DEPARTAMENTO DE BIOLOGÍA CELULAR, FISIOLOGÍA E INMUNOLOGÍA

Novel molecular insights on the role of splicing dysregulation in cancer.

Nuevos conocimientos moleculares sobre el papel de la desregulación del proceso de splicing en cáncer.

Emilia M Alors Pérez

Córdoba, 12 mayo 2022

TITULO: *Novel molecular insights on the role of splicing dysregulation in cancer*

AUTOR: *Emilia M. Alors Pérez*

© Edita: UCOPress. 2022
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

Novel molecular insights on the role of splicing dysregulation in cancer.

Nuevos conocimientos moleculares sobre el papel de la desregulación del proceso de splicing en cáncer.

Memoria de Tesis Doctoral presentada por **Emilia M. Alors Pérez**, Graduada en Biología, para optar al grado de **Doctora en Ciencias**.

Los Directores,

Dr. Justo P. Castaño Fuentes
Catedrático de Biología Celular
Universidad de Córdoba
IMIBIC

Dr. Raúl M Luque Huertas
Catedrático de Biología Celular
Universidad de Córdoba
IMIBIC

En Córdoba, a 12 de mayo de 2022

D. Justo P. Castaño Fuentes y D. Raúl M. Luque Huertas, Catedráticos del Departamento de Biología Celular, Fisiología e Inmunología de la Universidad de Córdoba,

INFORMAN

Que D. Emilia M Alors Pérez, Graduada en Biología, ha realizado bajo nuestra dirección el trabajo titulado **"Novel molecular insights on the role of splicing dysregulation in cancer"** 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 12 de mayo de 2022

CASTAÑO FUENTES JUSTO PASTOR - 30508275V

Firmado digitalmente por CASTAÑO FUENTES JUSTO PASTOR - 30508275V
Fecha: 2022.05.12 17:00:15 +02'00'

Fdo: D. **Justo P. Castaño Fuentes**

LUQUE HUERTAS RAUL MIGUEL - 30813254Q

Firmado digitalmente por LUQUE HUERTAS RAUL MIGUEL - 30813254Q
Fecha: 2022.05.12 21:24:56 +02'00'

Fdo.: D. **Raúl M. Luque Huertas**

TITULO DE LA TESIS: Nuevos conocimientos moleculares sobre el papel de la desregulación del proceso de splicing en cáncer.

DOCTORANDO: Emilia M Alors Pérez

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

Durante el desarrollo de la presente Tesis Doctoral, la doctoranda Emilia M Alors Pérez ha alcanzado e incluso superado notablemente los objetivos planteados al comienzo de la misma, al tiempo que ha desarrollado técnicas experimentales de gran utilidad para el grupo de investigación, todo lo cual le ha permitido obtener resultados muy relevantes en el estudio de los procesos de splicing alternativo asociados a diferentes patologías tumorales. Concretamente, como fruto de su trabajo durante este periodo, ha publicado un trabajo directamente relacionado con su tesis doctoral, en la revista Journal of Experimental and Clinical Cancer Research, 40(1):382, 2021; doi: 10.1186/s13046-021-02153-9, de gran relevancia dentro de nuestra área de investigación, situada en el Decil 1 del área de Oncología (25/313), factor de impacto: 11.161, cumpliendo así el criterio de calidad del programa de doctorado en Biomedicina de la Universidad de Córdoba. Asimismo, ha elaborado otros tres artículos originales que se están sometiendo actualmente para su publicación y se presentan como capítulos de esta tesis, así como una revisión bibliográfica sobre splicing y cáncer de páncreas, también preparada para ser sometida para su publicación. Además, la doctoranda ha realizado una estancia de 3 meses en la Universidad de Toulouse, Francia, donde amplió y perfeccionó los métodos de trabajo, cumpliendo así el criterio para optar a la mención internacional. Por último, la doctoranda ha presentado sus resultados en diferentes congresos de ámbito nacional e internacional, de los que han derivado varios capítulos de libro.
Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 12 de mayo de 2022.

Firma del/de los director/es

CASTAÑO FUENTES JUSTO PASTOR - 30508275V
Firmado digitalmente por CASTAÑO FUENTES JUSTO PASTOR - 30508275V
Fecha: 2022.05.12 16:59:25 +02'00'

LUQUE HUERTAS RAUL MIGUEL - 30813254Q
Firmado digitalmente por LUQUE HUERTAS RAUL MIGUEL - 30813254Q
Fecha: 2022.05.12 21:25:18 +02'00'

Fdo.: **D. Justo P. Castaño Fuentes**

Fdo.: **D. Raúl M. Luque Huertas**

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. Justo P. Castaño Fuentes y Raúl M Luque Huertas.

Los estudios realizados han sido subvencionados mediante los proyectos PI16/00264 (Instituto de Salud Carlos III, cofinanciado por la Unión Europea; ERDF/ESF, "Investing in your future"), BFU2016-80360-R, PID2019-105564RB-I00 y PID2019-105201RB-I00 (Ministerio de Ciencia e Innovación; anteriores MICIU y MINECO), Junta de Andalucía (BIO-0139) y CIBERobn. CIBER es una iniciativa del Instituto de Salud Carlos III. Por su parte, la financiación para el contrato de la doctoranda ha sido facilitado mediante una ayuda de formación en investigación en salud, pFIS (FI17/00282, asociada al proyecto PI16/00264). Durante el transcurso de la presente Tesis Doctoral, la doctoranda ha realizado dos estancias de tres meses de duración cada una de ellas. Una de estas estancias la realizó en un centro internacional, el "Centre de Recherches en Cancérologie de Toulouse (Francia)" bajo la supervisión de la Profesora Corinne Bousquet (noviembre 2019-febrero 2020), financiada por la ayuda complementaria para estancias breves de la Universidad de Córdoba, justificando la obtención de la Mención Internacional en el Título de Doctora de la Universidad de Córdoba.

List of abbreviations

ADH: Adherent

ADT: Androgen Deprivation Therapy

AR: Androgen receptor

AR-v7: Androgen receptor variant 7

ASO: Antisense Oligonucleotide

AUC: Area under the curve

BMI: Body mass index

BP: Binding Protein

BPE: Bovine Pituitary Extract

cDNA: complementary DNA

CgA: Chromogranin A

CHT: Caudal Hematopoietic Tissue

CRPC: Castration-Resistant PCa

CSC: Cancer Stem Cells

CT: Computerized Tomography

DAPI: 4′,6-diamidino-2-phenylindole

DM: Diabetes Mellitus

DMSO: Dimethyl sulfoxide

DPI: Days Post Injection

EGF: Epidermal Growth Factor

EJC: Exon Junction Complex

ES: Exon Skipping

ESE: Exonic Splicing Enhancer

FBS: Foetal Bovine Serum

FFPE: Formalin-Fixed Paraffin-Embedded

GEP-NETs: Gastroenteropancreatic NETs

GS: Gleason Score

GSEA: Gene Set Enrichment Analysis

H&E: Haematoxylin-Eosin stained

HNRNP: Heterogenous Nuclear Ribonucleoproteins

HS: Horse Serum

IHC: Immunohistochemistry

IPMN: Intraductal Papillary Mucinous Neoplasm

IR: Intron Retention

ISE: Intronic Splicing Enhancer

ISS: Intronic Splicing Silencer

ISUP: International Society of Urological Pathology

ITPN: Intraductal Tubulopapillary Neoplasm

MAPK: Mitogen-Activated Protein Kinase

MCN: Mucinous Cystic Neoplasm

MRI: Magnetic Resonance Imaging

mTOR: Mammalian target of rapamycin

NETs: Neuroendocrine Tumors

NIH: National Institutes of Health

NonSigPCa: Non-significant Prostate Cancer

NTAT: Non-tumor Adjacent Tissue

PanIN: Pancreatic Intraepithelial Neoplasia

PanNETs: Pancreatic NETs

PBS: Phosphate Buffered Saline

PCa: Prostate Cancer

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

Pd: Pladienolide-B

PDAC: Pancreatic Ductal Adenocarcinoma

PET: Positron Emission Tomography

PFA: Paraformaldehyde

PI3K: Phosphatidylinositol 3-Kinase

PSA: Prostate-Specific Antigen

PSI: Position-Specific Iterative

qPCR: quantitative Polymerase Chain Reaction

RBPs: RNA-Binding Proteins

ROC: Receiver Operating Characteristic

RRM: RNA binding domain

RT: Room Temperature

SCLC: Small Cell Lung Cancer

SE: Splicing Enhancers

SEM: Standard Error of The Mean

SEOM: Sociedad Española de Oncología Médica

SF3B1: Splicing Factor 3B Subunit 1

SFs: Splicing Factors

SigPCa: Significant Prostate Cancer

siRNA: small interference RNA

snRNP: small nuclear RNP

SPH: Sphere

sPLSDA: Sparse Partial Least Squares Discriminant Analysis

SR: Serine/arginine-rich

SRPKs: SR protein kinases

SRSF: Serine/arginine-rich splicing factors

SS: Splicing Silencers

SSOs: Splice-switching antisense oligonucleotides

T2DM: Type 2 diabetes mellitus

TNM: Tumor, Node and Metastasis staging

TPM: Transcripts Per Million

WHO: World Health Organization

Table of contents

Index of tables and figures

Index of Tables

Index of Figures

Figures of Introduction

Figures of Results

Appendix

Resumen

El cáncer y las enfermedades tumorales representan uno de los problemas de salud pública más graves y complicados, siendo una de las principales causas de muerte en el mundo. La notable variedad entre los tipos de cáncer y la heterogeneidad entre los pacientes con la misma patología tumoral es una de las limitaciones más relevantes en la investigación del cáncer. Esta heterogeneidad es el resultado de la interacción de múltiples elementos que residen en diferentes niveles, desde la localización anatómica del tumor hasta las características clinicopatológicas, la funcionalidad, los marcadores celulares-moleculares y las alteraciones genético-epigenéticas. Aunque los recientes avances en la caracterización molecular de diferentes patologías tumorales han permitido mejorar su taxonomía y explorar sus bases genéticas, el impacto traslacional de estos nuevos conocimientos sigue siendo limitado. Este es el caso de los tres tipos distintos de cáncer que han centrado la atención de la presente Tesis Doctoral: los tumores neuroendocrinos pancreáticos (PanNETs) y dos tipos de adenocarcinomas, el adenocarcinoma ductal pancreático (PDAC) y el cáncer de próstata (PCa).

Los tumores neuroendocrinos (NETs), a pesar de ser considerados una enfermedad rara, constituyen una patología emergente y única, con una incidencia creciente en todo el mundo. Los NETs se originan en células que comparten características de células neurales y endocrinas, y pueden encontrarse prácticamente en cualquier órgano, surgiendo con mayor frecuencia en el páncreas (PanNETs), donde suelen aparecer como tumores de bajo grado, en estadios localizados. En general, los PanNETs se detectan en un estadio avanzado, ya que la falta de marcadores precisos y de síntomas clínicos específicos complican el diagnóstico precoz, lo que conlleva tiempos de diagnóstico de entre cinco y siete años, dificultando la pronta aplicación de terapias eficaces y específicas.

El PDAC es un cáncer muy agresivo, con alta invasión perineural y vascular, y metástasis en estadios tempranos, lo que hace que la tasa de supervivencia de los pacientes sea muy baja. El mal pronóstico de este tipo de cáncer se debe en parte a la falta de biomarcadores específicos y al consiguiente diagnóstico tardío y difícil, pero también a su elevada resistencia a los tratamientos disponibles, lo que se traduce en un difícil manejo. Por ello, es necesario identificar nuevas herramientas moleculares que puedan proporcionar biomarcadores más sensibles y precisos para el diagnóstico precoz y la predicción del pronóstico, así como nuevas dianas terapéuticas y tratamientos eficaces de este cáncer.

El PCa es el segundo cáncer más frecuentemente diagnosticado y la quinta causa de muerte relacionada con el cáncer entre los hombres. Una limitación en el tratamiento del PCa es que el diagnóstico precoz del PCa se basa principalmente en los niveles plasmáticos de PSA, un biomarcador que presenta muchas deficiencias, entre las que destacan su poca sensibilidad, especificidad y valor predictivo. Esto conduce a la realización de biopsias innecesarias y la reducción de la calidad de vida del paciente. Asimismo, el tratamiento clínico del PCa agresivo (es decir, el PCa metastásico y resistente a la castración o CRPC) también se enfrenta a importantes limitaciones, como la falta de respuesta de los pacientes y el desarrollo de resistencia a las terapias hormonales y químicas. Por tanto, existe aún una importante necesidad clínica por lograr: obtener nuevos biomarcadores, fiables y específicos, para el diagnóstico temprano, la predicción del pronóstico, la selección del tratamiento, etc., lo cual mejoraría la calidad de vida de los pacientes.

Estas patologías tumorales constituyen ejemplos típicos de la complejidad del cáncer y de cómo los inicialmente escasos conocimientos moleculares han crecido hasta proporcionar herramientas útiles. De hecho, los tratamientos dirigidos molecularmente

han tenido un notable éxito clínico en el tratamiento de una amplia gama de cánceres, como el PanNET, el PDAC y el PCa. Sin embargo, los conocimientos y los tratamientos disponibles siguen siendo claramente insuficientes para abordar estas enfermedades y es necesario explorar nuevos mecanismos moleculares para encontrar herramientas adicionales. En este escenario, la alteración del splicing alternativo está emergiendo como un nuevo *hallmark* transversal del cáncer, que impregna todos los *hallmark* establecidos y proporciona atractivas dianas terapéuticas.

El splicing es un proceso esencial en la compleja regulación de la expresión génica, ya que permite el procesamiento adecuado del ARN y, por tanto, mantiene el correcto desarrollo y la homeostasis del individuo. El spliceosoma, la maquinaria molecular que lleva a cabo y regula el proceso de splicing, está compuesto por un conjunto discreto de ribonucleoproteínas que interactúan estrechamente con un gran grupo de proteínas, los factores de splicing. El proceso de splicing y sus sistemas reguladores asociados son puntos clave en el control de funciones celulares esenciales, y su alteración puede tener consecuencias patológicas. La importancia fisiopatológica primordial del splicing alternativo y sus procesos asociados está respaldada por la evidencia emergente que vincula diversas anomalías del splicing (elementos spliceosómicos mutados o desregulados: componentes del spliceosoma, factores de splicing o variantes de splicing) con múltiples patologías, desde enfermedades raras hasta cánceres comunes, en las que se ha demostrado la función patogénica de los elementos spliceosómicos y su valor como objetivos terapéuticos. En particular, el creciente número y la relevancia de los estudios que vinculan los defectos en los mecanismos relacionados con el splicing y el cáncer son abrumadores, apoyando inequívocamente su papel clave en la oncogénesis y la agresividad tumoral. De hecho, se ha propuesto la alteración del proceso de splicing como un nuevo *hallmark* transversal del cáncer, ya que aparece en

todos los cánceres estudiados hasta la fecha e influye e interactúa con cada uno de los *hallmarks* distintivos del cáncer.

Sobre la base de toda la información mencionada anteriormente, el objetivo general de esta Tesis Doctoral fue explorar la potencial desregulación de los componentes de la maquinaria de *splicing* en el cáncer, con el propósito final de descubrir nuevos biomarcadores y dianas farmacológicas con potencial para mejorar los enfoques diagnósticos y terapéuticos en estas patologías tumorales. Para lograr este objetivo, hemos dividido esta Tesis en tres secciones experimentales.

La primera sección de esta tesis doctoral se centró en los tumores neuroendocrinos pancreáticos (PanNETs), donde nuestro anterior trabajo desveló las propiedades oncogénicas de las variantes de splicing. Nuestro objetivo era evaluar la desregulación y el papel funcional del factor de splicing CELF4 en los PanNETs, así como evaluar su potencial papel como nuevo marcador de diagnóstico y objetivo de tratamiento en esta patología. Se evaluó la expresión de *CELF4* en una cohorte de 20 pacientes con PanNET, comparando el tejido tumoral y el adyacente no tumoral, lo que reveló una marcada sobreexpresión, que se confirmó mediante un análisis biocomputacional en un conjunto de datos de RNA-Seq, en el que investigamos más a fondo CELF4, explorando sus relaciones con las características clínicas, la expresión génica y los perfiles de eventos de splicing. Además, se emplearon dos líneas celulares modelo de PanNET, BON-1 y QGP-1, para evaluar la función de CELF4 *in vitro*, incluyendo un array de fosforilación de proteínas de la ruta mTOR, e *in vivo* en ratones xenógrafos con la línea celular BON-1. La sobreexpresión de *CELF4* se asoció estrechamente con características relevantes de malignidad, expresión específica de actores tumorales clave (por ejemplo, *TP53*) y perfiles distintos de eventos de splicing. En consecuencia, se observaron desregulaciones funcionales tras la modulación de *CELF4* en las líneas

celulares de PanNET en términos de proliferación *in vitro*, y su silenciamiento *in vivo* también redujo el crecimiento tumoral del xenoinjerto de BON-1. Curiosamente, la modulación de la expresión de *CELF4* en las células PanNET influyó en su respuesta al inhibidor de mTOR everolimus, una vía cuyos intermediarios se vieron alterados de manera notable tras el silenciamiento de CELF4.

La segunda sección se dividió en dos partes, con el objetivo de explorar diferentes aspectos de la desregulación de la maquinaria de splicing en dos adenocarcinomas distintos pero relacionados. En primer lugar, nos propusimos determinar el estado de la maquinaria de splicing y su relación con las características clínicas y moleculares de la agresividad del PDAC. Para ello, medimos la expresión de 45 componentes seleccionados de la maquinaria de splicing en tejidos tumorales frente a sus tejidos adyacente no tumorales de un conjunto de 79 muestras de PDAC fijadas en formol e incluidas en parafina. Los análisis de la supervivencia de los pacientes y los parámetros clínicos en relación con la expresión de estos elementos de splicing condujeron a la identificación de un subconjunto de 2 factores de splicing desregulados, PRPF8 y RBMX, cuya expresión estaba estrechamente relacionada con el mal pronóstico y los parámetros de malignidad, e incluso asociada con la presencia de mutaciones en genes clave implicados en el desarrollo y la progresión del PDAC. Estos resultados se confirmaron en cohortes independientes de bases de datos públicas. La modulación experimental de estos factores de splicing en líneas celulares de PDAC (Capan-2, y BxPC-3) revirtió su expresión a niveles no tumorales, y dio lugar a una disminución de las características clave relacionadas con el tumor, incluyendo la proliferación celular, la migración y la formación de colonias. En el segundo estudio, investigamos el factor de splicing SRSF2 en el adenocarcinoma de próstata (PCa) y el adenocarcinoma de páncreas (PDAC), como potencial biomarcador de diagnóstico y diana terapéutica. Para

ello, medimos los niveles de SRSF2 en muestras de tumores humanos de PCa y PDAC y en líneas celulares, y modificamos su expresión para evaluar diferentes criterios funcionales. Curiosamente, *SRSF2* estaba sobreexpresado en las muestras tumorales con respecto al tejido no tumoral tanto en el PCa como en el PDAC, donde los niveles de *SRSF2* se relacionaron de forma similar con las características clínico-moleculares que sugieren un potencial peor pronóstico y capacidad oncogénica. En relación con lo anterior, el silenciamiento de SRSF2 en líneas celulares modelo representativas de ambos adenocarcinomas inhibió la migración celular, la proliferación, la invasión y la formación de colonias en las líneas celulares de PCa, mientras que sólo causó una reducción igualmente pronunciada en la formación de colonias en las líneas celulares de PDAC. Estos hallazgos sugieren la interesante posibilidad de que SRSF2 pueda estar implicado en la regulación de la capacidad de iniciación tumoral, que está relacionada con las células madre del cáncer (CSC) dentro de la población celular.

La última sección de esta Tesis se centró en la exploración de la presencia y el papel funcional del factor central de splicing SF3B1 en PDAC y se interrogó su potencial como un objetivo accionable. Se analizó SF3B1 en los tejidos de PDAC, en un conjunto de datos de RNA-Seq y en bases de datos disponibles públicamente, examinando las asociaciones con las alteraciones de splicing y las características/genes clave. Los ensayos funcionales en líneas celulares de PDAC y CSCs derivadas de PDX sirvieron para probar los efectos del tratamiento con Pladienolide-B en líneas celulares modelo *in vitro*, e *in vivo* en dos modelos preclínicos (pez cebra y ratón). Descubrimos que *SF3B1* se sobreexpresa en PDAC humano, donde se correlacionó con el grado del tumor y la afectación de los ganglios linfáticos. Los niveles de *SF3B1* se asociaron estrechamente con distintos perfiles de eventos de splicing y con la expresión de genes clave del PDAC (*KRAS, TP53*). En las células de PDAC, Pladienolide-B aumentó la apoptosis y

disminuyó múltiples características relacionadas con el tumor, incluyendo la proliferación celular, la migración y la formación de colonias/esferas, alterando la señalización de AKT y JNK, y favoreciendo la expresión de las variantes de *splicing* proapoptóticas (*BCL-XS/BCL-XL, KRASa/KRAS, Δ133TP53/TP53).* Es importante destacar que el Pladienolide-B perjudicó de forma similar a las CSC, reduciendo su capacidad de crecimiento y aumentando su sensibilidad a la quimioterapia. El Pladienolide-B también redujo el crecimiento tumoral del xenoinjerto de las líneas modelo de PDAC/CSCs *in vivo* en el pez cebra y en ratones. Así pues, podemos concluir que, la sobreexpresión de SF3B1 representa una vulnerabilidad terapéutica en el PDAC que permite dirigir el splicing con Pladienolide-B no sólo en las células cancerosas sino también en las CSCs.

Por todo lo anterior, las principales conclusiones de esta Tesis son:

1. El factor de splicing *CELF4* se sobreexpresa en los PanNETs, donde sus niveles se asocian con características de malignidad y perfiles de splicing distintos. La modulación de los niveles de CELF4 influye de manera predecible en múltiples características oncológicas *in vitro* en líneas celulares de PanNETs, y su silenciamiento inhibe el crecimiento tumoral de los xenoinjertos. La expresión de CELF4 influye en las vías y mediadores vinculados a la vía mTOR, lo que probablemente explique cómo perjudica la respuesta de las células de PanNETs al tratamiento con everolimus.
2. La maquinaria de splicing está severamente desregulada en el PDAC, donde identificamos dos componentes específicos, PRPF8 y RBMX, que muestran una expresión disminuida que está estrechamente relacionada con una peor supervivencia y con marcadores clínicos y moleculares de mal pronóstico. La expresión de *PRPF8* y *RBMX* se asocia claramente con perfiles de splicing alterados, y la restauración de sus niveles de expresión rescató su capacidad supresora de tumores *in vitro* en modelos celulares de PDAC. Estos factores representan por tanto dos dianas prometedoras que

merecen una mayor investigación como nuevos biomarcadores potenciales y como instrumentos moleculares para combatir el PDAC.

3. Dos adenocarcinomas distintos, el PCa y el PDAC, comparten una sobreexpresión de SRSF2 previamente desconocida, que está vinculada a características clínico-moleculares que sugieren un potencial pronóstico y capacidad oncogénica. El silenciamiento de SRSF2 influyó de forma diferencial en las características funcionales del tumor indicativas de agresividad tumoral en el PCa (más sensible) y en el PDAC (menos sensible). Por el contrario, el silenciamiento de SRSF2 redujo de forma similar la formación de colonias en modelos celulares de ambos cánceres, lo que sugiere un posible papel de este factor en el control de la capacidad de iniciación tumoral por parte de las células cancerosas y/o las células madre del cáncer.
4. SF3B1 está sobreexpresado en PDAC, donde sus niveles se asocian con características clínicas, histológicas y moleculares clave. Además, la inhibición de la actividad de SF3B1 con Pladienolide-B reduce múltiples características del cáncer en las células de PDAC al alterar las vías de señalización y los eventos de splicing relevantes. Además, el tratamiento con Pladienolide-B reduce la capacidad de crecimiento de las CSCs, haciéndolas más sensibles al tratamiento con quimioterapia.

Corolario global

En conjunto, los estudios desarrollados en la presente Tesis aportan nuevas pruebas para avanzar en el conocimiento molecular del papel de la maquinaria de splicing y sus alteraciones en diferentes cánceres. En particular, identificamos factores de splicing específicos que están alterados en PanNETs, PDAC y PCa, y que parecen desempeñar un papel funcional relevante en estos tumores, donde podrían servir como herramientas útiles para el desarrollo de nuevos biomarcadores y podrían ser la diana de nuevos fármacos dirigidos al splicing, como Pladienolide-B y/o sus derivados. Estos

hallazgos refuerzan la relevancia de examinar el splicing alternativo, sus elementos y alteraciones para abrir nuevas vías para la medicina de precisión en cánceres sólidos.

Summary

Cancer and tumoral diseases represent one of most serious and complicated public health problems, being one of the leading causes of death in the world. The remarkable variety across cancer types, and the heterogeneity among patients with the same tumoral pathology is one of the most relevant limitations in cancer research. This heterogeneity results from the interaction of multiple elements that reside at different levels, from anatomical tumor location to clinicopathological features, functionality, cellular-molecular markers, and genetic-epigenetic alterations. Although recent advances in the molecular characterization of different tumoral pathologies has enabled to improve their taxonomy and explore their genetic basis, the translational impact of this novel knowledge is still limited. This is the case in the three distinct types of cancer that have focused the attention of the present Doctoral Thesis: pancreatic neuroendocrine tumors (PanNETs) and two kinds of adenocarcinomas, pancreatic ductal adenocarcinoma (PDAC) and prostate cancer (PCa).

Neuroendocrine tumors (NETs), despite being considered a rare disease, comprise an emerging and unique pathology, with increasing incidence worldwide. NETs originate in cells that share characteristics of neural and endocrine cells and may be found virtually in any organ, arising more frequently in the pancreas (PanNETs), where they tend to appear as low-grade tumors, in localized stages. PanNETs are often detected in an advanced stage, as the lack of precise markers and specific clinical symptoms complicate early diagnosis, leading to diagnostic times between five and seven years, which hinders the prompt application of effective and specific therapies.

PDAC is a highly aggressive cancer, with high perineural and vascular invasion and metastasis in early stages, which makes the survival rate of patients very low. The poor prognosis of this type of cancer is partly due to the lack of specific biomarkers and the consequent late and difficult diagnosis, but also to its high resistance to available

treatments, thus resulting in a difficult management. Accordingly, it is necessary to identify new molecular tools that can provide more sensitive and accurate biomarkers for the early diagnostic and prognostic prediction, as well as novel therapeutic targets and effective treatments of this cancer.

PCa is the second most frequently diagnosed cancer and the fifth leading cause of cancer-related death among men. A limitation in PCa management is that early PCa diagnosis is mainly based on plasma PSA levels, a biomarker that has severe drawbacks including poor sensitivity, specificity, and predictive value. This leads to unnecessary biopsies and reduced patient quality of life. Likewise, clinical management of aggressive PCa (i.e., the metastatic and castration resistant PCa or CRPC) also faces major limitations, including unresponsive patients and development of resistance to hormonal and chemical therapies. Therefore, there is an important unmet clinical need for new, reliable, and specific biomarkers for early diagnosis, prediction of disease prognosis, selection for treatment, etc., which would improve life quality of patients.

These tumoral pathologies comprise typical examples of cancer complexity and how the initially scarce molecular knowledge has grown to provide useful tools. In fact, molecularly targeted treatments have been a notable clinical success in treating a wide range of cancers, including PanNETs, PDAC and PCa. However, the knowledge and the treatments available are still clearly insufficient to tackle these diseases and there is a need to explore new molecular mechanisms to find additional tools. In this scenario, alteration of alternative splicing is emerging as a novel transversal hallmark of cancer, that pervades all the established cancer hallmarks and provides attractive therapeutic targets.

Splicing is an essential process in the complex regulation of gene expression, as it enables the appropriate processing of RNA and thereby maintains the proper

development and homeostasis of the individual. The spliceosome, which carries out and regulates the splicing process, is composed of a discrete set of ribonucleoproteins that interact closely with a large group of proteins, the splicing factors. Splicing process and its associated regulatory systems are key players in the control of essential cellular functions, and their derangement can bear pathological consequences. The prime pathophysiological importance of alternative splicing and its associated processes is supported by the emerging evidence linking diverse splicing anomalies (mutated or dysregulated spliceosomic players: spliceosome components, splicing factors or splice variants) with multiple pathologies, from rare diseases to common cancers, wherein pathogenic function of spliceosomic players and their value as therapeutic targets has been proven. In particular, the rising number and relevance of studies linking defects in splicing-related mechanisms and cancer is overwhelming, unequivocally supporting their key role in oncogenesis and tumor aggressiveness. In fact, an altered splicing process has been proposed as a new, transversal hallmark of cancer as it appears in all cancers studied to date and influences and interacts with each of the hallmarks of cancer.

