig. 1C). A more detailed analysis demonstrated that alterations in the expression levels of the components of this machinery were not associated with the grade of steatosis since a progressive dysregulation of these elements in individuals with different steatotic grades was not observed (Fig. 2).

Dysregulation of splicing machinery in steatotic condition could be associated with specific clinical alterations

Non-supervised hierarchical analysis of steatotic patients, according to the expression pattern of the spliceosome components and splicing factors, identified the existence of three distinct molecularly defined groups of steatotic patients (Clusters A, B and C), which were grouped together by the presence of similar alterations in certain spliceosome components and splicing factors (Fig. 3A). These three molecularly defined clusters of steatotic patients were characterized by alterations in the expression levels of key splicing machinery components, which were altered in one particular molecular cluster and that could define a specific fingerprint (Fig. 3B). Indeed, patients in molecular Cluster A are defined by low levels of the spliceosome components *RBM22*, *SNRNP200*, *SF3B1 tv1*, *PRPF40A* and *PRPF8*, and the splicing factors *SRSF4*, *TRA2B*, *TIA1*, *SNW1*, *CELF1* and *RBM3*; patients in molecular Cluster B by high levels of the splicing factors *RBM45* and *TRA2A;* and patients in molecular Cluster C by high levels of the spliceosome components *RNU6ATAC*, *RNU4*, *RNU11*, *RNU2*, *RNU12*, *U2AF2*, *TCERG1*, *RNU4ATAC*, *SF3B1* and *RNU6*, and the splicing factors *SND1*, *RAVER1*, *SRSF2*, *ESRP2*, and low levels of, *SRSF1*, *NOVA1* and *ESRP1*. In this sense, a combination of the expression levels of seven of these factors [two of them characterizing the cluster A (*SRSF4* and *TRA2B*), two characterizing the cluster B (*RBM45* and *TRA2A*) and three characterizing the cluster C (*SND1*, *RAVER1* and *RNU6ATAC*)] was able to perfectly classify patients in one of the three Clusters using the Random Forest algorithm followed by a cross-validation analysis (AUC=1, $p<0.001$).

Remarkably, the patients within each cluster were characterized by certain hepatic and clinical-metabolic alterations (Fig. 4). Thus, patients grouped in Cluster A (n=4) exhibited elevated glucose and haptoglobin levels compared to the rest of steatotic patients (Clusters B+C) and non-steatotic controls. Cluster B (n=17) presented high triglycerides and gamma-glutamyl transferase (GGT) plasma levels, as well as low alkaline phosphatase levels, in comparison with Clusters A+C and non-steatotic patients. Finally, patients in Cluster C (n=11) showed increased insulin, ALT and aspartate aminotransferase (AST) levels. In addition, patients included in Cluster C seemed to present a worsened biochemical and hepatic response to bariatric surgery after one year. These patients presented a differential response to bariatric surgery after one year of follow-up, inasmuch as only 70% of the patients included in Cluster C (7 out of 10) presented a total recovery from hepatic steatosis (determined by liver ultrasound) compared to 86% of patients included in Clusters A and B (12 out of 14), although this difference did not reach statistical significance (Fig. 5). In addition, patients included in cluster C exhibited a differential response one-year post-surgery in terms of normalization of GGT, glucose, triglycerides, alkaline phosphatase and HDL

plasma levels, compared to patients included in Clusters A and B, although only GGT and glucose differences reached statistical significance (Fig. 5). It should be noted that all patients similarly decreased BMI and WHR without differences between clusters.

Alterations in splicing machinery components influenced lipid accumulation and *de novo* lipogenesis in HepG2 cells

Several spliceosome components and splicing factors were selected for functional analysis in the HepG2 cell line based on the altered expression in the whole population of steatotic patients (*PTBP1*) or in a particular Cluster of steatotic patients [*SRSF4*, *RBM22* (Cluster A), *RBM45* (Cluster B), *SND1* and *RAVER1* (Cluster C)]. Remarkably, lipid accumulation in HepG2 cells induced by OA overload, which was validated by Oil Red O absorbance (Fig. 6A), did not alter the expression of any of the factors selected for functional analysis, except for a slight decrease in *SND1* after 10h of treatment (Fig. 6B). In contrast, silencing of the selected splicing factors using specific siRNAs (Fig. 6C), significantly altered lipid accumulation after treatment with OA (Fig. 6D). This approach demonstrated that the modulation (silencing) of certain selected factors, altered lipid accumulation, since the percentage of fat accumulation after OA treatment was lower in those cells compared with percentage of OA accumulation in negative control cells (Fig. 6D), without any obvious influence of cell viability [Supplementary Figure 1 (41)]. In addition, the silencing of *RBM22* and *SND1* produced a decrease in the expression levels of *FASN* and *SCD*, with no major significant changes in the expression levels of other enzymes implicated in *de novo* lipogenesis [Supplementary Figure 2 (41)].

Metabolic factors modulated the expression of key splicing factors

In vitro analysis in HepG2 cells showed that the expression of the key splicing machinery components selected above could be modulated in response to metabolic factors such as glucose, leptin, PA or IGF1. Specifically, glucose treatment significantly decreased the expression of *SND1* and leptin decreased the expression of *RBM22*, while IGF1 treatment significantly increased the expression of *RAVER1*. However, insulin did not modulate the expression of any of these selected splicing factors. Finally, while OA treatment could not modulate the expression levels of any of these factors (as shown before), PA produced an increase in the expression of *PTBP1* and *RBM22* [Supplementary Figure 3 (41)].

DISCUSSION

Our study represents a novel and comprehensive approximation for understanding the molecular dysregulations underlying the development of NAFLD in obese patients as compared to non-steatotic obese patients. Indeed, we have shown that the livers of steatotic obese patients exhibit an overt but differential (patient-dependent) alteration of the cellular machinery responsible for the regulation of the splicing process (spliceosome components and splicing factors). The clinical correlations together with the *in vitro* studies suggest that the dysregulation of certain spliceosome components and splicing factors could be causally linked to the development of hepatic steatosis and would be associated with specific NAFLD-related clinical complications (hyperinsulinemia, ALT levels, etc.), as well as with hepatic and biochemical improvements after bariatric surgery.

A growing body of evidence suggests that aberrant pre-mRNA splicing processes in the liver plays a significant role in the pathogenesis of NAFLD and NASH. Indeed, previous studies using microarray approaches have found an association between these liver disorders and changes in splicing factors expression, which could, thus, contribute to alterations in normal RNA splicing (6,10,28,47,48). Nevertheless, most studies have been focused on the consequences of changes in specific mRNA splicing variants (17,25). In the present study we have provided a proof-of-concept that alterations in the expression of spliceosome

components and splicing factors are associated to the development of hepatic steatosis in obese patients, including the main components of the spliceosome core (*RNU2*, *RNU6*, *SF3B1*, *RNU6ATAC* or *RNU4ATAC*) and a representative set of splicing factors (*PTBP1*, *SRRM1*, *SND1* or *SRSF2*) involved in the process. Some of these altered splicing factors have been previously related to liver development and metabolism, as it is the case of the polypyrimidine tract binding protein *PTBP1* (*HNRNPI*), which has been shown herein to be increased in the liver of steatotic patients, and has been reported to regulate the splicing of the fatty acid desaturase genes 2 and 3 (*FADS2* and *FADS*3), which are implicated in fatty acid elongation and desaturation (49), and multiple genes involved in cholesterol synthesis and uptake including *LDLR*, *MVK*, *HMGCS1*, and *PSCK9* (50). Also, studies in staphylococcal nuclease and tudor domain containing 1 (*SND1*) protein in rats have indicated that steatogenic conditions promote its action on low density lipid droplets (51). However, we describe herein, a dysregulation in some elements of the spliceosome core (snRNPs), which are essential for the appropriate recognition of intron sequences and the assembly of spliceosome to initiate the splicing process (16). These results show that a spliceosome-associated molecular fingerprint is able to discriminate between steatotic and non-steatotic livers in obese patients (with AUC>0.96). The fact that spliceosome alterations were not associated with the grade of hepatic steatosis suggests their role as triggers for initiating fat deposition within the liver. Indeed, the *in vitro* results suggest that alterations in spliceosome components and/or splicing factors could precede the development of hepatic steatosis in obese patients since the modulation of the expression of *PTBP1*, *SRSF4*, *RBM22*, *RBM45*, *SND1* and *RAVER1* lowered lipid accumulation in a well-established hepatocyte cell line model (HepG2 cells) (52), while an exogenous overload of lipids (OA) was not able to alter the expression of the mentioned spliceosome components and/or splicing factors. Remarkably, our results also indicate that the dysregulations in the expression of certain spliceosome components and/or splicing factors could have an impact on other key hepatic processes, inasmuch as silencing of *RBM22* and *SND1* decreased the expression of *FASN* and *SCD* enzymes, indicating a possible connection between the dysregulation of some of the splicing factors studied herein and *de novo* lipogenesis in the liver. In fact, it has been described that the inhibition of *SRPK2*, a regulator of the splicing factors RNA-binding SR proteins, results in intron retention and mRNA instability of lipogenic genes (53). It is also worth noticing that, although OA treatment did not alter the expression of these splicing machinery components, other metabolic factors, such as glucose, leptin, IGF1 or PA, modulated the expression of some of these key factors, suggesting a possible association between these metabolic factors, the dysregulation of specific spliceosome components and the development of NAFLD in obese patients. In this sense, several studies have previously shown a regulation of specific splicing factors by metabolic factors. It has been postulated that nutrients can modulate processes required for cell homeostasis through the alteration of gene expression and the splicing of pre-mRNAs encoding key regulatory proteins (e.g., insulin receptor, leptin receptor) (54). For example, it has been described that glucose can regulate the expression of the splicing factor *MBNL1* (55). Furthermore, the splicing of the fatty acid desaturase 3 has been observed to be modulated in the liver of baboons in response to different diets and in HepG2 cells after treatment with polyunsaturated fatty acids (56). This is especially important because the role of spliceosome in early NAFLD makes it attractive to further investigate derived screening strategies and preventive therapeutic targets.

Our results also pinpoint a putative association between changes in the expression of certain spliceosome components and splicing factors and NAFLD-associated comorbidities (hyperinsulinemia, ALT/AST levels, etc.) and with the biochemical response to bariatric surgery. Indeed, the expression pattern of the spliceosome components and splicing factors

analyzed herein seemed to identify three discrete molecularly defined subpopulations of steatotic obese patients characterized by the dysregulated expression of certain spliceosome components and splicing factors, associated to specific clinical characteristics and also with a differential response to bariatric surgery after one year. In particular, patients included in the Cluster A, which were mainly characterized by altered expression of *SRSF4*, *RBM22* and *TRA2B* compared to non-steatotic and Clusters B and C steatotic patients, presented an increase in glucose plasma levels, indicating a possible relationship between this splicing factors and glucose homeostasis. Indeed, one of the splicing factors dysregulated in this cluster of patients (*TRA2B*, also named *SRFS10*) has been previously described in relation to obesity, wherein *TRA2B* seems to be downregulated in liver samples from insulin resistant humans with obesity (26). Similarly, patients in Cluster B were characterized by altered hepatic expression of *RBM45* and *TRA2A* and presented higher triglycerides and GGT levels at the diagnosis. This represents a very interesting finding inasmuch as hepatic *TRA2A* has been found to be modulated in response to estrogens in order to control the alternative splicing of class B scavenger receptors BI (SR-BI) and BII (SR-BII) (57), which are crucial player in the hepatic uptake of TGs (58). Finally, obese patients included in Cluster C presented elevated levels of insulin and aminotransferases and a differential evolution of biochemical parameters and of the recovery in the grade of steatosis after bariatric surgery. Interestingly, *SND1*, one of the splicing factors whose expression alteration characterized Cluster C, is a splicing factor associated to the physiological function and pathological transformation of the liver (51,59). Indeed, hepatic *SND1* has been shown to be associated to different aspects of lipid metabolism in the liver and with the development of hepatic steatosis (51). Hence, these results, together with the *in vitro* results discussed above, suggest that the dysregulation of certain spliceosome components and/or splicing factors could also be involved in promoting the development of NAFLD and predisposing the patients to specific NAFLD-associated comorbidities.

Altogether, our present results demonstrate a novel relationship between the dysregulation of splicing machinery and the development of NAFLD and its associated metabolic co-morbidities, as well as with the biochemical improvement after bariatric surgery. These findings shed light to the possible underlying molecular mechanisms responsible for the development of hepatic steatosis in obese patients and, thus, provide novel information to explore the development of efficient screening strategies, diagnostic, prognostic or therapeutic tools for obesity-related NAFLD, which is becoming one of the main sources of morbidity and mortality in developed countries.

Financial support: This work was funded by Instituto de Salud Carlos III, co-funded by European Union (ERDF/ESF, “Investing in your future”) [PI17/002287, PI16/00264, PIE14/00005, CP15/00156], MINECO (BFU2016-80360-R, TIN2017-83445-P), Junta de Andalucía (BIO-0139, CTS-1406, PI-0541-2013) CIBERehd and CIBERobn. CIBER is an initiative of Instituto de Salud Carlos III, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain. Development Merit Award BX001114; and National Institutes of Health grant R01DK088133 (to R.D.K.).

Instituto de Salud Carlos III (ES), PI17/002287, Manuel D. Gahete; Instituto de Salud Carlos III http://dx.doi.org/10.13039/501100004587, PI16/00264, Raul M Luque; Instituto de Salud Carlos III http://dx.doi.org/10.13039/501100004587, PIE14/00005, José López-Miranda; Instituto de Salud Carlos III http://dx.doi.org/10.13039/501100004587, CP15/00156, Manuel D. Gahete; Ministerio de Economía, Industria y Competitividad, Gobierno de España (ES), BFU2016-80360-R, Raul M Luque; Ministerio de Economía, Industria y Competitividad, Gobierno de España (ES), TIN2017-83445-P, Sebastián Ventura;

Junta de Andalucía, BIO-0139, Raul M Luque; Junta de Andalucía , CTS-1406, Raul M Luque; Junta de Andalucía , PI-0541-2013, Manuel D. Gahete; Development Merit Award, BX001114, Rhonda D. Kineman; National Institutes of Health , R01DK088133, Rhonda D. Kineman

Corresponding Authors: Raul M. Luque and Manuel D. Gahete. Edificio IMIBIC, Hospital Universitario Reina Sofia. Av. Menendez Pidal s/n. 14004-Córdoba, Spain. E-mail address: raul.luque@uco.es or bc2gaorm@uco.es. Reprint requests should be addressed to Raul M. Luque.

Disclosure summary:
the authors have nothing to declare.

REFERENCES

1. Younossi Z, Tacke F, Arrese M, Sharma BC, Mostafa I, Bugianesi E, Wong VW, Yilmaz Y, George J, Fan J, Vos MB. Global Perspectives on Non-alcoholic Fatty Liver Disease and Non-alcoholic Steatohepatitis. *Hepatology*. 2018.
2. Masuoka HC, Chalasani N. Nonalcoholic fatty liver disease: an emerging threat to obese and diabetic individuals. *Annals of the New York Academy of Sciences*. 2013;1281:106-122.
3. Asrih M, Jornayvaz FR. Metabolic syndrome and nonalcoholic fatty liver disease: Is insulin resistance the link? *Molecular and cellular endocrinology*. 2015;418 Pt 1:55-65.
4. Baffy G, Brunt EM, Caldwell SH. Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace. *J Hepatol*. 2012;56(6):1384-1391.
5. Masuzaki R, Karp SJ, Omata M. NAFLD as a risk factor for HCC: new rules of engagement? *Hepatol Int*. 2016;10(4):533-534.
6. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, Murphy SK, Ashley-Koch AE, Choi SS, Michelotti GA, Hampton DD, Chen Y, Tillmann HL, Hauser MA, Abdelmalek MF, Diehl AM. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology*. 2014;59(2):471-482.
7. Fabbrini E, Magkos F. Hepatic Steatosis as a Marker of Metabolic Dysfunction. *Nutrients*. 2015;7(6):4995-5019.
8. Koutsari C, Lazaridis KN. Emerging genes associated with the progression of nonalcoholic fatty liver disease. *Hepatology*. 2010;52(3):807-810.
9. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology*. 2010;51(2):679-689.
10. Ye H, Liu W. Transcriptional networks implicated in human nonalcoholic fatty liver disease. *Molecular genetics and genomics : MGG*. 2015;290(5):1793-1804.
11. Yu J, Marsh S, Hu J, Feng W, Wu C. The Pathogenesis of Nonalcoholic Fatty Liver Disease: Interplay between Diet, Gut Microbiota, and Genetic Background. *Gastroenterology research and practice*. 2016:2862173.
12. Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol*. 2017;23(1):1-12.
13. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19-32.
14. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine*. 2012;18(8):472-482.
15. Havens MA, Duelli DM, Hastings ML. Targeting RNA splicing for disease therapy. *Wiley interdisciplinary reviews RNA*. 2013;4(3):247-266.

16. Matera G, Wang Z. A day in the life of the spliceosome. *Nature reviews molecular cell biology*. 2014;15:108-121.
17. Kaminska D, Hamalainen M, Cederberg H, Kakela P, Venesmaa S, Miettinen P, Ilves I, Herzig KH, Kolehmainen M, Karhunen L, Kuusisto J, Gylling H, Laakso M, Pihlajamaki J. Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. *Diabetologia*. 2014;57(2):347-351.
18. Peterfy M, Phan J, Reue K. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *J Biol Chem*. 2005;280(38):32883-32889.
19. Kulseth MA, Berge KE, Bogsrud MP, Leren TP. Analysis of LDLR mRNA in patients with familial hypercholesterolemia revealed a novel mutation in intron 14, which activates a cryptic splice site. *J Hum Genet*. 2010;55(10):676-680.
20. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320-323.
21. Kojima S, Kuo TF, Tatsukawa H. Regulation of transglutaminase-mediated hepatic cell death in alcoholic steatohepatitis and non-alcoholic steatohepatitis. *Journal of gastroenterology and hepatology*. 2012;27 Suppl 2:52-57.
22. Valerio Pazienza MB, Tommaso Mazza, Fareeba Sheedfar4, Concetta Panebianco, Roger Williams, Gianluigi Mazzoccoli, Angelo Andriulli, Tomoko Nakanishi, and Manlio Vinciguerra. SIRT1-metabolite binding histone macroH2A1.1 protects hepatocytes against lipid accumulation. *Aging*. 2014;6(1).
23. Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG. RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules*. 2015;5(2):893-909.
24. Dorota Kaminska MHmli, Henna Cederberg, Pirjo Käkelä, Sari Venesmaa, Pekka Miettinen, Imre Ilves, Karl-Heinz Herzig, Marjukka Kolehmainen, Leila Karhunen, Johanna Kuusisto, Helena Gylling, Markku Laakso, Jussi Pihlajamäki. Adipose tissue INSR splicing in humans associates with fasting insulin level and is regulated by weight loss. *Diabetologia*. 2014;57:347-351.
25. Kaminska D, Pihlajamaki J. Regulation of alternative splicing in obesity and weight loss. *Adipocyte*. 2013;2(3):143-147.
26. Pihlajamaki J, Lerin C, Itkonen P, Boes T, Floss T, Schroeder J, Dearie F, Crunkhorn S, Burak F, Jimenez-Chillaron JC, Kuulasmaa T, Miettinen P, Park PJ, Nasser I, Zhao Z, Zhang Z, Xu Y, Wurst W, Ren H, Morris AJ, Stamm S, Goldfine AB, Laakso M, Patti ME. Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. *Cell Metab*. 2011;14(2):208-218.
27. Brosch M, von Schonfels W, Ahrens M, Nothnagel M, Krawczak M, Laudes M, Sipos B, Becker T, Schreiber S, Rocken C, Schafmayer C, Hampe J. SFRS10--a splicing factor gene reduced in human obesity? *Cell Metab*. 2012;15(3):265-266; author reply 267-269.
28. Lopez-Vicario C, Gonzalez-Periz A, Rius B, Moran-Salvador E, Garcia-Alonso V, Lozano JJ, Bataller R, Cofan M, Kang JX, Arroyo V, Claria J, Titos E. Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. *Gut*. 2014;63(2):344-355.
29. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, Grimaldi A, Capron F, Poynard T. Sampling Variability of Liver Biopsy in Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2005;128(7):1898-1906.