Based on all the information mentioned above, the general aim of this Doctoral Thesis was to explore the potential dysregulation of splicing machinery components in cancer, with the final purpose of discovering novel biomarkers and pharmacologic targets with potential to improve the diagnostic and therapeutic approaches in these tumoral pathologies. To achieve this objective, we divided this Thesis in three experimental sections.

The first section of this Doctoral thesis was focused on Pancreatic neuroendocrine tumors (PanNETs), where our earlier work unveiled oncogenic properties of splicing variants. We aimed to evaluate the dysregulation and functional role of the splicing factor *CELF4* in PanNETs as well as to assess its potential role as a novel diagnostic marker

and treatment target in this pathology. Expression of *CELF4* was evaluated in a cohort of 20 PanNET patients, comparing tumor and non-tumoral adjacent tissue, revealed a marked overexpression, which was confirmed through biocomputational analysis in a RNA-Seq dataset, where we further investigated CELF4, exploring its interrelations with clinical features, gene expression, and splicing event profiles. Moreover, two PanNET model cell lines, BON-1 and QGP-1, were employed to assess CELF4 function *in vitro*, including a detailed mTOR phospho-antibody array, and *in vivo* in BON-1-xenografted mice. Upregulated *CELF4* was closely associated with relevant malignancy features, specific expression of key tumor players (e.g., TP53), and distinct splicing event profiles. Accordingly, functional consequences of *CELF4* modulation in PanNET cell lines were observed in terms of proliferation *in vitro,* and its silencing *in vivo* also reduced BON-1 xenograft tumor growth. Interestingly, modulating *CELF4* expression in PanNET cells influenced their response to the mTOR inhibitor everolimus, a pathway whose intermediaries were strikingly altered under CELF4 silencing.

The second section was divided in two parts, with the aim to explore different aspects of splicing machinery dysregulation in two distinct but related adenocarcinomas. First, we aimed to determine the status of splicing machinery and its relationship with clinical and molecular features of PDAC aggressiveness. To this end, we measured the expression of 45 selected splicing machinery components in tumor *vs* non-tumor adjacent tissues in a set of 79 formalin-fixed paraffin-embedded PDAC samples. Analyses of patient survival and clinical parameters in relation with the expression of these splicing elements led to the identification of a subset of 2 dysregulated splicing factors, PRPF8 and RBMX, whose expression was closely linked to poor prognosis and malignancy parameters, and even associated with the presence of mutations in key-genes involved in PDAC development and progression. These results were confirmed in independent cohorts of public

databases. Experimental modulation of these splicing factors in PDAC cell lines (Capan-2, and BxPC-3) reverted their expression to non-tumor levels and resulted in decreased key tumor-related features, including cell proliferation, migration, and colony formation. In the second study, we investigated SRSF2 splicing factor in prostate (PCa) and pancreatic adenocarcinoma (PDAC), as a potential diagnosis biomarker and therapeutic target. To this end, we measured SRSF2 levels in PCa and PDAC human tumor samples and cell lines and modified it expression to evaluate different functional endpoints. Interestingly, SRSF2 was overexpressed in tumoral samples with respect to non-tumoral tissue in both PCa and PDAC, where SRSF2 levels were similarly linked to clinical-molecular features suggestive of worse prognostic potential and oncogenic capacity. In line with this, SRSF2 silencing in representative model cell lines of both adenocarcinomas inhibited cell migration, proliferation, invasion, and colony formation in PCa cell lines, whereas it only caused a similarly pronounced reduction in colony formation in PDAC cell lines. These findings suggest the interesting possibility that SRSF2 could be involved in the regulation tumor initiation capacity, which is related to cancer stem cells (CSCs) within the cell population.

The last section of this Thesis was focused on exploring the presence and functional role of the core splicing factor SF3B1 in PDAC and interrogated its potential as an actionable target. SF3B1 was analyzed in PDAC tissues, a RNA-Seq dataset, and publicly available databases, examining associations with splicing alterations and key features/genes. Functional assays in PDAC cell lines and PDX-derived CSCs served to test Pladienolide-B treatment effects in model cell lines *in vitro*, and *in vivo* in two preclinical models (zebrafish and mice). We discovered that *SF3B1* was overexpressed in human PDAC, wherein it was associated with tumor grade and lymph-node involvement. *SF3B1* levels closely associated with distinct splicing event profiles and

expression of key PDAC players (*KRAS, TP53*). In PDAC cells, Pladienolide-B increased apoptosis and decreased multiple tumor-related features, including cell proliferation, migration, and colony/sphere formation, altering AKT and JNK signaling, and favoring proapoptotic splicing variants (*BCL-XS/BCL-XL, KRASa/KRAS, Δ133TP53/TP53*). Importantly, Pladienolide-B similarly impaired CSCs, reducing their stemness capacity and increasing their sensitivity to chemotherapy. Pladienolide-B also reduced PDAC/CSCs xenograft tumor growth *in vivo* in zebrafish and in mice. Thus, *SF3B1* overexpression represents a therapeutic vulnerability in PDAC that enables the targeting of splicing with Pladienolide-B not only in cancer cells but also in CSCs.

For all the above mentioned, the main conclusions of this Thesis are:

1. The splicing factor *CELF4* is overexpressed in PanNETs, where its levels associate with malignancy features and distinct splicing profiles. Modulation of *CELF4* levels predictably influences multiple cancer features *in vitro* in PanNETs cell lines, and its silencing inhibits xenograft tumor growth. *CELF4* expression impacts pathways and mediators linked to mTOR pathway, which likely explains how it impairs the response of PanNET cells to everolimus treatment.

2. The splicing machinery is severely dysregulated in PDAC, where we identified two specific components, *PRPF8* and *RBMX*, that display a downregulated expression, which is closely linked to poorer survival and clinical and molecular markers of bad prognosis. *PRPF8* and *RBMX* expression is distinctly associated to altered splicing profiles and restoring their expression levels rescued their tumor suppressor ability *in vitro* in PDAC

cell models. These factors represent two promising targets that deserve further research as new potential biomarkers and molecular instruments to tackle PDAC.

3. Two distinct adenocarcinomas, PCa and PDAC, share a previously unrecognized overexpression of *SRSF2*, which is linked to clinical-molecular features suggestive of worse prognostic potential and oncogenic capacity. *SRSF2* silencing differentially influenced functional tumor features indicative of tumor aggressiveness in PCa (more responsive) and PDAC (less responsive). In contrast, *SRSF2* silencing similarly reduced colony formation in cell models of both cancers, suggesting a possible role of this factor in the control of tumor initiation capacity by cancer cells and/or CSCs.

4. *SF3B1* is overexpressed in human PDAC, wherein its levels associate with key clinical, histological, and molecular features. Furthermore, targeting SF3B1 activity with Pladienolide-B reduces multiple cancer features in PDAC cells by altering relevant signaling pathways and splicing events. Importantly, Pladienolide-B treatment reduces CSCs stemness, making CSCs more sensitive to chemotherapy treatment.

Global corollary

Taken together, the studies developed in the present Thesis provide novel evidence to advance in the molecular knowledge of the role of the splicing machinery and its alterations in different cancers. In particular, we identify specific splicing factors that are altered in PanNETs, PDAC and PCa, and seem to play a relevant functional role in these tumors, where they could serve as useful tools for the development of new biomarkers and could be the target for newly developed splicing-directed drugs, like Pladienolide-B and/or its derivatives. These findings reinforce the relevance of examining alternative

splicing, its players, and abnormalities to open up novel avenues for precision medicine in solid cancers.

Introduction

1. Cancer

Cancer is one of the leading causes of death in the world. This complex disease is a critical cause of morbidity and mortality worldwide, in every world region and regardless of the level of human development. In 2020, around 19.3 million of new cancer cases were detected and 10 million cancer deaths took place in the world [1]. Europe represents the 22.8 % of these new cancer cases and 19.6 % of the cancer deaths, even though it accounts for 9.7 % of the global population [1]. Specifically in Spain, 277,394 cases were diagnosed in 2020, and estimates indicate that the number of mortality cases over the next two decades will increase from 113,000 (2020) to 160,000 (2040) (Source: SEOM).

Cancer cannot be considered a single pathology, as its extraordinary diversity integrates over a hundred of distinct diseases depending on multiple factors, such as the cell of origin, which can reside anywhere in the human body. In spite of this, all cancer cells share a common characteristic, their abnormal capacity to proliferate and spread to other parts of the body. The main types of cancer are usually termed based on the tissue or organ where they derive and can be classified by the specific cell that gives rise to them, as carcinoma (epithelial cells), sarcoma (from bone and soft tissues), neuroendocrine tumors and carcinoids (hormone-producing cells), leukemia and lymphoma (blood cells), and melanoma (melanocytes). However, multiple subtypes of tumors exist within each of these major cancer types, making the remarkable heterogeneity that characterizes tumor pathologies the main obstacle to find a global, effective cure for cancer. Thus, despite the valuable advances in understanding its origin and variability, which have led to great progress in improving its detection and treatment, there is still much to be learned in order to eradicate this pathology. In this sense, in the present century, molecular targeted therapies have emerged as a remarkable clinical tool for the successful treatment of many types of cancer.

Hanahan and Weinberg revolutionized the established ways to understand this set of indecipherable pathologies by introducing the concept of “hallmarks of cancer” in 2000 [2], and subsequently expanding them in 2011 and, recently (D. Hanahan), in 2022 [3, 4]. Through this novel conceptual approach, they described that the vast majority of cancer cells share common, distinctive features and complementary capabilities that enable tumor growth and metastatic dissemination. Their proposal provided a logical framework to understand the remarkable diversity of neoplastic diseases, which would share the following “Hallmarks of Cancer” (2011): sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis, resisting cell death and constitution and signaling interactions of the tumor microenvironment crucial to cancer phenotypes [3] which were complete this year (2022) with four new emerging hallmarks and enabling characteristics: Unlocking phenotypic plasticity, non-mutational epigenetic reprogramming, senescent cells and polymorphic microbiomes [4] **(Figure I1**).

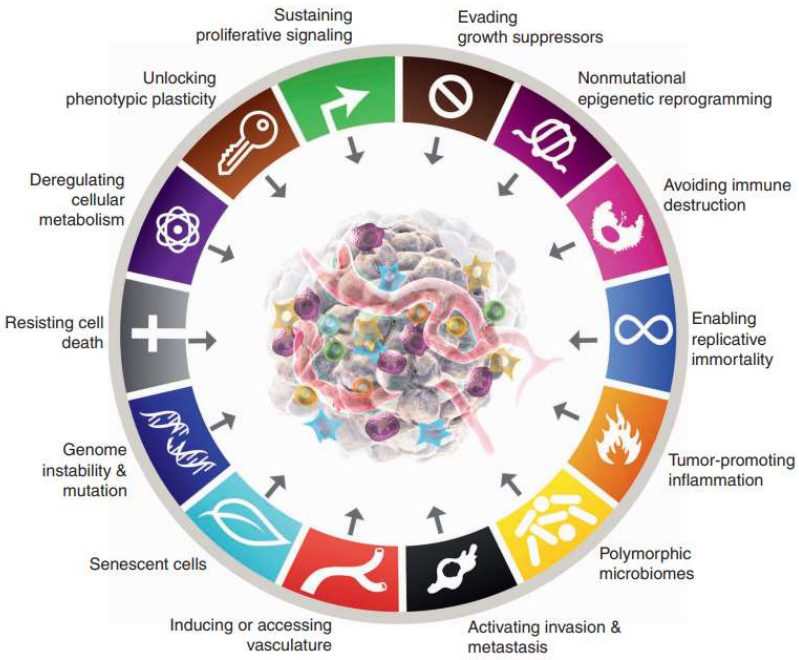


Figure I1. Hallmarks of cancer. Graphic representation of the hallmarks of cancer. Source: *Hanahan*, *Cell Reports. 2022*.

The proposed hallmarks of cancer represent complex phenotypic characteristics that can be acquired at a distinct point of tumorigenesis for different tumors. Elucidation of the underlaying mechanism responsible of the appearance of cancer cells and their "specific hallmarks" is crucial to advance towards new therapeutic approaches. The available knowledge on different types of cancer is nowadays far superior to that available few decades ago, but it is still incomplete and does not allow to satisfactorily explain this complex group of neoplastic diseases in a complete manner. Therefore, it is necessary to continue opening new avenues of research to fill knowledge gaps which can help, ultimately, to bring this disease closer to precision medicine based on molecular knowledge.

1.1. Neuroendocrine Tumors

Neuroendocrine tumors (NETs) are a diverse and heterogeneous group of neoplasms arising from neuroendocrine cells throughout the body. NETs share a common origin: the cells of the diffuse neuroendocrine system, but can differ markedly in several features, including their anatomical location, their secretory activity or inactivity, their degree of development and differentiation, the presence or absence of metastases, etc. (**Figure I2**).

Histologically, NETs are characterized by the presence of endocrine tissue-specific markers such as chromogranin A, synaptophysin or enolase, or by their in-/ability to produce serotonin, histamine, prostaglandins, substance P, insulin, gastrin, glucagon, or vasoactive intestinal polypeptide. The most prominent groups of NETs are gastroenteropancreatic NETs (GEP-NETs), which comprise gastrointestinal (GI) and pancreatic NETs (PanNETs), and pulmonary NETs (LungNETs) [5].

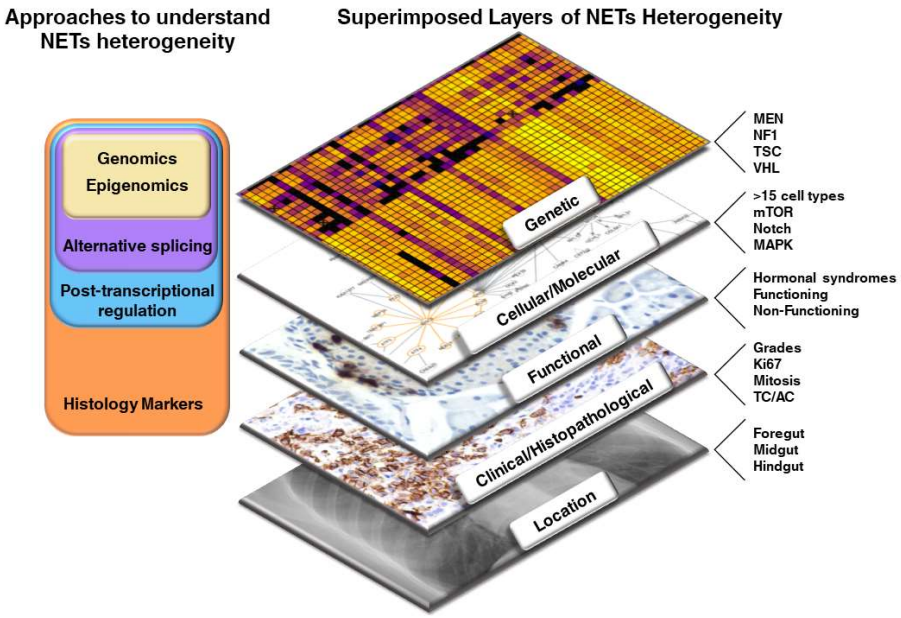


Figure I2. Levels of heterogeneity of NETs. Representation of the different layers of this heterogeneity interrelated and affecting each other. The left panel underscores some of the different approaches to address it. Source: *Pedraza-Arévalo et al., Rev Endocr Metab Disord 2018*.

The incidence of these tumors varies from 1.51 to 6.98 per 100,000 population annually [6, 7] (**Figure I3**); overall, it is estimated that more than 12,000 people in the United States are diagnosed with a NET each year, and approximately 175,000 people are living with this diagnosis [8]. A higher incidence of NETs has been identified in patients aged 65 years and older, with an increase of 8 to 25.3 times per 100,000 people [6]. An incidence increase has been noted in recent years, which varies in all organs, stages, and grades, being the lung the organ with the highest incidences identified (1.49 per 100,000 people) followed by GEP-NETs (3.56 per 100,000 people). Due to the increasing incidence and indolent nature of NETs, the prevalence of these tumors has also increased considerably, being estimated from 0.006 % to 0.048 % ($p < 0.001$) [9].

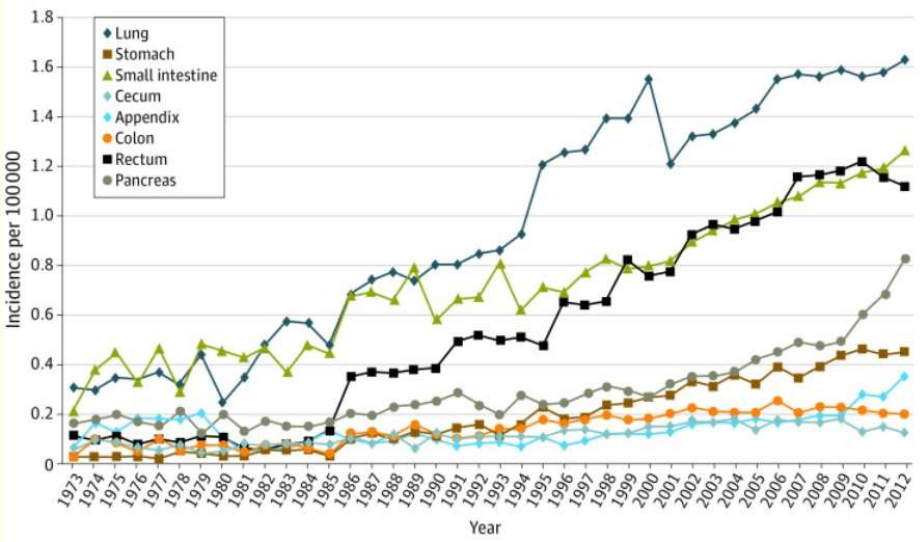


Figure I3. NETs incidence by body location. Incidence of different types of NETs by location, showing their increase over the last decades and the recent rise in PanNETs (grey circles). Source: *Dasari et al., JAMA Oncol. 2017.*

The multiple nonspecific symptoms typical of NETs cause severe delays in patient diagnosis. Indeed, almost 60 % of NETs are at an advanced stage at diagnosis [10]. The survival rate varies considerably from 6 months to more than 30 years, reducing quality of life due to the frequent interventions required during this time [6, 11]. The 5-year survival rate for people with NETs depends on several factors as NET type or its localization [8]. The median overall survival time for NETs patients is 9.3 years (112 months). NETs have a better median overall survival in localized stages (over 30 years) compared to NETs with local metastases (10.2 years) and with distant metastases (12 months) [6].

Due to their great heterogeneity, the anatomical classification of NETs must be completed with their cellular and molecular characterization, which facilitates a better classification and a more detailed and precise understanding of these tumors. In this context, important advances have been made in the identification of genetic mechanisms involved in the pathogenesis of NETs over the last years [7, 12, 13]. Although the intrinsic heterogeneity of the different NETs precludes the existence of a single common

characteristic molecular profile, it is possible to distinguish some specific alterations linked to each subtype. A limited number of molecular alterations, associated with chromosomal losses and/or gains, and point gene mutations, had been identified as initial events in the development of NETs, which are mainly associated with different inherited syndromes related to mutations in a discrete set of genes (such as *MEN1*, *RET*, *VHL*, *NF1*, or *TSCs*) [12, 14, 15]. In contrast, the molecular alterations underlying the development of sporadic NETs are currently being characterized. Thus, recent evidence identifies specific potential driver genes and the alteration of certain key pathways [7], such as genes involved in cell cycle control, chromatin remodelling, methylase or suppressors of the mTOR pathway (mammalian Target of Rapamycin). Actually, 14 % of sporadic PanNETs have mutations in genes of the phosphatidylinositol 3-kinase (PI3K)/Akt/mTOR pathway, a signaling cascade that regulates multiple key cellular functions, such as cell growth and proliferation, metabolism, motility, and apoptosis [16-18]. Indeed, the PI3K/Akt/mTOR pathway, particularly through mTOR, contributes to control those cell functions and, thereby, crucial processes like angiogenesis [19, 20].

1.1.1. Pancreatic Neuroendocrine Tumors

Pancreatic neuroendocrine tumors (PanNETs) represent 62 % of all diagnosed gastroenteropancreatic neuroendocrine tumors [5], with an increasing incidence over the past few years, and a 90 % of 5-year Relative Survival Rate [6]. The most frequently used cancer staging system is based on the WHO classification, which categorizes tumors into three grades based on proliferative activity and morphology [7]. PanNETs are often detected in an advanced stage, as the lack of precise markers specific clinical symptoms complicate early diagnosis, leading to diagnostic times between 5 – 7 years, which hinders the prompt application of effective and specific therapies [8].

PanNETs mainly arise from the endocrine cells of the pancreas, located in the Langerhans islets, and are classified based on their secretory activity as functioning and non-functioning tumors. In turn, since functioning tumors can exhibit the characteristics of their cell of origin, they can cause a clinical syndrome produced by excess hormonal release. Thus, according to their secreted hormone, functioning PanNETs are divided in:

- Insulinomas (insulin producing cells), which result in hypoglycaemia.
- Glucagonomas (glucagon producing cells), causing glucose intolerance, necrolytic migratory erythema, stomatitis/glossitis, hypoaminoacidemia.
- Gastrinomas (gastrin producing cells), which result in severe peptic ulceration.
- Somatostatinomas (somatostatin producing cells), which cause diarrhoea, diabetes, gallstones, anaemia, weight loss.
- VIPomas (vasoactive intestinal peptide (VIP) producing cells), which result in watery diarrhoea, hypokalaemia, achlorhydria (WDHA syndrome).

In contrast, non-functioning tumors do not produce any functional hormones, and are consequently diagnosed in more advanced stages, due to their size and/or because metastatic state (mostly in liver) has already been reached [9, 10].

Despite their intrinsic heterogeneity, PanNETs share some distinctive characteristics, such as high expression of somatostatin receptors (particularly SST_2 and SST_5), high vascularization, and alteration in different signaling pathways (as mTOR or PI3K/AKT). In fact, these features represent the main targets for medical treatment when the primary (an only curative) approach, surgery, cannot be applied or is not effective. Even though the treatments directed to SSTs (e.g. somatostatin analogues), mTOR pathway (e.g. everolimus) or angiogenesis (e.g. sunitinib) can effectively decrease hormone hypersecretion, and reduce tumor size or vascularization, in a high number of

cases, tumors reduce or lose their response, often leading to greater aggressiveness, hypervascularization or even an increase in tumor metastasis [11-13].

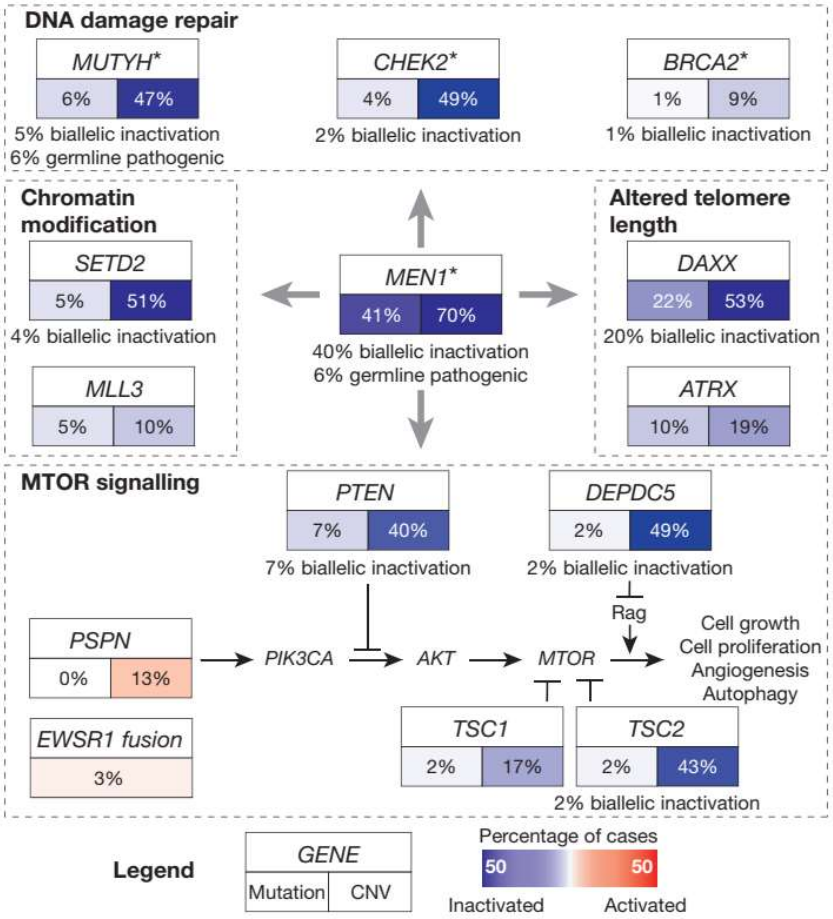


Figure I4. Representative core of PanNETs pathways. Graphic representation of somatic mutations and copy number changes for the key genes altered in PanNETs. Source: *Scarpa A., et al. Nature 2007.*

PanNETs are usually sporadic, while a minor proportion (less than a 10 %) can be associated with part of three hereditary syndromes characterized by well-known mutations in *MEN1*, *VHL*, and *NF1* [8, 21]. Available data indicate that PanNETs are highly heterogeneous in terms of genetic alterations and epigenetic modifications [15-

18], being *DAXX/ATRX* the most frequently mutated genes. This diversity, coupled to the current lack of effective systemic treatments to cure this disease, underscores the necessity of further exploring the molecular basis of PanNETs to find new therapeutic avenues [11].

1.2. Adenocarcinoma

Carcinomas constitute around 80 – 90 % of all cancer types, and can be broadly classified into squamous cell carcinoma, derived from the squamous epithelium, and adenocarcinomas. Adenocarcinoma is a cancer type that develops from epithelial cells generally producing fluids or mucus, being these tissues commonly described as glandular tissues. Most cancers diagnosed are adenocarcinoma, including breast, pancreas, prostate, lung, or colon.

In general terms, adenocarcinoma can be classified into different types depending on their ability to invade surrounding structures and their differentiation grade:

- *In situ* adenocarcinoma, corresponding to initial tumor stages, is localized just in the tissue of origin where the tumor develops.
- Invasive adenocarcinoma, where cancer cells reach to other tissues/organs.
- Well-differentiated adenocarcinoma: cancer cells that still resemble the original tissue and are growing slowly.
- Poorly differentiated adenocarcinoma: tumor cells have very different characteristics from the original tissue, which may indicate greater potential for malignancy and poor treatment response.
- Moderately differentiated adenocarcinoma: an intermediate level between well and poorly differentiated.

A precise characterization of the adenocarcinoma is crucial, in that treatment choice varies according to tumor location, type and classification. In general, adenocarcinomas are difficult to treat, due to their aggressiveness, and have poor prognosis. The main

treatment options include surgery, radiation therapy, and chemotherapy, while targeted treatments are currently emerging.

1.2.1 Pancreatic ductal adenocarcinoma

Pancreatic ductal adenocarcinoma (PDAC) is the most common type of tumor of the pancreas (90 %) and the most aggressive one. PDAC mortality rate is one of the highest of all cancers worldwide, being higher in Europe and Northern America (**Figure I5**). It comprises 4.5 % of all deaths by cancer, having almost the same number of deaths as new cases. Low survival is associated with poor prognosis, late diagnosis, and development of drug resistance [22]. PDAC risk factors include genetics (in around 10 % of the cases), tobacco use, alcohol use, pancreatitis, and obesity, among others, which generally increase inflammatory pancreatic damage [23].

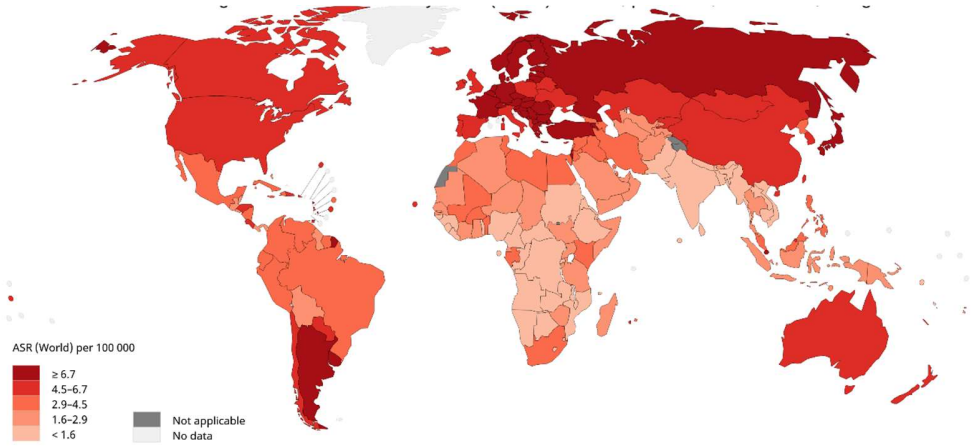


Figure I5. Estimated pancreatic cancer mortality rates in 2020. Source: Globocan 2020.