30. Vuppalanchi R, Unalp A, Van Natta ML, Cummings OW, Sandrasegaran KE, Hameed T, Tonascia J, Chalasani N. Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol*. 2009;7(4):481-486.
31. Tsai E, Lee TP. Diagnosis and Evaluation of Nonalcoholic Fatty Liver Disease/Nonalcoholic Steatohepatitis, Including Noninvasive Biomarkers and Transient Elastography. *Clin Liver Dis*. 2018;22(1):73-92.
32. Castera L, Friedrich-Rust M, Loomba R. Non-Invasive Assessment of Liver Disease in Patients with NAFLD. *Gastroenterology*. 2019.
33. Lupsor-Platon M, Stefanescu H, Muresan D, Florea M, Szasz ME, Maniu A, Badea R. Noninvasive assessment of liver steatosis using ultrasound methods. *Med Ultrason*. 2014;16(3):236-245.
34. Li Q, Dhyani M, Grajo JR, Sirlin C, Samir AE. Current status of imaging in nonalcoholic fatty liver disease. *World J Hepatol*. 2018;10(8):530-542.
35. Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Lonardo A. NAFLD as a Sexual Dimorphic Disease: Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk. *Adv Ther*. 2017;34(6):1291-1326.
36. Kwekel JC, Desai VG, Moland CL, Branham WS, Fuscoe JC. Age and sex dependent changes in liver gene expression during the life cycle of the rat. *BMC Genomics*. 2010;11:675.
37. Hormaechea-Agulla D, Gahete MD, Jimenez-Vacas JM, Gomez-Gomez E, Ibanez-Costa A, F LL, Rivero-Cortes E, Sarmento-Cabral A, Valero-Rosa J, Carrasco-Valiente J, Sanchez-Sanchez R, Ortega-Salas R, Moreno MM, Tsomaia N, Swanson SM, Culler MD, Requena MJ, Castano JP, Luque RM. The oncogenic role of the In1-ghrelin splicing variant in prostate cancer aggressiveness. *Mol Cancer*. 2017;16(1):146.
38. Hormaechea-Agulla D, Jimenez-Vacas JM, Gomez-Gomez E, F LL, Carrasco-Valiente J, Valero-Rosa J, Moreno MM, Sanchez-Sanchez R, Ortega-Salas R, Gracia-Navarro F, Culler MD, Ibanez-Costa A, Gahete MD, Requena MJ, Castano JP, Luque RM. The oncogenic role of the spliced somatostatin receptor sst5TMD4 variant in prostate cancer. *FASEB J*. 2017;31(11):4682-4696.
39. Rincon-Fernandez D, Culler MD, Tsomaia N, Moreno-Bueno G, Luque RM, Gahete MD, Castano JP. In1-ghrelin splicing variant is associated with reduced disease-free survival of breast cancer patients and increases malignancy of breast cancer cells lines. *Carcinogenesis*. 2018;39(3):447-457.
40. Gahete MD, delRio-Moreno M, Camargo A, Alcala-Diaz JF, Alors-Perez E, Delgado-Lista J, Reyes O, Ventura S, Perez-Martínez P, Castaño JP, Lopez-Miranda J, Luque RM. Changes in Splicing Machinery Components Influence, Precede, and Early Predict the Development of Type 2 Diabetes: From the CORDIOPREV Study. *EBioMedicine*. 2018;In press.
41. delRio-Moreno M, Alors-Perez E, Gonzalez-Rubio S, Ferrin G, Reyes O, Rodriguez-Peralvarez M, Sanchez-Frias ME, Sanchez-Sanchez R, Ventura S, Lopez-Miranda J, Kineman RD, Mata Mdl, Castaño JP, Gahete MD, Luque RM. Data from: Dysregulation of the splicing machinery is associated to the development of non-alcoholic fatty liver disease (Supplementary Information). Figshare. Deposited 7 March 2019. 10.6084/m9.figshare.7811867.v2.
42. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*. 2002;3(7):RESEARCH0034.

43. Sarmento-Cabral A, L-Lopez F, Gahete MD, Castano JP, Luque RM. Metformin Reduces Prostate Tumor Growth, in a Diet-Dependent Manner, by Modulating Multiple Signaling Pathways. *Mol Cancer Res*. 2017;15(7):862-874.
44. L-Lopez F, Sarmento-Cabral A, Herrero-Aguayo V, Gahete MD, Castano JP, Luque RM. Obesity and metabolic dysfunction severely influence prostate cell function: role of insulin and IGF1. *J Cell Mol Med*. 2017;21(9):1893-1904.
45. Sarmento-Cabral A, Peinado JR, Halliday LC, Malagon MM, Castano JP, Kineman RD, Luque RM. Adipokines (Leptin, Adiponectin, Resistin) Differentially Regulate All Hormonal Cell Types in Primary Anterior Pituitary Cell Cultures from Two Primate Species. *Sci Rep*. 2017;7:43537.
46. Xia J, Wishart DS. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr Protoc Bioinformatics*. 2016;55:14 10 11-14 10 91.
47. Murphy SK, Yang H, Moylan CA, Pang H, Dellinger A, Abdelmalek MF, Garrett ME, Ashley-Koch A, Suzuki A, Tillmann HL, Hauser MA, Diehl AM. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2013;145(5):1076-1087.
48. Webster NJG. Alternative RNA Splicing in the Pathogenesis of Liver Disease. *Front Endocrinol (Lausanne)*. 2017;8:133.
49. Reardon HT, Park WJ, Zhang J, Lawrence P, Kothapalli KS, Brenna JT. The polypyrimidine tract binding protein regulates desaturase alternative splicing and PUFA composition. *J Lipid Res*. 2011;52(12):2279-2286.
50. Medina MW, Krauss RM. Alternative splicing in the regulation of cholesterol homeostasis. *Curr Opin Lipidol*. 2013;24(2):147-152.
51. Garcia-Arcos I, Rueda Y, Gonzalez-Kother P, Palacios L, Ochoa B, Fresnedo O. Association of SND1 protein to low density lipid droplets in liver steatosis. *J Physiol Biochem*. 2010;66(1):73-83.
52. Ricchi M, Odoardi MR, Carulli L, Anzivino C, Ballestri S, Pinetti A, Fantoni LI, Marra F, Bertolotti M, Banni S, Lonardo A, Carulli N, Loria P. Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in cultured hepatocytes. *Journal of gastroenterology and hepatology*. 2009;24(5):830-840.
53. Lee G, Zheng Y, Cho S, Jang C, England C, Dempsey JM, Yu Y, Liu X, He L, Cavaliere PM, Chavez A, Zhang E, Isik M, Couvillon A, Dephoure NE, Blackwell TK, Yu JJ, Rabinowitz JD, Cantley LC, Blenis J. Post-transcriptional Regulation of De Novo Lipogenesis by mTORC1-S6K1-SRPK2 Signaling. *Cell*. 2017;171(7):1545-1558 e1518.
54. Ravi S, Schilder RJ, Kimball SR. Role of precursor mRNA splicing in nutrient-induced alterations in gene expression and metabolism. *J Nutr*. 2015;145(5):841-846.
55. Malakar P, Chartarifsky L, Hija A, Leibowitz G, Glaser B, Dor Y, Karni R. Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. *Sci Rep*. 2016;6:31222.
56. Reardon HT, Hsieh AT, Park WJ, Kothapalli KS, Anthony JC, Nathanielsz PW, Brenna JT. Dietary long-chain polyunsaturated fatty acids upregulate expression of FADS3 transcripts. *Prostaglandins Leukot Essent Fatty Acids*. 2013;88(1):15-19.
57. Zhang X, Moor AN, Merkler KA, Liu Q, McLean MP. Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors. *Endocrinology*. 2007;148(11):5295-5304.
58. Thuahnai ST, Lund-Katz S, Williams DL, Phillips MC. Scavenger receptor class B, type I-mediated uptake of various lipids into cells. Influence of the nature of the donor particle interaction with the receptor. *J Biol Chem*. 2001;276(47):43801-43808.
59. Jariwala N, Rajasekaran D, Mendoza RG, Shen XN, Siddiq A, Akiel MA, Robertson CL, Subler MA, Windle JJ, Fisher PB, Sanyal AJ, Sarkar D. Oncogenic Role of SND1 in

Development and Progression of Hepatocellular Carcinoma. *Cancer Res*. 2017;77(12):3306-3316.

Figure 1. Differential expression of splicing machinery components in the liver of steatotic and non-steatotic obese patients. A) Study design and pattern of dysregulation of spliceosome components and splicing factors in the liver of steatotic and non-steatotic obese patients. A graphical summary of the study design is shown in the left panel. A schematic representation of fold change levels of spliceosome components and splicing factors between the liver of steatotic and non-steatotic obese patients, represented in red (increase) or blue (decrease), is depicted in the right panel. **B) Expression levels and ROC curves of significantly altered spliceosome components and splicing factors in the liver of steatotic compared with non-steatotic obese patients.** mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *HPRT* and *ACTB*] of the different spliceosome components (first row) and splicing factors (second row) significantly altered in the liver of obese women with (ST) and without steatosis (NON ST). Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-steatosis patients (t-test: *p<0.05, **p<0.01, ***p<0.001). Receiver operating characteristic (ROC) curve analysis, to determine the accuracy of the components of the splicing machinery and splicing factors to discriminate between patients with or without liver steatosis, are included below each graph. **B) ROC curves of subsets of spliceosome components and splicing factors generated by Random Forest computational algorithm followed by cross validation analysis, considering the expression of a selection of the most relevant and discriminatory splicing machinery components.** Specifically, three subset of specific splicing machinery components are included: *PTBP1, RBM45, SRRM1, RNU4* and *RNU6ATAC; PTBP1, RBM22, SRSF1, SRRM1, SNRNP70 and RNU6ATAC; CELF1, PTBP1, RBM22, RBM3, SRRM1* and *RNU6*. ***AUC***: Area under curve. ***p***: p value.

Figure 2. Pattern of dysregulation of spliceosome components and splicing factors in the liver of non-steatotic and steatotic obese patients according to the grade of lipid accumulation. Above, fold change expression levels among patients with different levels of steatosis compared to non-steatotic livers, represented in red (increase) or blue (decrease). Below, mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *HPRT* and *ACTB*] of the different spliceosome components and splicing factors in the liver of obese women without steatosis (NON ST) and with three different levels of steatosis (mild, moderate and severe). Values represent the mean ± SEM. Asterisks indicate values that significantly differ from non-steatosis patients (t-test: *p<0.05, **p<0.01, ***p<0.001).

Figure 3. The expression of splicing machinery components is differentially altered in the liver of steatotic individuals. A) Unsupervised clustering analysis of the expression levels of the splicing machinery in patients with steatosis. This bioinformatic approach identified three molecularly-defined populations of steatotic patients (CLUSTER A, B and C). B). Specific changes of certain components of the splicing machinery defined each cluster of steatotic patients. The three molecularly defined clusters of steatotic patients were associated with the alteration in the expression of certain spliceosome components and splicing factors, compared to patients without steatosis or included in the other clusters. The alteration of selected factors (within the frame) was able to classify patients in the three Clusters with an AUC of 1, using the classification algorithm Random Forest. Data indicate mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *HPRT* and *ACTB*] in each cluster (A, B and C) compared to the rest of

patients with and without steatosis (NON ST). Values represent the mean ± SEM. Asterisks indicate values that significantly differ between groups (t-test *p<0.05, **p<0.01, ***p<0.001).

Figure 4. Each molecularly defined population of steatotic patients (CLUSTER A, B and C) was characterized by certain hepatic and clinical-metabolic alterations. Relevant clinical parameters associated to each cluster of patients grouped according to their expression levels of the spliceosome components and splicing factors. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from control cells (t-test: *p<0.05, **p<0.01). ***GGT:*** Gamma-Glutamyltransferase; ***ALT:*** Alanine Transaminase, ***AST:*** Aspartate Transaminase.

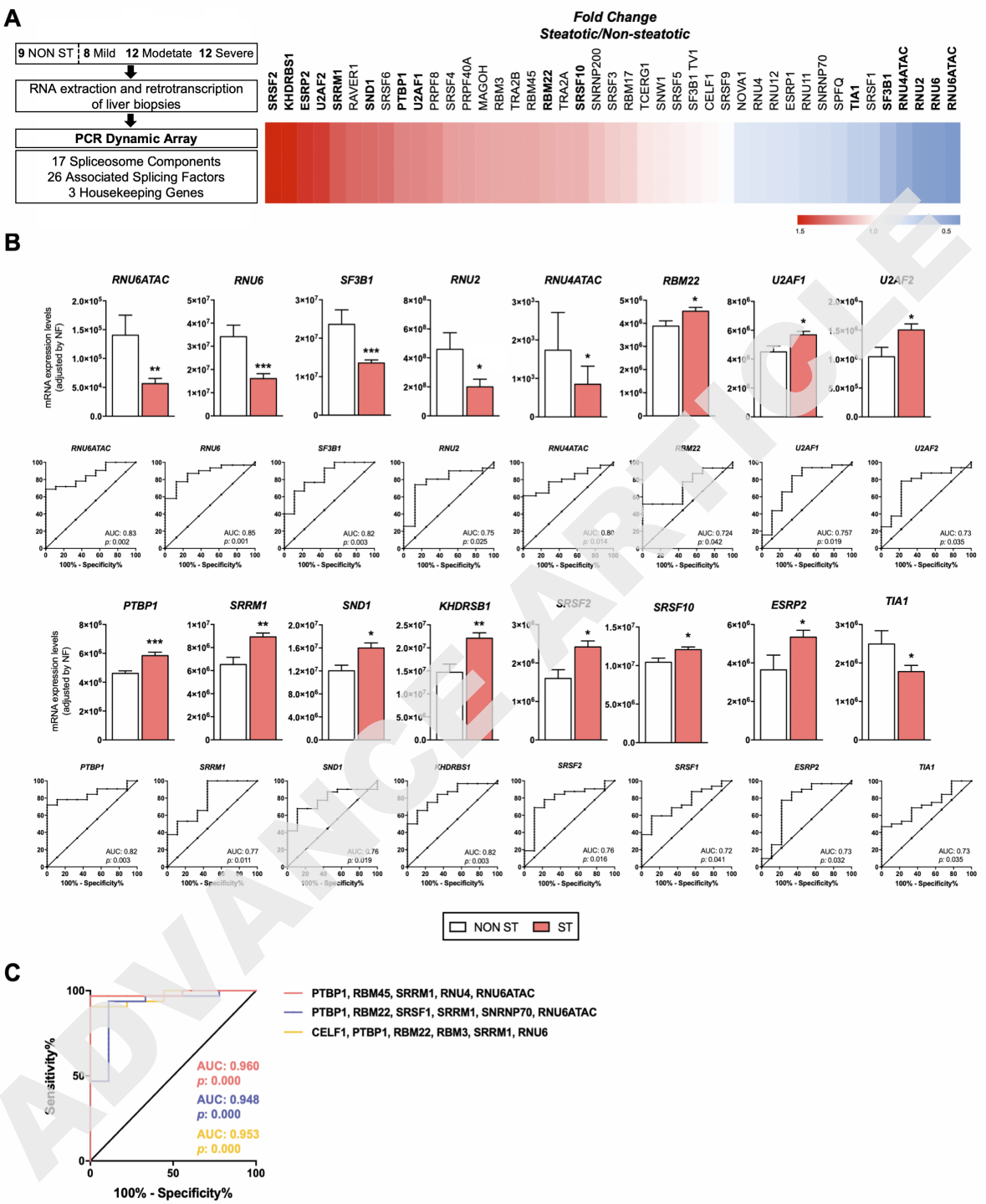
Figure 5. Cluster C presented a worst response to bariatric surgery after one year of follow-up. Recovery from hepatic steatosis and evolution of BMI, WHR, GGT, Glucose, Triglycerides, Alkaline phosphatase and HDL levels one year after bariatric surgery in patients from Cluster C (CL C) vs. Clusters A+B (CL A+B). The data are expressed as percentage of the value before surgery (normalized to 100%). Values represent the mean ± SEM. Asterisks indicate values that significantly differ from control cells (t-test: *p<0.05, **p<0.01). ***BMI:*** Body Mass Index, ***WHR:*** Waist-Hip Ratio, ***GGT:*** Gamma-Glutamyltransferase; ***ALT:*** Alanine Transaminase, ***AST:*** Aspartate Transaminase, ***HDL:*** High Density Lipoprotein.

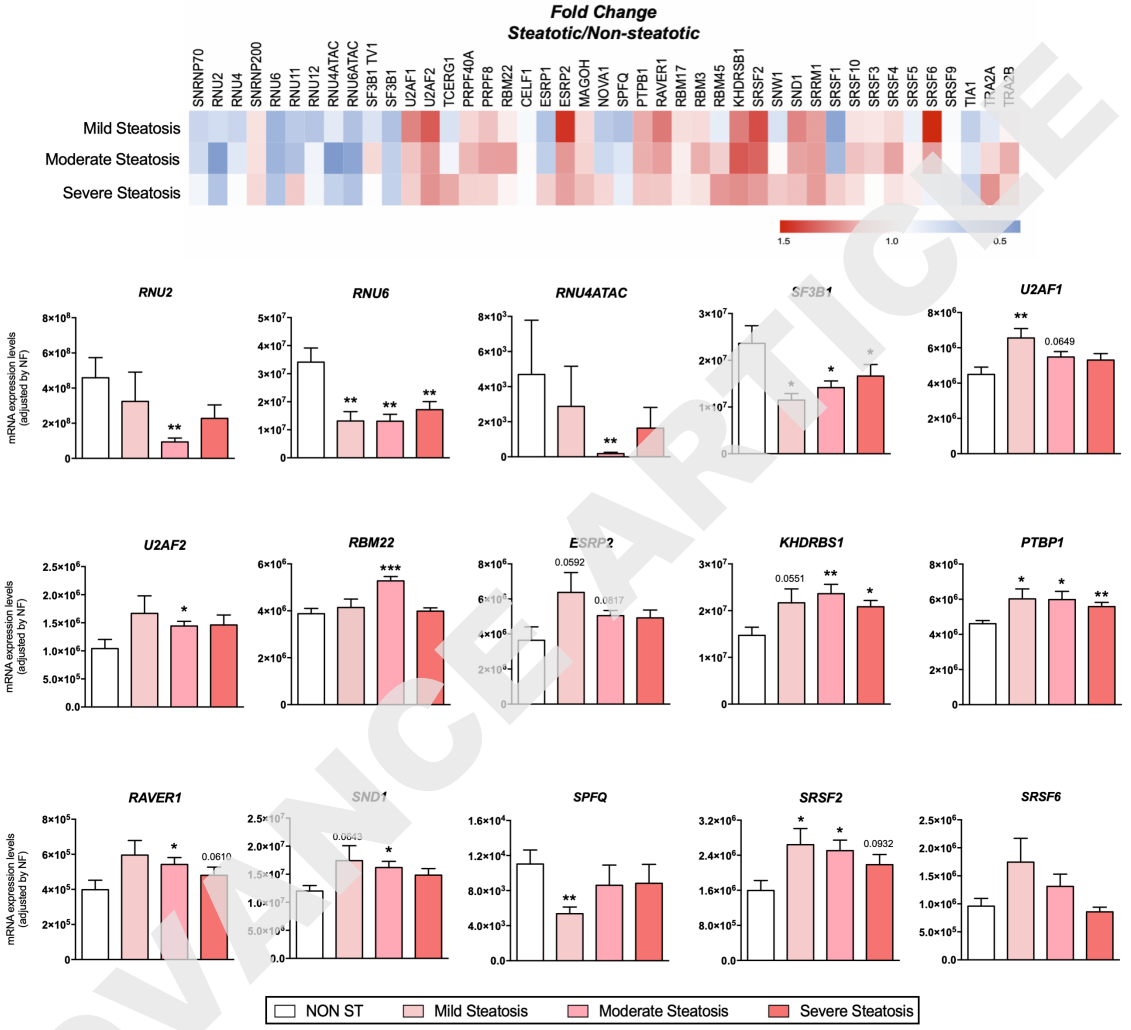
Figure 6. Modulation of splicing machinery components influenced lipid accumulation in HepG2 cells. A) Validation of lipid accumulation in HepG2 cell lines by Oil Red O absorbance at 520 nm. Data are expressed as a percentage of the control (normalized to 100%). **B) Effect of lipid accumulation on the expression of certain spliceosome components and splicing factors at 10, 24 and 48h.** mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *HPRT* and *GAPDH*] of the different spliceosome components and splicing factors in HepG2 cells treated with 500µM of OA. The data are expressed as percentage of the control (normalized to 100%). Values represent the mean ± SEM (n=5). Asterisks indicate values that significantly differ from control cells (t-test: *p<0.05, **p<0.01, ***p<0.001). **C) qPCR validation of the silencing with specific siRNAs.** mRNA expression levels adjusted by the expression level of *ACTB*. Data are expressed as percentage of control random siRNAs (Scramble; set at 100%). **D) Effect of silencing of certain splicing factors in HepG2 cells on lipid accumulation determined by Oil Red O absorbance at 520 nm.** Data are expressed as percentage of the control (normalized to 100%). Values represent the mean ± SEM (n=5). ***OA***: oleic acid.

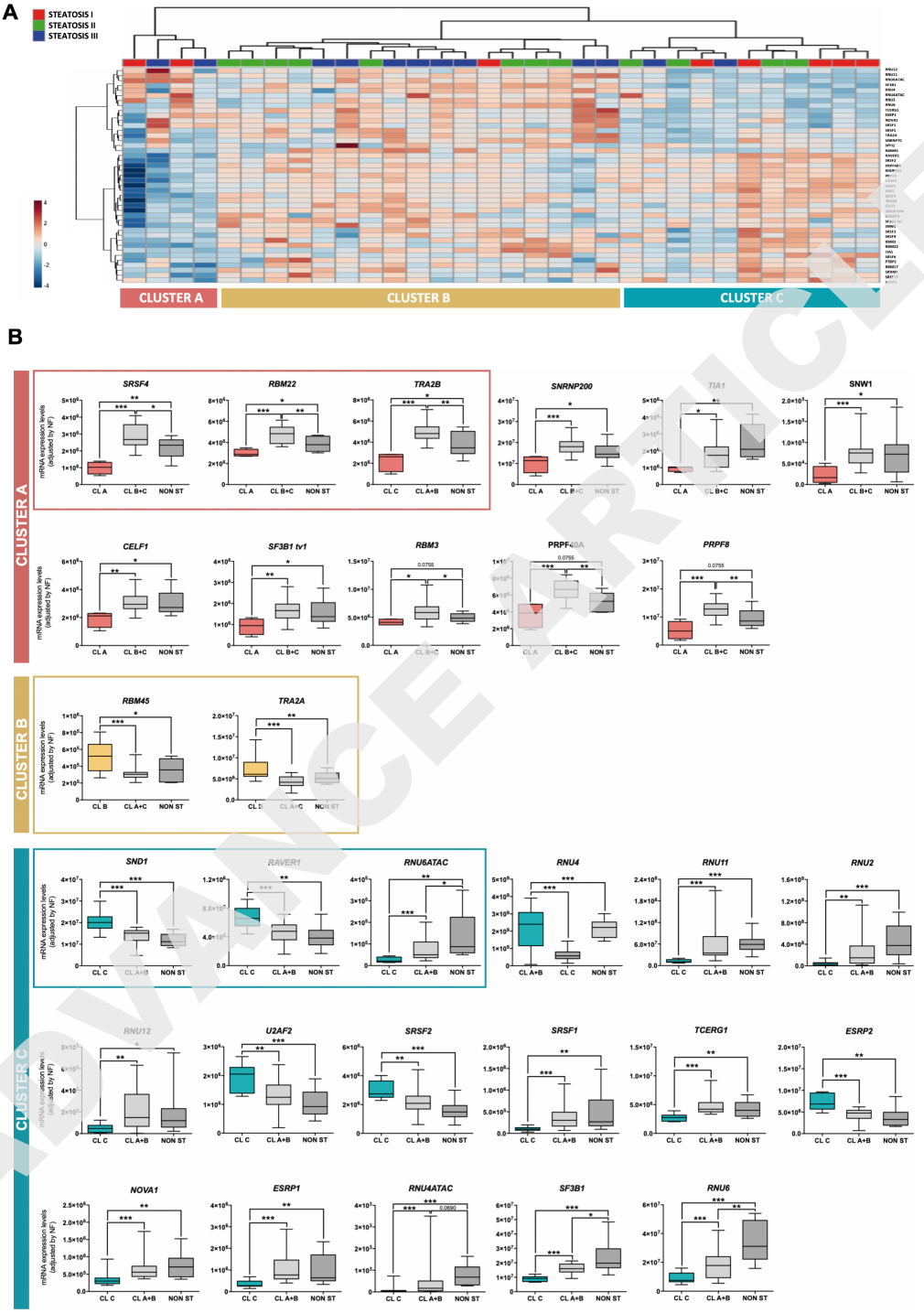
Table 1. Demographic and clinical characteristics of the patients included in this study. ***BMI:*** Body Mass Index, ***WHR:*** Waist-Hip Ratio, ***GGT:*** Gamma-Glutamyltransferase; ***ALT:*** Alanine Transaminase, ***AST:*** Aspartate Transaminase, ***HDL:*** High Density Lipoprotein, ***LDL:*** Low Density Lipoprotein, ***CRP:*** C-Reactive Protein.

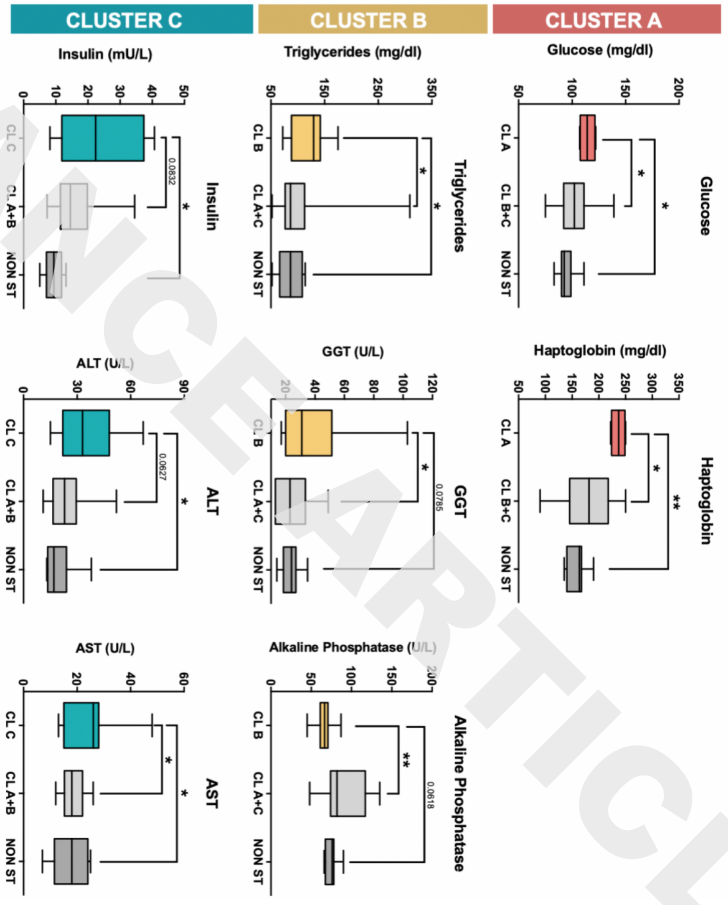
	Non Steatosis	Steatosis	*p*-value
Age (years)	37.00 ± 5.00	40.41 ± 1.77	0.425
	n=9	n=32	
Body Weight (kg)	125.60 ± 3.07	133.1 ± 3.58	0.292
	n=9	n=32	
BMI (kg/m2)	48.96 ± 0.92	50.63 ± 1.31	0.513
	n=9	n=32	
WHR	0.86 ± 0.02	0.91 ± 0.01	0.059

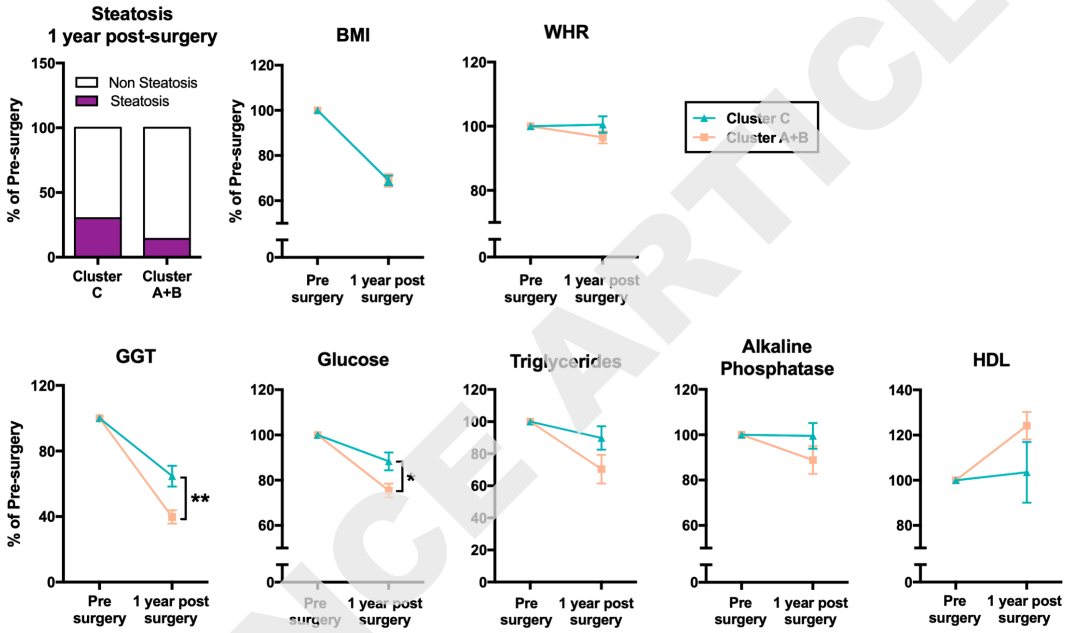
	n=9	n=32	
HOMA-IR	2.00 ± 0.48	5.00 ± 0.71	0.049
	n=5	n=22	
Glucose (mg/dl)	101.70 ± 7.62	106.60 ± 3.33	0.125
	n=9	n=32	
Insulin (mU/L)	9.44 ± 1.32	19.59 ± 2.29	0.049
	n=5	n=22	
Glycated Haemoglobin (%)	6.01 ± 0.44	6.181 ± 0.13	0.175
	n=8	n=32	
GGT (U/L)	23.33 ± 2.11	38.94 ± 5.05	0.181
	n=9	n=33	
ALT (U/L)	19.44 ± 2.76	28.19 ± 2.35	0.041
	n=9	n=32	
AST (U/L)	17.11 ± 2.17	21.84 ± 1.76	0.202
	n=9	n=32	
Alkaline Phosphatase (U/L)	74.67 ± 2.54	78.75 ± 4.03	0.602
	n=9	n=32	
Bilirubin (mg/dl)	0.60 ± 0.07	0.66 ± 0.05	0.835
	n=8	n=27	
Direct Bilirubin (mg/dl)	0.30 ± 0.04	0.27 ± 0.016	0.644
	n=6	n=28	
Indirect Bilirubin (mg/dl)	0.38 ± 0.05	0.37 ± 0.03	0.735
	n=6	n=28	
HDL (mg/dl)	43.78 ± 4.89	41.50 ± 1.82	0.596
	n=9	n=32	
LDL (mg/dl)	126.60 ± 11.38	136.10 ± 6.34	0.245
	n=9	n=31	
Triglycerides (mg/dl)	103.30 ± 18.72	136.60 ± 19.81	0.215
	n=9	n=32	
Cholesterol (mg/dl)	191.60 ± 11.97	204.20 ± 6.62	0.374
	n=9	n=32	
Creatinin (mg/dl)	0.82 ± 0.16	0.69 ± 0.02	0.470
	n=8	n=28	
CRP (mg/L)	8.45 ± 1.28	12.62 ± 1.89	0.476
	n=9	n=27	
Total Protein (g/dl)	6.95 ± 0.27	7.00 ± 0.08	0.817
	n=9	n=32	
Albumin (g/dl)	3.98 ± 0.16	4.08 ± 0.06	0.451
	n=9	n=32	
Haptoglobin (mg/dl)	158.50 ± 6.50	190.20 ± 8.24	0.065
	n=8	n=31	

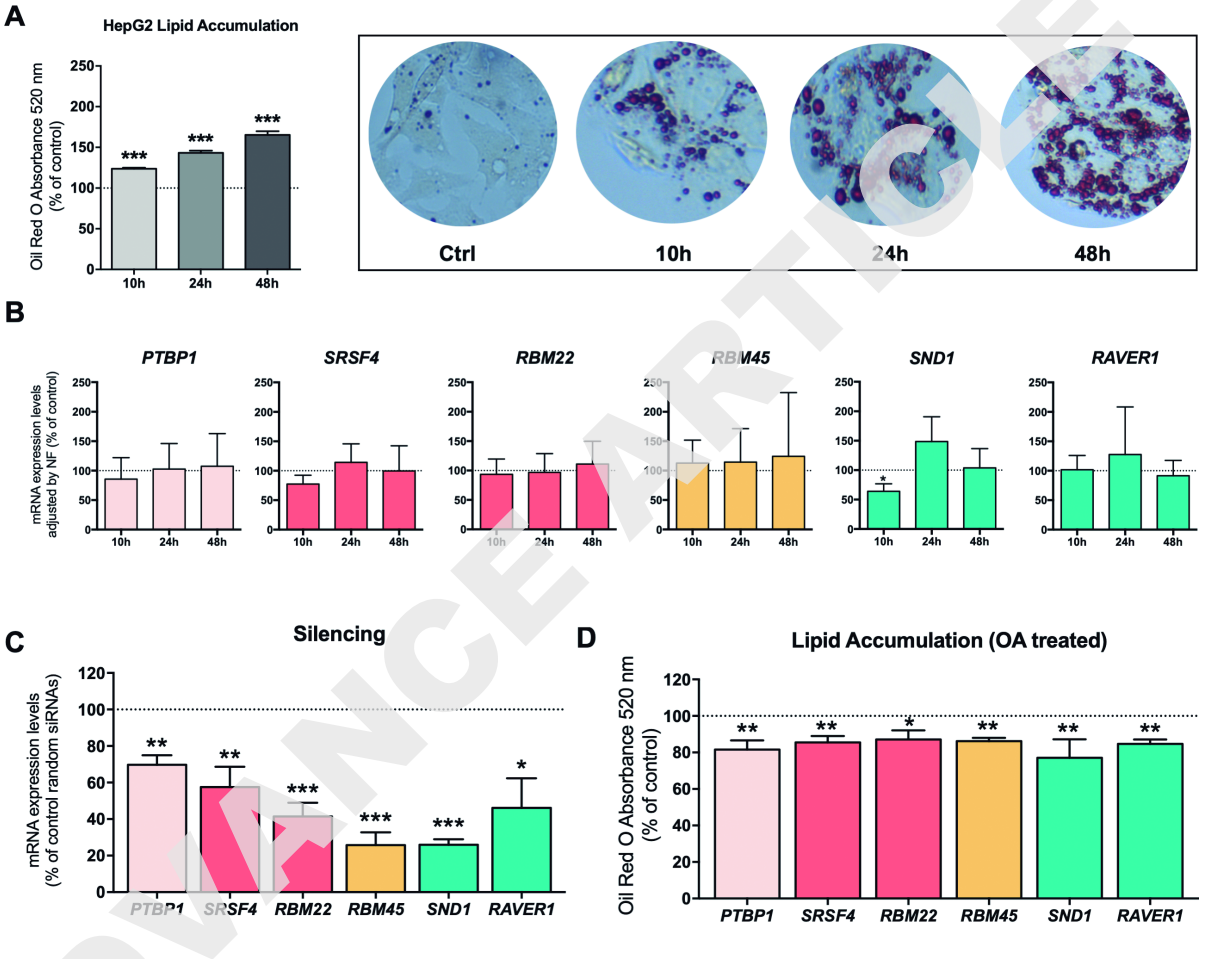












Supplementary Information

Supplementary Material and Methods

Additional notes on the analysis of splicing machinery components by qPCR dynamic array based on microfluidic technology

A 48.48 Dynamic Array (Fluidigm, San Francisco, CA, USA) was implemented to determine the expression of 48 transcripts in 48 samples, simultaneously. Following manufacturer's instructions, 1.25ng of cDNA of each sample was pre-amplified using 1µL of PreAmp Master Mix (Fluidigm) and 0.5 µL of all primers mix (500nM) in a T100 Thermal-cycler (Bio-Rad, Hercules, CA, USA), using the following program: 1) 2min at 95ºC; 2) 15sec at 94ºC and 4min at 60ºC (14 cycles). After pre-amplification, each sample was treated with 2 µL of a 4U/µL Exonuclease I solution (New England BioLabs, Ipswich, MA, USA) for 30min at 37ºC followed by 15min at 80ºC to remove unincorporated primers. Then, samples were diluted with 18µL of Buffer TE (Thermo Scientific, Madrid, Spain), and 2.7µL were mixed with 3µL of EvaGreen Supermix (Bio-Rad) and 0.3µL of DNA Binding Dye Sample Loading Reagent (Fluidigm). Similarly, primers were diluted to 5µM with 2X Assay Loading Reagent (Fludigm). Then, control line fluid was charged in the chip and Prime script program was run into the IFC controller MX (Fluidigm). Finally, 5µL of each primer and 5µL of each sample were pipetted into their respective inlets on the chip and the Load Mix script in the IFC controller software was run. After this program, the qPCR was run in the Biomark System (Fluidigm) following the thermal cycling program: (1) 95ºC for 1min; (2) 35 cycles of denaturing (95ºC for 5sec) and annealing/extension (60ºC for 20sec); and (3) a last cycle where final PCR products were subjected to graded temperature-dependent dissociation (60ºC to 95ºC, increasing 1ºC/3 sec). Data were processed with Real-Time PCR Analysis Software 3.0 (Fluidigm).