PDAC progresses from pre-invasive lesions, which include cystic [intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm (MCN) and intraductal tubulopapillary neoplasm (ITPN)] and a non-cystic lesion which is the most common one [pancreatic intraepithelial neoplasia (PanIN)] (**Figure I6**). Cystic lesions can be diagnosed by image methods, but PanINs cannot be detected early or other than microscopically [24]. PanINs have different grades, progressing to a major dysplasia, characterized by

cell polarity loss and increase in nuclei size. This transformation of pancreatic tissue has been associated to different mechanisms like genomic instability and mutations. Several studies have evidenced that some of the most frequent mutations of PDAC appear in pre-invasive lesions like PanIN [25]. For example, it has been shown that *KRAS* mutation rate increases as it reaches a major grade of dysplasia, appearing even in early stages of the lesion. *KRAS* is the most frequently mutated gene in PDAC (95 % of patients) and it is considered as the main driver of tumorigenesis [26]. Frequent mutations in genes other than *KRAS* have been characterized in PDAC, including genes encoding tumor suppressor proteins *CDKN2A*, *TP53* or *SMAD4*. Other mutations have also been reported in genes involved in essential cell processes like chromatin remodeling and DNA damage repair, which likely contribute to boost genomic instability [27]. One of the principal causes of the damage of the genome, which appears to be an early event in pancreatic lesion, is telomere shortening.

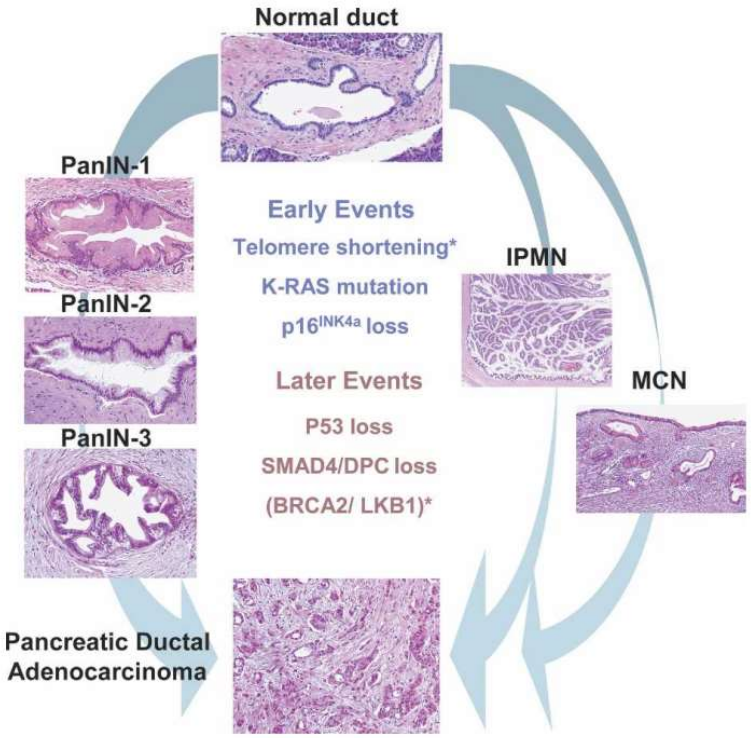


Figure I6. Pancreatic precursor lesions and genetic events involved in PDAC progression. *Source: Hezel AF et a.,Genes Dev.2006.*

Nevertheless, although the genomic landscape of PDAC is well characterized and its most frequent mutations have been proposed as therapeutic targets for PDAC on multiple occasions, there are currently no drugs which can effectively tackle them to cure this cancer [28, 29].

Along with the described genetic alterations, there are additional factors that reside at different hierarchical levels and contribute to PDAC malignancy. The histological/tissue level comprise a critical level in this type of tumors because of the nature, volume, and cellular composition of the stromal compartment. Likewise, the inflammatory component that accompanies pancreatic damage keeps developing when it becomes adenocarcinoma. Besides cancer cells, PDAC contains several relevant cell types that comprise a true microenvironment, including different classes of fibroblasts, pancreatic stellate cells, cancer stem cells, macrophages, infiltrated lymphocytes, etc. [30] (**Figure I7**). These cells have been shown to communicate with cancer cells, and their interaction is necessary for tumor progression, favoring tumor growth, angiogenesis, metastasis and driving drug resistance [31].

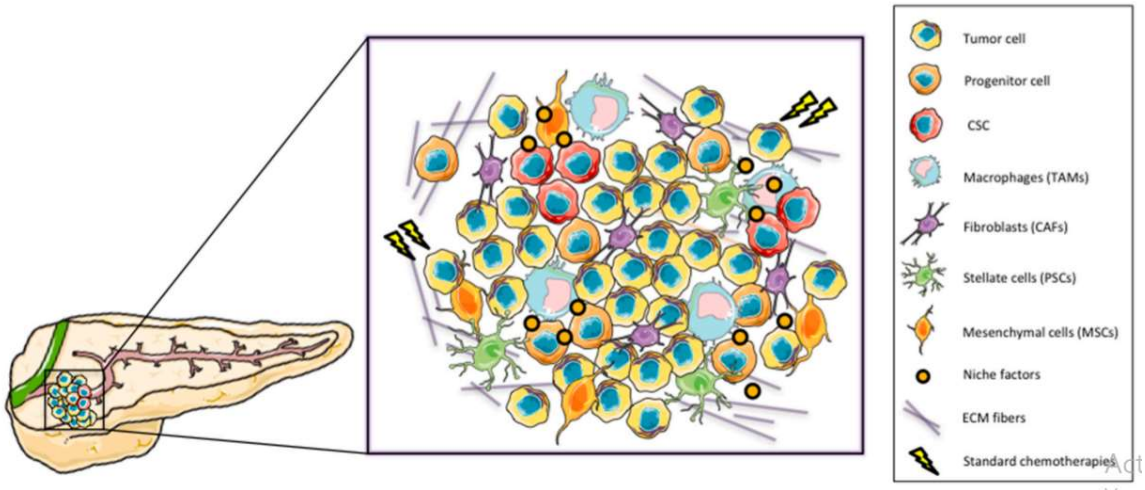


Figure I7. Pancreatic precursor lesions and genetic events involved in PDAC progression. Source: *Valle S. et al., Cancers (Basel). 2018.*

Despite the remarkable advances achieved in the molecular knowledge of PDAC, the number of patients who survive this pathology has not improved relevantly in the last years, with a dismal 5-year survival rate of 10 %. This is mainly due to the lack of effective treatments, being surgery the most effective and only curative approach. Radiotherapy is recommended in specific cases. Current pharmacological treatments include the chemotherapeutics gemcitabine, 5-fluorouracil, paclitaxel, or combinations of some of them, like the combination FOLFIRINOX, which is the main treatment for advanced metastatic PDAC [32].

1.2.2 Prostate Cancer

Prostate cancer (PCa) is the second common type of cancer in men (14.1 %), just after lung cancer, being the fifth leading cause of cancer-related death among men (6.8 %) (**Figure I8**). PCa is one of the most frequently diagnosed cancer, estimating 1,414,259 of new cases in 2020. In Europe, is the cancer with highest incidence among men (33.5 %), with a mortality rate of 28.8 % and 5-year prevalence of 37.8 % [1].

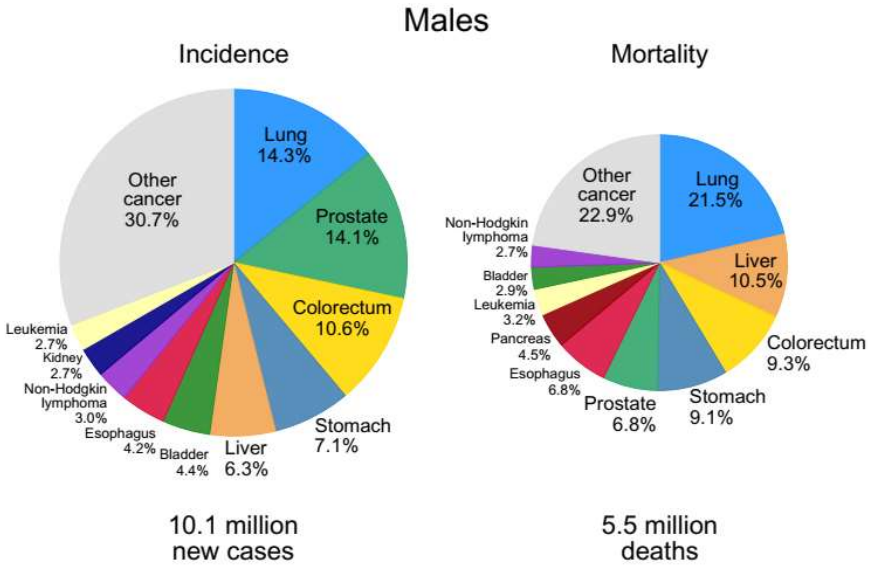


Figure I8. Cancer incidence and mortality in 2020 in males. Distribution of cases and deaths for the top 10 most common cancer in 2020 where prostate cancer is represented in green. Source: *Globocan 2020.*

Death rates for PCa have been decreasing in many countries due to early diagnosis because of screening and improved treatment. On the one hand, advanced imaging techniques, including Computerized Tomography (CT scan), Magnetic Resonance Imaging (MRI scan), and Positron Emission Tomography (PET scan), are used to detect and follow-up patients, with the aim of avoiding unnecessary treatments and ensuring an appropriate diagnostic and monitoring, this strategy being applied to patients with low-graded tumors [33]. On the other hand, in high-grade PCa patients, surgery for radical prostatectomy, removing all the urinary system, is the most employed approach to eradicate the tumor. In addition, there are other possible therapeutic tools for PCa patients, including radiotherapy, chemotherapy, or androgen deprivation therapy, which are selected based on diagnostics, such as TNM stage, prostate-specific antigen PSA levels, or GS [33].

These complex tumors are primarily classified based on their response to androgen deprivation therapy into hormone sensitive PCa, and castration resistant PCa (CRPC) [34]. Histologically, the gold standard for PCa grading in routine clinical practice is the Gleason Score (GS). GS is based on the histopathology of the cells, from the most to the least differentiated [35]. When GS classification is applied, PCa can be divided in non-significant PCa ($GS < 7$) and significant PCa ($GS \geq 7$), where GS higher than 6 corresponded to prostate biopsy and radical prostatectomy specimens [36] (**Figure I9**).

Several external factors have been associated with higher possibilities to suffer PCa, including poor eating, physical activity, or smoking habits; however, the main risk factors are age, family history, genetic factors, and ethnicity [37, 38]. PCa has a complex aetiology and despite increased understanding of the underlying molecular basis of this disease, its accurate diagnosis and targeted treatment remains a major challenge. PCa is usually suspected when patients present with an elevated serum level of PSA and/or after an abnormal digital rectal examination [39]. However, PSA levels are not specific

for PCa and high levels could be associated with other pathologies as benign prostatic hyperplasia and prostatitis [40] or activities as bike cycling. For all these reasons, the search for new, more sensitive, and specific non-invasive biomarkers that facilitate the diagnosis of PCa, as well as novel actionable therapeutic targets are necessary.

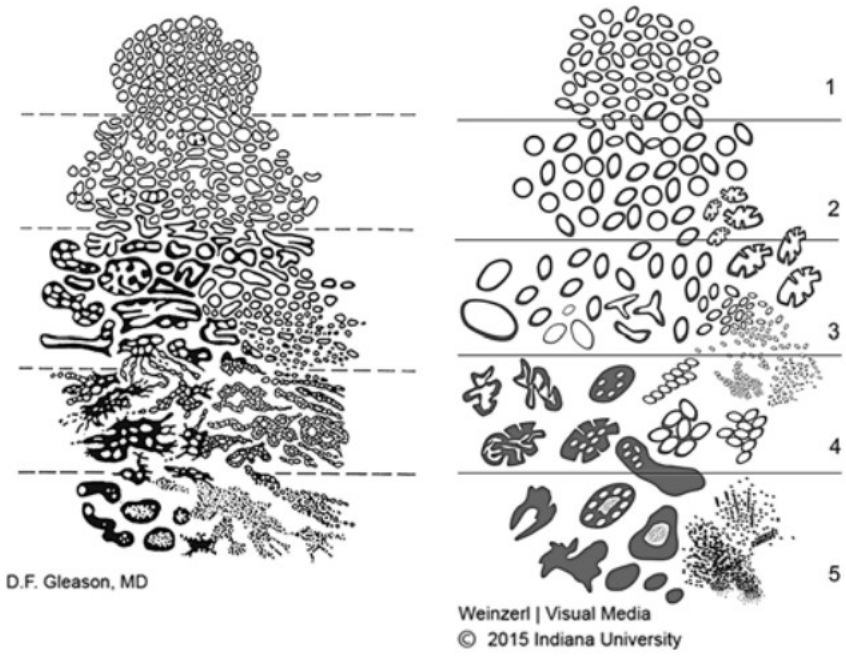


Figure I9. Comparison between the original (left panel) and the 2015 modified ISUP (right panel). Gleason Score schematic diagrams of PCa histologic patterns. Source: *Epstein et al., Am J Surg Pathol 2016.*

2. The spliceosome and the splicing process

Splicing is a complex cellular mechanism by which the immature or precursor RNA is processed, removing the sequences that will not be part of the final RNA, or introns, and binding the exons, that form the mature RNA [41, 42]. However, most of the genes (> 95 %) do not undergo this simple cut and paste process, or canonical splicing, but an intricately regulated process called alternative splicing [41, 43, 44]. This phenomenon allows the generation of different combinations of sequences through the inclusion and exclusion of concrete groups of exons, which results in a variety of mature RNA

transcripts from the same precursor, named as splicing variants, that may carry out different or even opposite functions [45]. This is an essential process in the cell to ensure an appropriate regulation of gene expression because it enables an increase in the variety of gene products and thereby enhances the versatility of the genome [44]. For all these reasons, the accurate regulation of the splicing process is crucial for the correct development and homeostasis of the cell and the organism. The process of splicing and its delicate regulation is carried out by the spliceosome, a ribonucleoproteic complex that recognizes specific RNA sequences to precisely localize the introns and cut them, and subsequently bind the adjacent exons [46]. In mammals, there are two different spliceosomes that act separately: the major spliceosome, that processes more than 99 % of the introns, and the minor spliceosome, that acts over a small and specific set of introns [47]. Accordingly, introns are classified in U2-type (or -dependent, GT-AT) and U12-type (or -dependent, AT-AC) following the spliceosome that processes them or the flanking sequences [48]. Both spliceosomes consist of a main core of small nuclear RNAs (snRNAs), named as RNU1, RNU2, RNU4, RNU5 and RNU6 for the major spliceosome, and RNU11, RNU12, RNU4ATAC and RNU6ATAC (RNU5 is present in both), which are joined together to proteins to form small nuclear ribonucleoproteins (snRNP; U1-U6), forming an association and working together in a dynamic and coordinated manner [46, 47]. In addition, the spliceosomes closely interact with the splicing factors, a diverse set of over 300 molecules that complete the splicing machinery, helping the snRNPs to select and process the precise sequences, and taking part dynamically in every step of the process, participating in both general tasks as well as very specific events [49, 50].

The splicing process has been classically investigated in simple research models, easier to study than human, like yeast, but the key steps are very well conserved in mammals (**Figure I10**). Summarizing the explanation by Matera and Wang in 2014 [51]

and other studies [48, 52], U1 and U2 recognize and bind to 5' and 3' splice sites of the pre-mRNA, respectively. Next, U2 recognizes sequences in the so-called branch point and interacts with U1, forming the pre-spliceosome. Then, preassembled U4-U5-U6 complex is recruited, and several conformational changes take place to form a catalytically active complex, resulting in U2/U6 structure that catalyses the splicing reaction. In this step, U1 and U4 are released from the complex. At this point, the first catalytic step is carried out, cutting the binding between the first exon and intron-exon lariat intermediate. Finally, after some further conformational changes, the second catalytic step leads to the separation of intron and second exon and the binding of both exons, leaving the post-spliceosomal complex with the intron lariat free. Finally, U2, U5 and U6 are released. All the described steps are firmly regulated by several spliceosome proteins, which ensure that the cuts and bindings are correct, making possible the sequence recognition and putting together and separating the other components.

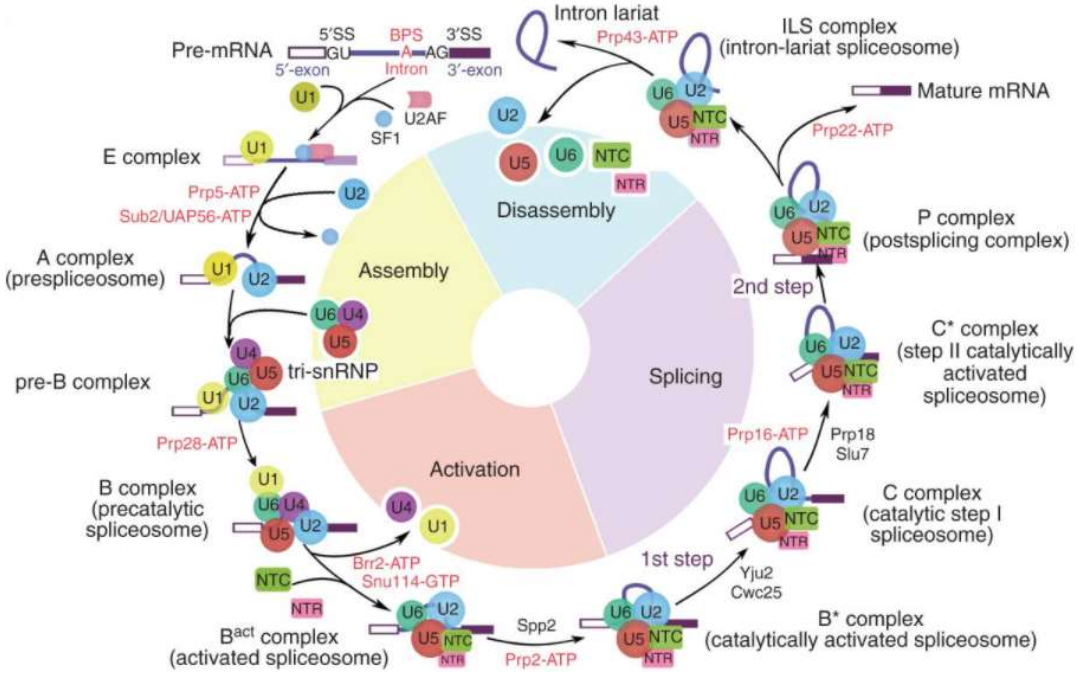


Figure I10. Splicing process. The assembled spliceosome sequentially assumes eight different compositional states during each cycle of the splicing reaction: pre-B, B, activated (Bact), catalytically activated (B*), step I spliceosome (C), step II activated (C*), postcatalytic (P), and intron lariat spliceosome (ILS). Source: *C.Yan, et al Cold Spring Harb Perspect Biol. 2019.*

Typically, the introns of mammals are long and present several decoy splice sites that must not be spliced [43]. As mentioned earlier, alternative splicing is based on the inclusion/exclusion of selected sequences; therefore, a precise regulation is needed to correctly splice each sequence. To this end, cis-regulatory elements are distributed through the RNA, known as splicing regulatory elements, and, depending on their function and location, are classified in exonic/intronic enhancers/silencers (ESE, ISE, ESS, and ISS, respectively) [43, 44, 53]. Those sequences recruit trans-regulatory elements, the splicing factors, that will suppress or activate steps of the splicing process. However, these events are completely dependent on the context, since the same factor may be a splicing enhancer and a splicing silencer if it binds to an enhancer or silencer element [44, 48, 53].

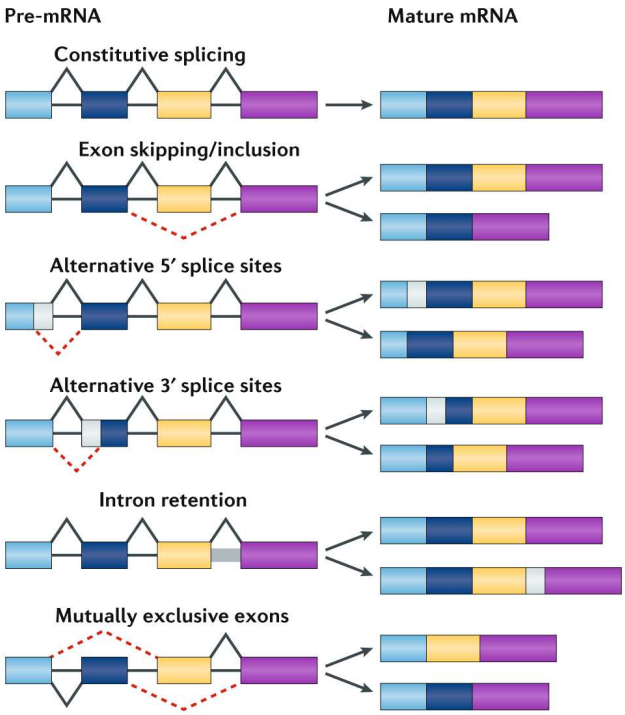


Figure I11. Schematic representation of alternative splicing events. Splicing reactions are represented by a red dotted line. Cis-regulatory elements ESE, ISE, ESS and ISS are included as well as the Branch site. Source: *Frankiw L. et al., Nature. 2019.*

Furthermore, there are additional possibilities for splicing regulation. Thus, the structure of the precursor RNA may alter the accessibility to regulatory domains or even

the spliceosome complexes [43]. In addition, the activity of the splicing machinery is finely regulated through modulation in its components, including regulation of their genetic expression by transcription factors, miRNAs or epigenetics [54-56], or even posttranslational modifications, such as phosphorylation or acetylation [57-59], that may affect their location or activity.

This complex regulation allows the correct progression of the splicing process, including the variations that cause alternative splicing. Specifically, there are five different types of alternative splicing: 1) cassette exon skipping, an exon is excluded together with the two flanking introns; 2) mutually exclusive exons, two exons that cannot be included together, one of each is excluded in two different isoforms; 3) intron retention, there is no cutting in the intron and this is included in the mature RNA; 4) alternative 3' splice site, and 5) alternative 5' splice site, the exon is not fully included in the final RNA, but it is cut in a different site [51, 60, 61] (**Figure I11**). Taken together, all these data demonstrate the great complexity of the splicing process and underscore its relevance in the control of normal functioning of the cell.

3. Splicing dysregulation in Cancer

Despite the impressive and growing list of alterations in genes and regulatory mechanisms that have been described to date, these are still insufficient to provide an effective therapeutic strategy to battle cancer. Interestingly, recent studies have demonstrated that a novel, common hallmark shared by tumoral pathologies is the alteration in the normal gene expression pattern, associated to an incorrect functioning of the machinery that regulates the splicing process. In fact, the splicing process appears altered in all cancers studied to date, where it influences and interacts with each of the hallmarks of cancer [62, 63].

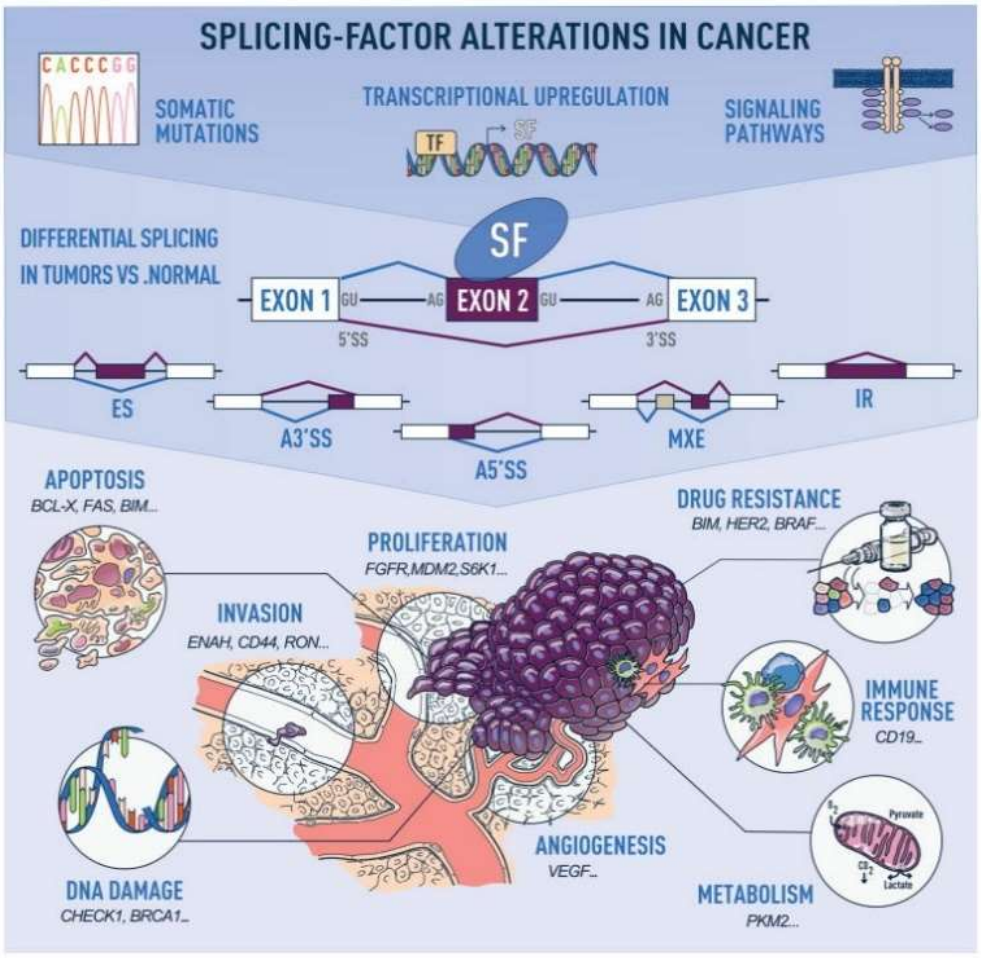


Figure I12. Alternative splicing defects in cancer. Source: *Urbanski L, et al Wiley Interdiscip Rev RNA. 2018.*

The pathophysiological importance of alternative splicing and its associated processes is supported by the emerging evidence linking diverse splicing anomalies with multiple pathologies, from rare diseases [64] to leukaemia [65], wherein pathogenic function of splicing variants and their value as therapeutic targets has been proven [66]. Dysregulation of the splicing machinery leads to the expression of aberrant RNAs and/or proteins that can contribute to the development or progression of diverse pathologies, including cancer. Specifically, current studies revealed that this “spliceosomic pathway” may play important roles in cancer progression [67] (**Figure I12**), for example through mutations and alterations in the expression levels of several components of the splicing

machinery. Actually, earlier reports already revealed the appearance of certain cancer-associated oncogenic variants derived from splicing machinery defects, which led to the use of components of the splicing machinery as diagnostic or prognostic markers in various cancers [52, 63, 67], as they provide relevant clinical information, such as resistance to treatment [52]. Nevertheless, the precise mechanisms underlying the contribution of splicing alterations to oncogenesis are still incompletely understood.

3.1. Splicing dysregulation in Neuroendocrine Tumors

Splicing is a poorly explored subject in NETs, where just a few cases of abnormal variants had been described [68, 69]. However, mounting evidence suggests that splicing is dysregulated in NETs and that this is linked to pathological features [70, 71]. These dysregulations could be associated to mutations or altered expression of specific components of the splicing machinery. Compelling evidence for this notion is offered by the recent discovery of recurrent mutations in the spliceosome U1 snRNA, which change their preferential base-pairing targeting, creating novel splice junctions and altering the splicing pattern of multiple genes, including known cancer drivers [72]. Likewise, frequent mutations in a key splicing factor, SF3B1, drive malignancy in melanoma by causing missplicing of *BRD9* and thereby disrupting non-canonical BAF [73].

In line with this, other relevant component of the splicing machinery, NOVA1, has been found to be overexpressed in PanNETs, it is associated to significant clinical parameters (DOI:10.1101/2022.02.09.479525), whereas knockdown of another splicing factor, SRRM4, supressed tumor growth in small cell lung cancer (SCLC) models [74]. There are also two RBM splicing factors related with SCLC tumor suppressor activities: RBM5 and RBM10 [75]. Specifically, *RBM10* has been related with major cell proliferation and transformation-associated processes in SCLC. Additionally, the splicing factor *ESRP1* has been also found downregulated in chemo-resistant SCLC cells [76].

Defects in alternative splicing function might underlie the appearance of abnormal splicing variants which can carry out an oncogenic function as is the case of the truncated somatostatin receptor subtype 5 (SST_5TMD4)[77] and intron 1 retaining ghrelin (In1-ghrelin)[78]. Previous work from our research group has demonstrated that these variants impart oncogenic properties in NETs, enhancing tumor aggressiveness and malignancy features, by altering signaling pathways and basic cellular processes [77], in line with that found also in other cancers [79, 80]. Of note, a splice variant of the read-through fusion transcript *INS-IGF2* is known to be expressed in insulinomas, while it is not expressed in normal pancreatic tissue [81]. Also, a splice variant of 5 transmembrane domains of *CCK2R* gene, which diminishes CCK2R membrane density and its activity, is expressed in insulinomas [82].

Nevertheless, and despite the potential translational value of the discoveries on splicing dysregulation in NETs, no studies have explored in detail this regulatory system in these tumors. Hence, it is both timely and necessary to open new avenues of research, like alternative splicing, to bring this disease closer to precision medicine based on molecular knowledge [83].

3.2. Splicing dysregulation in Prostate Cancer

A paradigm of the importance of correct splicing is seen in PCa, where Androgen Receptor (AR) splicing variants have been related with PCa progression. Specifically, early studies described truncated variants of the AR lack a ligand-binding domain and, consequently, are constitutively active, providing them the ability to confer PCa cells resistance to anti-androgen therapies. Actually, the truncated variant AR-V7 is the most expressed of all AR variants and underlies the most aggressive form of the disease, CRPC [84, 85]. Even though this variant was described in 2008 and that some splicing

factors involved in the processing of this truncated variant are now known, no effective treatment to successfully block its action has yet been achieved.

In line with this notion, an increasing list of (abnormal) splicing variants are being identified as key oncogenic players in PCa, offering new diagnostic and therapeutic opportunities. This is the case, for example, of the truncated variant of SST_5, SST_5TMD4, which is overexpressed in PCa and increases its malignancy [86], as well as the splice variant of ghrelin, In1-ghrelin, which is also detected in plasma and could be used as non-invasive biomarker for PCa [80]. Similarly, the tumor suppressor KLF6 possesses a splicing variant, KLF6-SV1, that is present in PCa, increasing its aggressiveness and being associated with a poor prognosis [87], similar to that found for other genes and splice variants, like FGFR2-IIIb [88], PKM2 [21], *VEGF165b* [89] or *CCND1b* [90], which also play an oncogenic role in PCa.

Furthermore, in recent years, it has been demonstrated that the splicing machinery is severely dysregulated in PCa and can thereby contribute to tumor aggressiveness and treatment resistance. Indeed, a clear overexpression of most core components of the dysregulated splicing machinery have been found in PCa tumor tissue compared with non-tumoral adjacent tissue. Specifically, alterations in SNRP200, SRRM1, and SRSF3 were associated with relevant clinical and molecular parameters linked to aggressiveness [91]. As well, SRRM4 [92, 93], and ESRP2 [94] are involved in PCa malignancy. Moreover, blockade of the activity of the splicing factor SF3B1 exerts strong antitumoral effects in PCa cells [95]. These results shed new light on the mechanisms underlying the contribution of splicing alterations to PCa oncogenesis are invited to further explore this subject to complete our understanding of this disease.

3.3. Splicing dysregulation in Pancreatic Ductal Adenocarcinoma

Splicing dysregulation also appears to be involved in PDAC. Thus, a decade ago, while evaluating the expression levels of multiple genes in human pancreatic cancer cell lines, *Carrigan et al.* observed a clear decrease of the expression of several splicing machinery components [96]. More recently, some of these changes were validated in PDAC human samples generating a spliceosomal signature based on upregulated and downregulated genes, which could discriminate between tumoral and control samples [97].

Those alterations in the expression of the splicing machinery commonly consist in the overexpression of spliceosome components and/or splicing factors. This is the case of the overexpression of genes related with proliferation and apoptosis, like *SRPK1* [98, 99], *CLK1 [100]*, *HNRNPK* [101]; associated to metastasis and invasion alterations, *PRPF40A* [102], *ESPR1* [103] and *SRSF6* [104]; and those linked to acquisition of chemotherapy resistance, *SRPK1* [98, 99], *SRSF1* and *PTBP1* [105].

Whole-exome sequencing of PDAC revealed a large number of mutations in key oncogenes and tumor suppressor genes as *KRAS* or *TP53 [106]*, which were known to be among the most commonly mutated genes in PDAC [25]. Interestingly, this study also identified mutations with lower frequency in some genes involved in essential processes, like splicing, and which conferred a higher tumor heterogeneity, as *SF3B1* [107], *U2AF2* [97, 108], or *RBMX* [109].

These pioneering studies prompted further analysis on alternative splicing in PDAC, which have provided valuable information on the pattern of splicing events and signatures in PDAC cell lines [96] and human tumor tissue [97, 110]. The landscape of alternative splicing in PDAC shows that the most common alterations in the protein-coding genes are skipped exon and alternative 1st exon, followed by intron retention. As in other tumors, in PDAC there are several genes that generate splicing variants that specifically appear

in tumor tissue and confer advantages to cancer cells. This is the case of the CD44 variants *CD44v2* and *CD44v6*, which can be detected in human PDAC tissue by immunohistochemistry, where their expression is related to an increase of mortality rate [111-113], and are linked to an EMT phenotype and higher invasiveness and chemoresistance features [114]. The isoform of the prolactin receptor (PRLR) PRLR-SF plays and important role in metabolism reprograming and prevents PDAC tumor progression [115]. Likewise, the role of splice variants of FGFR1 has been explored, revealing that FGFR1-IIIb and IIIc promote tumorigenesis by modulating key tumoral conditions like cell proliferation, adhesion, and movement [116-120]. Moreover, the dysregulation of *BCL2* expression has been related with apoptosis resistance, which may be explained, at least in part, by the overexpression of antiapoptotic *BCL2L1* isoforms in human PDAC tissue compared with normal tissue [121].

In spite of these insightful studies, much remains to be known about the expression pattern of spliceosome components and associated factors in this PDAC, as well as on the implications of their dysregulation in the development and progression of PDAC. Further elucidation of these alterations could pave the way for the discovery of new molecular targets and the implementation of better diagnostic and therapeutic methods.

4. Splicing modulation for therapeutic benefit

Splicing alterations can play important roles in the development of multiple pathologies. In particular, the dysregulation of alternative splicing is attracting an increasing interest due to its impact in cancer development and progression. The importance of this area resides in that altered splicing can promote the expression of aberrant variants and modify cell functioning, inducing not only the development of cancer but also insensitivity to targeted therapies. Consequently, novel strategies are being

designed and developed to mitigate the effect of splicing errors and to employ splicing dysregulation as an actionable therapeutic target.

4.1. Targeting splicing core

The widespread alteration of splicing in cancer, among other pathologies, has prompted the development and testing of different types of molecules capable to interact with specific elements of the core spliceosome complex and modulate their functioning. Three of these compounds, representative of different chemical natures, are Spliceostatin, Pladienolides or Herboxidienes. Some of these compounds share a similar mechanism aimed at inhibiting SF3B1 and, consequently, interfering with snRNP U2, destabilizing it and preventing the transition of the spliceosome complex. The potential clinical utility of these molecules and their derivates has been demonstrated in several studies. In fact, the first clinical study conducted with a splicing inhibitor tested the compound E7107, a Pladienolide-B derivative, demonstrating with a 40 cancer-patients cohort the safety, tolerability, and effective pharmacokinetic behaviour of the compound [122]. Another compound known as H3B-8800, a SF3b complex modulator able to kill spliceosome-mutant epithelial and hematologic tumor cells, has been tested in preclinical assays [123], and is currently under clinical evaluation in a phase I study to evaluate the safety, tolerability, pharmacokinetics, pharmacodynamics, and preliminary activity of H3B-8800 in participants with Myelodysplastic Syndromes, Acute Myeloid Leukaemia, or Chronic Myelomonocytic Leukaemia (NCT02841540). However, further clinical trial efforts will be required to confirm the toxicology, safety and potential benefit of compounds targeting the splicing machinery.

4.2. Targeting splicing regulatory elements

An alternative approach to manipulate and/or reverse splicing alterations, without blocking spliceosome machinery core, is based on targeting regulatory proteins that modulate splicing. Use of these spliceosomal regulators could be directed, for example, to mutated, altered, or overexpressed molecules involved in pathological processes.

Phosphorylation and dephosphorylation of proteins by kinases are key for multiple biological processes like metabolism, transcription, cell cycle progression, cell movement, apoptosis, and differentiation [124-127]. Thus, the potential of kinases as therapeutic targets has received considerable attention. Alternative splicing is also regulated by kinases that phosphorylate/dephosphorylate splicing factors, like the SR proteins [128], which serves as a signal of nuclear localization and facilitates interaction with other splicing factors. These phosphorylation can be performed by SR protein kinases (SRPKs), topoisomerase 1 (TOP1), protein kinase B (PKB/AKT), NIMA-related kinases (NEK2), PRP4 kinase (PRP4K), dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) [129, 130], cAMP-dependent protein kinase (PKA) [131, 132], and by the family of cdc-like kinases (CLKs). Many studies reveal that dysregulation of splicing kinases has an important role in tumorigenesis and therapeutic response [133], as it is the case of SM08502, which has been shown to reduce Wnt pathway by inhibiting CLK activity plus inhibiting SR splicing factor phosphorylation, leading to the disruption of the spliceosome activity. It is currently in a Phase 1 clinical trial to evaluate the safety and pharmacokinetics of orally administered SM08502 in patients with advanced solid tumors [134].

These studies show the great potential of modulating splicing regulators. Although these kinds of drugs are already being proved in solid tumors, more research needs to

be conducted to assess their therapeutic capacity to treat different types of tumoral pathologies, as NETs, PCa or PDAC.

4.3. Oligonucleotides

One promising approach to solve alterations in splicing process is based on the use of short antisense oligonucleotides (ASO) that can act by blocking the interaction between proteins and RNAs or between two RNAs. Splice-switching antisense oligonucleotides (SSOs) are nucleotides composed by 15-30 synthetic nucleotides or nucleotide analogues, chemically modified to avoid enzymatic degradation of the target RNA, which may specifically bind to a target complementary sequence and thereby block the binding between splicing factors and pre-RNA. SSOs modify the splicing in the nucleus blocking splice sites and sequences that must be recognized by splicing factors [135]. ASOs therapy has already been applied to treat certain diseases such as Spinal Muscular Atrophy and have been approved to treat clinical disorders [136]. Their use in cancer is under evaluation [137], after promising preclinical showing the potential of these oligonucleotides in various tumoral pathologies [138, 139]. Thus, ASOs have been employed in models of SCLC [74, 140], in CRPC in [141], even focused on targeted inhibition of splicing isoforms key genes in PDAC as *KRAS* [142] or *BCL-X* [143-145].

Aims of the study

The **general hypothesis** of this Thesis derives from the emerging notion that, during the development and progression of cancer, there are splicing-related molecular dysregulations that substantially influence tumor behaviour. In this context, we propose that a careful screening and characterization focused in spliceosomic processes and elements, particularly in the spliceosome molecular machinery, can provide discoveries of key importance to understand the normal physiological regulation of the cells, to elucidate alterations that contribute to tumor development and aggression, and to identify specific components and mechanisms that can serve as novel tools to devise potential diagnostic and prognostic biomarkers, and actionable therapeutic targets.

To test this general hypothesis, we proposed the following **objectives**:

Objective 1: To quantitatively assess and analytically examine the expression profile of the core components of the splicing machinery, the associated splicing factors and the most relevant subsidiary markers in representative cohorts of clinically well-characterized pancreatic neuroendocrine tumors, and prostate and pancreatic adenocarcinomas in comparison with non-tumoral samples used as reference-controls.

Objective 2: To identify key specific altered spliceosomic-related components (based on the results of objective 1) and determine the association between their dysregulation and the clinical characteristics of the patients (e.g. disease progression, evolution, survival, etc.), assessing the possible pathophysiological implications of these elements and associations in different cancer types.

Objective 3: To explore the potential pathophysiological role of a set of the spliceosomic-related molecules previously selected on the basis of their altered profile and clinically relevant associations (results of objectives 1 and 2), by examining their contribution to functional processes linked with tumor progression and aggressiveness, and to the response to drugs currently used in or being tested for the treatment of the

cancers under study, using and ample set of functional assays on in vitro and in vivo preclinical models.

Methodology

1. HUMAN SAMPLES

1.1 Patient samples

In the present Thesis, we studied samples obtained from various cohorts of patients with three different tumoral pathologies (PanNETs, PDAC, and PCa). The Ethics Committee of the Reina Sofia University Hospital (Córdoba, Spain) approved the studies, which were conducted in accordance with the Declaration of Helsinki. Patient clinical parameters were collected. Written informed consent was signed by every patient. FFPE samples were obtained from the Andalusian Biobank.

First, Formalin-fixed paraffin-embedded (FFPE) samples from 20 primary PanNETs were collected, tumor tissue and non-tumor adjacent tissue (used as reference-control) from the same piece were separated by expert pathologists and extracted. Clinical parameters were summarised in **Table 1**. The mean age of PanNETs patients was 55-year-old.

Table 1. Summary of clinical parameters of PanNETs patients.

Number of samples	**20**
Age (mean; years)	55 ± 14
Body Mass Index	28.00 ± 3.48 kg/m^2
Gender (female)	57.1 %
Smoking	68.8 %
Family history of neoplasia	12.5 %
Comorbidities	83.3 %

Secondly, for PDAC studies, 75 formalin-fixed paraffin-embedded (FFPE) samples were collected, each one containing PDAC tumor tissue and its corresponding non-tumor adjacent tissue (NTAT). In addition, 24 fresh PDAC samples were obtained from surgical resections. The histopathological classification of tissues as tumor or NTAT and additional sample studies were performed separately by two experienced pathologists. Clinicopathological data of the FFPE cohort are described in **Table 2**. Briefly, individuals

in the study exhibited a median of age of 65-years-old at diagnosis, being in 71 % male. Regarding the cohort, the samples are classified in T2 and T3 stage with absent of metastasis.

Table 2. Summary of clinical parameters of PDAC patients.

Characteristics	Samples (n = 75)
Age, years median (range)	65 (32-76)
Sex, n (%)	
Female	22 (29.3)
Male	53 (70.7)
T stage, n (%)	
T1	5 (6.7)
T2	14 (18.79)
T3	43 (57.39)
T4	9 (12)
NA	4 (5.3)
N stage, n (%)	
N0	25 (33.3)
N1	46 (61.3)
NA	4 (5.3)
M stage, n (%)	
M0	62 (82.7)
M1	7 (9.3)
MX	2 (2.7)
NA	4 (5.3)

Regarding PCa, 42 samples were included, obtained by core needle biopsies, following National Comprehensive Cancer Network (NCCN) guidelines [146]. To use as control, non-tumoral prostate samples were collected from patients after cystoprostatectomy due to bladder cancer but without PCa (n = 12). The appropriate classification of the samples as tumor or non-tumor was confirmed by expert pathologists and it is summarized in **Table 3**. Additionally, demographic, and clinical parameters regarding tumor aggressiveness and metabolic status were collected. Briefly, included individuals exhibited a median of age of 76 years old at diagnosis. Regarding PCa cohort, all the samples had at least a Gleason score of 7, with a 65 % of higher grade. Additionally, a 33 % of the patients suffered extraprostatic extension and a 52 % perineural infiltration.

Table 3. Summary of clinical parameters of PCa patients.

Patients	42
Age, years [median (IQR)]	75 (69 - 81)
PSA levels, ng/mL [median (IQR)]	62.0 (36.2 - 254.5)
Sig PCa	42 (100 %)
Metastasis	28 (66.7 %)

1.1.1 Online data-sets

Several datasets were employed to compare/corroborate results from our samples, contrast the hypothesis of this study, and thereby generate stronger conclusions.

In these databases, we mostly explored information about spliceosome and splicing factors expression, splicing variants dysregulation in tumor pathologies and their relationships with clinical parameters. The selection of the databases was based on the availability of data to compare the expression level of the splicing machinery components in tumor *vs.* control tissue and/or to study their behaviour in response to different clinical parameters.

Specifically, for PDAC studies, gene expression data were downloaded from the public Array Express database E-MTAB-1791 [147], GSE15471 [148], “The Cancer Genome Atlas” (TCGA) databases using cBioPortal (PanCancer Atlas) [149], and GSE79670 [150]. In the case of PCa samples, expression, and clinical data of interest for this study were downloaded from “The Cancer Genome Atlas” (TCGA) database using cBioPortal [5, 6].

1.2 Cell lines

To explore the functional relevance of the selected molecules under study in this Thesis, several human cell lines were employed as models for the pathologies of interest.

Specifically, for PanNET studies BON-1 and QGP-1 cell lines were used. The pancreatic neuroendocrine BON-1 cell line was stablished in 1986 from a peripancreatic lymph node metastasis of a 28-year-old male with PanNET, while QGP-1 cell line derived from primary PanNET tumor tissue obtained from a 61-year-old male [151]. BON-1 cells were kindly provided by Dr. M.C. Zatelli (University of Ferrara, Italy), whereas QGP-1 cells were kindly provided by Dr. K. Öberg (University of Uppsala, Sweden).

To perform PDAC functional assays, we employed the non-tumoral pancreas-derived HPDE E6E7 cell line, generously provided by Dr. F.X. Real, Spanish National Cancer Centre (CNIO), Madrid, Spain, used as a control. In addition, three PDAC model cell lines were acquired from commercial suppliers: Capan-2, BxPC-3, and MIA PaCa-2 (ATCC, Barcelona, Spain). Capan-2 derived from a PDAC primary tumor from a 56-year-old male [152], BxPC-3 cell line was obtained from a non-metastatic 61-year-old woman's adenocarcinoma derived from the body of the pancreas [153], and MIA PaCa-2 cell line (considered the most aggressive of the three studied) is derived from a infiltrative tumor that involved the body and tail of the pancreas from a 65-year-old male [154, 155].

To perform PCa assays we employed two model cell lines: PC-3, androgen-independent with AR and AR-v7 expression, and LNCaP, androgen independent without AR expression, respectively. PC-3 derived from a human prostatic adenocarcinoma metastatic to bone from an 62-year-old male [156] and LNCaP was obtained from a 50-year-old male affected by metastatic prostate carcinoma [157].

These cell lines grew in a constant humidified 37 ºC atmosphere with 5.0 % CO_2 and were monthly checked for mycoplasma contamination by PCR as previously reported [158].

1.2.1 Culture of cell lines

Cell lines were cultured according to manufacturer instructions as detailed below.

The carcinoid-like BON-1, cultured in Dulbecco's Modified Eagles Medium complemented with F12 (DMEM-F12; Life Technologies, Barcelona, Spain) and the somatostatinoma-derived QGP-1 cells were grown in RPMI 1640 (Lonza, Basel, Switzerland), both supplemented with 10 % foetal bovine serum (FBS; Sigma-Aldrich, Madrid, Spain), 1 % glutamine (Sigma-Aldrich) and 0.2 % antibiotic (Gentamicin/ Amphotericin B; Life Technologies).

The HPDE E6E7 cell line was cultured in Keratinocyte Serum Free Medium (Gibco, Madrid, Spain) containing two mandatory additives [(bovine pituitary extract (BPE) and human recombinant epidermal growth factor (EGF)] and 1 % antibiotic-antimycotic (Gentamicin/ Amphotericin B). Capan-2 were cultured in McCoy's 5A Medium (Gibco) supplemented with 10 % FBS (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich) and 0.2 % antibiotic-antimycotic. BxPC-3 were cultured in RPMI 1640 medium (Lonza, Basel, Switzerland) with 2 mM L-glutamine, and 0.2 % antibiotic-antimycotic. MIA PaCa-2 were cultured in Dulbecco's Modified Eagle's Medium with 4,500 mg/L of glucose (DMEM 4.5 g/l glucose) supplemented with 10 % FBS, 2.5 % Horse Serum [107], 2 mM L-glutamine and 0.2 % antibiotic-antimycotic.

PC-3 and LNCaP were cultured in RPMI 1640 (Lonza), both supplemented with 10 % FBS (Sigma-Aldrich), 1 % glutamine (Sigma-Aldrich) and 0.2 % antibiotic (Gentamicin/ Amphotericin B).

1.2.2 PDX-derived tumor cell lines and CSC-Enriching Culture

PDAC patient-derived xenografts (PDAC PDX) were obtained from Dr. Manuel Hidalgo under a Material Transfer Agreement with the CNIO, Madrid, Spain (Reference no. I409181220BSMH) and were originally described and genetically characterized [159]. To establish primary A6L, 215, 253 and 354 PDX-derived cultures, PDXs were enzymatically digested, resuspended and cultured in RPMI 1640 medium supplemented with 10 % FBS and 50 U/mL penicillin/streptomycin. All cultures were tested for mycoplasma at least every 4 weeks.

To enrich for CSCs, 1,000 cells from each cell line were seeded in 24-well Corning Costar ultra-low attachment plate (Merck, Madrid, Spain) to avoid cell attachment and differentiation. Cells were cultured in DMEM-F12 (Thermo Fisher Scientific, Madrid, Spain) supplemented with B-27 (Gibco) and FGF (PreproTech EC, London, U.K.). Numbers of spheres were determined by microscopy using an inverted EVOS FL microscope (Thermo Fisher Scientific) using a 10X objective with phase contrast.

1.2.3 Freezing and Thawing cell lines

During this Thesis, several cell lines were employed, which required a correct maintenance. To freeze cell lines, one million cells were resuspended in its appropriate complete medium with 10 % of FBS, and 5 % of dimethyl sulfoxide (DMSO) was added until 1 mL in a cryotube. After their maintenance in an isopropanol chamber store at 80ºC for 24-, 48- h, cryotubes were stored under liquid nitrogen conditions until use. For

thawing cells, cryotubes were fast warmed in a water bath and 5 mL of complete medium was added to dilute and block DMSO activity. Cells were centrifuged, resuspended with their corresponding medium, and seeded in a 25 cm^2 flask.

1.2.4 Reagents

The following drugs were used to achieve the goals of different studies within this Thesis:

- Everolimus (Sigma-Aldrich) was used at a concentration of 100 nM.
- Lanreotide (Sigma-Aldrich) was used at concentration of 100 nM.
- Sunitinib (Sigma-Aldrich) was used at concentration of 100 nM.
- Pladienolide-B (Santa Cruz Biotechnology, Bergheimer, Germany) was initially used in the 0.01 - 100 nM range.
- Gemcitabine (Santa Cruz Biotechnology) was used at a concentration of 100 nM.

1.2.5 Transfections with siRNA and plasmids

Several genetic alterations were performed through different transfection assays to achieve the goals of our studies, as it is described below.

To overexpress selected genes (detailed below), 150,000 cells were seeded in 6-well plates and transfected using Lipofectamine 2000 Transfection Reagent (Invitrogen, Madrid, Spain) according to manufacturer's instructions. Empty plasmid was used as negative control in all experiments. On the other hand, to achieve the silencing of selected genes specific commercial siRNAs were employed and reduction in mRNA expression and functional consequences were compared with a commercial scramble siRNA. Briefly, 150,000 cells were seeded in 6-well culture plates and transfected with siRNA using

Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, Thermo Fisher), following the manufacturer's instructions. Cells were detached after 48 h of transfection to seed them for transfection validation (qPCR and Western blot) and carrying out functional assays. The experiments were performed at least in triplicate per cell line on independent days.

- Specific plasmid (SC111360, Origene, Rockville, MD, USA) was used to increase *CELF4* expression levels at final concentration of 1 µg cells (BON-1, and QGP-1); empty pCDNA3.1 plasmid ("mock") was used as control.
- BON-1, and QGP-1 cells were transfected with 100 nM of *CELF4* siRNA (SR311214, Origene).
- 1 µg of plasmids was used to overexpress *PRPF8* (#SC116070, Origene) and *RBMX* (#RC200777, Origene) in Capan-2, and BxPC-3 cell lines, empty pCMV6-XL4 plasmid was used as negative control in *PRPF8* experiments, and pCMV6-Entry empty plasmid in *RBMX* experiments.
- HPDE E6E7, Capan-2, BxPC-3, and MIA PaCa-2 cells were transfected with *SF3B1* specific siRNA previously validated in our laboratory (s23851; Thermo Fisher) at 75 nM [95, 160].
- PC-3, LNCaP, Capan-2, and MIA PaCa-2 cells were transfected with two different *SRSF2* specific siRNAs (ID:12628 and 12444; Thermo Fisher) at 30 nM.

1.3 FUNCTIONAL ASSAYS

1.3.1 Proliferation assay

Cell proliferation/viability in response to gene modulations and/or drug administrations was evaluated using Alamar-Blue fluorescent assays as previously reported [86]. Briefly, cells were seeded in 96-well plates at a density of 3,000–5,000 cells/well, reduction of Alamar-Blue Reagent (Bio-Source International, Camarillo, CA, USA) was measured at 0-, 24-, 48-, and 72 h with 10 % Alamar-Blue after 24 h starvation

with 0 % FBS complete medium, by measurement of fluorescent signal exciting at 560 nm and reading at 590 nm (Flex Station 3, Molecular Devices, Sunnyvale, CA, USA).

1.3.2 Wound-Healing assay

Cell migration was evaluated by wound-healing assay. This experiment is based on cell capacity to migrate in two dimensions. Briefly, 50,000 cells were seeded in a 96-well Essen ImageLock plate (Essen BioScience, Ann Arbor, Michigan, USA) and grown to confluence. Scratches were then made in the plate using 96-pin WoundMaker (Essen BioScience). An inverted microscope with a digital camera was used to take wound photos at 40x magnifications in the moment of scratching and after 24 h. Images were analyzed and evaluated using ImageJ-1.51s software.

1.3.3 Colony formation assay and sphere formation

In order to analyze the percentage of tumor initiating cells in the different studies performed, colony and sphere formation were evaluated.

Colony formation assay is based on the ability of a cell to growth when it is isolated. We evaluated this feature on different tumor cell lines and PDX-derived cell lines in response to gene overexpression or silencing, or drug treatment. To this end, 5,000 (cells lines) or 2,000 (PDX-derived cell lines) cells were seeded in 6-well plates and incubated for 10-days, changing medium every 3-days. In the case of Pladienolide-B experiments, cells were treated for 24, 48 and/or 72 h with vehicle or Pladienolide-B before the seeding. After incubation, cells were fixed with Crystal Violet. Cell lines-derived colony numbers were evaluated using ImageJ-1.51s software. PDX-derived cell lines were washed and incubated with 500 µL PBS containing 10 % SDS. Colonies lysates were examined at

520 nm (Synergy™-HT-Multi-Mode Microplate-Reader; BioTek, Winooski, Vermont, USA).

The study of sphere formation was based on the ability of survival of cancer cells in a low adherence condition growing in cluster or “spheres”. To assess sphere formation, 1,000 cells were seeded in a 24-well Corning Costar ultra-low attachment plate (Sigma-Aldrich) in DMEM F-12 medium (Gibco) supplemented with EGF (20 ng/ml) and FGF (20 ng/ml) for PDAC cells (MIA PaCa-2, BxPC-3 and Capan-2) and with 20 ng/mg EGF in case of PCa cells (PC-3 and LNCaP) for 10 days when sphere numbers were determined in all the cases. Treatments were added at the moment of plating and refreshed every 3 days.

1.3.4 Apoptosis assay

To evaluate the apoptotic rate for PDAC cell lines, 5,000 cells/well were seeded in white 96-well plates and cultured for 24 h with Pladienolide-B treatment or vehicle, and apoptotic rate was measured using Caspase-Glo 3/7 Reagent (Promega), following the manufacturer’s instructions [161]. For Annexin-V staining, floating and attached cells were pooled and resuspended in 1X Annexin-V staining buffer containing Annexin-V-FITC diluted 1:20 (Cat no. 29001, Biotium, Freemont, CA) and then, incubated for 20 min at room temperature prior to flow cytometric analysis. Cytometry data was acquired with an Invitrogen™ Attune™ NxT 4-laser cytometer with software version 3.1.1.

1.3.5 Flow Cytometry

Primary pancreatic cells (monolayers and spheres) were trypsinized and resuspended in Sorting Buffer (3 µM EDTA, and 3 % FBS in 1X PBS). To identify CD133

positive CSC, the following conjugated antibodies were used: anti-CD133/1-APC or PE; (Miltenyi), and appropriate isotype-matched control antibodies. For autofluorescent detection, cells were excited with blue laser 488 nm and selected as intersection with the filters 530/30 (BL1) and 590/40 (BL2)[159]. For all assays, 2 mg/mL DAPI (Cat no. D9564, Sigma) was used to exclude dead cells with laser VL1. Data were analyzed with FlowJo 9.3 software (Tree Star Inc., Ashland, OR). Cytometry data was acquired with an Invitrogen™ Attune™ NxT 4-laser cytometer with software version 3.1.1.