Bioinformatical analysis

1. Data preprocessing

Before conducting the computational study, the dataset was preprocessed as follows: 1) An univariate analysis of outliers was conducted and those values that lie outside 1.5 times the inter-quartile range were removed; 2) All genes with more than 70% of missing values were eliminated; 3) All missing values and outliers were replaced by the median of the variable in each group of patients; 4) All variables with zero variance were removed; 5) All variables were centered by means of subtracting the original values by the mean, and then, they were scaled by means of dividing by the standard deviation. By this way, all variables had the same impact and, therefore, the posterior estimation of the variables' relevance was not biased by those variables with extreme values.

2. Estimation of the variables' relevance

It is well-known that predictive performance can be enhanced if redundant, noisy and interacting variables are removed at the time of constructing models [1,2]. In this work, several well-known feature-weighting algorithms [3] were used to compute the relevance of variables. The variables were ordered from higher to lower relevance according to a weight that represents the ability to distinguish between steatotic and non-steatotic patients. To avoid possible biases in the process of estimating the variables' relevance, five well-known feature estimation methods were used, namely Correlation-Feature-Selection (CFS), Correlation-Attribute-Evaluation (CAE), Gain-Ratio (GR), Information-Gain (IG), and Relief-F (RF) [4,5]. These algorithms are filter methods that evaluate the usefulness of a variable (or a set of variables) through measures of distance, dependency, information or correlation on data [1,2], so they are not influenced by classification algorithms in the feature estimation process. The use of these supervised feature weighting methods can lead to a superior estimation of the variables' relevance, having as main advantages: (I) consideration of the expert knowledge unlike of several traditional approaches, such as Principal Component Analysis (PCA), that do not exploit the a priori classification of patients; (II) detection of redundant information; (III) detection of interacting features.

In this work, the five feature-weighting methods were assessed by leave one-out cross validation, and the process was implemented in the R language. In the case of RF method, it was executed with a set of number of nearest neighbors equal to [1]. The five estimation methods returned ranking of variables and, therefore, an average ranking can be computed as presented in Figure 1.

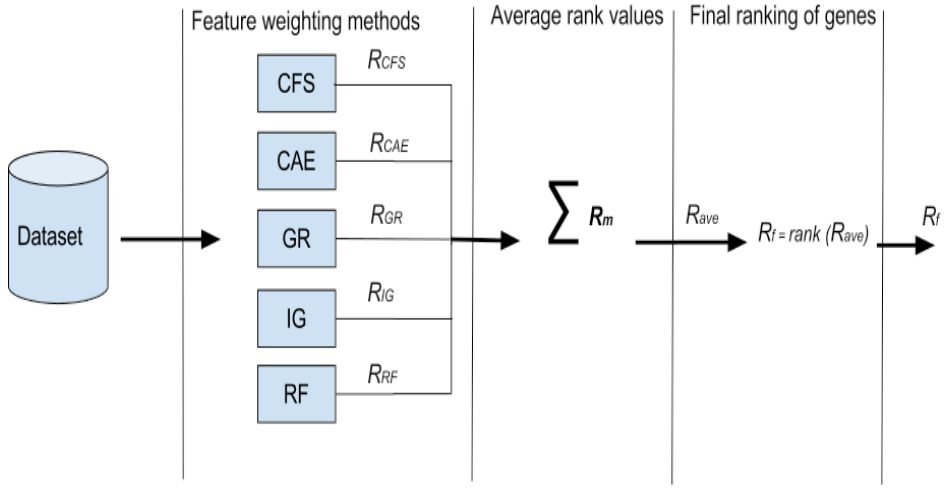


Figure 1. Workflow describing the five estimation methods used for ranking of variables and their integration to obtain the average ranking

3. Construction of predictive models with subsets of relevant variables

Subsequently, it was also determined the subset of variables that can best discriminate between steatotic and non-steatotic patients. To determine the best subset of variables by means of a final ranking of variables, we implemented a method previously published by our group [2], which provides a heuristic for searching on feature rankings. The number of possible subsets of variables is of exponential size but through a series of steps that are performed for each sub-ranking (Figure 2), the best subset of variables is selected as this one that produces the best classifier at discriminating between patients that have steatosis or not.

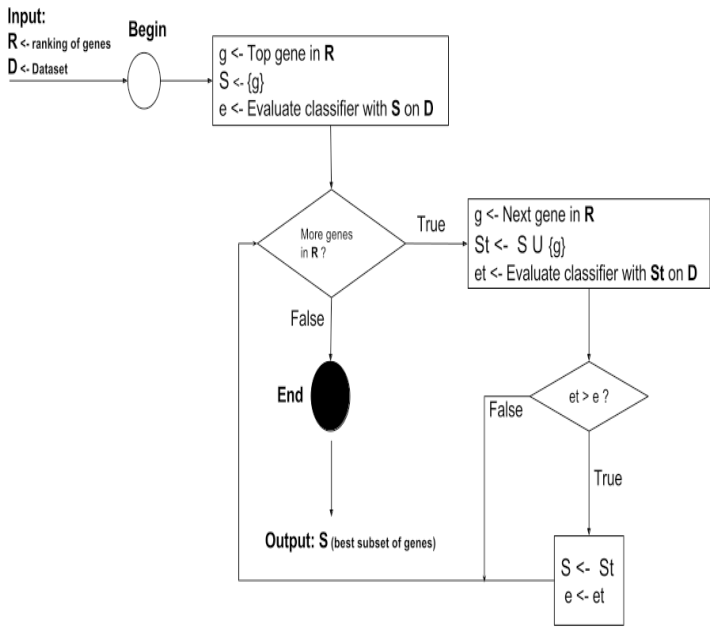


Figure 2. Workflow describing the series of steps that are performed for each sub-ranking

2. Cross-validation of predictive models

The capacity of generalization of the models was studied by means of assessing the models in a data partition that has not been seen previously (Cross-validation studies). This scenario was performed in R language by leave one-out cross-validation to construct the models, using the same seed to partition the database, so allowing the replication of the computational study.

References

1. Reyes O, Morell C, Ventura S. Evolutionary feature weighting to improve the performance of multi-label lazy algorithms. *Integr Comput-Aid E* 2014; **21**(4): 339-54.
2. Reyes O, Morell C, Ventura S. Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. *Neurocomputing* 2015; **161**: 168-82.
3. Wettschereck D, Aha DW, Mohri T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artif Intell Rev* 1997; **11**(1-5): 273-314.
4. Garcia S, Luengo J, Herrera F. Data Preprocessing in Data Mining. *Intel Syst Ref Libr* 2015; **72**: 1-320.
5. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* 2003; **53**(1-2): 23-69.

Supplementary Tables

Supplementary Table 1

	Gene	Accession Number	Primer Sequence (Sense. Se)	Primer Sequence (Antisense. As)	Product Size	Nucleotide Position
Spliceosome Components	*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
Splicing Factors	*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 genes	*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

Supplementary Table 1. Specific primers for human transcripts used in this study, including components of the major and minor spliceosomes, associated splicing factors and three housekeeping genes (HK) that were specifically designed and used in qPCR-based microfluidic assays. NCBI accesion number, primers sequences, expected product sizes and nucleotide positions for the genes studied are included.

Supplementary Table 2

Gene	Accession Number	Primer Sequence (Sense. Se)	Primer Sequence (Antisense. As)	Product Size	Nucleotide Position
SREBP1	NM_001005291.2	TGAGGACAGCAAGGCAAAG	GACAGGCAGAGGAAGACGAG	105	Se 1676; As 1780
ACLY	NM_001303274.1	AACTTTCTGATCGAGCCCTTC	ACATCACCCACGTCCACAC	125	Se 480; As 604
ACAC	NM_198834.3	TTTCTTCCATCTCCCCCTCT	CCATGCCAATCTCATTTCCT	118	Se 5509; As 5626
FASN	NM_004104.5	CTACGACTACGGCCCTCATTT	TCCATGAAGCTCACCCAGTT	99	Se 3117; As 3215
SCD	NM_005063.5	ACGTGGCTTTTTCTTCTCTCAC	GTACCTCCTCTGGAACATCACC	130	Se 794; As 923

Supplementary Table 2. Specific primers for human transcripts associated to *de novo* lipogenesis process. NCBI accession number, primers sequences, expected product sizes and nucleotide positions for the genes studied are included.

Supplementary Table 3

	Mild	Moderate	Severe	*ANOVA*
Age (years)	35.50 ± 3.82 n=8	39.67 ± 2.79 n=12	44.42 ± 2.54 n=12	0.140
Body Weight (kg)	132.40 ± 9.15 n=8	132.50 ± 5.28 n=12	134.20 ± 5.70 n=12	0.975
BMI (kg/m^2)	50.34 ± 2.80 n=8	50.68 ± 1.74 n=12	50.77 ± 2.54 n=12	0.992
WHR	0.86 ± 0.02 n=8	0.92 ± 0.01 n=12	0.92 ± 0.03 n=12	0.097
HOMA-IR	3.73 ± 0.70 * n=7	4.06 ± 0.82* n=8	7.75 ± 1.55 n=7	0.03
Glucose (mg/dl)	97.75 ± 4.71 n=8	105.70 ± 4.80 n=12	113.30 ± 6.46 n=12	0.192
Insulin (mU/L)	15.69 ± 2.96 n=7	16.24 ± 3.34 n=8	27.33 ± 4.37 n=7	0.062
Glycated Haemoglobin (%)	5.80 ± 0.22 n=8	6.17 ± 0.13 n=12	6.45 ± 0.26 n=12	0.135
GGT (U/L)	29.63 ± 10.62 n=8	37.25 ± 7.72 n=12	46.83± 8.54 n=12	0.418
ALT (U/L)	20.75 ± 2.09 n=8	29.08 ± 3.53 n=12	32.25 ± 4.70 n=12	0.159
AST (U/L)	17.13 ± 1.94 n=8	22.00 ± 2.70 n=12	24.83 ± 3.49 n=12	0.242
Alkaline Phosphatase (U/L)	73.00 ± 4.45 n=8	82.08 ± 8.69 n=12	79.25 ± 5.93 n=12	0.694
Bilirubin (mg/dl)	0.70 ± 0.10 n=6	0.65 ± 0.08 n=11	0.63 ± 0.08 n=10	0.875
Direct Bilirubin (mg/dl)	0.27 ± 0.03 n=7	0.27 ± 0.03 n=11	0.27 ± 0.02 n=10	0.997
Indirect Bilirubin (mg/dl)	0.34 ± 0.05 n=7	0.390 ± 0.06 n=11	0.38 ± 0.06 n=10	0.859
HDL (mg/dl)	40.38 ± 3.66 n=8	41.83 ± 3.05 n=12	41.92 ± 3.12 n=12	0.941
LDL (mg/dl)	134.90 ± 16.87 n=8	139.00 ± 11.02 n=12	133.70 ± 6.94 n=11	0.937
Triglycerides (mg/dl)	126.30 ± 37.12 n=8	121.10 ± 19.55 n=12	159.10 ± 43.69 n=12	0.375
Cholesterol (mg/dl)	202.30 ± 18.97 n=8	205.30 ± 11.29 n=12	204.30 ± 6.99 n=12	0.984
Creatinin (mg/dl)	0.69 ± 0.02 n=6	0.65 ± 0.02 n=11	0.72 ± 0.06 n=11	0.581
CRP (mg/L)	11.42 ± 3.77 n=6	10.85 ± 2.30 n=10	14.88 ± 3.71 n=11	0.627
Total Protein (g/dl)	7.19 ± 0.18* n=8	7.18 ± 0.10* n=12	6.70 ± 0.09 n=12	0.007
Albumin (g/dl)	4.26 ± 0.16 n=8	4.08 ± 0.08 n=12	3.68 ± 0.07 n=12	0.122
Haptoglobin (mg/dl)	181.10 ± 17.64 n=8	176.40 ± 13.92 n=11	209 ± 11.49 n=12	0.193

Supplementary Table 3. **Demographic and clinical characteristics of the three group of steatotic grade patients included in this study.** ***BMI:*** Body Mass Index, ***WHR:*** Waist-Hip Ratio, ***GGT:*** Gamma-Glutamyltransferase; ***ALT:*** Alanine Transaminase, ***AST:*** Aspartate Transaminase, ***HDL:*** High Density Lipoprotein, ***LDL:*** Low Density Lipoprotein, ***CRP:*** C-Reactive Protein.

Supplementary Table 4

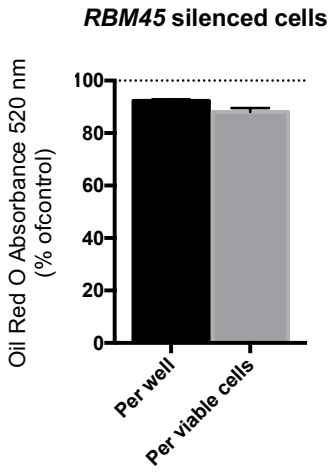
	RNU6ATAC	*RNU6*	*SF3B1*	*RNU2*	*RNU4ATAC*	*RBM22*	*U2AF1*	*U2AF2*
Sensitivity	75.76	84.83	77.19	71.88	77.42	51.61	84.85	75.76
Specificity	66.67	75	77.78	87.5	71.43	77.78	66.67	77.78

	PTBP1	*SRRM1*	*SND1*	*KHDRBS1*	*SRSF2*	*SRSF10*	*ESRP2*	*TIA1*
Sensitivity	75.76	63.64	75.76	78.79	75.76	66.67	71.88	69.7
Specificity	88.89	66.67	66.67	66.67	77.78	66.67	77.78	66.67

Supplementary Table 4. Sensitivity and Specificity values of ROC Curves of altered spliceosome components and splicing factors in patients with hepatic steatosis *vs.* patients without steatosis.

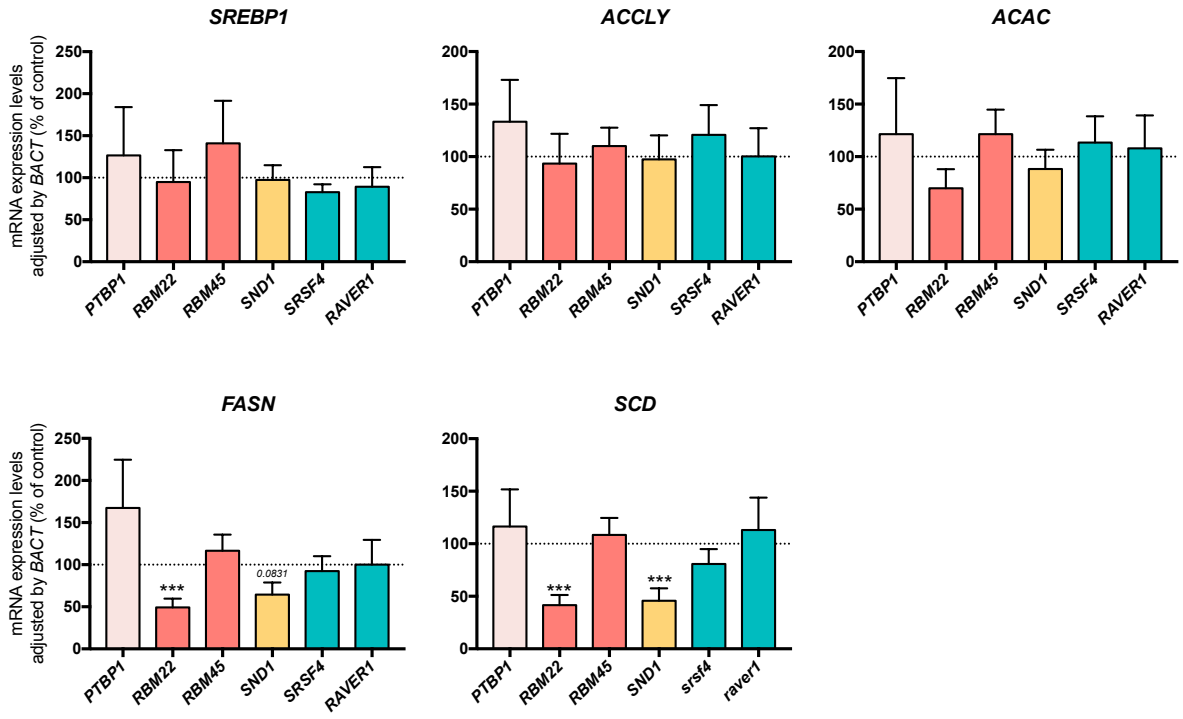
Supplementary Figures

Supplementary Figure 1



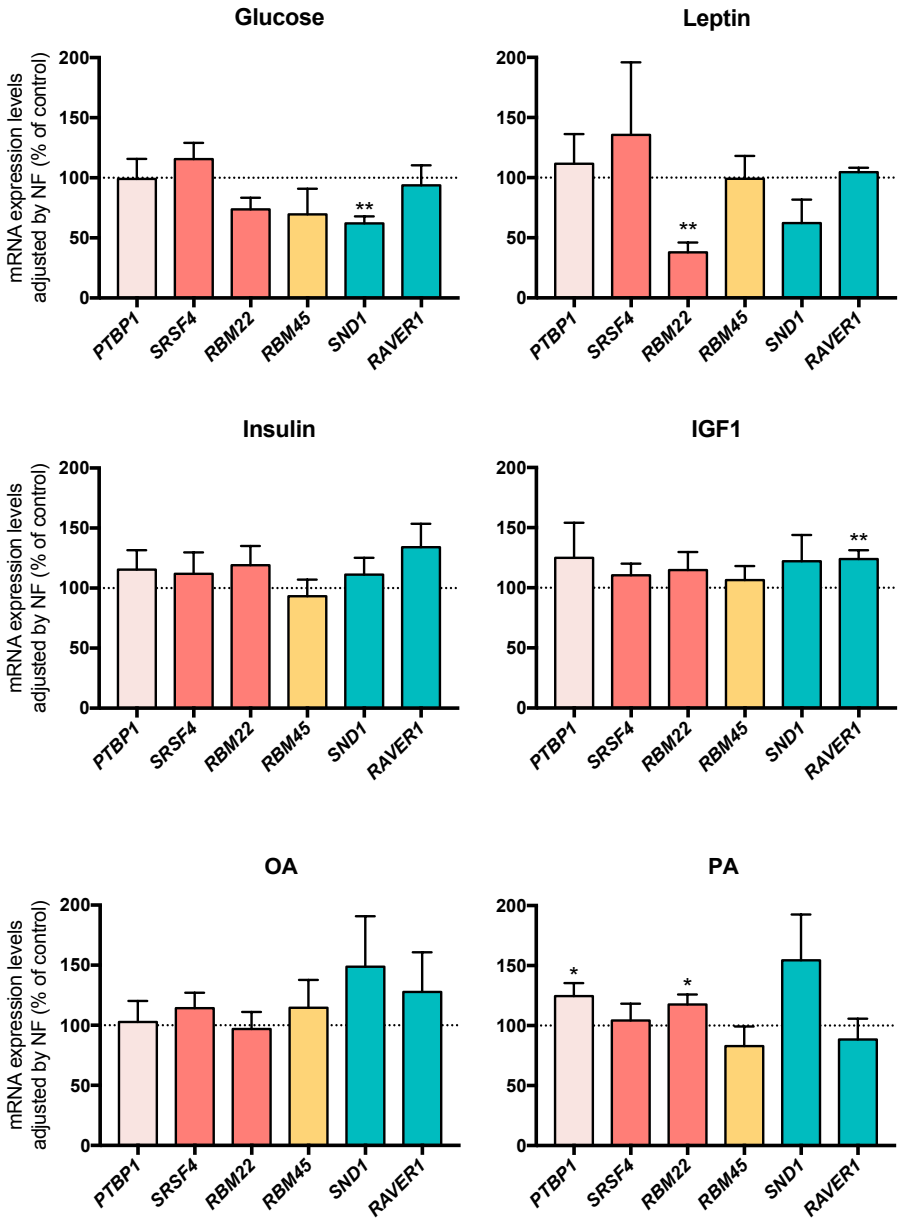
Supplementary Figure 1. Cell viability in response to the insults did not alter the observed results. Fat accumulation in *RBM45* silenced HepG2 cells after OA treatment was lower compared with percentage of OA accumulation in control cells, as shown also in Fig.4C. This result was similar when considering viable cells per well, since the mean of percentage of lipid accumulation in RBM45 silenced cells compared with negative control cells was 92% analyzing ORO absorbance per well, and 88% when dividing Oil Red O absorbance into viable cells per well.