1.4 MOLECULAR ASSAYS

1.4.1 RNA extraction

Total RNA from FFPE was extracted using Maxwell MDx 16 Instrument (Promega, Madrid, Spain) with the Maxwell 16 LEV RNA FFPE Kit (Promega, Madison, USA) according to the manufacturer's instructions which is designed to obtain optimal purification of RNA from FFPE tissue samples in an easy and safe manner. Lysates are placed into the cartridges, and, after proteinase K digestion, RNA was obtained isolated in RNase-free water.

Total RNA was isolated from PanNET, PDAC, PCa and PDX-derived PDAC cell lines using TRIzol Reagent (Invitrogen, Barcelona, Spain) following the manufacturer's instructions, and was treated with DNase (Promega, Barcelona, Spain). Quickly, cells were incubated in 6-well plates and were washed with 1X PBS before its collection with TRIzol. 600 µL was added and collected, phenol-chloroform protocol was followed to obtain the aqueous phase which was concentrated and purified with 2-propanol precipitation and 70 % ethanol washing steps. Samples were dried and resuspended in RNase-free water.

Tumors from xenografted mice models were isolated by using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen) following the manufacturer's instructions with the objective of have the maximal recovery of these limited samples. Briefly, samples were lysed using a pestle and homogenized in Buffer RLT. Two different columns were provided to have an efficient isolate of genomic DNA, RNA, and proteins. Samples were resuspended in DNase-, and RNase-free water, respectively.

In every case, the amount of RNA recovered and its purity (before and after DNase treatment) was determined using the NanoDrop 2000 (Thermo Fisher Scientific).

1.4.2 Total RNA retrotranscription to cDNA

After RNA extraction, 1 µg was reverse transcribed to cDNA using random hexamer primers [First Strand Synthesis (MRI Fermentas, Hanover, MD)] in a 20 µL volume. To this end, 1 µL of random hexamer primers was mixed with 1 µg of each RNA and was incubated for 5 min at 65 ºC. Then, 4 µL of appropriate buffer, 2 µL of dNTPs, 1 µL of Ribolock, and 1 µL of reverse transcriptase (all of them provided in the specific kit) were added to each sample. Samples were maintenance for 1 h at 42 ºC and finally reaction was stopped by 70 ºC for 5 min.

1.4.3 Conventional PCR

cDNA was amplified with the ThermoFisher PCR Kit (Thermo Fisher Scientific). For each reaction, 9.5 µL of master mix, 12.5 µL of water, 1 µL of each primer (forward and reverse; Table 4) and 1 µL of cDNA were used. A SuperCycler Trinity (Kyratec, Australia) thermocycler was used, with the following template: 1) 95 ºC for 2 min followed by 40 cycles of 2) 95 ºC for 15 s, 60 ºC for 15 s and 72 ºC for 30 s. 3) In the last step, PCR

product elongation was extended for an additional 5 min at 72 ºC. PCR products were mixed with 5 µL of loading buffer and resolved in a 2 % agarose gel via electrophoresis with 100 bp molecular-weight size marker (Thermo Fisher Scientific).

1.4.4 Quantitative real-time PCR (qPCR)

Quantitative real-time PCR (qPCR) was performed to measure RNA expression levels in human and mice samples, and cell lines using the Brilliant III SYBR Green qPCR Master Mix (Stratagene, La Jolla, CA). Each reaction was made using 10 µL of SYBER Green, 8.4 µL of Water, 0.3 µL of Forward Primer, 0.3 µL of Reverse Primer and 1 µL of cDNA (50 ng/µL). The reactions were done using the Stratagene Mx3000p system and using previously reported thermal profile [24] with the following template: an initial denaturalization of 3 min at 95 ºC, followed by 40 cycles of 95 ºC for 20 s denaturalization and 60 ºC for 20 s annealing/extension. Specific primers for transcripts studied were designed with Primer3 and Primer Blast software are shown in **Table 4**.

Table 4. Summary of primers used for qPCR and sequencing experiments. Size (bp) = length of the sequences amplified by each pair of primers.

Transcript	Sense	Antisense	Size
ACTB	ACTCTTCCAGCCTTCCTTCCT	CAGTGATCTCCTTCTGCATCCT	176
BCL-XL	GATGGCCACTTACCTGAATGA	TGCTGCATTGTTCCCATAGA	94
BCL-XS	GAGCTTTGAACAGGATACTTTTGTG	GAAGAGTGAGCCCAGCAGAA	97
CASP3	GTCTCAATGCCACAGTCCAGT	TTTTTCAGAGGGGATCGTTG	97
CELF1	AACAGAAGAGAATGGCCCAGC	TGCTGAAGGAGTGCTAAATACTG	121
ERB2	CTGTGTTCCATCCTCTGCTG	TGCCTGTCCCTACAACTACCTT	97
ESRP1	TTTTGGGATCACTGCTGGGG	TGTCCCACCTTCTTGTTGGC	108
ESRP2	AGAGCCCAGCAGTCAATTGTT	GTCTCACTGTCCACCACATCAG	96
GAPDH	AATCCCATCACCATCTTCCA	AAATGAGCCCCAGCCTTC	122
HPRT1	CTGAGGATTTGGAAAGGGTGT	TAATCCAGCAGGTCAGCAAAG	157
KHDRBS1	GAGCGAGTGCTGATACCTGTC	CACCAGTCTCTTCCTGCAGTC	106
KLF4	ACCCACACAGGTGAGAAACC	ATGTGTAAGGCGAGGTGGTC	170
KRAS	CTTGGATATTCTCGACACAGCA	AAAGAAAGCCCTCCCCAGT	83
KRAS4A	ACAGAGAGTGGAGGATGCTTTTT	AGCCAGGAGTCTTTTCTTCTTTG	92

MAGOH	GCCAACAACAGCAATTACAAGA	TTATTCTCTTCAGTTCCTCCATCAC	88
MIK67	GACATCCGTATCCAGCTTCCT	GCCGTACAGGCTCATCAATAAC	139
NANOG	TGAACCTCAGCTACAAACAGGTG	AACTGCATGCAGGACTGCAGAG	162
NFKB1	CCTGTCCTTTCTCATCCCATC	TGCCAGAGTTTCGGTTCACT	85
NODAL	AGCATGGTTTTGGAGGTGAC	CCTGCGAGAGGTTGGAGTAG	160
NOVA1	TACCCAGGTACTACTGAGCGAG	CTGGTTCTGTCTTGGCCACAT	124
OSCT3/4	CTTGCTGCAGAAGTGGGTGGAGGAA	CTGCAGTGTGGGTTTCGGGCA	169
PRPF40A	GCTCGGAAGATGAAACGAAA	TGTCCTCAAATGCTGGCTCT	130
PRPF8	TGCCCACTACAACCGAGAA	AGGCCCGTCCTTCAGGTA	139
PTBP1	TGGGTCGGTTCCTGCTATT	CAGATCCCCGCTTTGTAC	111
RAVER1	GTAACCGCCGCAAGATACTG	CGAAGGCTGTCCCTTTGTATT	126
RBM17	CAAAGAGCCAAAGGACGAAA	TACATGCGGTGGAGTGTCC	107
RBM22	CTCTGGGTTCCAACACCTACA	GGCACAGATTTTGCATTCCT	137
RBM3	AAGCTCTTCGTGGGAGGG	TTGACAACGACCACCTCAGA	98
RBM45	CCCATCAAGGTTTTCATTGC	TTCCCGCAGATCTTCTTCTG	123
RNU11	AAGGGCTTCTGTCGTGAGTG	CCAGCTGCCCAAATACCA	108
RNU12	ATAACGATTCGGGGTGACG	CAGGCATCCCGCAAAGTA	106
RNU2	CTCGGCCTTTTGGCTAAGAT	TATTCCATCTCCCTGCTCCA	116
RNU4	TCGTAGCCAATGAGGTCTATCC	AAAATTGCCAGTGCCGACTA	103
RNU4ATAC	GTTGCGCTACTGTCCAATGA	CAAAAATTGCACCAAAATAA	85
RNU6	CGCTTCGGCAGCACATATA	AAAATATGGAACGCTTCACGAA	101
RNU6ATAC	TGAAAGGAGAGAAGGTTAGCACTC	CGATGGTTAGATGCCACGA	112
SF3B1	CAGTTCCGTCTGTGTGTTCG	GCTGCCTTCTTGCCTTGA	101
SF3B1	CAGTTCCGTCTGTGTGTTCG	GCTGCCTTCTTGCCTTGA	101
SF3B1 tv1	GCAGACCGGGAAGATGAATA	TTTTCCCTCCATCTGCAAAA	88
SFPQ	TGGTAGGGGGTGAAAGTG	TTAAAAACAAGAAATGGGGAAATG	125
SND1	ACTACGGCAACAGAGAGGTCC	GAAGGCATACTCCGTGGCT	101
SNRNP200	GGTGCTGTCCCTTGTTGG	CTTTCTTCGCTTGGCTCTTCT	103
SNRNP70	TCTTCGTGGCGAGAGTGAAT	GCTTTCCTGACCGCTTACTG	114
SNW1	ATGCGTGCCCAAGTAGAGAG	TCCCCATCCTCTTTTTCCA	134
SOX2	AGAACCCCAAGATGCACAAC	CGGGGCCGGTATTTATAATC	154
SRRM1	GTAGCCCAAGAAGACGCAAA	TGGTTCTGTGACGGGGAG	108
SRRM4	CCTTCACCACCTCCTCAC	TTCGGCACATTCCAGACA	113
SRSF1	TGTCTCTGGACTGCCTCCA	TGCCATCTCGGTAAACATCA	98
SRSF10	CTACACTCGCCGTCCAAGAG	CCGTCCACAAATCCACTTTC	103
SRSF2	TGTCCAAGAGGGAATCCAAA	GTTTACACTGCTTGCCGATACA	113
SRSF3	TAACCCTAGATCTCGAAATGCATC	CATAGTAGCCAAAAGCCCGTT	117
SRSF4	GGAACTGAAGTCAATGGGAGAA	CTTCGAGAGCGAGACCTTGA	110
SRSF5	GCAAAAGGCACAGTAGGTCAA	TTTGCGACTACGGGAACG	92
SRSF6	AGACCTCAAAAATGGGTACGG	CTTGCCGTTCAGCTCGTAA	82
SRSF9	CCCTGCGTAAACTGGATGAC	AGCTGGTGCTTCTCTCAGGA	87
TCERG1	GAGGAGCCCAAAGAAGAGGA	CACCAGTCCAAACGACACAC	112
TIA1	TAAATCCCGTGCAACAGCAGA	TATGCAGGAACTTGCCAACCA	124
TP53	AAGGAAATTTGCGTGTGGAG	CCAGTGTGATGATGGTGAGG	180

TP53Δ133	ACTCTGTCTCCTTCCTTACAG	GTGTGGAATCAACCCACAGCT	132
TRA2A	TCAAAGGAGGCTATGGAAAGG	TGTGTGCGCTCTCTTGGTTA	90
TRA2B	GATGATGCCAAGGAAGCTAAAG	AGGTAGGTCTCCCCATGTAAATTC	130
U2AF1	GAAGTATGGGGAAGTAGAGGAGATG	TTCAAGTCAATCACAGCCTTTTC	120
U2AF2	CTTTGACCAGAGGCGCTAAA	TACTGCATTGGGGTGATGTG	130

1.4.5 qPCR dynamic array based on microfluidic technology

A quantitative PCR dynamic array based on microfluidic technology was used to simultaneously measure the expression of 96 genes in 96 samples of PDAC tumor samples and non-tumor adjacent samples, as previously reported by our group [162]. Biomark System and FluidigmVR Real-Time PCR Analysis Software v.3.0.2 and Data Collection Software v.3.1.2 (Fluidigm) were used to obtain RNA expression levels in these samples. Primers for specific human genes were designed with Primer3 and Primer Blast software (see **Table 4**). This technique is based on the microfluidic technology, minimizing sample handling, and thus reducing errors. All the steps were performed following the manufacturer's instructions and advice.

Briefly, to perform the primer mix, 1 µL of all the primers (forward and reverse) of the genes that were to be measured was collected, and up to 200 µL of final volume completed with DNA Suspension Buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA; TEKnova, PN T0221). The final concentration of each assay was 500 nM.

To prepare the sample pre-mix, after reverse transcription of the samples, a preamplification protocol to increase the number of copies of the target DNA was done. The components of the Pre-mix reaction are detailed in **Table 5**.

Table 5. Summary of pre-amplification reaction components.

Reaction components	Volume (µL)
PreAmp Master Mix	52.8
Pooled DELTAgene Assay Mix (500 nM)	26.4
Water	118.8
Total	198

In a PCR plate, 3.75 µL of pre-mix was added with 1.25 µL of cDNA for each sample, making a total volume of 5 µL. These samples were processed by the following template: 2 min at 95 ºC, following the 10 cycles of 15 s at 95 ºC for denaturalization, and 4 min at 60 ºC (annealing/extension). Next, an exonuclease treatment was carried out in T100 Thermal-cycler (BioRad) for 30 min at 37 ºC to digest the cDNA following the 15 min at 80 ºC to inactivate the enzyme. To this end, 2 µL of Exonuclease I (4 U/µL) (Biolabs) to each sample was added. Finally, samples were diluted with 1x TE Buffer at pH 8.3 (Molecular Probes) in a 5-fold dilution.

The control line fluid was injected into the accumulator of the chip, and it was placed into the IFC controller. When the script was finished, 5 µL of each assay and 5 µL of each sample (3.3 µL of the preamplificated sample was mixed with 2.5 µL of EvaGreen Supermix with Low ROX (Bio-Rad, PN 172- 5211) and 0.25 µL DNA Binding Dye Sample Loading Reagent 20X (Fluidigm, PN 100- 3738) were loaded into their respective inlets on the chip, and the Load Mix script in the IFC controller software was run. After this program, the chip is put in the Biomark System following the manufacturer's protocol (Fluidigm). Data were processed with Real-Time PCR Analysis Software 3.0 (Fluidigm).

1.4.6 Western Blotting

Cells were cultured (250,000/well, 12-well plates) for 24 h before collecting them. In Pladienolide-B studies, cells were treated during 24 h with the specific concentration of

Pladienolide-B or vehicle. For all the cases, medium was removed and 300 µL of pre-warmed SDS-DTT buffer (62.5 mM Tris-HCl, 2 % SDS, 20 % glycerol, 100 mM DTT and 0.005 % bromophenol blue) at 65 ºC was added to lyse the cells. Samples were sonicated for 10 seconds and boiled for 5 min at 95 ºC. Extracted protein samples were separated in 12.5 % polyacrylamide gels by SDS-PAGE, transferred to a nitrocellulose membrane (Ref. 1704270, Millipore) and blocked with 5 % non-fat dry milk in Tris-buffered saline with 0.05 % Tween-20 (Ref. 93773, Sigma-Aldrich). Membranes were then incubated with the following primary antibodies: CELF4 (Ref. ab171740, abcam), SF3B1 (Ref. ab172634, abcam), SRSF2 (Ref. ab204916, abcam), phospho-ERK1/2 (Ref. 4370S, Cell Signaling Technology; Danvers, MA), phospho-AKT (Ref. 9271S, Cell Signaling Technology), phospho-JNK (Ref. AF1205, R&D Systems; Minneapolis, MN), total ERK1/2 (SC-154, Santa Cruz Biotechnology; Santa Cruz, CA), total AKT (Ref. 9272S, Cell Signaling Technology), total JNK (Ref. AF1387, R&D Systems). Then, horseradish peroxidase-conjugated goat antirabbit IgG (Ref. 7074, Cell Signaling Technology) was used to incubate the membranes for 1 h. Bond antibodies were visualized using Clarity 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 and ponceau staining was used to normalize with total protein loading.

1.4.7 mTOR phospho-antibody array

To study the potential changes in mTOR pathway after *CELF4* silencing, we carried out an antibody phosphoarray based on fluorescent detection. Two slides were employed to measure mTOR activity under siCELF4 and Scramble (used as control) conditions. All procedures required to perform protein extraction, biotinylation of proteins, its conjugation

to antibody array, and detection by Dye-Streptavidin were performed following the manufacturer's instructions with the reagents provided by the assay kit. A total of 5 x 10^6 of QGP-1 and BON-1 cells were seeded in 25 cm^2 flasks and after 24 h of transfection, the culture was washed with 1X PBS for 5 times and cells were collected using a scraper and 200 µL of extraction buffer to prevent protein degradation and dephosphorylation. After cell silencing, cells were lysed with a non-denaturing extraction buffer provided in the Antibody Array Assay Kit. Lysate samples protein concentration was measure by UV absorbance A280 using the NanoDrop 2000 (Thermo Fisher Scientific). Biotinylation of protein samples was perform with Biotin/DMF solution and was detected by Cy3-streptavidin. The conjugation was scanning at the Laboratory of Genetics at SCAI (UCO) using Axon GenePix 4000B. The information regarding specificity, detectability and reproducibility for the assay can be accessed at the company website.

1.4.8 Immunohistochemistry (IHC) analysis

IHC analysis of key proteins was performed on formalin-fixed, paraffin-embedded (FFPE) sample blocks, comprising tumor and non-tumor adjacent tissues using ImmPRESS UNIVERSAL REAGENT Anti-Mouse/Rabbit IgG PEROXIDASE (Vector Laboratories, Maravai LifeSciences, Barcelona, Spain). CELF4 antibody was used at 1:50, SF3B1 was used at 1:250. Staining was evaluated by assessing a combined score comprising the percentage of positive cells (0%=0, 1-25%=1, 26-50%=2, 51-75%=3, 76-100%=4) multiplied by the intensity (no staining = 0, weak staining = 1, moderate = 2, strong = 3), ranging from 0 to 12. Two expert pathologists performed IHC analyses of samples using a blinded protocol to determine protein staining intensities in samples.

1.4.9 Confocal microscopy

SF3B1 was analyzed in HPDE E6E7 and MIAPaCa-2 cell lines after 24 h of treatment with vehicle or Pladienolide-B, and after SF3B1 silencing. Briefly, cell lines were grown in glass coverslips and fixed with 4 % PFA. SF3B1 (1:250), Wheat Germ Agglutinin, Alexa Fluor 647 Conjugate (W32466, Thermo Fisher) was used to label membrane cells (1:300), Alexa Fluor 488 Donkey anti-Rabbit secondary antibody (1:500) (A-21206, Thermo Fisher), and nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich). Samples were visualized with a LSM710 confocal laser-scanning microscope (Carl Zeiss, Jena, Germany), images were processed using the Huygens Essential software package (version 2.4.4; SVI, Hilversum, The Netherlands), and analyzed with ImageJ to study SF3B1 cell distribution.

1.5 ANIMAL MODELS

1.5.1 Zebrafish breeding, *in vivo* xenograft assays and image analysis

Zebrafish embryos were obtained by crossing adults (*Danio rerio*, wild type). Zebrafish adults were maintained in 30 L aquaria with a ratio of 1 fish/liter of water, a 14:10 day/night cycle and a water temperature of ≈ 28.5 ºC, according to published procedures [163]. All procedures used in the experiments, fish care and treatment were performed in agreement with the Animal Care and Use Committee of the University of Santiago de Compostela and the standard protocols of Spain (Directive 2012-63-DaUE). At the final point of the experiments, zebrafish embryos were euthanized by tricaine overdose.

Collection of the zebrafish embryos occurred at 0 h post fertilization (hpf). After that, eggs were incubated at 28.5 ºC until 48 hpf. At this point, hatched embryos were anesthetized with 0.003 % of tricaine (Sigma) and injected with MIA PaCa-2 or A6L cells, stably infected with an mCherry-H2B expressing lentivirus as previously described [164], under different treatment conditions (control and Pladienolide-B treated; 1 nM). MIAPaCa-2-mCherry-H2B and A6L-mCherry-H2B cells were incubated at 37 ºC and 5 % CO_2 before injection until they reached a confluence of 70 %. MIAPaCa-2-mCherry-H2B and A6L-mCherry-H2B cell preparations consisted of cell trypsinized and concentrated in a vial at a rate of 10^6 cells per tube for each condition and resuspended with 10 µL of PBS with 2 % of polyvinylpyrrolidone (PVP) to avoid cellular aggregation. For cell injection, borosilicate needles (1 mm O.D. x 0.75 mm I.D.; World Precision Instruments) were used. Between 100-200 cells were injected into circulation in each embryo (Duct of Cuvier) using a microinjector (IM-31 Electric Microinjector, Narishige) with an output pressure of 15 kPA and 10 ms of injection time per injection. Afterwards, embryos were incubated for 6 days post injection (dpi) at 34 ºC in 30 mL Petri dishes with SDTW (Salt Dechlorinate Tap Water). Imaging of the injected embryos were performed using a fluorescence stereomicroscope (AZ-100, Nikon) at 1, 4 and 6 dpi to measure the spreading and proliferation of the injected cells in circulation in the zebrafish for each of the conditions assayed. Quantifish software [165] was used to perform the image analysis of the photographs taken of the embryos at 1, 4 and 6 dpi. Quantifish measures, in each of the images provided, the intensity of the fluorescence and the area of the positive pixel above a certain threshold of the cells. With these parameters, integrated density is obtained allowing for the comparison of different times between images to obtain a proliferation ratio of the cells in the region of the caudal hematopoietic tissue (CHT) of the embryos, where the cells metastasize.

1.5.2 Xenograft mice model

All experimental procedures were carried out following the European Regulations for Animal Care, in accordance with guidelines and regulations, and under the approval of the University of Córdoba Research Ethics Committee.

For CELF4 experiments, 2 x 10^6 BON-1 cells were injected in each flank of 7-week-old male athymic BALB/cAnNRj-Foxn1nu mice (Janvier Labs, Le Genest-Saint-Isle, France; n = 6 mice), resuspended in 100 µL of basement membrane extract. *CELF4* siRNA and it scramble (used as control) were injected into one flank by using AteloGene® reagent (KOKEN Co, #KKN1394) to transfect the siRNA molecule into cells by local administrations according to manufacturer's instructions, when tumors were measurable. Tumor growth was monitored twice per week for 4 weeks, by using a digital caliper and were sacrificed after 15 days of silencing.

For Pladienolide-B experiments, 2 x 10^6 of MIA PaCa-2 cells, resuspended in 100 µL of basement membrane extract, were injected in each flank of 7-week-old male athymic BALB/cAnNRj-Foxn1nu mice (Janvier Labs; n=5 mice). Tumor growth was monitored twice/week for 7-weeks. At the fourth week of grafting, mice were injected intratumorally with 100 µL of Pladienolide-B or with vehicle.

After euthanasia of mice, each tumor was dissected, fixed, processed, and sectioned for histopathologic examination of necrosis after Hematoxylin & Eosin (H&E) staining by expert pathologists. A piece from each tumor was frozen for RNA and protein extraction whenever feasible.

1.6 BIOINFORMATIC ANALYSES

1.6.1 Gene expression and splicing variants analysis

1.6.1.1 CELF4 RNA-Seq analysis

We analyzed 33 transcriptomes of human non-functional PanNETs deposited in NCBI (GSE118014) [166]. Raw paired-end FASTQ files were downloaded to generate a mean of 180 million paired-end reads per sample that were aligned to the human genome (hg19) using Tophat [167]. Differential expressed genes (DEG) were identified using HTSeq and DESeq [168, 169]. Differentially expression of RNA transcript levels was performed with R packages and a minimum of 3 counts per gene in more than two independent samples were required. To perform a clustering for *CELF4* expression, we generated four groups in terms of Q1 (high) or Q4 (low) expression using mclust [170]. A fold change of >1.5 with a q-value<0.05 were used to call DEG. Signaling pathway enrichment was analyzed using the Gene set Enrichment Analysis (GSEA) tool [171] and DAVID Resources [172].

To detect splicing variants, we quantified transcripts using Salmon [173] with the v34 release of human GENCODE transcriptome [174]. To calculate the relative abundances of splicing events, the same high and low expression groups of *CELF4* expression used above were applied to detect a differential splicing (p value<0.1) of the Percent Spliced In index (PSI or Ψ) using SUPPA2 software [175]. Classification of splicing events profiling was established into 7 types of events according to their splicing pattern: skipped exon, mutually exclusive exons, alternative 5' splice site, alternative 3' splice site, retained intron, alternative first exon and alternative last exon (as illustrated in **Figure I11**).

1.6.1.2 RBMX, PRPF8 and SF3B1 RNA-Seq analysis

RNA-Seq data of an additional cohort of 94 PDAC patient samples were included to explore *RBMX, PRPF8* and *SF3B1* expression and splicing profile. Briefly, all samples were fresh frozen, and RNA was isolated using the miRNeasy Mini Kit (Qiagen, Milan, Italy), sample quantification was performed using a Qubit and Bioanalyzer to confirm quantitation and quality, respectively. RNA-Seq libraries were generated using RiboZero rRNA depletion followed by RNA library prep using NEBNext Ultra RNA Directional kit. IIlumina HiSeq2500 v4 was used and libraries were sequenced using PE 75 cycles at 7 samples per lane (> 50 million reads per sample). Subsequently, we performed a similar biocomputational analysis protocol than that used for *CELF4*. To perform a clustering for *RBMX, PRPF8* and *SF3B1* expression, the Salmon quant-files were imported to R [176] and summarized to gene-level using Tximeta [177]. The gene abundances were imported to EdgeR [178, 179] and normalized by the trimmed mean of M-values (TMM) method [180]. TMM-normalized expression values of *RBMX, PRPF8* and *SF3B1* were used to classify the patients according to their expression using mclust [170] into groups using mclust E model (univariate, equal variance), which generated three groups labelled as low, intermediate, and high expression. Subsequently, PSI and TPM values for the low and high *RBMX, PRPF8* and *SF3B1* expression groups were used with SUPPA2 to perform the differential splicing analysis with local events, then splicing differences using delta PSI (ΔΨ) were calculated. The difference in average adjusted PSI from each group, and p value < 0.05 (for *SF3B1)* and p value < 0.01 (for *RBMX,* and *PRPF8)* were considered significant.

PSI values were used to calculate the relative frequency of each splicing event per sample [$Relative\ Frecuency\ (event\ i) = \frac{\Sigma\ PSI\ (event\ i)}{\Sigma\ PSI\ (total\ events)}$] and estimate the splicing event composition per sample. The comparison between the *RBMX, PRPF8* and *SF3B1*

high and low expression groups was tested by Wilcoxon test and Kolmogorov-Smirnov test with significance cutoff at $p<0.05$. Classification of SE profiling was established into the same 7 types of events according to their splicing pattern as detailed above (**Figure I11**).

1.6.1.3 mTOR phospho-antibody array analysis

Results from the measurement of the mTOR phospho-antibody array were provided by the Laboratory of Genetics at Core Research Support Service (SCAI) of the University of Cordoba as a matrix data array, and its analysis was performed with R packages in our lab. Differential expression between samples with CELF4 silencing and its scramble was analyzed using an empirical Bayesian method (limma R package) [181]. A fold change of < 1.5 and p value < 0.05 were used to detect phosphosites. Statistical modeling/machine learning technique provided a way to classify phophosite and identify relevant underlying biomarkers. In that context, the R package PHONEMeS [182] was used, which employs boolean logic models of signaling networks downstream of a perturbed kinase to detect signaling mechanisms and drug modes of action. This package needs CPLEX by randomly downsampling the measurements and retrieving one solution for each interaction. Molecular signaling classification was profiled by Cytoscape [183].

1.7 STATISTICAL ANALYSIS

Samples from all groups were processed at the same time. Statistical differences between two variables were calculated according to normality, assessed by Kolmogorov-Smirnov test, using parametric t-test or non-parametric Mann Whitney U test. For groups

with three or more variables, One-Way ANOVA analysis or Kruskal-Wallis test were performed. To normalize values within treatment and control and minimize intragroup variations in the different experiments, the values obtained were compared with controls (set at 100 %). The receiver operating characteristic (ROC) curves were used to evaluate the suitability of genes to distinguish different groups of samples. Heatmaps, principal component analysis and VIP score were performed through Metaboanalyst software (https://www.metaboanalyst.ca; McGill University, Montreal, Canada). Results from *in vitro* studies were obtained from at least 3 separate independent experiments carried out on different days with different cell preparations. Data were expressed as mean ± SEM, $p<0.05$ was considered statistically significant, asterisks indicate significant differences (* $p<0.05$; ** $p<0.01$; *** $p<0.001$). Analyses were performed with SPSS v.22 (IBM SPSS Statistics Inc., Chicago, IL, USA) and GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA).

Results

Chapter I

Splicing Dysregulation in Neuroendocrine Tumors

The splicing factor CELF4 enhance aggressiveness features in pancreatic neuroendocrine tumors

Alternative splicing is an essential process that enhances biological versatility in eukaryotes by modulating the generation of different RNAs from the same pre-RNA [41, 44]. However, inappropriate functioning of the splicing machinery (spliceosome and splicing factors) generates aberrant splicing variants that can play oncogenic roles. In fact, dysregulation of alternative splicing is being increasingly regarded as a new epigenetic cancer hallmark associated with multiple dysfunctions in tumor cells [3, 66]. In this context, we have previously demonstrated that overexpression of aberrant alternative splicing variants of somatostatin receptor 5 (SST_5TMD4) and ghrelin (In1-ghrelin) are directly associated to malignant features in PanNETs, where they alter signaling pathways and basic cellular processes, thereby enhancing tumor aggressiveness [77, 78]. This led us to recently explore the status of the splicing machinery and its potential role in tumorigenesis in these tumors. Initial results revealed a broad alteration of the splicing machinery and disclosed a plausible role of NOVA1 in PanNETs (DOI:10.1101/2022.02.09.479525). Here, we aim to evaluate the dysregulation and functional role of the splicing factor *CELF4* in PanNETs as well as to assess its potential role as a novel diagnostic marker and treatment target in this pathology.

1. *CELF4* is dysregulated in PanNETs and associated with clinical parameters

CELF4 expression levels were measured in a cohort of 20 primary tumors from patients with PanNETs (DOI:10.1101/2022.02.09.479525), comparing tumor with non-tumoral adjacent tissue, used as reference. This showed that *CELF4* was drastically upregulated in tumor tissues compared to their non-tumor adjacent matching ones (**Fig.**

R1A). Specificity and sensitivity comparisons using ROC curve analysis of risk score showed a high predictive accuracy of the classifying *CELF4* diagnostic, with an area under the curve of 0.892 (p = 0.001) (**Fig. R1B**). Analysis of clinical parameters revealed that *CELF4* expression was associated with lower abdominal pain and metastasis, two relevant malignancy features in PanNETs (**Fig. R1D,E**). Higher levels of CELF4 in tumoral than non-tumoral adjacent tissue were also observed at protein level by immunohistochemistry (IHC) (**Fig. R1C**).

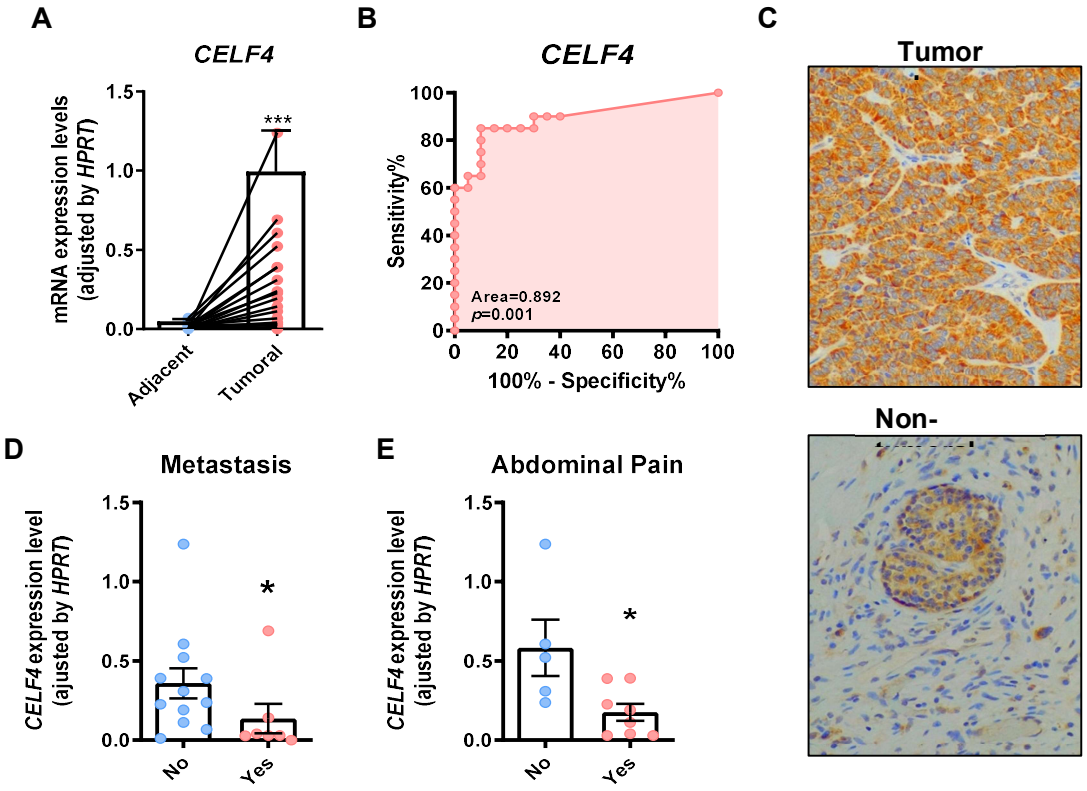
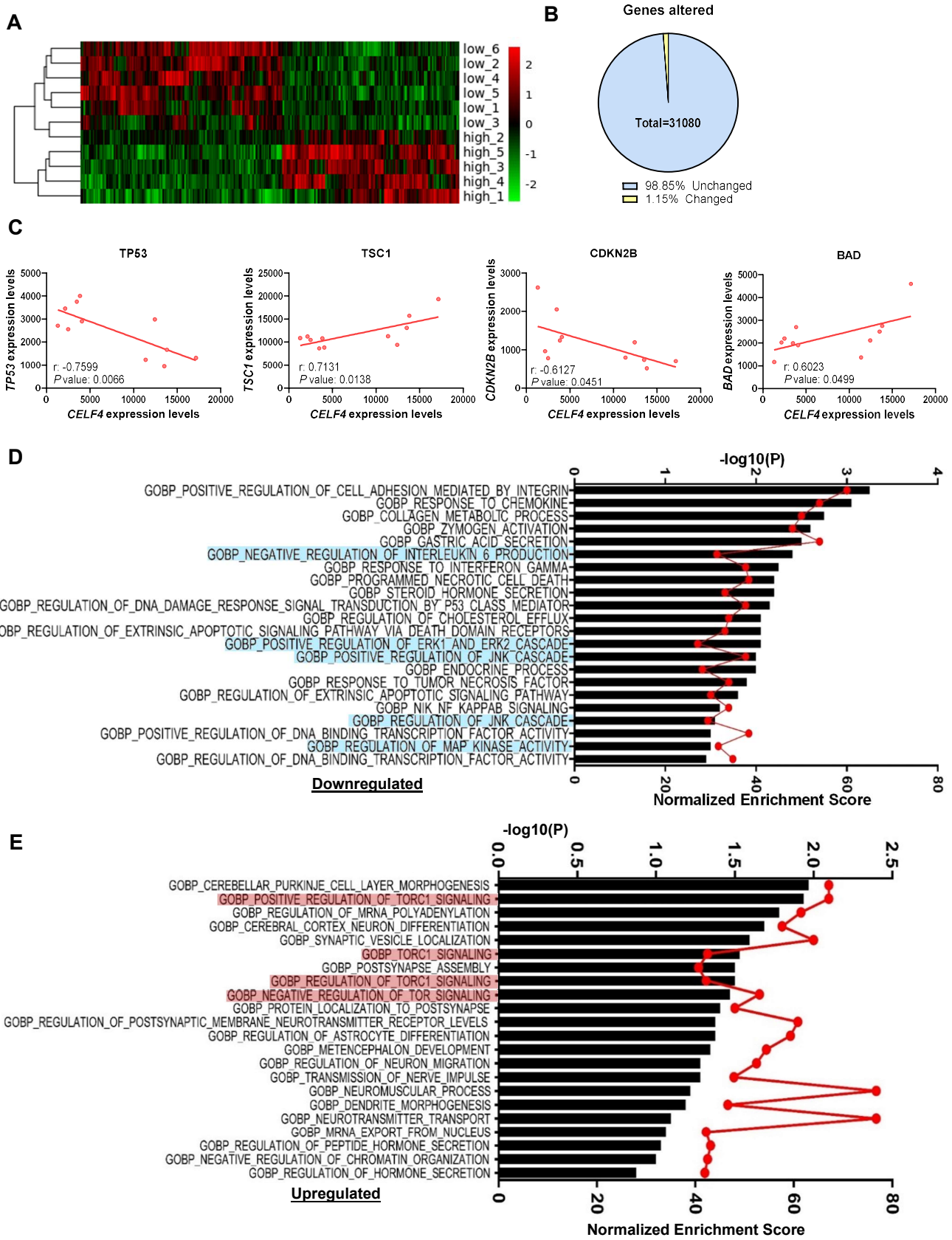


Figure R1. CELF4 dysregulation in PanNETs. **A)** *CELF4* expression levels in FFPE cohort of 20 PanNETs patient cohort, tumor tissue is compared with controls (non-tumor adjacent tissue). Data are represented by absolute mRNA levels normalized by *HPRT* expression levels. **B**) *CELF4* ROC curve in FFPE cohort of PanNETs tissue compared with non-tumor adjacent tissue (used as control). Data are represented by absolute mRNA levels normalized by *HPRT* expression levels. **C**) CELF4 IHC analysis in PDAC FFPE samples *vs*. NTAT. Representative IHC 20X-image. **D, E**) *CELF4* expression levels in tumor tissue FFPE cohort association with clinical parameters (metastasis and abdominal pain). Data are represented by absolute mRNA levels normalized by *HPRT* expression levels. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from control (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

2. Genetic alterations associated to *CELF4* expression levels

To explore the putative significance of CELF4 alteration in PanNETs, we explored a previously published RNA-Seq dataset corresponding to 11 PanNET patients (mean age patients 52.7 years-old; 54 % Males; 90.9 % Low Grade Tumors; GSE118014), which were divided into two groups based on CELF4 expression levels: high (n=5) and low levels (n=6). Unsupervised analysis revealed that low and high CELF4 expressing tumors were clearly segregated according to gene expression (**Fig. R2A**). A total of 357 genes (1.15 %) were differentially expressed according to the expression of CELF4, suggesting that CELF4 may act as a global transcriptional activator in PanNETs. From these, 46.78 % were upregulated and 53.22 % downregulated (**Fig. R2B, Appendix 1**). Specifically, we observed an inverse correlation with the tumor suppressors TP53 and CDKN2B and direct with TSC1 and BAD (**Fig. R2C**). To get further insights into the biological functions affected by differentially expressed genes, we used DAVID software and GSEA to perform KEGG analysis, comparing with the different expression of CELF4. Among the top significant KEGG enriched hits in the low *CELF4* expression of group, relevant relationships were found with IL-6, ERK1 and ERK2, JNK or MAPK activity (**Fig. R2C**). In contrast, high CELF4 expression was closely associated with TORC1 signaling and regulation of mRNA, aside from neuronal-related pathways (**Fig. R2D, Supplemental Figure 1**).

Figure R1 (Next Page). Gene expression profiling in PanNETs with low vs. high *CELF4* expression. **A**) Heatmap diagram depicting the most differentially expressed genes. The colour bar (red high-green low) codifies the gene expression level in fold change. **B**) Genes number representative scheme of genes differentially expressed in low (right) vs. high (left) *CELF4* expression samples. A total of 357 genes were differentially expressed, of which 167 (46.78 %) were upregulated and 190 (53.22 %) were downregulated in low vs. high *CELF4* expression sample groups (right). **C**) Correlation of *CELF4* mRNA levels with key genes. Data represents mean ± SEM. **D, E**) Statistically significant GO biologic functions identified using DAVID Resources of the genes differentially downregulated (**D**) or upregulated (**E**) in low vs. high *CELF4* expression groups, ranked by enrichment score. Enrichment score, black bars (down x-axis); −log P value, filled red circle with red line (up x-axis).


A
low_6
low_2
low_4
low_5
low_1
low_3
high_2
high_5
high_3
high_4
high_1
2
1
0
-1
-2
B
Genes altered
Total=31080
98.85% Unchanged
1.15% Changed
C
TP53
TP53 expression levels
r: -0.7599
P value: 0.0066
CELF4 expression levels
TSC1
TSC1 expression levels
r: 0.7131
P value: 0.0138
CELF4 expression levels
CDKN2B
CDKN2B expression levels
r: -0.6127
P value: 0.0451
CELF4 expression levels
BAD
BAD expression levels
r: 0.6023
P value: 0.0499
CELF4 expression levels
D
-log10(P)
GOBP_POSITIVE_REGULATION_OF_CELL_ADHESION_MEDIATED_BY_INTEGRIN
GOBP_RESPONSE_TO_CHEMOKINE
GOBP_COLLAGEN_METABOLIC_PROCESS
GOBP_ZYMOGEN_ACTIVATION
GOBP_GASTRIC_ACID_SECRETION
GOBP_NEGATIVE_REGULATION_OF_INTERLEUKIN_6_PRODUCTION
GOBP_RESPONSE_TO_INTERFERON_GAMMA
GOBP_PROGRAMMED_NECROTIC_CELL_DEATH
GOBP_STEROID_HORMONE_SECRETION
GOBP_REGULATION_OF_DNA_DAMAGE_RESPONSE_SIGNAL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR
GOBP_REGULATION_OF_CHOLESTEROL_EFFLUX
GOBP_REGULATION_OF_EXTRINSIC_APOPTOTIC_SIGNALING_PATHWAY_VIA_DEATH_DOMAIN_RECEPTORS
GOBP_POSITIVE_REGULATION_OF_ERK1_AND_ERK2_CASCADE
GOBP_POSITIVE_REGULATION_OF_JNK_CASCADE
GOBP_ENDOCRINE_PROCESS
GOBP_RESPONSE_TO_TUMOR_NECROSIS_FACTOR
GOBP_REGULATION_OF_EXTRINSIC_APOPTOTIC_SIGNALING_PATHWAY
GOBP_NIK_NF_KAPPAB_SIGNALING
GOBP_REGULATION_OF_JNK_CASCADE
GOBP_POSITIVE_REGULATION_OF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY
GOBP_REGULATION_OF_MAP_KINASE_ACTIVITY
GOBP_REGULATION_OF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY
Downregulated
Normalized Enrichment Score
E
-log10(P)
GOBP_CEREBELLAR_PURKINJE_CELL_LAYER_MORPHOGENESIS
GOBP_POSITIVE_REGULATION_OF_TORC1_SIGNALING
GOBP_REGULATION_OF_MRNA_POLYADENYLATION
GOBP_CEREBRAL_CORTEX_NEURON_DIFFERENTIATION
GOBP_SYNAPTIC_VESICLE_LOCALIZATION
GOBP_TORC1_SIGNALING
GOBP_POSTSYNAPSE_ASSEMBLY
GOBP_REGULATION_OF_TORC1_SIGNALING
GOBP_NEGATIVE_REGULATION_OF_TOR_SIGNALING
GOBP_PROTEIN_LOCALIZATION_TO_POSTSYNAPSE
GOBP_REGULATION_OF_POSTSYNAPTIC_MEMBRANE_NEUROTRANSMITTER_RECEPTOR_LEVELS
GOBP_REGULATION_OF_ASTROCYTE_DIFFERENTIATION
GOBP_METENCEPHALON_DEVELOPMENT
GOBP_REGULATION_OF_NEURON_MIGRATION
GOBP_TRANSMISSION_OF_NERVE_IMPULSE
GOBP_NEUROMUSCULAR_PROCESS
GOBP_DENDRITE_MORPHOGENESIS
GOBP_NEUROTRANSMITTER_TRANSPORT
GOBP_MRNA_EXPORT_FROM_NUCLEUS
GOBP_REGULATION_OF_PEPTIDE_HORMONE_SECRETION
GOBP_NEGATIVE_REGULATION_OF_CHROMATIN_ORGANIZATION
GOBP_REGULATION_OF_HORMONE_SECRETION
Upregulated
Normalized Enrichment Score

Further analysis revealed that 430 changes in spliceosomic events were associated to *CELF4* expression (Volcano plot **Fig. R3A**) (**Appendix 2**). These splicing pattern alterations were mainly attributable to exon skipping, alternative 5' splice site and alternative first exon splicing events, which were the most altered as compared with normal-overall event pattern (considering *CELF4* expression) (**Fig. R3B**). GO-David analysis of the altered genes unveiled the presence of splicing related-functions, like regulation of transcription and spliceosomal complex assembly. Interestingly, MAPK cascade and MAPKK activity were also notably represented (**Fig. R3C**).

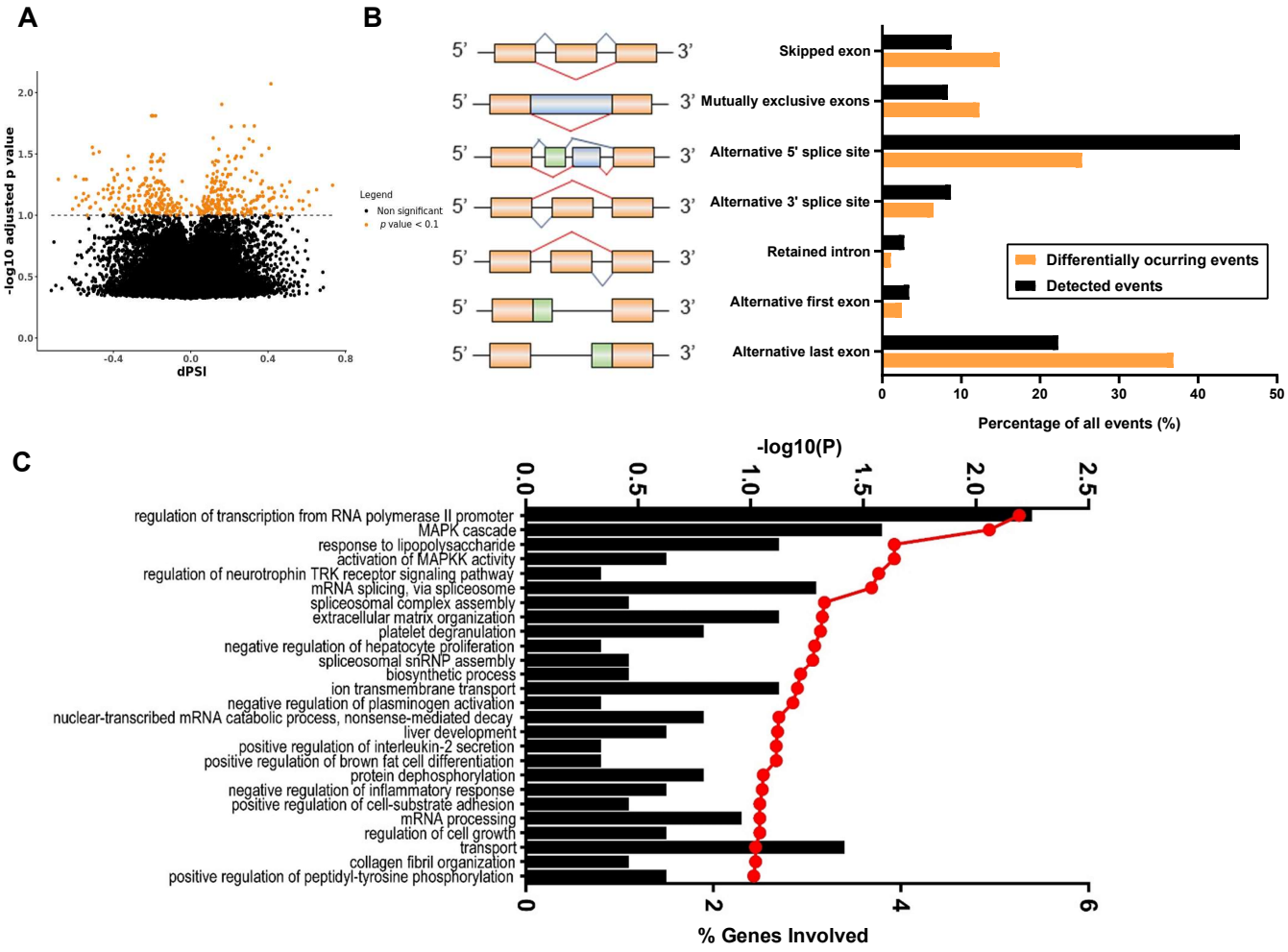


Figure R2. Analysis of splicing events according to CELF4 expression levels in PanNETs. **A**) Volcano-plot where calculated ΔΨ of total events is plotted against the –log P-value of the Fisher's Exact Test to assay differential splicing events between high and low CELF4 expression groups of samples. **B**) Alternative Splicing events characterization of RNA-Seq samples. **C**) Statistically significant GO biologic functions identified using DAVID Resources in relation to CELF4 expression, ranked by −log P-value. Percentage of genes involved, black bars (down x-axis); −log P value, filled red circle with red line (up x-axis).

4. *CELF4* modulation *in vitro* and *in vivo* unveils a possible role in aggressiveness of PanNETs

Having found quantitative associations between *CELF4* expression levels and relevant splicing and signaling pathways linked to key cancer cell functions, we next aimed to explore the role of *CELF4* in PanNET aggressiveness and its potential as therapeutic target. To this end, two widely employed PanNET cell models (BON-1 and QGP-1) were employed. First, *CELF4* expression levels were assessed in the two cell lines (**Fig. R4A**), which showed that both cell lines have appreciable mRNA levels amenable to manipulation through genetic alterations. *CELF4* silencing by specific siRNAs decreased its expression levels by 20 % and 40 % in BON-1 and QGP-1 cells, respectively, as compared to scramble siRNA (used as negative control) (**Fig. R4B**)**.** On the other hand, *CELF4* was overexpressed in both cell lines with a specific plasmid, obtaining substantial increases of mRNA levels (**Fig. R4C**). Interestingly, *CELF4* silencing with the specific siRNA significantly reduced the proliferation rate in both cell lines (**Fig. R4C, D**). In BON-1 cells, a significant reduction was observable at 24 and 48 h; whereas in QGP-1 cells, the effect was long-lasting (48 h and 72 h) and appeared quantitatively more prominent (at 24 h cells had not grown enough after starving) (**Fig. R4C**). Consistent with these results, *CELF4* overexpression resulted in the opposite effect, an increase in proliferation in both cell lines, being most prominent in BON-1 after 48 h (**Fig. R4D**). Furthermore, the antitumoral effects exerted by *CELF4* silencing *in vitro*

were closely reproduced in an *in vivo* xenograft mice model. Specifically, xenograft tumors generated by inoculated BON-1 cells followed for two weeks drastically slowed down their growth after an intratumoral injection with *CELF4* silencing siRNA but not when scrambled siRNA was injected (**Fig. R4F**).

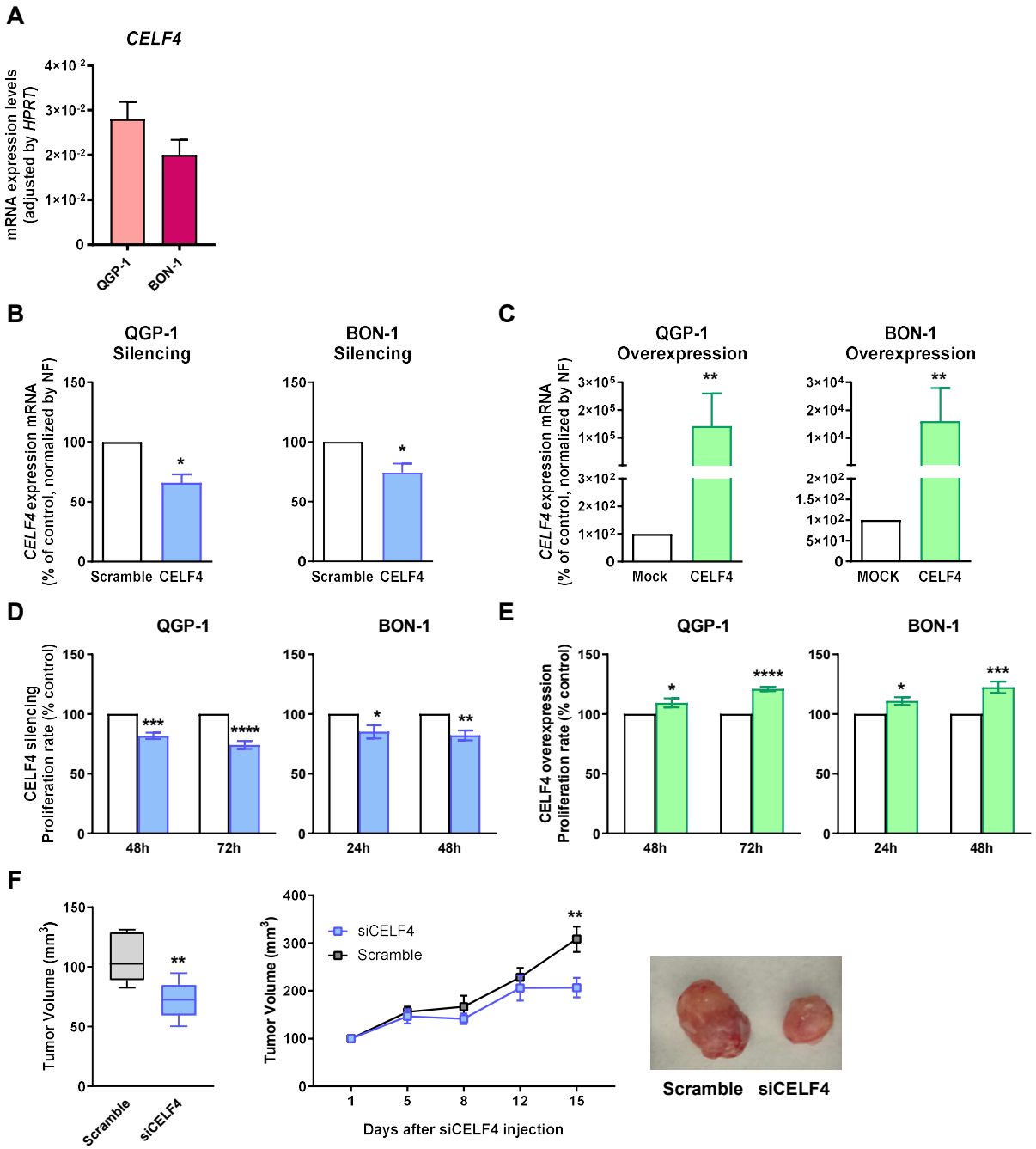


Figure R3. Functional effects of CELF4 expression modulation in QGP-1 and BON-1 cell lines. **A**) CELF4 mRNA expression levels in BON-1 and QGP-1 cell lines. Data are represented by absolute mRNA levels normalized by HPRT expression levels. Data are expressed as a percentage of control (Scramble; set at 100%). **B, C**) mRNA expression levels of CELF4 silencing and overexpression with specific siRNAs (**B**) and plasmid (**C**) in QGP-1 and BON-1. Changes in cell proliferation at 24, 48 and/or 72 h of QGP-1 and BON-1 cell lines in response to **D**) CELF4 silencing or **E**) CELF4 overexpression. Control (scramble or mock plasmid, respectively) was set at 100 %. **F**) Left panel. Development of tumor volume growth in BON-1 xenografted mice after CELF4 siRNA injection; tumor volume is expressed as mm^3. Middle panel. Relative tumor volume of BON-1 xenografted mice in CELF4 siRNA-injected mice compared with scramble-injected mice at time of euthanasia; tumor volume is expressed as mm^3. Right panel. Representative picture of paired xenografted tumors with CELF4 downregulation (right) compared with Scramble (left). Values represent the mean ± SEM. Asterisks indicate values that significantly differ from control (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

5. Cancer therapies effectiveness can be triggered by modulation of *CELF4*

We next asked whether *CELF4* expression levels could influence the response of PanNET cells to the currently available pharmacological treatment for these tumors: mTOR inhibitors (e.g., everolimus), somatostatin analogues (e.g., lanreotide), and antiangiogenic drugs (e.g., sunitinib). To answer this question, we tested the *in vitro* effects of everolimus, lanreotide, and sunitinib in BON-1 and QGP-1 cells where *CELF4* was either overexpressed or silenced (**Fig. R5A, B, C**). Results from this experimental approach revealed a markedly distinct responsiveness of both cell lines at the three drugs, and intriguingly differential interaction of CELF4 with each of the drugs. Specifically, in both cell types, silencing of *CELF4* expression seemed to enhance the antiproliferative action of everolimus, whereas, in contrast, *CELF4* overexpression did not interfere with the response to everolimus, which clearly overrode the enhanced proliferation caused by overexpression of the gene (**Fig. R5A**). In clear contrast, cells were poorly responsive to lanreotide treatment, which reduced proliferation only in BON-1 cells (and not consistently), and paradoxically increased it at long term (72 h) in QGP-

1 cells, while these marginal effects did not seem to be influenced by *CELF4* silencing or overexpression (**Fig. R5B**). Interestingly, BON-1 and QGP-1 cells were unresponsive to sunitinib treatment under *in vitro* basal culture conditions, whereas this kinase inhibitor significantly decreased the enhanced proliferation rate in BON-1 cells overexpressing *CELF4* (**Fig. R5C**). Thus, the PanNET cell models tested showed a limited, barely informative response to lanreotide or sunitinib, but displayed a robust responsiveness to everolimus, which appeared to be clearly influenced by *CELF4* expression levels.