Supplementary Figure 2



Supplementary Figure 2. Impact of the silencing of specific splicing factors on the expression of de novo lipogenesis enzymes in HepG2 cells. mRNA expression levels, adjusted by the expression level of *ACTB*, of different *de novo* lipogenesis enzymes in response to the silencing of key spliceosome components and splicing factors in HepG2 cells. Data are expressed as percentage of the control (normalized to 100%). Values represent the mean ± SEM (n=5). Data are expressed as percentage of control random siRNAs (Scramble; set at 100%). Asterisks indicate values that significantly differ from scramble-treated HepG2 cells (t***p<0.001).

Supplementary Figure 3



Supplementary Figure 3. Effect of treatment with several metabolic factors on the expression of selected splicing factors in HepG2 cells. mRNA expression levels [adjusted by a normalization factor (NF) calculated from the expression level of *HPRT* and *GAPDH*] of the different spliceosome components and splicing factors in response to different metabolic stimuli in HepG2 cells. Data are expressed as percentage of the control (normalized to 100%). Values represent the mean ± SEM (Glucose and leptin n=3; Insulin, IGF1, Oleic acid and Palmitic acid n=5). Data are expressed as percentage of control random siRNAs (Scramble; set at 100%). Asterisks indicate values that significantly differ from non-treated HepG2 cells (t-test: **p<0.01, ***p<0.001). ***OA***: oleic acid. ***PA***: palmitic acid.

Article IV

Peptides derived from the extracellular domain of the somatostatin receptor splicing variant SST5TMD4 increase malignancy in multiple cancer cell types

Q1 MERCEDES DEL RIO-MORENO, EMILIA ALORS-PEREZ, PATRICIA BORGES DE SOUZA, MARIA E. PRADOS-GONZALEZ, JUSTO P. CASTAÑO, RAÚL M. LUQUE, and MANUEL D. GAHETE

CÓRDOBA, SPAIN

Q2 **Extracellular fragments derived from plasma membrane receptors can play relevant roles in the development/progression of tumor pathologies, thereby offering novel diagnostic or therapeutic opportunities. The truncated variant of somatostatin receptor subtype-5, SST5TMD4, is an aberrantly spliced receptor with 4 transmembrane domains, highly overexpressed in several tumor types, whose C-terminal tail is exposed towards the extracellular matrix, and could therefore be the substrate for proteolytic enzymes. In silico analysis implemented herein predicted 2 possible cleavage sites for metalloproteases MMP2, 9, 14, and 16 in its sequence, which could generate 3 releasable peptides. Of note, expression those MMPs was directly correlated with SST5TMD4 in several cancer-derived cell lines (ie neuroendocrine tumors and prostate, breast, and liver cancers). Moreover, incubation with SST5TMD4-derived peptides enhanced malignancy features in all cancer cell types tested (ie proliferation, migration, etc.) and blunted the antiproliferative response to somatostatin in QGP-1 cells, acting probably through PI3K/AKT and/or MEK/ERK signaling pathways and the modulation of key cancer-associated genes (eg MMPs, MKI67, ACTR2/3, CD24/44). These results suggest that SST5TMD4-derived peptides could contribute to the strong oncogenic role of SST5TMD4 observed in multiple tumor pathologies, and, therefore, represent potential candidates to identify novel diagnostic, prognostic, or therapeutic targets in cancer. (Translational Research 2019; ■■:■■-■■)**

Abbreviations: CSCs = cancer stem cells; ECM = extracellular matrix; GPCR = G-protein coupled receptor; HCC = hepatocellular carcinoma; IGF1 = insulin-like growth factor 1; MMP = matrix metalloproteinase; NET = neuroendocrine tumor; SS14 = somatostatin 14; SSA = somatostatin analog; SST1-5 = somatostatin receptors 1-5; SST5TMD4 = somatostatin receptor 5 variant with 4 transmembrane domains; TMD = transmembrane domain

From the Maimonides Institute for Biomedical Research of Cordoba (IMIBIC), Córdoba, Spain; Universidad de Córdoba, Córdoba, Spain; Reina Sofia University Hospital, Córdoba, Spain; CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn), Córdoba, Spain; Agrifood Campus of International Excellence (ceiA3), Córdoba, Spain.

Submitted for Publication October 4, 2018; received submitted February 6, 2019; accepted for publication February 26, 2019.

Reprint requests: Manuel D. Gahete, Raúl M. Luque and Justo P. Castaño. Edificio IMIBIC. Av. Menéndez Pidal s/n, Córdoba 14004, Spain. e-mail: justo@uco.es raul.luque@uco.es bc2gaorm@uco.es.

1931-5244/$ - see front matter

https://doi.org/10.1016/j.trsl.2019.02.013

AT A GLANCE COMMENTARY

del Rio-Moreno M, et al.

Background

Extracellular fragments derived from membrane receptors can play relevant roles in tumor pathologies. The truncated variant of somatostatin receptor subtype 5, SST5TMD4, is an aberrantly spliced receptor highly overexpressed in several tumor types, whose C-terminal tail is exposed towards the extracellular matrix and could be the substrate of matrix metalloproteinases to induce the release of soluble peptides.

Translational Significance

Peptides derived from the SST5TMD4 increase in vitro tumorigenic potential of different tumor-derived cell lines, indicating that they could have important pathological implications and could be potential candidates to identify novel diagnostic, prognostic, and/or therapeutic tools in endocrine-related tumor pathologies.

INTRODUCTION

Somatostatin, a neuropeptide mainly produced in central nervous system and gastrointestinal tract, exerts a broad range of biological functions in endocrine and nonendocrine tissues, including inhibition of hormone secretions or cellular growth.[1,2] Somatostatin binds to 5 G-protein coupled receptors subtypes with 7-transmembrane-domain, named somatostatin receptors SST1-5, to modulate diverse downstream pathways.[3] The 5 SSTs are encoded by 5 separated genes and exert distinct effects depending on the particular constellation of SSTs available on cell surface, as they can interact with themselves or with other receptors.[4,5] Somatostatin binding to SST1-5 can trigger antitumoral effects, inhibiting cell proliferation, angiogenesis and hormone secretion, and/or inducing apoptosis.[3,6] Consequently, somatostatin synthetic analogues (SSA), like octreotide and lanreotide, were developed and are used as medical therapy for different tumor pathologies including pituitary or neuroendocrine tumors (NETs).[3,6] Nevertheless, although SST1-5 is abundantly expressed in other cancers, such as prostate, breast, or hepatocellular carcinoma (HCC), their potential clinical value in these pathologies is still to be defined.[7-9]

In this scenario, several pieces of evidence have demonstrated that the somatostatin system is more complex than originally envisioned. Indeed, recent studies have revealed the existence of diverse mechanisms that increase variability of G-protein coupled receptors, including the SST1-5, as is the case of alternative splicing processes that could generate noncanonical truncated variants with less than 7-transmembrane-domain.[10-13] These truncated forms are functionally active by modulating the physiology of their canonical isoforms or by exerting separate, independent functions.[12,13] Moreover, presence of these truncated receptors is commonly associated to development/progression of tumor pathologies, as is the case of the growth hormone-releasing hormone receptor,[10,11] cholecystokinin receptor,[14] or adrenergic receptors.[15] In the somatostatin family, our group has also identified a noncanonical truncated splicing variant of SST5 that only harbors 4TMDs and was therefore named SST5TMD4, which is generated by cryptic splice sites in the coding sequence and the distal, non-coding 3'UTR of the *SST5* gene.[4] The SST5TMD4 is barely expressed in normal tissues but is overexpressed in different tumor pathologies as pituitary and NETs, as well as in breast, prostate, and thyroid cancer.[4,16-21] More importantly, SST5TMD4 acts as an oncogene in these pathologies, wherein its presence is associated to resistance of SSA and with malignancy features, as it correlates with clinical parameters of aggressiveness and promotes cell proliferation, migration, invasion, and exacerbated hormone secretion.[4,16-22]

Remarkably, a unique feature of this truncated SST5TMD4 receptor is the presence of 4TMDs, which implies that its C-terminal tail is exposed to the extracellular region.[4,23] In this regard, an increasing number of studies suggest that extracellular fragments derived from shedding of plasma membrane receptors can play relevant functional roles in the development/progression of tumor pathologies and might, therefore, provide novel diagnostic/therapeutic tools for these pathologies.[24-27] Therefore, since SST5TMD4 presents only 4 TMDs and its C-terminal tail is directed towards the extracellular matrix (ECM) instead of the cytoplasm,[16,23] the extracellular region of SST5TMD4 may be susceptible to the action of proteases confined in the ECM such as metalloproteinases (MMPs), a group of zinc- and calcium-dependent proteolytic enzymes capable to degrade the majority of ECM proteins, such as collagen and elastin, as well as to regulate the activity of other proteinases, growth factors, and cell receptors.[28,29]

Consequently, we hypothesized that the action of MMPs on the SST5TMD4 extracellular domain, which is exposed to the ECM, could induce the release of soluble peptides that may exert some of the pathological actions previously ascribed to SST5TMD4, and therefore, could provide useful tools in the diagnostic,

prognostic, and/or therapeutic management of patients harboring SST5TMD4-related tumor pathologies. Thence, the aim of the present work was to determine if SST5TMD4 C-terminal domain has potential MMPs sites that could generate derived peptides, and to evaluate if such peptides could be able to directly trigger functional responses in different cancer cells in which the somatostatin system plays a relevant role.

MATERIALS AND METHODS

In silico prediction of MMPs cleavage sites and synthesis of SST5TMD4-derived peptides. The complete sequence of SST5TMD4 was analyzed with different bioinformatic tools for prediction of the transmembrane regions: TMpred, THTMM (Technical University of Denmark), MobylePasteur, and DAS-TMfilter. Furthermore, an in silico prediction of putative cleavage sites for MMPs in the extracellular segment of SST5TMD4 was performed using CleavPredic. The resulting SST5TMD4-derived peptides (M7, M10, and M17) were chemically synthesized by Genosphere Biotechnologies (Paris, France) and used in successive experiments. Specifically, peptides were synthesized by solid phase synthesis using standard Fmoc-protecting group chemistry and purification was achieved through semipreparative C18 reversed-phase HPLC using TFA/CH3CN gradient.

Cell lines and treatments. The androgen-sensitive 22Rv1 (CRL-2505 and androgen-insensitive PC-3 (CRL-1435 prostate cancer cell lines, the SNU-387 (HB-8065) HCC cell line, and the invasive ER-negative MDA-MB-231 (HTB-26 and noninvasive ER-positive MCF-7 (HTB-22 breast cancer cell lines were purchased from ATCC (Manassas, VA). Two previously validated human NET cell lines, the carcinoid-like BON-1,[30] and the somatostatinoma-derived QGP-1[31] were also used. All of them were maintained according to manufacturer's instructions at 37°C and 5% CO_2, under sterile conditions, validated by short tandem repeat analysis (GenePrint 10 System, Promega, Barcelona, Spain), and routinely tested for mycoplasma contamination, as previously reported.[32,33] In each experiment, cells were treated with the 3 SST5TMD4-derived peptides at a final concentration of 10^{-7} M. This concentration was selected based on dose-response experiments carried out in QGP-1 cell line to evaluate the proliferative response to 5 different concentrations of the SST5TMD4-derived peptides (10^{-9} M to 10^{-5} M) after 24 hours. This approach revealed that 10^{-7} M was lower dose that caused a maximal proliferative response (Supplementary Fig 1, A). Furthermore, we validated that 10^{-7} M of the M17 peptide was more effective than the 2 lowest concentrations tested (10^{-9} M and 10^{-8} M) in BON-1, 22Rv1, and SNU-387 cells (Supplementary Fig 1, B). Somatostatin 14 (SS14; Biogenesis, Poole, UK), octreotide (GP-Pharm, Barcelona, Spain), lanreotide (provided by IPSEN Bioscience, Cambridge, Massachusetts), and pasireotide (Novartis Pharmaceuticals Corporation, East Hanover, New Jersey) were also used at a concentration of 10^{-7} M based on previous studies.[34,35]

Measurements of cell proliferation. Cell proliferation was determined by Alamar Blue-based assays (Thermo Scientific) in all cell lines as previously reported.[16,20,21,32,33,36] Briefly, cells were seeded and serum-starved for 24 hours. Cell proliferation was evaluated every 24 hours. At least 3 experiments were performed in independent days. In all the experiments, cells were seeded per quadruplicate per treatment.

Measurement of cell migration capacity. Cell migration capacity was evaluated on PC-3, MDA-MB-231, MCF-7, SNU-387, and BON-1 cells, by wound healing assay as previously described.[16,20,32,33,36] Cells were serum starved for 1 hour when confluence was reached, and a wound was made in the center of each well using a 100 μl pipette tip. Cells were treated with each SST5TMD4-derived peptides per triplicate in serum-free medium to avoid proliferation. At least 3 experiments were performed in independent days, in which 4 pictures along the wound were acquired per well at 0 and 24 hours. Wound healing was calculated as the area of a rectangle centered in the image 24 after the wound vs the area of the rectangle just after wounding, using Image J-1.51s software (NIH, Bethesda, Maryland), and expressed as percentage of recovered area compared to vehicle-treated control cells.

Tumorspheres formation. Tumorspheres formation was assessed to evaluate the proliferation of cancer stem-like cells (CSCs) present in the population of tumor cell lines. These cells constitute a subpopulation that exhibit initiation and metastasis properties. In order to enrich the CSCs population to analyze its self-renewal capability, cells were grown under nonadherent serum-free conditions and supplemented with growth factors, as only CSCs can proliferate in this environment. Thereby, tumorspheres formation was assessed using Ultra-Low Attachment Multiwell Plates (Sigma-Aldrich, Madrid, Spain) as previously reported.[19,33] Briefly, cells were seeded in DMEM/F12 (Thermo Scientific, Madrid, Spain) serum-free medium supplemented with 20ng/mg EGF (Sigma-Aldrich; Madrid, Spain) for prostate cells; 20 ng/ml bFGF (Sigma Aldrich) and 20 μl/ml B27 supplement vitamin A (Thermo Scientific) for HCC cells; 20 ng/ml EGF and 20 μl/ml B27 supplement vitamin A for breast

cancer cell lines; and 20 ng/mg IGF1 (Sigma-Aldrich) for NET cells lines. Treatments were added in the moment of plating and refreshed every 2 days. The number of tumorspheres was determined after 7 days of incubation. At least 3 experiments were performed in independent days. In all the experiments, cells were seeded per triplicate per treatment.

Western blot. Phosphorylation levels of ERK1/2, AKT and JNK were analyzed by western blot as previously reported.[19,21,32,33] Briefly, cells were serum starved for 1 hour, treated with SST5TMD4-derived peptides for 5, 10, or 15 minutes and total protein was collected in SDS-DTT buffer. At least 3 experiments were performed in independent days. Proteins were separated by SDS-PAGE and transferred to a nitrocellulose membrane (EMD Millipore, Darmstadt, Germany). Antibodies for phospho-ERK1/2 (#4370S, Cell Signaling Technology; Danvers, Massachusetts), phospho-AKT (#9271S, Cell Signaling Technology), phospho-JNK (AF1205, R&D Systems; Minneapolis, Minnesota), total ERK1/2 (SC-154, Santa Cruz Biotechnology; Santa Cruz, California), total AKT (#9272S, Cell Signaling Technology), total JNK (AF1387, R&D Systems) and horseradish peroxidase-conjugated goat antirabbit IgG (#7074, Cell Signaling Technology) were used. Blots were exposed to Clatirty Western-ECL Blotting Substrate (Bio-Rad Laboratories, Madrid, Spain) and scanned using ImageQuant Las 4000 system (GE Healthcare Europe GmbH). Images were analyzed using ImageJ-1.51s software.

RNA isolation and real-time quantitative RT-PCR. Details regarding RNA extraction, quantification, and reverse transcription have been previously reported elsewhere by our group.[37] Briefly, total RNA was isolated from cells using TRI Reagent (Sigma-Aldrich, Madrid, Spain), followed by DNase treatment. The amount and purity of RNA recovered was determined using the NanoDrop2000 spectrophotometer (Thermo Scientific, Madrid, Spain). One μg of RNA was reverse transcribed to cDNA, using random hexamer primers and RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific), and amplified by quantitative real-time PCR using specific primers for human transcripts (Supplementary table 1). qPCR reactions were carried out using the Stratagene M×3000p system with the Brilliant III SYBR Green Master Mix (Stratagene, La Jolla, California). The expression level of each transcript was adjusted by a normalization factor obtained from the expression levels of 3 different housekeeping genes [beta-actin (*ACTB*), hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)] using Genorm 3.3.[38]

Statistical analysis. All statistical analyses were performed using GraphPad Prism (La Jolla, California). Proliferation assay and western blot analyses were analyzed by 2-way ANOVA followed by Bonferroni's posthoc test. Migration, tumorspheres formation, and gene expression assays were analyzed using 1-way ANOVA followed by Bonferroni's posthoc test. Correlations were studied by using the Pearson correlation test. All data were obtained from at least 3 independent experiments from different cellular passages and expressed as mean ± SEM. *P* values smaller than 0.05 were considered statistically significant.

RESULTS

SST5TMD4 C-terminal domain displays 2 putative MMPs cleavage sites. An analysis of the complete sequence of SST5TMD4 using different bioinformatic tools for the prediction of the transmembrane regions revealed that the C-terminal extracellular domain starts at the amino-acid position 179 (Fig 1, A). Subsequently, an in-silico prediction of putative cleavage sites for MMPs revealed 2 processing sites; 1 for MMP2, 9 and 14 (at position 200) and 1 for MMP16 (at position 210), which could lead to the generation of 3 putative peptides in response to the alternative and/or simultaneous action of such MMPs. These putative peptides have a length of 17, 10, and 7 amino acids (and consequently were named M17: HRERLSGHKSW-QEKGPG; M10: HRERLSGHKS; and M7: WQEKGPG) (Fig 1, B).