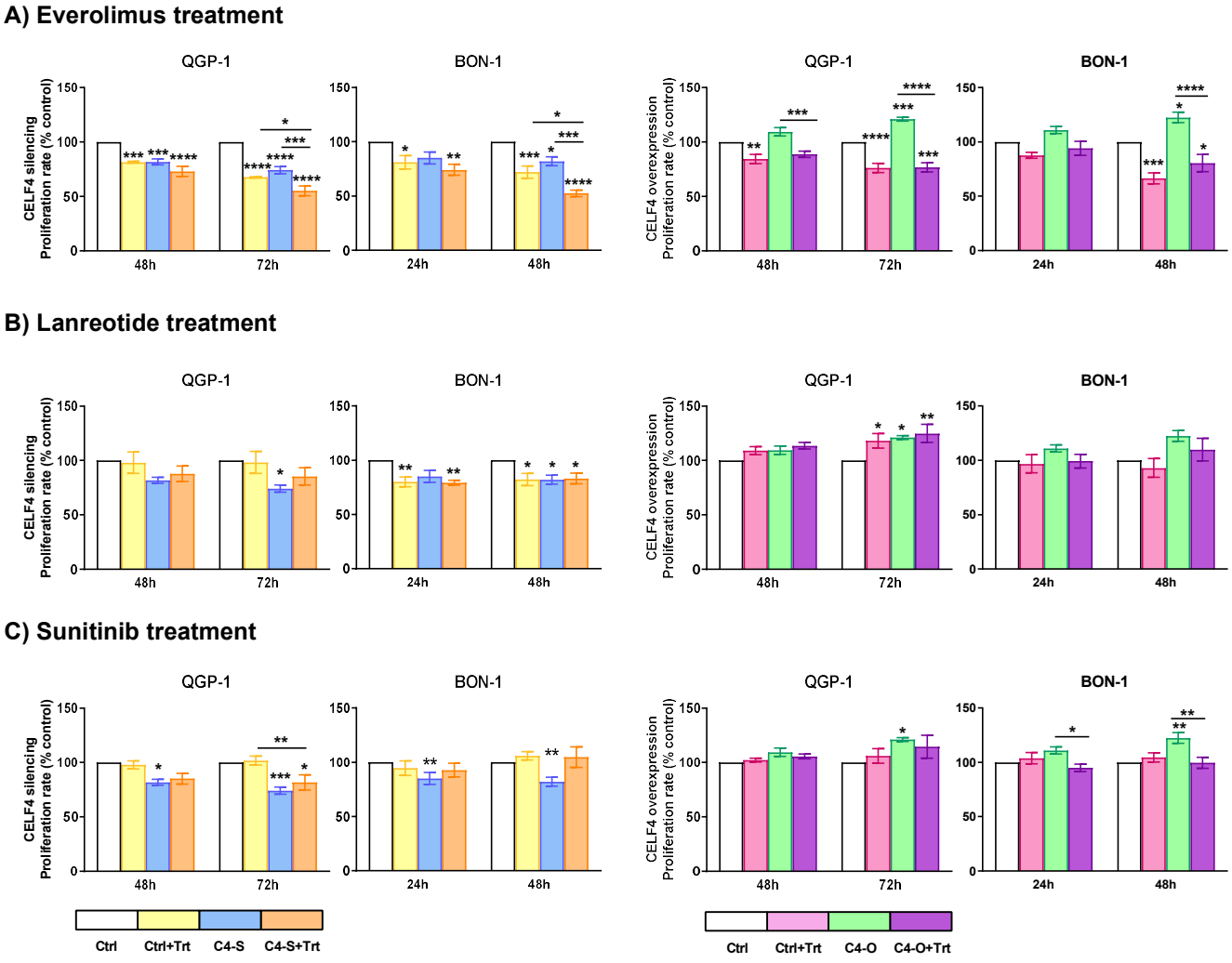


Figure R5. Functional effects of CELF4 expression modulation in QGP-1 and BON-1 cell lines after drug treatments. **A, B, C**) Changes in proliferation rate of BON-1 and QGP-1 cell lines, at 24, 48, and/or 72 h, in response to CELF4 silencing (left panels) or overexpressing (right

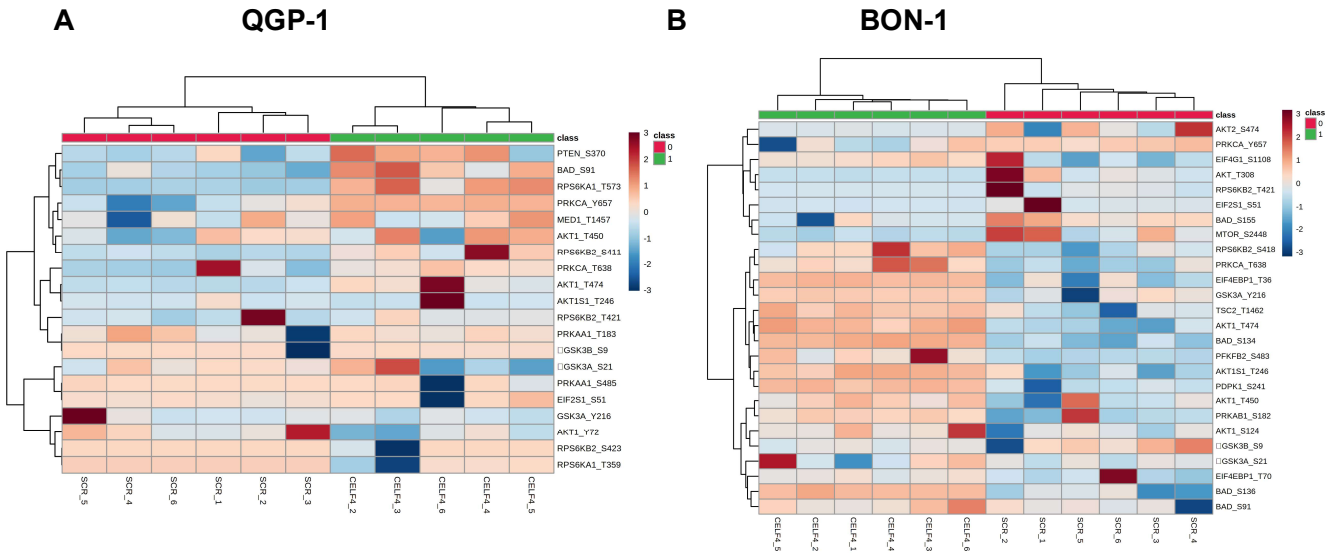
panels) CELF4 and after treatment with **A**) Everolimus, **B**) Lanreotide, or **C**) Sunitinib. Control (untreated scramble or mock plasmid transfected cells, respectively) was set at 100%. Values represent the mean ± SEM. Asterisks indicate values that significantly differ from control (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

6. Signaling pathways associated to *CELF4* genetic alteration

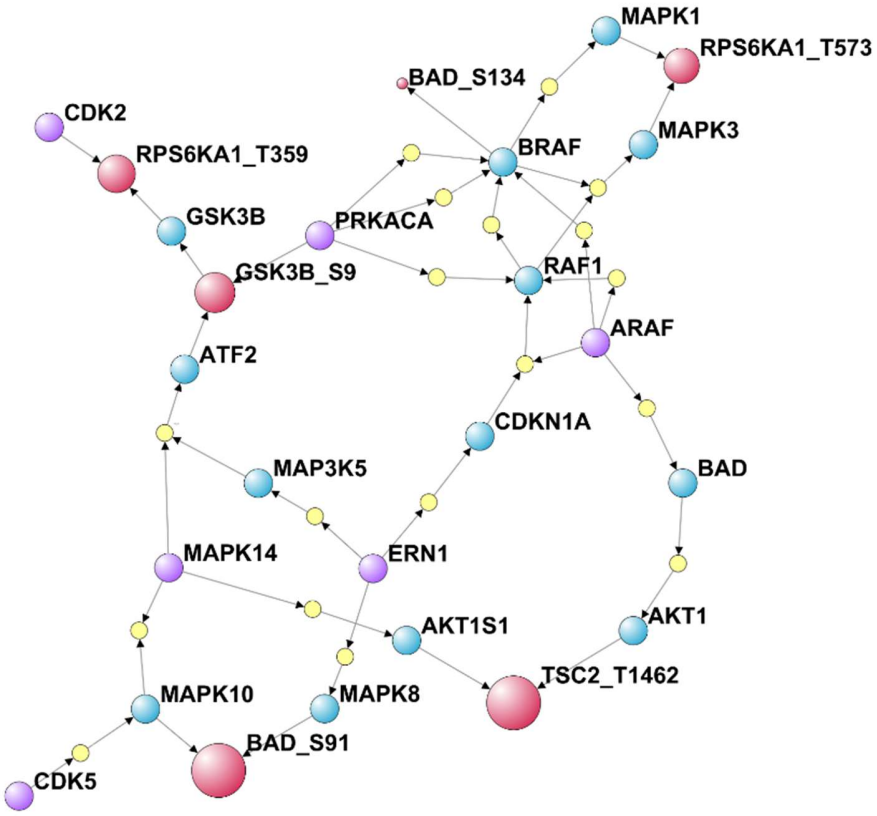
The functional interplay between CELF4 expression in PanNET cells and their response to everolimus prompted us to further investigate the relationship of this splicing factor with the mTOR pathway, as primary target of everolimus. To this end, we evaluated changes in phosphorylation in QGP-1 and BON-1 cell lines after *CELF4* silencing (or scramble transfection), assaying an ample panel of proteins that provide a complete collection of the molecular components of the mTOR pathway by means of a phospho-antibody array. Results from this assay enabled to identify a total of 17 proteins significantly altered by *CELF4* silencing. Of those, 11 (65 %) were selectively altered in BON-1, while 6 proteins (35 %) were altered in QGP-1 cells (**Fig. R6A, B**). Interestingly, only the protein BAD was similarly altered by lack of CELF4 in both cell lines, although the precise phosphosite affected was different in each one. To further delineate and understand these findings, we designed a signaling network model with altered phosphoproteins, which enables to predict interactions and detect possible intermediates altered in the pathway. Despite the differences observed in the phospho-assay, both cell lines rendered similar results in the resulting functional network. In BON-1, the signaling network model comprised 24 nodes and 42 edges (**Fig. R6C**) whereas in QGP-1 the model yielded 26 nodes and 46 edges (**Fig. R6D**). In both models, an expected downstream of phosphorylation of mTOR canonical pathway, CDK5 and ERN1 appeared to be mostly altered, followed by MAP3K5. Based on their predicted kinases, we were able to connect an additional 17 disrupted sites in QGP-1 and 14 in BON-1 cells to the alteration of mTOR (**Appendix 3 and 4**), which thereby comprise the main targets

responsible for BAD (Serine 134), and TSC2 (Threonine 1462) phosphorylation alteration in both cell lines (**Fig. R6C, D**).

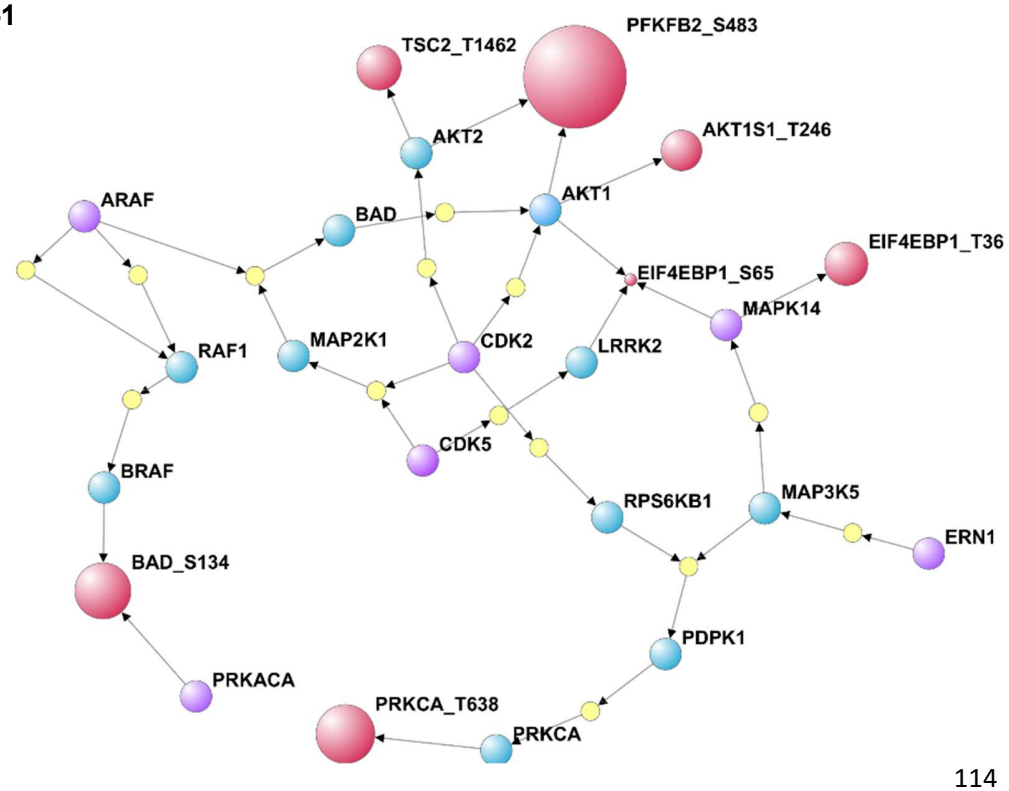
Figure R6. Influence of CELF4 expression on the functional profile of phosphoprotein of mTOR pathway. A, B) Unsupervised clustering analysis of phosphorylated protein levels of mTOR pathway components in CELF4 silencing QGP-1 (**A**), BON-1 (**B**) samples (1; green) compared with control untreated scramble (0; red). **C, D**) PHONEMeS solution model of signaling for mTOR phospho-antibody array after CELF4 silencing in QGP-1 (**A**) and BON-1 (**B**). Target proteins (purple circles) correspond to the highly regulated proteins, which were connected to its target phosphorylation sites (red circles) through intermediary kinases (blue circles). Central kinases, which were also identified by kinase activation prediction, are shown as intermediary kinases with small yellow circles.



C) QGP-1


MAPK1
RPS6KA1_T573
BAD_S134
CDK2
MAPK3
RPS6KA1_T359
BRAF
GSK3B
PRKACA
GSK3B_S9
RAF1
ARAF
ATF2
CDKN1A
MAP3K5
BAD
MAPK14
ERN1
AKT1S1
AKT1
TSC2_T1462
MAPK10
MAPK8
BAD_S91
CDK5

D) BON-1


PFKFB2_S483
TSC2_T1462
AKT2
AKT1S1_T246
ARAF
BAD
AKT1
EIF4EBP1_T36
EIF4EBP1_S65
MAPK14
RAF1
MAP2K1
CDK2
LRRK2
CDK5
BRAF
RPS6KB1
MAP3K5
ERN1
BAD_S134
PDPK1
PRKACA
PRKCA_T638
PRKCA

Chapter II

Splicing Alterations in Adenocarcinoma

Section I

Dysregulated splicing factor SRSF2 plays a similar oncogenic role in prostate and pancreatic cancer

Dysregulation of alternative splicing is becoming a novel hallmark in cancer [2, 3]. Alterations in the functioning of the splicing machinery due to mutations or altered expression of specific spliceosome components and splicing factors is a primary cause of splicing disruption [41, 44, 184]. Evidence supporting this notion was initially gathered in haematological malignancies but is also accumulating recently in solid tumors [62, 65, 67, 185]. In this context, our group and others have provided compelling evidence that altered expression of splicing factors can substantially contribute to oncogenesis or cancer aggressiveness through splicing dysregulation [91, 101, 186-188]. Specifically, we have recently described examples of such oncogenic splicing alterations in two adenocarcinomas, of the pancreas and the prostate, which display obvious differences but also share key molecular features that span from membrane receptors and transcription factors to signaling pathways and tumor suppressors [189]. Moreover, in line with this, we reported substantial parallelisms in the contribution of altered expression of the core splicing factor SF3B1 in these two cancers [190, 191]. Accordingly, we sought to further examine whether the adenocarcinomas of pancreas (PDAC) and prostate (PCa) could share additional alterations in the splicing machinery by comparing the status and potential role of SRSF2, a splicing factor with well-known widespread functions in splicing regulation and beyond [192].

Precise regulation of splicing is achieved by the combined interaction of *cis* and *trans* elements. *Cis* elements contained within the nucleotide sequence can promote silencing (splicing silencers: SS) or enhancing (splicing enhancers: SE) of the splicing process [43, 44]. Splicing factors (SFs) comprise a special group of RNA-binding proteins (RBPs) that

act in *trans*, i.e. by recognizing and binding to motifs in the sequence of the non-mature RNA, to ensure the correct and effective process of alternative splicing [50]. The serine/arginine-rich (SR) family is a particularly relevant group of SFs that contain at least one RNA recognition motif (RRM) at the N-terminus, which recognizes specific sequences in the RNA (Splicing Regulatory Elements; SREs), a glycine-rich spacer region, and a domain rich in arginine and serine residues (RS domain) at the C-terminus [193]. In general, SR proteins bind to Exonic Splicing Enhancers (ESEs) to enhance splicing by interacting with the spliceosome [194]. Among the SR, SRSF2 (formerly known as SC35) has some peculiar features [192, 195]. Unlike the other 11 SR SFs, which can translocate from the nucleus to the cytoplasm and *vice-versa*, SRSF2 is found exclusively in the nucleus and does not interact with cytoplasmic RNA molecules [196]. Also, its longer L3 loop region facilitates interactions with highly degenerated RNA sequences of ESEs, thus promoting global rather than target-specific actions on splicing [192]. Moreover, beyond its role in regulating splicing, recent reports indicate that SRSF2 can also participate in the control of genomic stability, gene transcription, mRNA stability, and translation [192]. In line with its core and widespread regulatory functions, various studies have examined the role of SRSF2 in multiple physiological and pathological settings, from immune cell function to neurodegeneration and even viral replication, where this factor can act as a hub for diverse mediators [192]. On the other hand, most studies on SRSF2 in the molecular and clinical oncology field have mainly focused on the effects of specific mutations in myelodysplastic syndromes [197, 198], whereas a limited number of reports have also shown that SRSF2 can play a relevant role in certain cancers such as hepatocellular carcinoma [199], colon cancer [200], renal cancer [201], and neuroendocrine tumors of the lung [202]. To date, there are no reports focused on the presence and potential role of SRSF2 in the pancreatic and prostate adenocarcinomas. Accordingly, in the present study we aimed to decipher and compare

the presence and putative dysregulation of SRSF2 in these cancers, and, if so, whether it could contribute to the development and/or progression of these cancer types, with the ultimate goal of exploring its potential as prognostic biomarker and actionable therapeutic target in these devastating pathologies.

1. SRSF2 overexpression correlates with tumor malignancy parameters.

To assess the expression levels of *SRSF2* in PCa and PDAC, we analyzed two cohorts of human samples collected in the Hospital Universitario Reina Sofía (Córdoba) using a qPCR array. Specifically, we found that *SRSF2* mRNA levels were significantly higher in PCa tissues compared to adjacent control tissues (n= 84 patients; **Fig. R7A**), being its expression positively correlated with ISUP (International Society of Urological Pathology) grade (**Fig. R7B**) and elevated in those patients displaying significant PCa (Gleason Score ≥ 7) compared to non-clinically significant PCa (**Fig. R7C**). Moreover, *SRSF2* expression was directly associated with additional malignancy features such as extraprostatic extension and lymphovascular invasion (**Fig. R7D and E**, respectively).

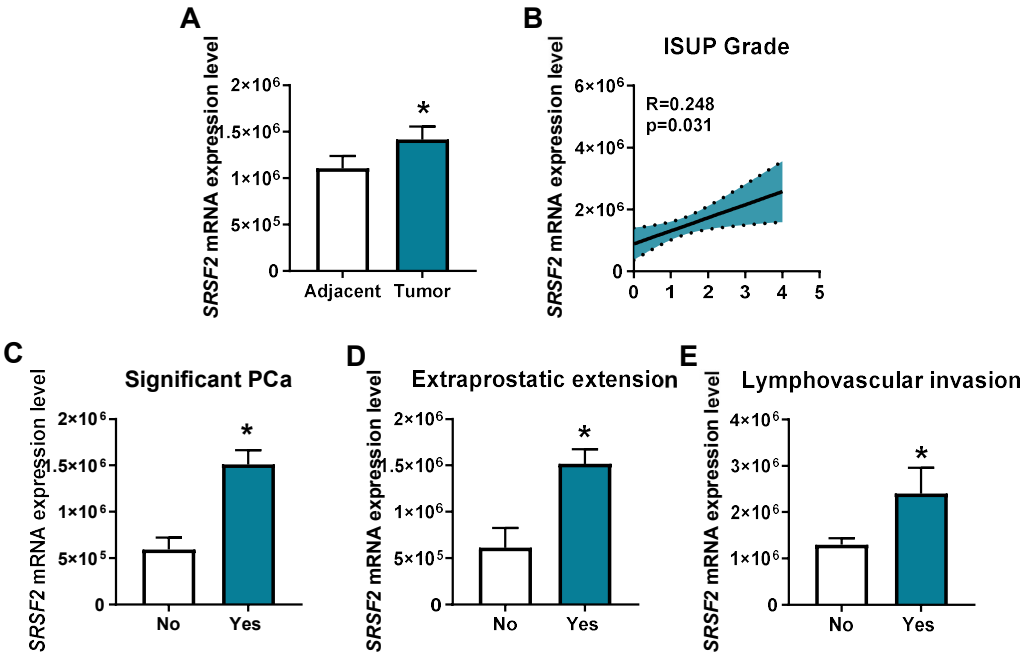


Figure R7. SRSF2 dysregulation in PCa. **A**) SRSF2 expression levels in PCa tissues compared to control adjacent tissue (n = 84). Correlation of SRSF2 mRNA levels with: ISUP grade (**B**), clinically significant PCa (significant GS ≥ 7) (**C**) extraprostatic extension (**D**) and lymphovascular invasion (**E**). Data represent mean ± SEM. Asterisks (*p < 0.05) indicate statistically significant differences.

Similarly, *SRSF2* expression was also significantly higher in PDAC tissues compared to their corresponding adjacent tissues used as reference (n= 75 patients; **Fig. R8A**). Likewise, *SRSF2* expression was directly associated with key malignancy features such as perineural invasion (**Fig. R8B**) and lymphovascular invasion (**Fig. R8C**). Moreover, increased *SRSF2* expression was observed as the tumor stage raised, while showing a biphasic relationship with the size and extent of the main tumor (T stage), with high levels in T1 and a sharp drop in T2 followed by a progressive increase in T3 and T4 (**Fig. R8D and E**).

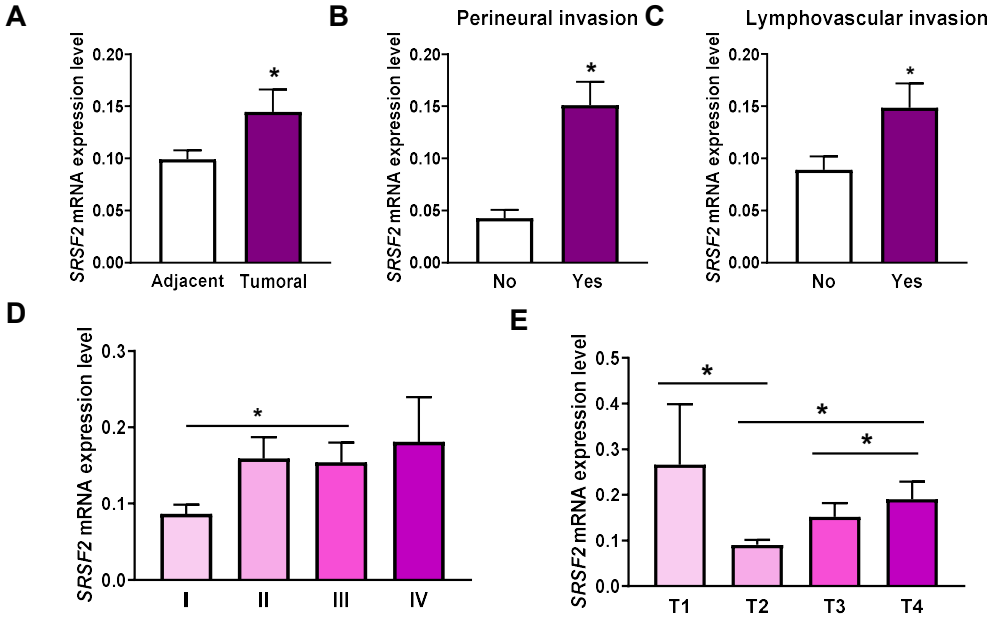


Figure R8. SRSF2 dysregulation in PDAC. **A**) *SRSF2* expression levels in PDAC Formalin-Fixed Paraffin-Embedded samples compared with non-tumoral adjacent tissue (n = 75). Expression levels of SRSF2 in relation to: Perineural invasion (**B**), lymphovascular invasion (**C**), tumor stage (**D**) and stage according to the TNM system (**E**). Data represent mean ± SEM. Asterisks (*) indicate statistically significant differences; p < 0.05.

An *in silico SRSF2* biocomputational study was carried out in different available databases with PCa patients (Wallace cohort [203]; 69 PCa tissues *vs.* 29 adjacent non-tumorous prostate tissues) and PDAC patients (PanCancer database (TGCA) [204]; 151 PDAC tissues). These analyses revealed that *SRSF2* expression levels significantly correlated with the expression of relevant genes in each cancer type. Specifically, *SRSF2* expression was directly correlated with the expression of key genes involved in the splicing process (i.e. with *SF3B1* or *SRSF1* in both pathologies, and with *SRSF6 and EIF4A3* in PCa), and also with tumor suppressor genes (*TP53* and *SMAD4*) in the case of PDAC (**Fig. R9A, B**).

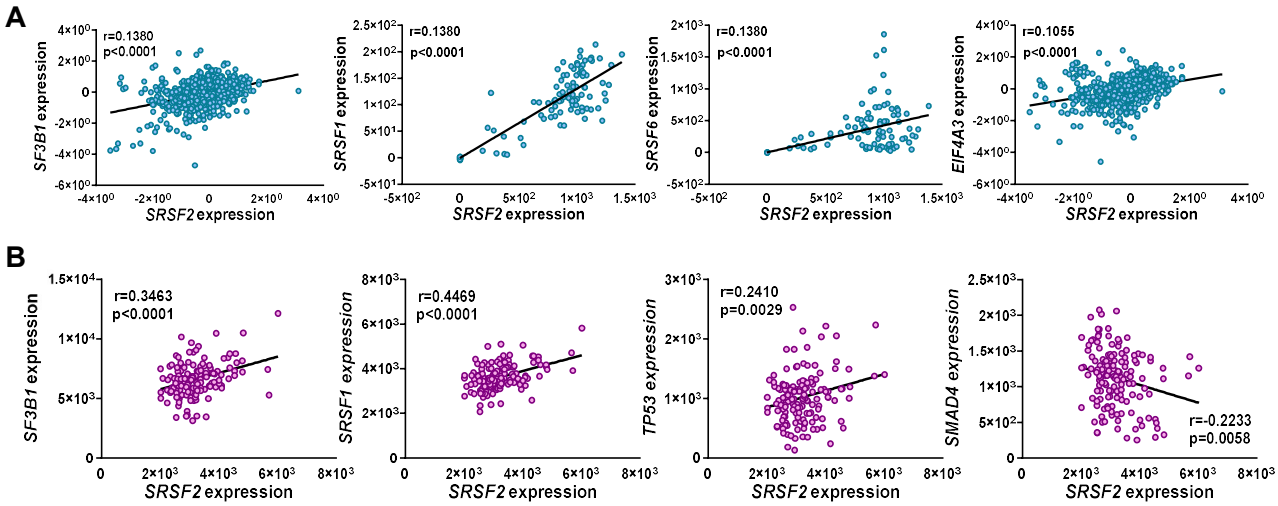


Figure R9. *SRSF2* expression levels in relation with expression of relevant genes in PCa and PDAC. Correlations between **A**) *SRSF2* mRNA levels and expression of splicing-related genes in PCa (Wallace cohort). **B**) *SRSF2* mRNA levels and expression of key genes in PDAC (PanCancer cohort).

2. *SRSF2* silencing alters key functional parameters of malignancy in PCa and PDAC derived cell lines.