SST5TMD4 is coexpressed with MMP2, 9, 14, and/or 16 in cancer-derived cell lines. Expression studies revealed a coexpression of *SST5TMD4* with *MMP2, 9, 14,* and/or *16* in these cancer cell-lines (Fig 1, C). Indeed, *SST5TMD4* was expressed, although at low levels, in all cell lines, while the expression patterns of *MMP2, 9, 14,* and *16* was cell line-dependent, wherein relevant expression levels of some *MMPs* were found in each cell line (Fig 1, C). Interestingly, although no correlation was found between *SST5TMD4* and each *MMP* (*MMP2*: $r^2 = 0.2792$, $P = 0.2228$; *MMP9*: $r^2 = 0.0001$, $P = 0.9814$; *MMP14*: $r^2 = 0.3033$, $P = 0.2001$; *MMP16*: $r^2 = 0.3916$, $P = 0.1328$), a direct correlation between the total expression of *MMPs* and *SST5TMD4* was found ($r^2 = 0.8896$, $P = 0.0014$). All these cells lines also exhibited considerable levels of *SST5*, but *SST5* expression was not correlated with that of *SST5TMD4* (PC-3: $r^2 = 0.1124$, $P = 0.5812$; 22Rv1: $r^2 = 0.0118$, $P = 0.8617$; SNU-387: $r^2 = 0.8341$, $P = 0.2671$; MDA-MB-231: $r^2 = 0.515$, $P = 0.2824$; MCF-7: $r^2 = 0.3256$, $P = 0.4293$; BON-1: $r^2 = 0.7248$, $P = 0.1487$; QGP-1: $r^2 = 0.01049$, $P = 0.8698$), indicating that the activity of the truncated and wild type variants might be not invariably linked.

SST5TMD4-derived peptides increased cell proliferation and migration of different cancer cell lines. All 3 SST5TMD4-derived peptides significantly enhanced cell proliferation rate compared to control cells in PC-3, 22Rv1, SNU-387, MDA-MB-231, BON-1, and QGP-1 cells lines at 24, 48, and/or 72 hours (Fig 2). Conversely, no such effect was observed in MCF-7 cells (data not shown). Additionally, all cell lines tested (PC-3, SNU-387, MDA-MB-231, MCF-7, and BON-1) exhibited a higher migration rate when treated with SST5TMD4-derived peptides (Fig 3).

SST5TMD4-derived peptides increased tumorspheres formation of breast cancer and NETs cell lines. SST5TMD4-derived peptides increased the number of tumorspheres formed in breast cancer (MDA-MB-231 and MCF-7) and NETs (BON-1 and QGP-1) cell lines compared to control cells, this stimulatory effect being higher in MCF-7 and BON-1 than in MDA-MB-231 and QGP-1 cells (Fig 4). Whereas, no similar changes were observed in prostate cancer and HCC cells in response to treatment with SST5TMD4-derived peptides (data not shown).

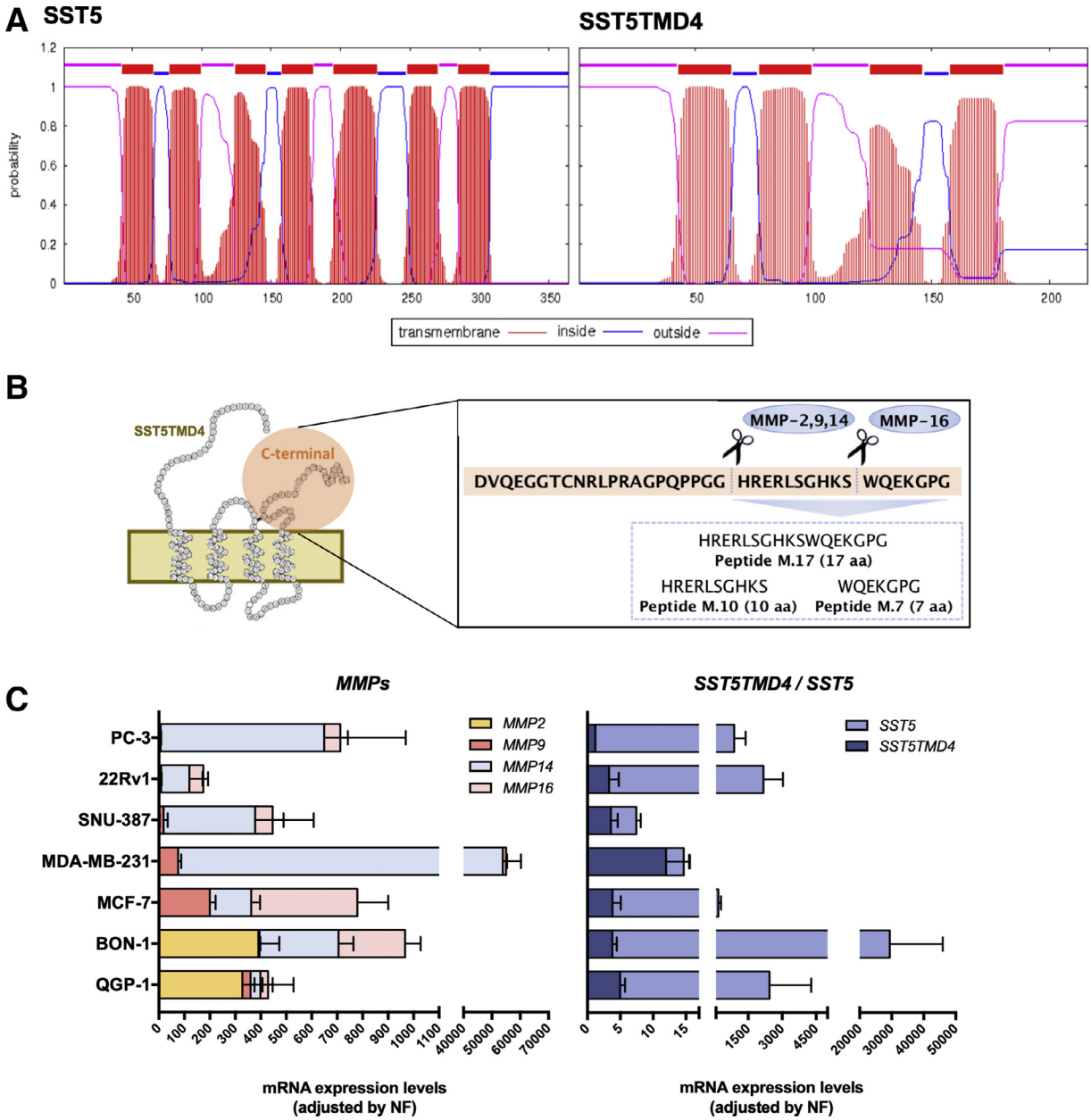


Fig 1. (A) Scheme representation of the predicted topology of SST5 and its truncated splicing variant SST5TMD4 with only 4 transmembrane domains, using the THTMM program. Areas that correspond to the transmembrane (red), intracellular (blue), and extracellular (pink) regions, calculated by probability analysis, are indicated. **(B) In silico determination of the 2 processing sites for metalloproteases (MMP2, 9, 14, and 16) in the extracellular domain of SST5TMD4 and prediction of the peptides derived from their processing (M7, M10, and M17).** The presence of possible cleavage sites for proteases in the extracellular segment of SST5TMD4 was analyzed using the CleavPredict software. **(C) Expression levels of *MMPs*, *SST5TMD4, and SST5* in all the human tumor-derived cell lines used in this study.** Data indicate mRNA expression levels of *MMP2, 9, 14, 16, and SST5TMD4* adjusted by a normalization factor (NF) (calculated from the expression level of *ACTB*, *HPRT*, and *GAPDH).* Values represent the mean ± SEM. ***aa***, amino acids.

SST5TMD4-derived peptides induced intracellular signaling activation in PC-3, MDA-MB-231, and NETs cells lines. ERK1/2 and AKT phosphorylation was clearly increased in some of the cancer cell lines analyzed in response to SST5TMD4-derived peptides, while no changes in JNK were observed. Indeed, PC-3, BON-1, and QGP-1 cells exhibited higher levels of pAKT and pERK after the treatment with the 3 SST5TMD4-derived peptides at 5, 10, or 15 minutes (Fig 5, A, C, and D). Identical effects were observed in pAKT levels in MDA-MB-231 cells (Fig 5, B). Conversely, phosphorylation levels of pAKT and pERK in response to M7, M10, and M17 showed no

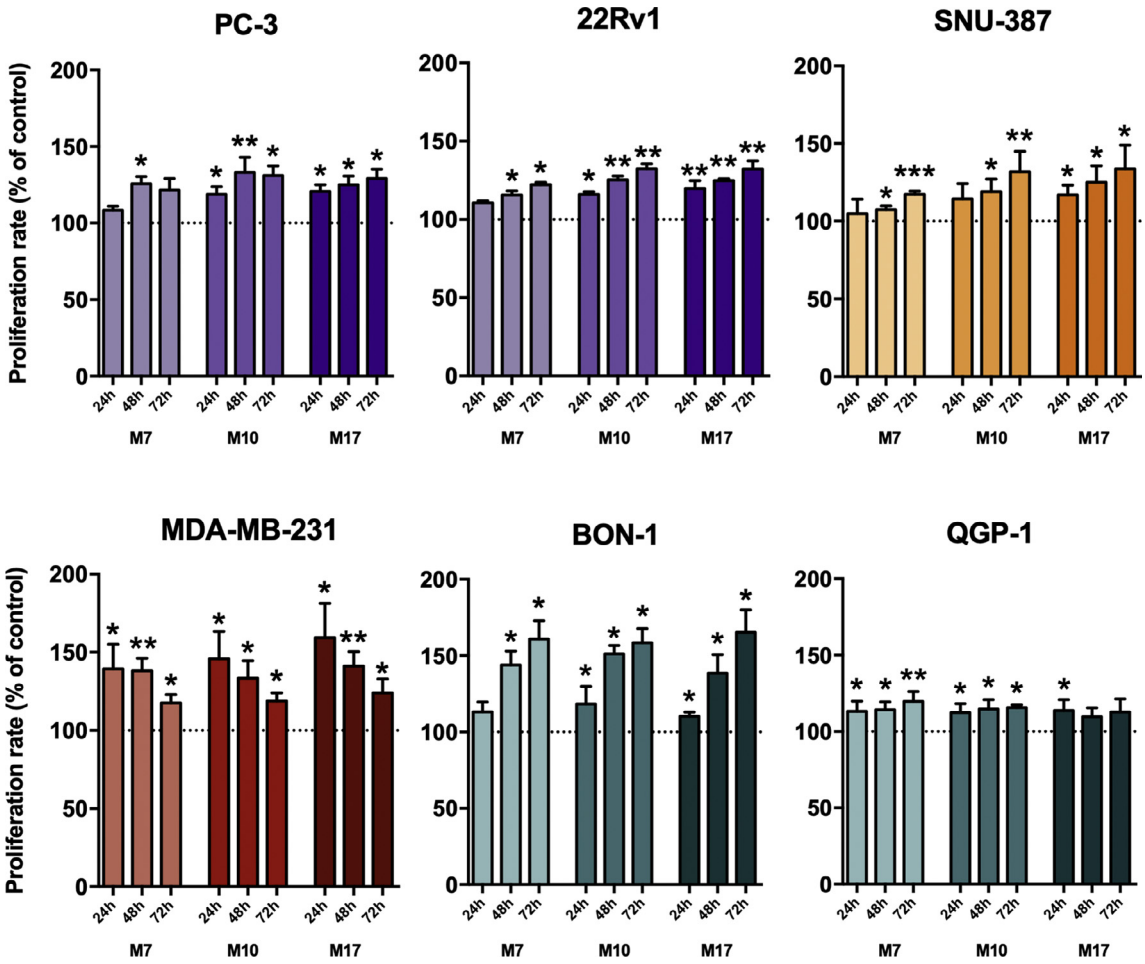


Fig 2. Effect of SST5TMD4-derived peptides M7, M10, and M17 on cell proliferation (24, 48, and/or 72 hours) of prostate cancer (PC-3 and 22Rv1), hepatocellular carcinoma (SNU-387), breast cancer (MDA-MB-231), and NET (BON-1 and QGP-1) cell lines. The data are expressed as percentage of proliferation rate compared to vehicle-treated control cells (set at 100%). Values represent the mean ± SEM of 3−7 experiments (in each experiment cells were seeded per quadruplicate per treatment). Asterisks indicate values that significantly differ from control cells (*t* test: **, $P < 0.01$; *, $P < 0.05$).

significant alteration in 22Rv1, MCF-7, and SNU-387 cell lines (data not shown).

SST5TMD4-derived peptides abrogated the antiproliferative effect of somatostatin in QGP-1 cell line. In order to explore the putative implication of the SST5TMD4-derived peptides in the cellular response to somatostatin or its analogs, we first evaluated the responsiveness of all the cell lines used herein to SS14 and SSAs (octreotide, lanreotide, and pasireotide) (data not shown), wherein only QGP-1 cells exhibited a robust and consistent inhibitory response to SS14, but not to SSAs, in terms of proliferation rate. For this reason, the putative implication of the SST5TMD4-derived peptides in the response to SS14 was tested in this cell line. After 48 hours of treatment, a decrease in cell proliferation was observed in SS14-treated QGP-1 cells; however, this effect was not observed when SS14 was combined with the SST5TMD4-derived peptides (Fig 6), suggesting that SST5TMD4-derived peptides could be playing a role in the blunted response to SSAs previously observed in SST5TMD4-expressing cells. However, further studies are required to substantiate these preliminary findings.

Expression of proliferation, migration, and stem cells associated genes in BON-1 cell line after treatment with SST5TMD4-derived peptides. Inasmuch as the 3 SST5TMD4-derived peptides exerted consistent effects on all the functional and signaling studies implemented in BON-1 cell line, we used this as a model cell line to further determine the expression levels of several genes involved in cancer development/progression (Fig 7). Specifically, M7 and M17 peptides increased the expression of the 2 proliferation markers analyzed (*CCND3* and *MKI67*). These 2 peptides also increased the expression of the components of the Arp2/3 complex (*ACTR2*, *ACTR3,* and *ARC*), and *MMP2*, whereas none of the 3 peptides altered the expression of *MMP9*. Similar tendencies in the modulation of *CCND3*, *MKI67, ACTR2*, and *ARC* were found for M10. Remarkably, all 3 SST5TMD4-derived peptides also enhanced the expression of *CD24* and *CD44*.

DISCUSSION

There is emerging evidence that receptor-derived peptide fragments are involved in cancer development

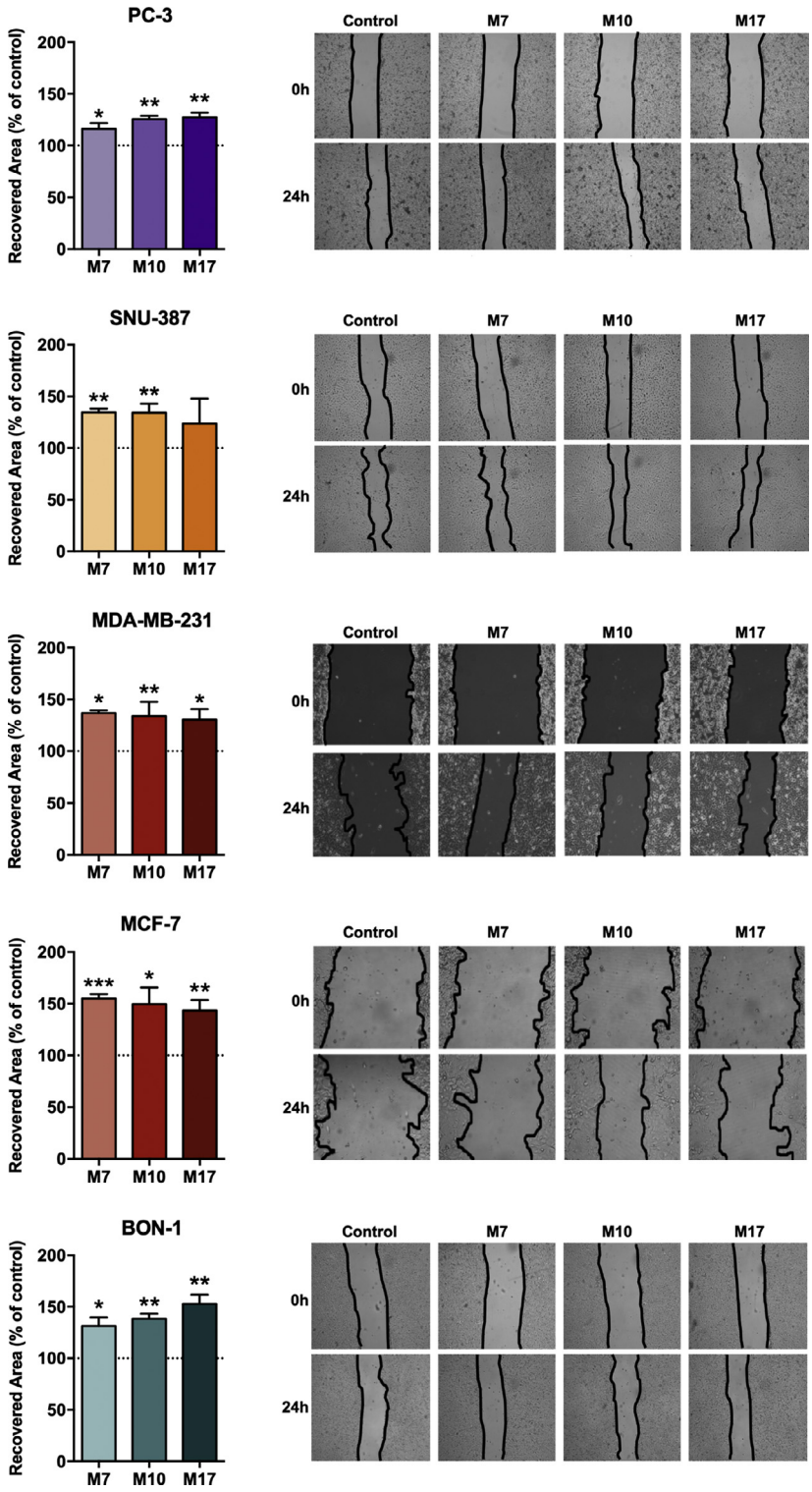


Fig 3. Effect of SST5TMD4-derived peptides M7, M10, and M17 on the migration capacity of PC-3, SNU-387, MDA-MB-231, MCF-7, and BON-1 cell lines after 24 hours of treatment. Representative images of the migration capacity are also included. The data are expressed as percentage of recovered area compared to vehicle-treated control cells (set at 100%). Values represent the mean ± SEM of 3−7 experiments (in each experiment cells were seeded per triplicate per treatment). Asterisks indicate values that significantly differ from control cells (*t* test: **, $P < 0.01$; *, $P < 0.05$).

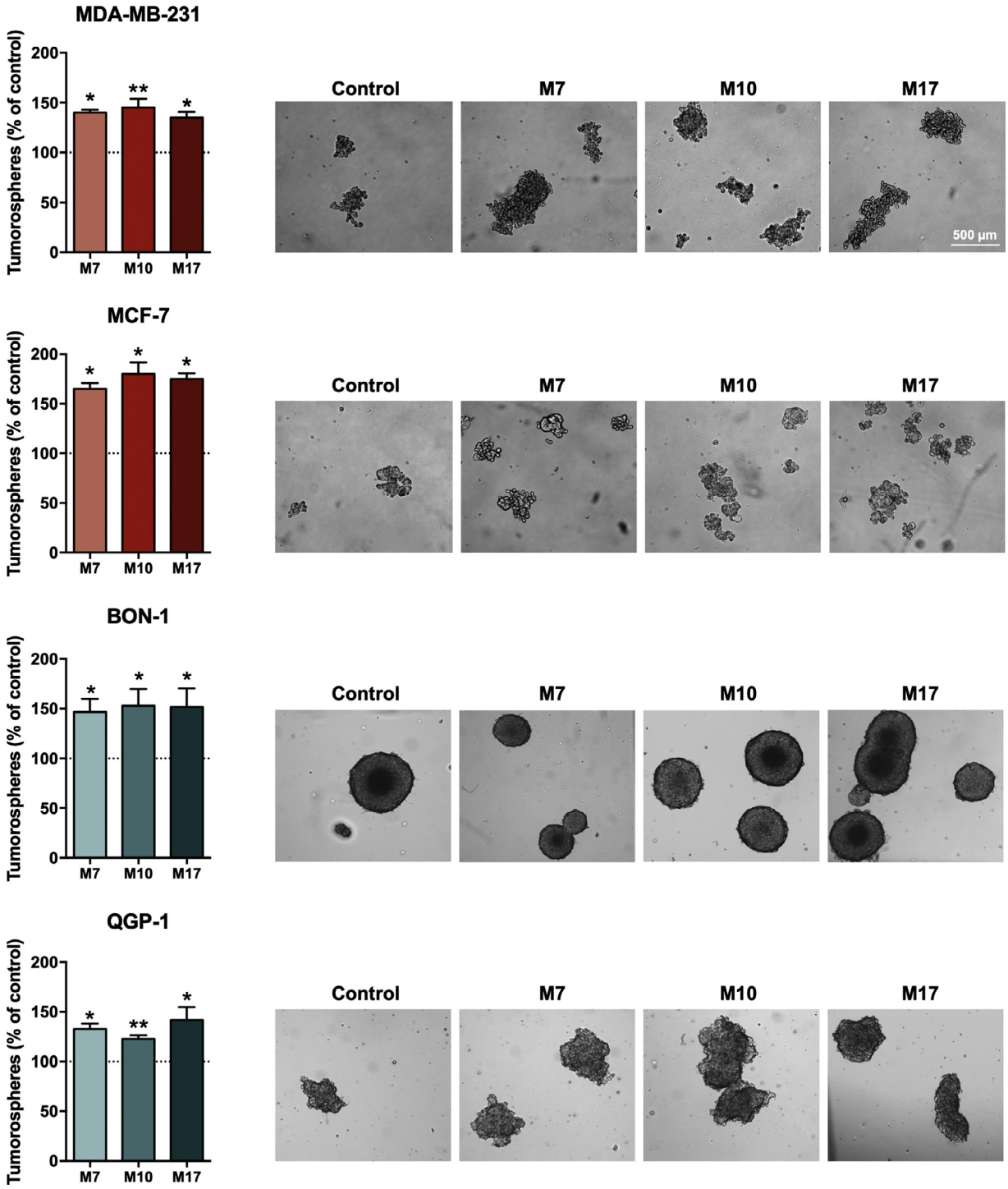


Fig 4. Effect of SST5TMD4-derived peptides M7, M10, and M17 on tumorsphere formation in breast cancer (MDA-MD-231 and MCF-7) and NET (BON-1 and QGP-1) cell lines. Representative images of tumorsphere formation are also included. Original magnification: ×10. Proliferation of cancer stem cells is quantified as numbers of tumorospheres respect to vehicle-treated control cells (set as 100%). Values represent the mean ± SEM of 3−6 experiments (in each experiment cells were seeded per triplicate per treatment). Asterisks indicate values that significantly differ from control (*t* test: *, $P < 0.05$; **, $P < 0.01$).

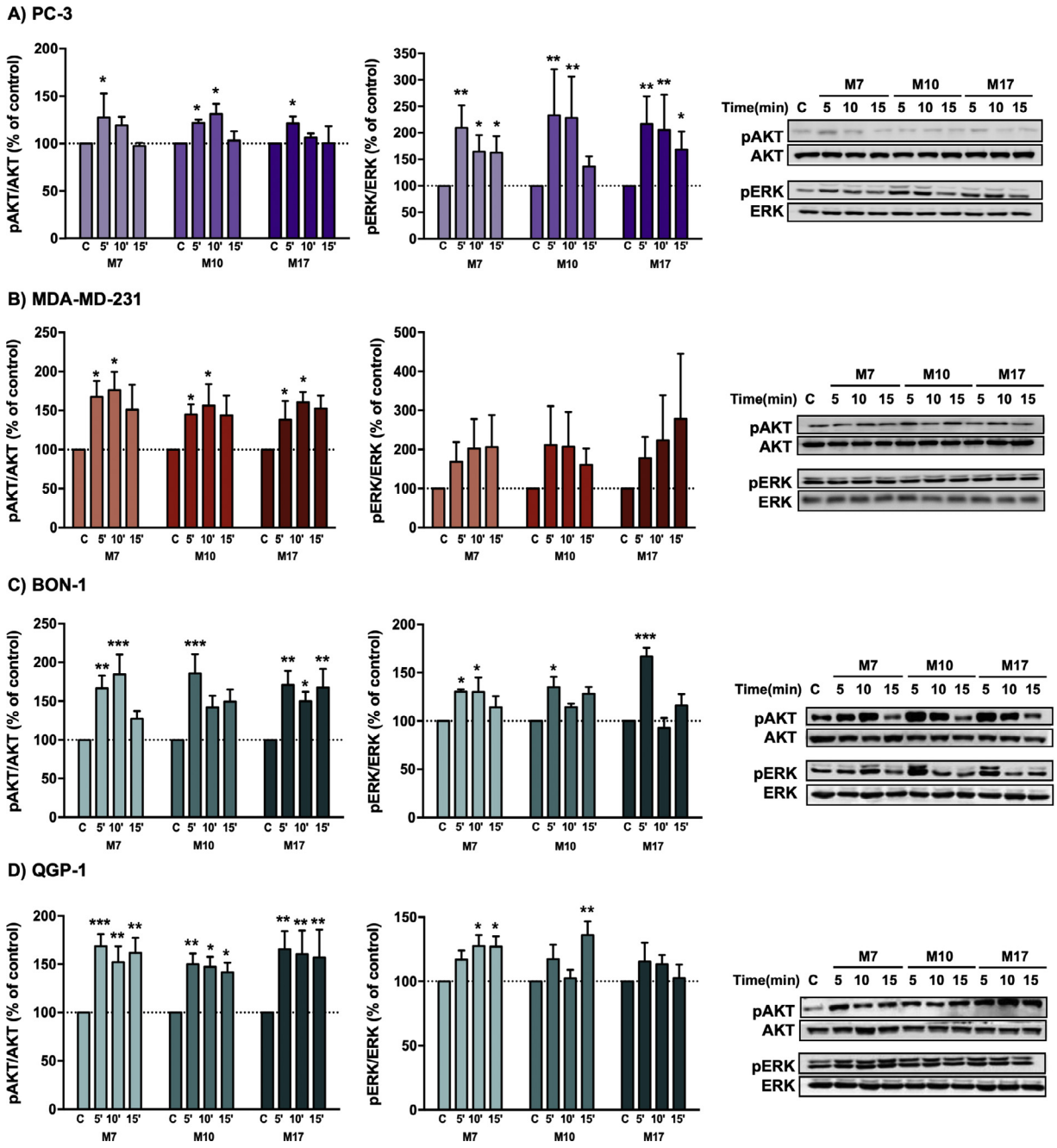


Fig 5. Effect of SST5TMD4-derived peptides M7, M10, and M17 on the activation of AKT and ERK signaling in PC-3 (A), MDA-MB-231 (B), BON-1 (C), and QGP-1 (D) cell lines. Data are expressed as percentage of phosphorylation after 5, 10, and 15 minutes of incubation with M7, M10, and M17 peptides compared to nontreated control cells (set at 100%). Values represent the mean ± SEM of 4 experiments. Asterisks indicate values that significantly differ from control (*t* test: * $P < 0.05$, ** $P < 0.01$, ***$P < 0.001$). ***C***, control.

and progression.[10,11,14,15,24-27,39] However, this issue has never been explored in SST5TMD4, a truncated receptor that is markedly overexpressed in several tumors, including pituitary, NETs, breast, prostate, and thyroid cancers, wherein its presence is associated to higher aggressiveness and/or resistance to medical treatment with SSAs.[4,16-19,21,22] Like other truncated receptors,[10-12] SST5TMD4 exhibits a preferential intracellular localization, where it disrupts the function of other SSTs, mainly SST2 and SST5,[4] and inhibits the ability of MCF-7 and SST2-transfected CHO-K1 cells to respond to somatostatin/SSAs.[16] However, a

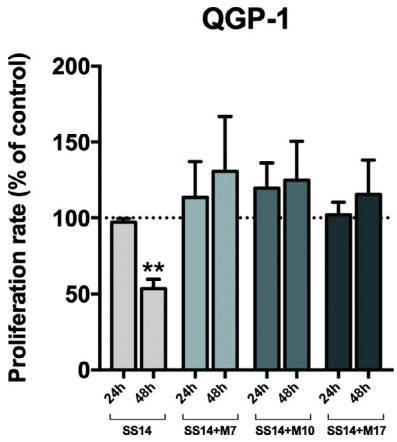


Fig 6. Effect of somatostatin-14 (SS14) alone or in combination with SST5TMD4-derived peptides on QGP-1 cells proliferation (24 and/or 48 hours). The data are expressed as percentage of proliferation rate compared to vehicle-treated control cells (set at 100%). Values represent the mean ± SEM of 3 experiments (in each experiment cells were seeded per quadruplicate per treatment). Asterisks indicate values that significantly differ from control cells (*t* test: **, $P < 0.01$; *, $P < 0.05$).

substantial proportion of SST5TMD4[4] also resides at the cell membrane, wherein it acts as a functional receptor capable to mediate ligand-induced intracellular responses to somatostatin and cortistatin, a natural somatostatin analog.[4,19,22] This unique location at the cell membrane, together with its distinctive feature of harboring only 4TMDs, confers SST5TMD4 one of its most relevant particularities, namely, that its C-terminal tail is exposed towards the ECM, rendering this extracellular region as a potential substrate for MMPs, which are tightly associated to cancer angiogenesis, invasiveness, and metastasis.[29] Here, we explored this novel aspect of SST5TMD4 biology to further understand the mechanisms of actions of this truncated receptor, and to identify novel putative diagnostic, prognostic, and therapeutic tools.

A detailed in silico analysis revealed, for the first time, the existence of 2 putative cleavage sites in the unique SST5TMD4 C-terminal extracellular domain, which could be used by MMP2, 9, and 14 and/or MMP16, respectively, to generate 3 derived peptides with 7, 10, and 17 aa. MMPs are able to degrade the majority of ECM proteins, including cell receptors,[28,29] and have been functionally linked to the aggressiveness of different tumors types.[29,40,41] Thus, some MMPs, especially MMP2, 9, and 14 play relevant roles in tumor pathologies, as they are present in tumor and stromal cells, including cancer-associated fibroblasts, macrophages, and endothelial cells,[42] wherein they can contribute to several tumor-associated processes, including cell invasion, growth, and survival.[43] These tumor-promoting features of MMPs may be, at least in part, mediated by the processing of key receptors present at the cellular surface, which can generate receptor-derived peptides with oncogenic capacity.[24-26] Actually, this could be the case of SST5TMD4, inasmuch as previous results have demonstrated that both, this truncated receptor and the MMPs potentially involved in its processing, are overexpressed in several tumor pathologies.[4,17,19-21,42,43] Importantly, this study provides 1 more piece of evidence to support this notion by demonstrating that the SST5TMD4 truncated variant is coexpressed with the MMPs implicated in the processing of its extracellular domain in all the cancer cell lines studied, wherein their expression levels are directly correlated.

This study also demonstrates that the SST5TMD4-derived peptides are capable to enhance the malignant characteristics (ie proliferation, migration, tumorspheres formation, and response to somatostatin) of multiple cancer cells derived from diverse tumor pathologies as NETs, breast, prostate, and liver cancer; however, some of these actions seem to exhibit different dynamics (ie proliferation rate) or, even, be cell-line dependent (ie migration or tumorosphere formation). Certainly, SST5TMD4-derived peptides were able to drastically increase the capacity of the BON-1 cells to proliferate, migrate, and induce tumorsphere formation; while in the QGP-1 cells, SST5TMD4-derived peptides clearly increased tumorsphere formation and hampered the cellular response to somatostatin, but exerted lesser effects on proliferation and had no effect on migration rate. In this sense, SST5TMD4 has been previously found overexpressed in gasteroenteropancreatic-NETs (GEP-NETs) and associated to aggressiveness, metastasis, and worse prognosis. Consistent with the present results, SST5TMD4 transfected BON-1 cells evidenced increased proliferation,[20] whereas SST5TMD4-transfected QGP-1 did not.[20] These differences could relate to the distinct nature of the 2 NET cell models,[20,44] since QGP-1 cells have constitutive high expression of somatostatin, which imparts a constant inhibition pattern that might hinder the stimulatory action of the SST5TMD4-derived peptides. Similar divergences were found in breast and prostate cancer cell models. Thus, whereas MDA-MB-231 cells exhibited increased cell proliferation, migration, and tumorspheres formation in response to the SST5TMD4-derived peptides, in MCF-7, a stimulatory effect of the 3 peptides was only found in cell migration capacity and tumorspheres formation, these differences being probably attributable to the distinct origin and aggressiveness of each cell line.[45] Consistent with these results, it was previously observed that forced overexpression of SST5TMD4 in MCF-7 and MDA-MB-231 cells

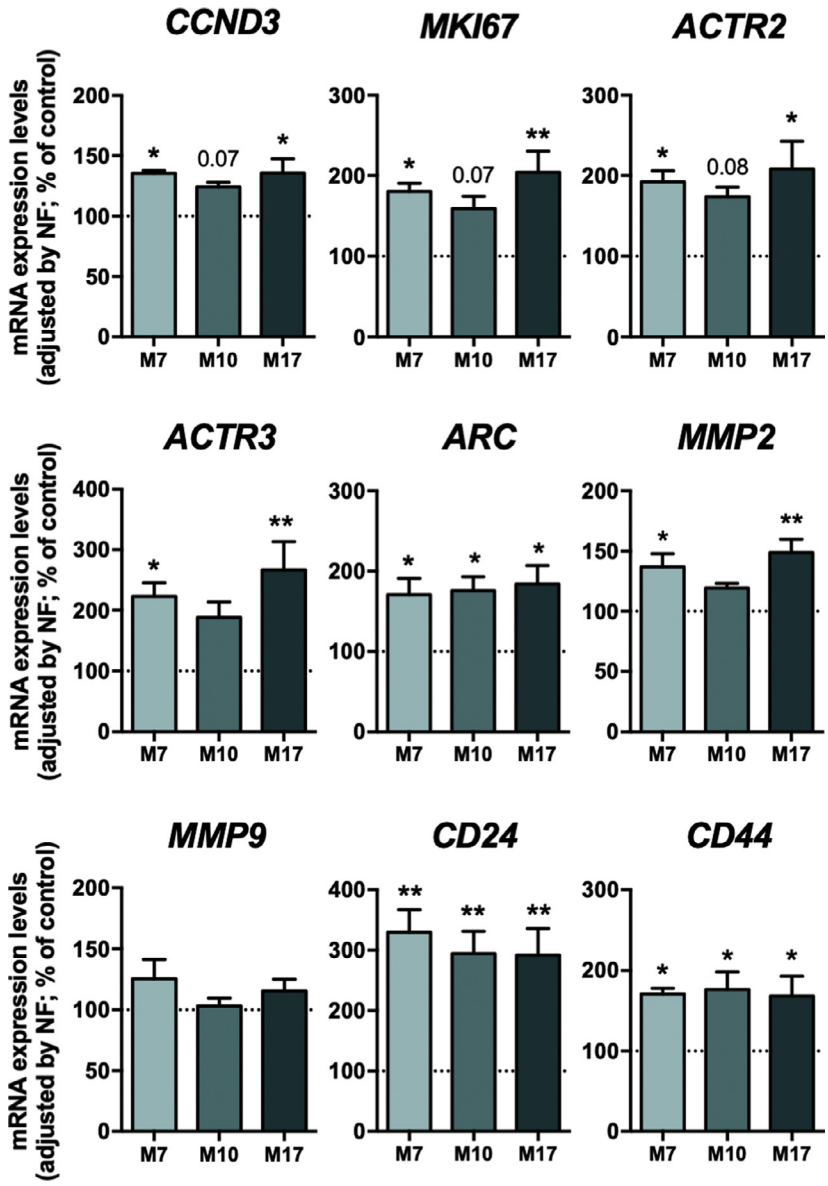


Fig 7. Effect of SST5TMD4-derived peptides M7, M10, and M17 on the expression of genes implicated in malignancy features in the NET cell line BON-1. Data indicate mRNA expression levels of *CCDN3, MKI67, ACTR2, ACTR3, ARC, MMP2, MMP9, CD24,* and *CD44* adjusted by a normalization factor (NF) (calculated from the expression level of *ACTB*, *HPRT*, and *GAPDH,)* and are expressed as percentage of expression compared to vehicle-treated control cells (set at 100%). Values represent the mean ± SEM of 3−4 experiments. Asterisks indicate values that significantly differ from control (*t* test: *, $P < 0.05$; **, $P < 0.01$).

increased proliferation, cell invasion and/or tumorsphere formation capacity.[16,19] In the case of prostate cancer cells, the bone metastasis-derived PC-3 cell line showed increased cell proliferation and migration in response to the treatment with the SST55TMD4-derived peptides. In contrast, in 22Rv1 cells, a cell line more representative of the early disease, exhibited a clear stimulatory effect of the 3 peptides only on cell proliferation.[46,47] It should be noted that SST5TMD4 is overexpressed in prostate cancer, especially in metastatic disease, and has been shown to hamper the response to SSAs.[21] Indeed, SST5TMD4 has been shown to play a functional role in prostate cancer cells, wherein its overexpression enhanced cell proliferation, migration, and promoted tumor growth in vivo.[21]

Of note, this report describes for the first time an increase in cell proliferation and migration in the HCC cell line SNU-387 in response to SST5TMD4-derived peptides. These results open a new research avenue in the study of this pathology in that, although the

presence and putative functional role of the SST5TMD4 has not yet been investigated in detail, it has been reported that HCC could present resistance to SSAs,[7,9] an important therapeutic limitation that has been shown to be associated, in several instances, with the presence of the SST5TMD4.[17,21,22] Moreover, these data on HCC cells demonstrate that the malignancy-prone effects of the SST5TMD4-derived peptides could have a wide range of action across different cancer types.

Finally, this study unveils some of the putative molecular mechanisms underlying the pathological effects exerted by the SST5TMD4-derived peptides in cancer cells. Precisely, SST5TMD4-derived peptides could exert their function through the activation of different cancer-relevant signaling pathways,[48-50] such as PI3K/AKT, MEK/ERK, but not JNK, in different cancer cell lines. In line with this, it has been also previously reported that SST5TMD4 exerts its functions via modulation of several pathways, including activation of PI3K/AKT[16] and MEK/ERK,[16,21] 2 signaling pathways associated to malignancy promotion in several tumor pathologies.[48-50] Interestingly, to further explore the effects of the SST5TMD4-derived peptides, the expression of different key genes associated to malignancy features was evaluated in BON-1 cell line, where the 3 SST5TMD4-derived peptides exerted consistent effects on all the functional and signaling studies tested. Interestingly, we observed a clear upregulation of the proliferation markers *CCND3* and *MKI67*, also used to assess the grade and differentiation of NETs,[51] the *MMP2*, involved in ECM degradation[29] and likely in the own production of SST5TMD4-derived peptides, and *ACTR2/3* complex, closely related to cell migration and invasion.[52] Furthermore, 2 well-accepted stem cells surface markers *CD24* and *CD44* were also increased in BON-1 cells in response to the 3 SST5TMD4-derived peptides.

Altogether, the results presented herein demonstrate, for the first time, the presence of MMPs cleavage sites in the sequence of the extracellular SST5TMD4 C-terminal tail and that the SST5TMD4 truncated variant is coexpressed in all the cell lines studied and directly correlated with the MMPs implicated in the processing of its extracellular domain. In addition, this study demonstrates that the 3 SST5TMD4-derived peptides are capable to enhance the malignant characteristics (proliferation, migration, and tumorspheres formation) of cancer cells derived from diverse tumor pathologies (ie NETs, breast, prostate, and liver cancer), and blunted the antiproliferative response to somatostatin in QGP-1 cells, likely through the activation of PI3K/AKT and/or of MEK/ERK pathways and by the modulation key pro-oncogenic genes. Indeed, these results invite to suggest the idea that the peptides derived from the SST5TMD4 extracellular domain could have important biological activities and pathological implications, for they could contribute to the strong oncogenic role of SST5TMD4 previously reported in multiple tumor pathologies. Therefore, this study suggests that the SST5TMD4-derived peptides could be potential candidates for future studies aimed to identify novel diagnostic, prognostic, and/or therapeutic tools in several cancer types.

ACKNOWLEDGEMENTS

Conflicts of Interest: All authors have read the journal's authorship agreement and have no conflicts of interest to declare.

This work was funded by Instituto de Salud Carlos III, co-funded by European Union (ERDF/ESF, "Investing in your future") [PI17/002287, PI16/00264, CP15/00156], MINECO (BFU2016-80360-R), Junta de Andalucía (BIO-0139, CTS-1406, PI-0541-2013) and CIBERobn. CIBER is an initiative of Instituto de Salud Carlos III, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain. Q3

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.trsl.2019.02.013.

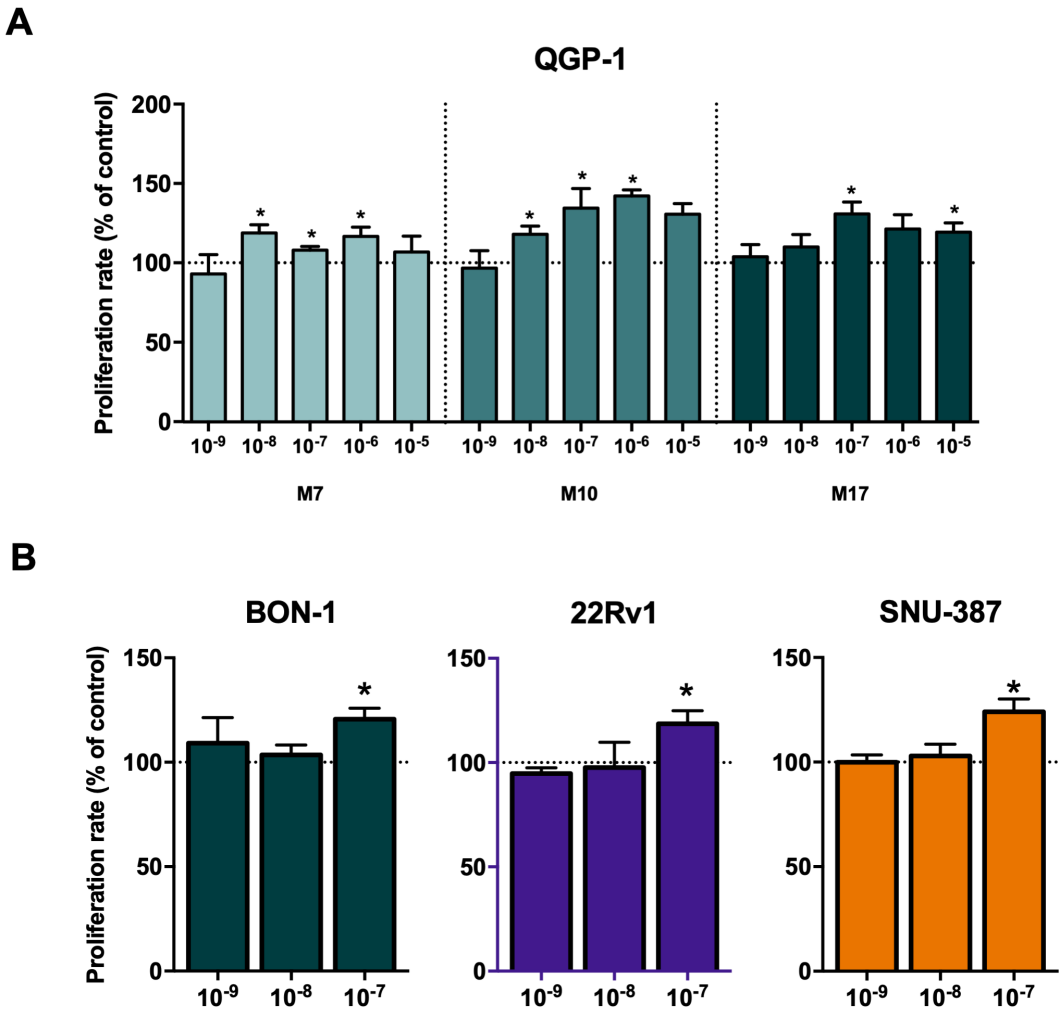
REFERENCES

1. Gahete MD, Cordoba-Chacon J, Duran-Prado M, et al. Somatostatin and its receptors from fish to mammals. Ann N. Y. Acad Sci 2010;1200:43–52.
2. Ruscica M, Arvigo M, Steffani L, Ferone D, Magni P. Somatostatin, somatostatin analogs and somatostatin receptor dynamics in the biology of cancer progression. Curr Mol Med 2013;13:555–71.
3. Theodoropoulou M, Stalla GK. Somatostatin receptors: from signaling to clinical practice. Front Neuroendocrinol 2013;34:228–52.
4. Duran-Prado M, Gahete MD, Martinez-Fuentes AJ, et al. Identification and characterization of two novel truncated but functional isoforms of the somatostatin receptor subtype 5 differentially present in pituitary tumors. J Clin Endocrinol Metab 2009;94:2634–43.
5. Günther T, T G, Dournaud P, et al. International union of basic and clinical pharmacology. somatostatin receptors: structure, function, ligands and new nomenclature. Pharmacol Rev 2018:in press. Q4
6. Chalabi M, Duluc C, Caron P, et al. Somatostatin analogs: does pharmacology impact antitumor efficacy? Trends Endocrinol Metab 2014;25:115–27.
7. Samonakis DN, Notas G, Christodoulakis N, Kouroumalis EA. Mechanisms of action and resistance of somatostatin analogues

for the treatment of hepatocellular carcinoma: a message not well taken. Dig Dis Sci 2008;53:2359–65.
8. Watt HL, Kharmate G, Kumar U. Biology of somatostatin in breast cancer. Mol Cell Endocrinol 2008;286:251–61.
9. Ji XQ, Ruan XJ, Chen H, Chen G, Li SY, Yu B. Somatostatin analogues in advanced hepatocellular carcinoma: an updated systematic review and meta-analysis of randomized controlled trials. Med Sci Monit 2011;17:RA169–76.
10. Rekasi Z, Czompoly T, Schally AV, Halmos G. Isolation and sequencing of cDNAs for splice variants of growth hormone-releasing hormone receptors from human cancers. Proc Natl Acad Sci U. S. A. 2000;97:10561–6.
11. Havt A, Schally AV, Halmos G, et al. The expression of the pituitary growth hormone-releasing hormone receptor and its splice variants in normal and neoplastic human tissues. Proc Natl Acad Sci U. S. A. 2005;102:17424–9.
12. Markovic D, Challiss RAJ. Alternative splicing of G protein-coupled receptors: physiology and pathophysiology. Cell Mol Life Sci 2009;66:3337–52.
13. Cordoba-Chacon J, Gahete MD, Duran-Prado M, Luque RM, Castano JP. Truncated somatostatin receptors as new players in somatostatin-cortistatin pathophysiology. Ann N. Y. Acad Sci. 2011;1220:6–15.
14. McWilliams DF, Watson SA, Crosbee DM, Michaeli D, Seth R. Coexpression of gastrin and gastrin receptors (CCK-B and delta CCK-B) in gastrointestinal tumour cell lines. Gut. 1998;42:795–8.
15. Hawrylyshyn KA, Michelotti GA, Coge F, Guenin SP, Schwinn DA. Update on human alpha1-adrenoceptor subtype signaling and genomic organization. Trends Pharmacol Sci 2004;25:449–55.
16. Duran-Prado M, Gahete MD, Hergueta-Redondo M, et al. The new truncated somatostatin receptor variant sst5TMD4 is associated to poor prognosis in breast cancer and increases malignancy in MCF-7 cells. Oncogene 2012;31:2049–61.
17. Puig-Domingo M, Luque RM, Reverter JL, et al. The truncated isoform of somatostatin receptor5 (sst5TMD4) is associated with poorly differentiated thyroid cancer. PloS One 2014;9:e85527.
18. Luque RM, Ibanez-Costa A, Neto LV, et al. Truncated somatostatin receptor variant sst5TMD4 confers aggressive features (proliferation, invasion and reduced octreotide response) to somatotropinomas. Cancer Lett 2015;359:299–306.
19. Gahete MD, Rincon-Fernandez D, Duran-Prado M, et al. The truncated somatostatin receptor sst5TMD4 stimulates the angiogenic process and is associated to lymphatic metastasis and disease-free survival in breast cancer patients. Oncotarget 2016;7:60110–22.
20. Sampedro-Nunez M, Luque RM, Ramos-Levi AM, et al. Presence of sst5TMD4, a truncated splice variant of the somatostatin receptor subtype 5, is associated to features of increased aggressiveness in pancreatic neuroendocrine tumors. Oncotarget 2016;7:6593–608.
21. Hormaechea-Agulla D, Jimenez-Vacas JM, Gomez-Gomez E, et al. The oncogenic role of the spliced somatostatin receptor sst5TMD4 variant in prostate cancer. FASEB J 2017;31:4682–96.
22. Duran-Prado M, Saveanu A, Luque RM, et al. A potential inhibitory role for the new truncated variant of somatostatin receptor 5, sst5TMD4, in pituitary adenomas poorly responsive to somatostatin analogs. J Clin Endocrinol Metab 2010;95:2497–502.
23. Domingo B, Gasset M, Duran-Prado M, et al. Discrimination between alternate membrane protein topologies in living cells using GFP/YFP tagging and pH exchange. Cell Mol Life Sci 2010;67:3345–54.
24. Ebben JD, Lubet RA, Gad E, Disis ML, You M. Epidermal growth factor receptor derived peptide vaccination to prevent lung adenocarcinoma formation: An in vivo study in a murine model of EGFR mutant lung cancer. Mol Carcinog 2016;55:1517–25.
25. Bifulco K, Longanesi-Cattani I, Liguori E, et al. A urokinase receptor-derived peptide inhibiting VEGF-dependent directional migration and vascular sprouting. Mol Cancer Ther 2013;12:1981–93.
26. Carriero MV, Bifulco K, Minopoli M, et al. UPARANT: a urokinase receptor-derived peptide inhibitor of VEGF-driven angiogenesis with enhanced stability and in vitro and in vivo potency. Mol Cancer Ther 2014;13:1092–104.
27. Craig SE, Brady-Kalnay SM. Tumor-derived extracellular fragments of receptor protein tyrosine phosphatases (RPTPs) as cancer molecular diagnostic tools. Anticancer Agents Med Chem 2011;11:133–40.
28. Gurevich LE. Role of matrix metalloproteinases 2 and 9 in determination of invasive potential of pancreatic tumors. Bull Exp Biol Med 2003;136:494–8.
29. Hadler-Olsen E, Winberg JO, Uhlin-Hansen L. Matrix metalloproteinases in cancer: their value as diagnostic and prognostic markers and therapeutic targets. Tumour Biol 2013;34:2041–51.
30. Evers BM, Townsend Jr. CM, Upp JR, et al. Establishment and characterization of a human carcinoid in nude mice and effect of various agents on tumor growth. Gastroenterology 1991;101:303–11.
31. Kaku M, Nishiyama T, Yagawa K, Abe M. Establishment of a carcinoembryonic antigen-producing cell line from human pancreatic carcinoma. Gan 1980;71:596–601.
32. Hormaechea-Agulla D, Gahete MD, Jimenez-Vacas JM, et al. The oncogenic role of the In1-ghrelin splicing variant in prostate cancer aggressiveness. Mol Cancer 2017;16:146.
33. Rincon-Fernandez D, Culler MD, Tsomaia N, et al. In1-ghrelin splicing variant is associated with reduced disease-free survival of breast cancer patients and increases malignancy of breast cancer cells lines. Carcinogenesis 2018;39:447–57.
34. van Hoek M, Hofland LJ, de Rijke YB, et al. Effects of somatostatin analogs on a growth hormone-releasing hormone secreting bronchial carcinoid, in vivo and in vitro studies. J Clin Endocrinol Metab 2009;94:428–33.
35. Ibanez-Costa A, Rivero-Cortes E, Vazquez-Borrego MC, et al. Octreotide and pasireotide (dis)similarly inhibit pituitary tumor cells in vitro. J Endocrinol. 2016;231:135–45.
36. Luque RM, Sampedro-Nunez M, Gahete MD, et al. In1-ghrelin, a splice variant of ghrelin gene, is associated with the evolution and aggressiveness of human neuroendocrine tumors: Evidence from clinical, cellular and molecular parameters. Oncotarget 2015;6:19619–33.
37. Taboada GF, Luque RM, Bastos W, et al. Quantitative analysis of somatostatin receptor subtype (SSTR1-5) gene expression levels in somatotropinomas and non-functioning pituitary adenomas. Eur J Endocrinol 2007;156:65–74.
38. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034. Q5
39. Burgoyne AM, Phillips-Mason PJ, Burden-Gulley SM, et al. Proteolytic cleavage of protein tyrosine phosphatase mu regulates glioblastoma cell migration. Cancer Res 2009;69:6960–8.
40. Shuman Moss LA, Jensen-Taubman S, Stetler-Stevenson WG. Matrix metalloproteinases: changing roles in tumor progression and metastasis. Am J Pathol 2012;181:1895–9.
41. Salem N, Kamal I, Al-Maghrabi J, et al. High expression of matrix metalloproteinases: MMP-2 and MMP-9 predicts poor survival outcome in colorectal carcinoma. Future Oncol 2016;12(3):323–31.

42. Turunen SP, Tatti-Bugaeva O, Lehti K. Membrane-type matrix metalloproteases as diverse effectors of cancer progression. Biochim Biophys Acta 2017;1864:1974–88.
43. Itoh Y. Membrane-type matrix metalloproteinases: their functions and regulations. Matrix Biol 2015;44−46:207–23.
44. Vandamme T, Peeters M, Dogan F, et al. Whole-exome characterization of pancreatic neuroendocrine tumor cell lines BON-1 and QGP-1. J Mol Endocrinol 2015;54:137–47.
45. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research. Breast Cancer Res 2011;13:215.
46. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). Invest Urol 1979;17:16–23.
47. Sramkoski RM, Pretlow II TG, Giaconia JM, et al. A new human prostate carcinoma cell line, 22Rv1. In Vitro Cell Dev Biol Anim 1999;35:403–9.
48. Lopez-Bergami P, Huang C, Goydos JS, et al. Rewired ERK-JNK signaling pathways in melanoma. Cancer Cell 2007;11:447–60.
49. McCubrey JA, Steelman LS, Chappell WH, et al. Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. Biochim Biophys Acta 2007;1773:1263–84.
50. De Luca A, Maiello MR, D'Alessio A, Pergameno M, Normanno N. The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. Expert Opin Ther Targets 2012;16(Suppl 2): S17–27.
51. Jernman J, Valimaki MJ, Louhimo J, Haglund C, Arola J. The novel WHO 2010 classification for gastrointestinal neuroendocrine tumours correlates well with the metastatic potential of rectal neuroendocrine tumours. Neuroendocrinology 2012;95:317–24.
52. Molinie N, Gautreau A. The Arp2/3 regulatory system and its deregulation in cancer. Physiol Rev 2018;98:215–38.

Supplementary Figure 1



Supplementary Figure 1. A) Dose-response of SST5TMD4-derived peptides M7, M10 and M17 (10^{-9}M to 10^{-5}M) on QGP-1 cell proliferation (24 h). B) Dose-response of M17 (10^{-9}M to 10^{-7}M) on SNU-387, 22Rv1 and BON-1 cell proliferation (24 h). The data are expressed as percentage of proliferation rate compared to vehicle-treated control cells (set at 100%). Values represent the mean ± SEM of 3 experiments. Only one experiment was done with SNU-387 cells. In each experiment cells were seeded per quadruplicate per treatment). Asterisks indicate values that significantly differ from control cells (t-test: *, $p<0.05$).

Supplementary Table 1

Gene	Accession Number	Primer Sequence (Sense)	Primer Sequence (Antisense)	Product Size
ACTB	NM_001101	ACTCTTCCAGCCTTCCTTCCT	CAGTGATCTCCTTCTGCATCCT	176
ACTR2	AF006082.1	ACAACTTTTGGATGACCCGACAA	ACCTTTCCAGTCAAAGGGCAGAG	200
ACTR3	AF006083.1	CCAATCCGCCATGGTATAGTTGA	CACAGCAATGTACAAGCCTGGAA	200
ARC	AF006088.1	TGGGTGAATACCACTGCCAAGTT	GATGTTCATGCCCAACAAGCTCT	205
CCND3	NC_000006.12	GCTACGTACCCCGCGCCTC	GCAGACCGCACCCAGGAGC	209
CD24	NC_000006.12	TGAAGAACATGTGAGAGGTTTGAC	GAAAACTGAATCTCCATTCCACAA	208
CD44	NC_000011.10	CAATAGCACCTTGCCCACAAT	AATCACCACGTGCCCTTCTATG	97
GAPDH	NM_002046	AATCCCATCACCATCTTCCA	AAATGAGCCCCAGCCTTC	122
HPRT	NM_000194.2	CTGAGGATTTGGAAAGGGTGT	TAATCCAGCAGGTCAGCAAAG	157
MKI67	NM_002417.4	GACATCCGTATCCAGCTTCCT	GCCGTACAGGCTCATCAATAAC	139
MMP16	NM_005941.4	ATAGCGACGGGAATTTTGTG	TCCACTTCCAAGGGTTATCAAG	107
MMP14	NM_004995.3	GGAATAACCAAGTGATGGATGG	CCCAATGCTTGTCTCCTTTG	132
MMP2	NC_000016.10	CTACGATGGAGGCGCTAATG	ACTCTTTGTCCGTTTTGGGG	150
MMP9	NC_000020.11	CAGTGCCATGTAAATCCCCA	CACCTCCACTCCTCCCTTTC	102
SST5	NM_001053	CTGGTGTTTGCGGGATGTT	GAAGCTCTGGCGGAAGTTGT	183
SST5TMD4	DQ448304	TACCTGCAACCGTCTGCC	AGCCTGGGCCTTTCTCCT	98

Supplemetary Table 1. Specific primers for human transcripts used in this study. NCBI accesion number, primers sequences and expected product sizes for the genes studied are included.