To assess the functional role of SRSF2 in tumor malignancy, we carried out different functional assays using two model cell lines of PCa, PC-3 cells, androgen-independent

and with a more aggressive phenotype, and LNCaP cells, androgen-sensitive and with a less aggressive phenotype, and another two PDAC cell lines, the more aggressive MIAPaCa-2 cell line, and the less aggressive BxPC-3 cell line. Firstly, we evaluated the expression levels of *SRSF2* in PCa-derived and PDAC-derived cell lines (**Fig. R10A**). This revealed that *SRSF2* is highly expressed in these cell lines, being its mRNA levels higher in cell lines displaying less aggressive phenotypes in PCa (LNCaP) and PDAC (BxPC-3). Then, a specific siRNA against SRSF2 (siSRSF2) was used to decrease its expression levels which were validated by qPCR and Western blot (**Fig. R10B, C**).

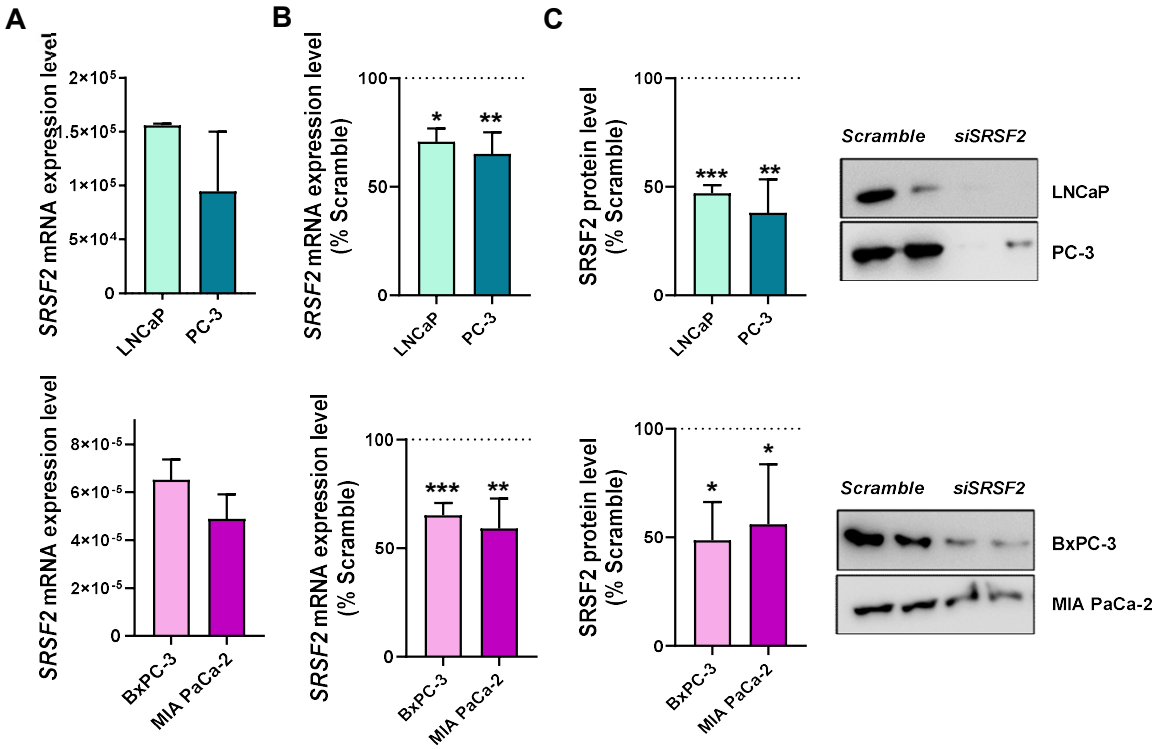


Figure R10. *SRSF2* mRNA levels in PCa and PDAC cell lines and validation of *SRSF2* silencing. A) Basal *SRSF2* mRNA levels in PCa and PDAC cell lines. Decrease of *SRSF2* levels after silencing with specific siRNA in PCa and PDAC cell lines at mRNA (**B**) and protein (**C**) levels. Data are expressed as a percentage of control (Scramble; set at 100%) of n=3-5 independent, experiments. Asterisks indicate significant differences (*p<0.05; **p<0.01; ***p<0.001).

To evaluate the effect of *SRSF2* silencing on tumor aggressiveness, we first evaluated cell viability as a reliable surrogate of tumor cell proliferation. This showed that the reduction in *SRSF2* expression caused a significant decrease in cell proliferation in both PCa cell lines, PC-3 cells and LNCaP cells, being this effect already observed at 24 h and long lasting (72 h) (**Fig. R11A**). In marked contrast, *SRSF2* silencing did not alter proliferation rates in PDAC BxPC-3 cells, and only caused a non-significant trend for reduction (p = 0.0655 at 48 h) in proliferation rate in the more aggressive PDAC MIA PaCa-2 cells (**Fig. R11B**).

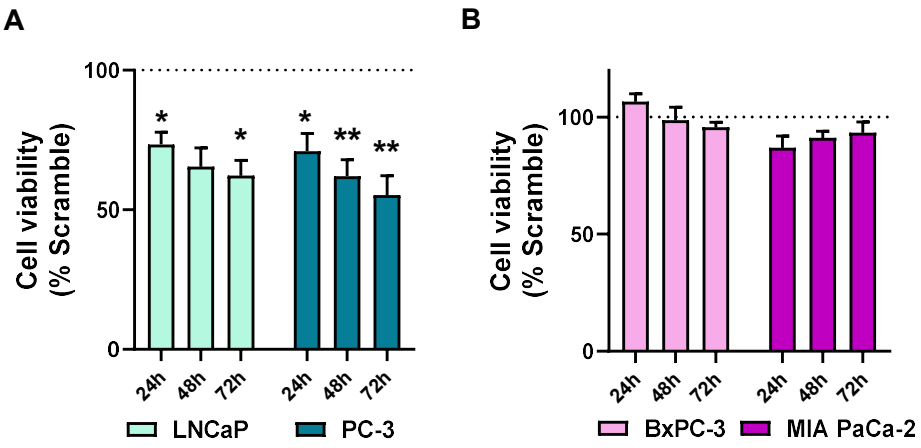


Figure R11. Cell viability as a surrogate of proliferation rate in PCa and PDAC cell lines after *SRSF2* silencing. A) PCa (LNCaP and PC-3) and, **B**) PDAC (BxPC-3 and MIA PaCa-2) cell lines. *SRSF2* expression was silenced with a specific siRNA and cell viability was measured at 0, 24, 48 and 72h. Data are expressed as percentage of Scramble (adjusted to 100%) and represent mean ± SEM of n=4 separate, independent experiments. Asterisks (*p < 0.05; **p < 0.01) indicate statistically significant differences.

Interestingly, *SRSF2* silencing significantly reduced (by 30 %) the migration capacity of PC-3 cell line (**Fig. R12A**). This functional parameter could not be measured on LNCaP cell line due to its inability to migrate. Likewise, *SRSF2* silencing significantly decreased the migration ability of BxPC-3 cells (by more than 50 %), while its effect on migration of MIA PaCa-2 did not reach statistical significance (**Fig. R12B**).

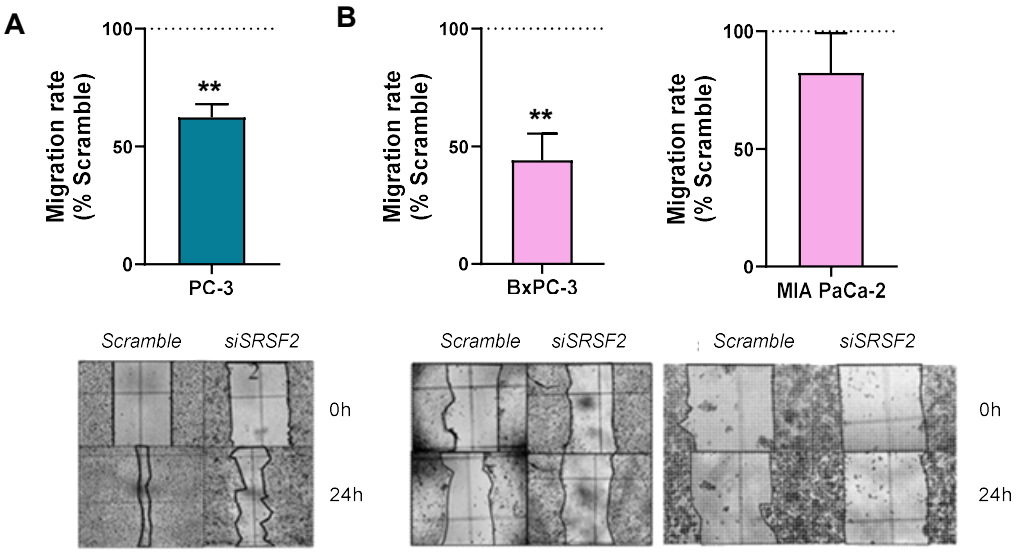


Figure R12. Migration rate in PCa and PDAC cell lines after *SRSF2* silencing with specific siRNA. A) PCa-derived cell lines (PC-3) and, **B**) PDAC-derived cell lines (BxPC-3 and MIAPaCa-2). Cell migration was measured by Wound healing assay at 24h. Representative images of cell migration are shown. Data are expressed as percentage of Scramble (adjusted to 100%) and represent mean ± SEM of n=3 independent, separate experiments. Asterisks (*p < 0.05; **p < 0.01) indicate statistically significant differences.

We also evaluated the effect of *SRSF2* silencing on colony formation by PCa and PDAC cell lines. Of note, colony formation was significantly reduced in response to *SRSF2* silencing in both PCa (**Fig. R13A**) and PDAC (**Fig. R13B**) derived lines, being this effect more apparent in PCa-derived lines (30 % reduction in LNCaP cells, 58 % in PC-3 cells, vs. 22 % in BxPC-3 cells and 33 % in MIA PaCa-2 cells).

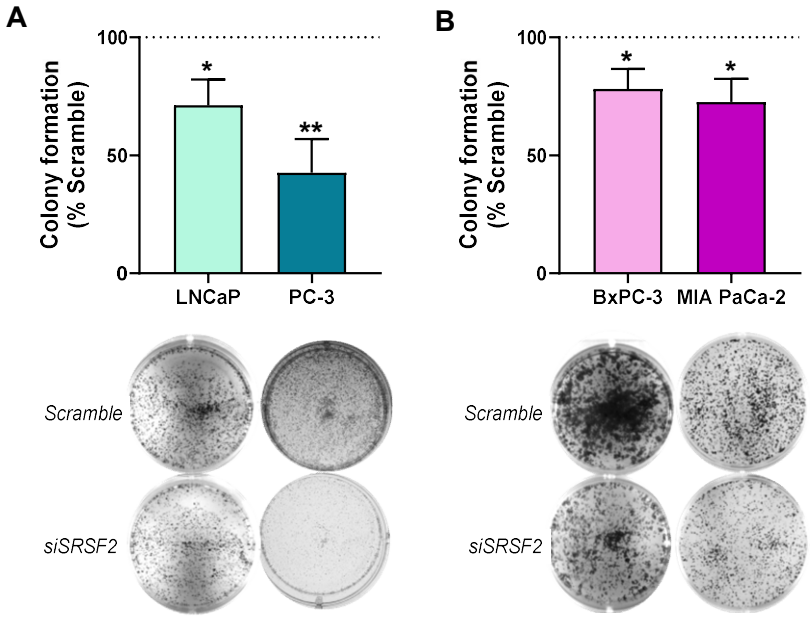


Figure R13. Colony formation capacity of PCa and PDAC cell lines after *SRSF2* silencing with specific siRNA. **A**) PCa-derived cell lines (LNCaP and PC-3) and, **B**) PDAC-derived cell lines (BxPC-3 and MIA PaCa-2). Data are expressed as percentage of Scramble (adjusted to 100%). Representative images of colony formation were included. Data represent mean ± SEM of n=3 independent, separate experiments. Asterisks (*p < 0.05; **p < 0.01) indicate statistically significant differences.

Section II

Dysregulation of the splicing machinery as a target for pancreatic ductal adenocarcinoma

Pancreatic adenocarcinoma accounts for 90% of pancreas neoplasms and is one of the most lethal cancers worldwide, with a dismal 10% survival rate 5 years after diagnosis [205] . Despite the profound knowledge acquired in recent years on the molecular basis of PDAC [101, 149], its translation to the patient is still very limited. Accordingly, opening novel areas of research is required to tackle this disease. A growing number of studies [62], including several from our group [91, 160, 186, 187], show that many different cancers share as a common hallmark the alteration of the splicing machinery, which drives to abnormal patterns of alternative splicing and gives rise to aberrant variants with oncogenic potential. Interestingly, PDAC was one of the first cancers where alternative splicing was explored, which disclosed mutations and alterations in the expression of several components of the splicing machinery, both spliceosome core elements and splicing factors, and led to identify dysregulated profiles of splice variants [96, 97, 206]. Thus, functional and bioinformatic studies in PDAC have provided evidence for the relevance of specific alterations in splicing machinery components, such as *SRPK1* and *SRSF1*, whose study in PDAC cell lines suggested their relation to tumor progression and gemcitabine resistance [98, 207]; and *RBM5*, which has been shown to be correlated to *KRAS* expression in PDAC and several clinical parameters, suggesting a role in tumor invasion and progression [208]. Likewise, *ESRP1* expression has been related to a better overall survival rate and lower grading tumors [103].

Taken this evidence together, we posited that the alterations found in individual factors may indicate that the splicing machinery is uniquely and profoundly dysregulated in PDAC, and that its systematic study could help to identify further elements susceptible

to serve as new biomarkers and operable tools. Accordingly, in the present study we devised a strategy to explore the expression of the components of the spliceosome core and a selected set of splicing factors in various cohorts of PDAC, assess their relation to clinical/molecular parameters and study key functional and pathological features.

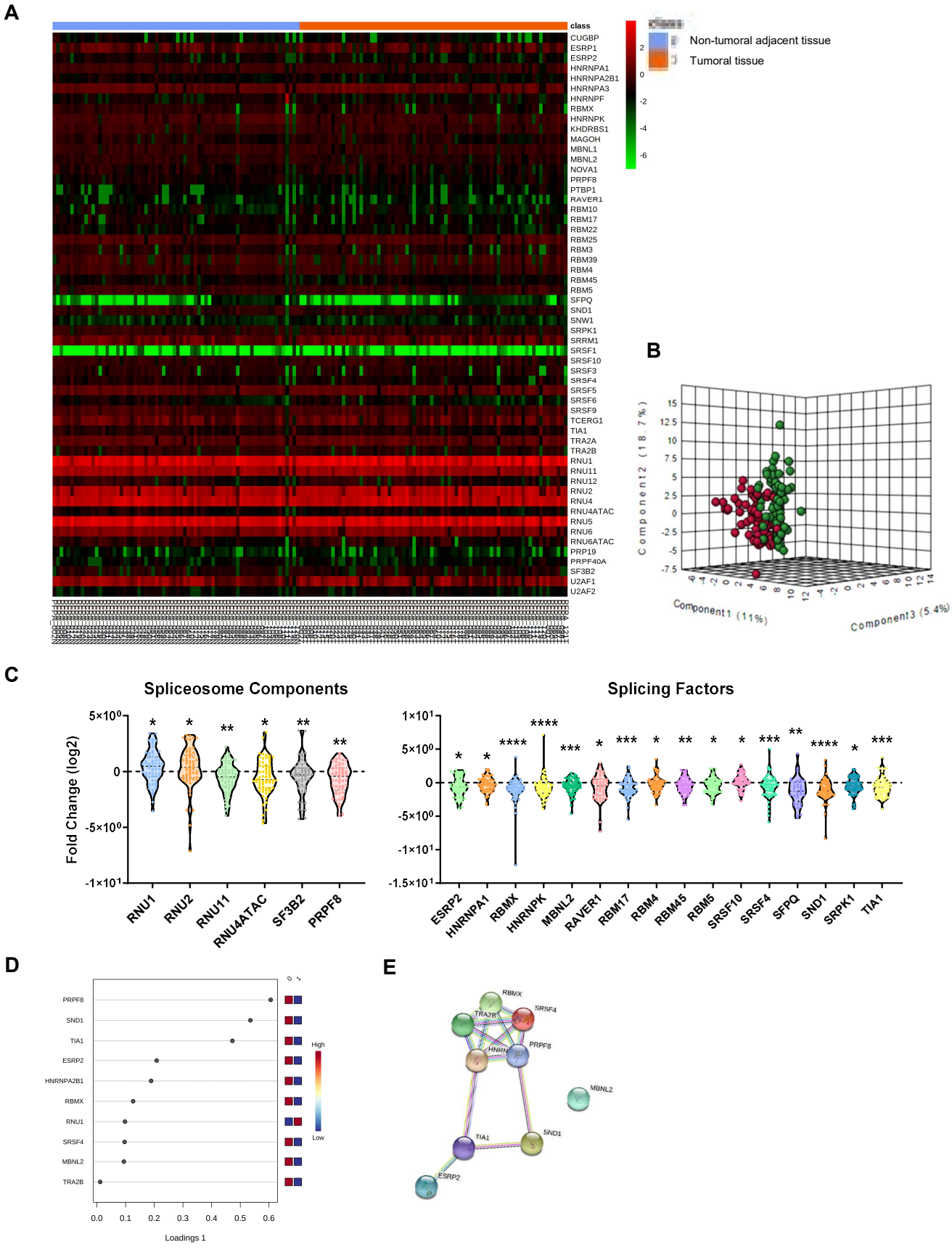
1.1. The pattern of expression of the splicing machinery is severely altered in PDAC.

Results from microfluidic qPCR dynamic array revealed a clear dysregulation of splicing machinery expression in tumor *vs.* non-tumor adjacent tissues in a set of 79 FFPE PDAC samples (**Fig. R14A**). Further analysis of these data was performed by applying a statistical method to select among them the best predictive or discriminative elements to help classifying the tumor *vs* non-tumor tissues. As illustrated by the data distribution in the Principal Components Analysis (PCA) plot (**Fig. R14B**), two separate groups emerged from gene expression levels, suggesting that both sample groups could be discriminated based on the expression pattern of the splicing machinery components. In fact, a relevant proportion of the 18 spliceosome components and 41 splicing factors measured (33 % and 39 %, respectively) were differentially expressed in tumor-vs non-tumor tissue, with a clear predominance of downregulation, as shown in the violin plots in **Figure R14C**. The statistical analysis of these results was refined using Sparse Partial Least Squares Discriminant Analysis (sPLSDA) and plotting the generated loadings, which portrayed the genes with the highest ability to discriminate between tumor *vs*. non-tumor adjacent tissues (**Fig. R14D**). As shown, when the variables were ranked by the absolute values of their loadings, the top 10 genes showing the most consistent and prominent differences between the expression in tumor and non-tumor adjacent tissues include: *PRPF8*, *SND1*, *TIA1*, *ESRP2*, *HNRNP2AB1*, *RBMX*, *RNU1*, *SRSF4*, *MBNL2*, and *TRA2B* (**Fig. R14D**). Interestingly, a simple STRING analysis exploring known

protein-protein interactions predicted a potential network of interrelationship among nearly all the selected genes, with a particularly tight putative cross-regulation between *PRPF8*, *RBMX*, *HNRNP2AB1*, *SRSF4*, and *TRA2B* (**Fig. R14E**).

To gain a better understanding of the top 10 dysregulated splicing machinery components in PDAC, we inspected them in further detail. As illustrated in **Fig. R15A**, in this discovery cohort, tumor tissue exhibited higher levels of expression than the corresponding non-tumor adjacent tissues in only one spliceosome component *RNU1*, whereas lower RNA levels were observed for *RBMX, PRPF8*, *SND1*, *TIA1*, *ESRP2*, *HNRNPA2B1*, *TRA2B*, *SRSF4* and *MBNL2.* Furthermore, an analysis based on ROC curves indicated that all the splicing machinery components selected had an Area Under the Curve (AUC) close to or higher than 0.6, supporting their high capacity to discriminate between tumor *vs.* non-tumor adjacent tissues. In particular, *SND1*, *RBMX*, and *TRA2B* had AUCs above 0.7, and therefore could hold a higher potential to discriminate between tumor and non-tumor samples (**Fig. R15B**). Furthermore, an integrated ROC curve combining the most significantly altered splicing machinery components (*PRPF8, SND1, TIA1, ESRP2, HNRNPA2B1, RBMX, RNU1, SRSF4, MBNL2,* and *TRA2B*) yielded an AUC of 0.823 and a 95% CI ranging 0.725-0.954 (**Fig. R15C**).

Figure R14. Splicing dysregulation in Pancreatic Ductal Adenocarcinoma. A) Unsupervised clustering analysis of mRNA expression levels of spliceosome components in PDAC FFPE samples (1; orange) compared with non-tumoral adjacent tissue (0; blue). **B**) Principal Components Analysis (PCA) of the splicing machinery components analyzed in PDAC FFPE samples cohort. **C**) Fold Change of mRNA levels expressions of significative Spliceosome Components and Splicing Factors of PDAC FFPE samples compared with non-tumoral adjacent tissue. Data are represented by Fold Change mRNA levels normalized by ACTB expression levels ± SEM. Asterisks indicates values that significantly differences between groups (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). **D**) sPLSDA analysis showing the most modified factors in our cohort. **E**) STRING analysis of relationships among altered components based on the top 10 genes showing the most differences between the expression in tumor and non-tumor adjacent tissues.


A
class
CUGBP
ESRP1
ESRP2
HNRNPA1
HNRNPA2B1
HNRNPA3
HNRNPF
RBMX
HNRNPK
KHDRBS1
MAGOH
MBNL1
MBNL2
NOVA1
PRPF8
PTBP1
RAVER1
RBM10
RBM17
RBM22
RBM25
RBM3
RBM39
RBM4
RBM45
RBM5
SFPQ
SND1
SNW1
SRPK1
SRRM1
SRSF1
SRSF10
SRSF3
SRSF4
SRSF5
SRSF6
SRSF9
TCERG1
TIA1
TRA2A
TRA2B
RNU1
RNU11
RNU12
RNU2
RNU4
RNU4ATAC
RNU5
RNU6
RNU6ATAC
PRP19
PRPF40A
SF3B2
U2AF1
U2AF2
Non-tumoral adjacent tissue
Tumoral tissue
B
Component2 (18.7%)
Component1 (11%)
Component3 (5.4%)
C
Spliceosome Components
Splicing Factors
Fold Change (log2)
RNU1
RNU2
RNU11
RNU4ATAC
SF3B2
PRPF8
ESRP2
HNRNPA1
RBMX
HNRNPK
MBNL2
RAVER1
RBM17
RBM4
RBM45
RBM5
SRSF10
SRSF4
SFPQ
SND1
SRPK1
TIA1
D
PRPF8
SND1
TIA1
ESRP2
HNRNPA2B1
RBMX
RNU1
SRSF4
MBNL2
TRA2B
High
Low
Loadings 1
E
RBMX
SRSF4
TRA2B
PRPF8
MBNL2
TIA1
SND1
ESRP2

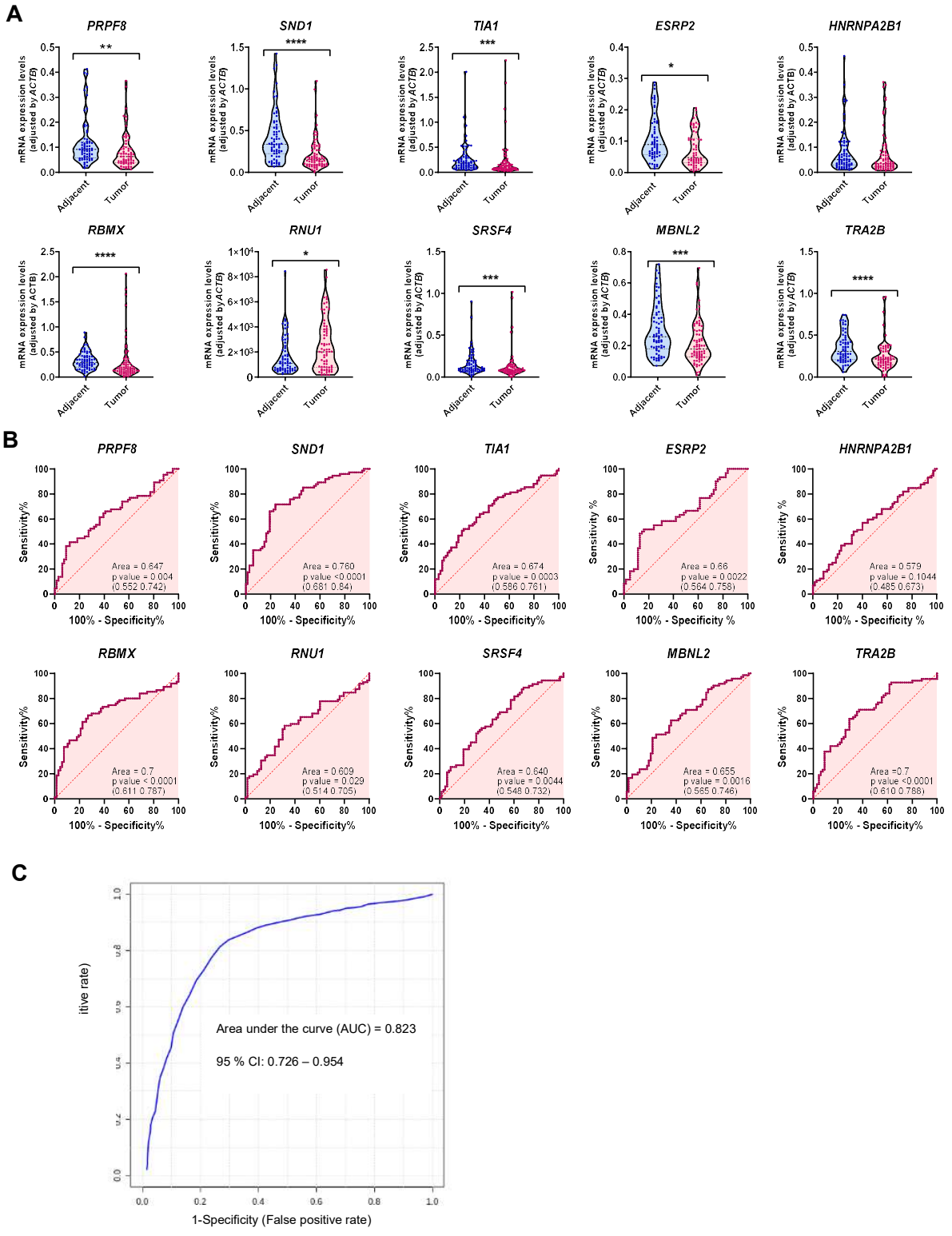


Figure R15. Top splicing factor expression profile in PDAC. A) mRNA expression levels of selected splicing machinery components in PDAC FFPE samples compared with non-tumoral

adjacent tissue. Data are represented by mRNA levels normalized by ACTB expression levels. **B**) ROC curve analysis of selected splicing machinery components in PDAC FFPE samples compared with non-tumoral adjacent tissue. **C**) Integrated ROC curve combining the most significant dysregulated splicing machinery components (*PRPF8, SND1, TIA1, ESRP2, HNRNPA2B1, RBMX, RNU1, SRSF4, MBNL2,* and *TRA2B*) Data represents mean ± SEM. Asterisks indicate values that significantly differ between groups (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001).

1.2. Splicing machinery dysregulation is associated with key clinical parameters and with distinct profiles of splicing events.

Results from expression studies, statistical relevance (loading plots and ROC curves) and predicted interactions, together with the lack of previous knowledge in PDAC and association to clinical features (see below), led us to select *PRPF8* and *RBMX* to explore them in further detail. These two genes displayed marked differences between tumor *vs*- non-tumor tissues and their possible role in PDAC has not been reported to date.

To validate the results obtained here in an external cohort, we first carried out an *in silico* analysis of a PDAC cohort including 195 tumors and 41 non-tumor tissue samples obtained from the public database "ArrayExpress" (E-MTAB-1791). In this case, reference tissue was obtained from healthy pancreas. Interestingly, results showed a neat parallelism with those found in our cohort, for both selected genes, *PRPF8* and *RBMX*, which showed lower levels in tumoral samples vs. normal pancreatic tissue (**Fig. R16A**). This concordance between cohorts reinforces the significance of our findings and, thus, invites to study in greater detail the possible functional and pathological relevance of these splicing factors. Actually, in this external cohort, these two splicing factors were the only ones that displayed an association with clinical parameters, which was not appreciable for the rest of genes explored. Specifically, the expression levels of both genes were associated to histological grade, although in a different manner. Thus, *PRPF8* expression levels were inversely correlated to histological grade, being

progressively lower in G1, G2, and G3/4 PDAC samples. Conversely, *RMBX* levels were higher in G2 than G1 samples, with no apparent differences in G3/G4. These results suggest that lower *PRPF8* levels, but not *RBMX* expression, are associated with more undifferentiated tumors (**Fig. R16B**).

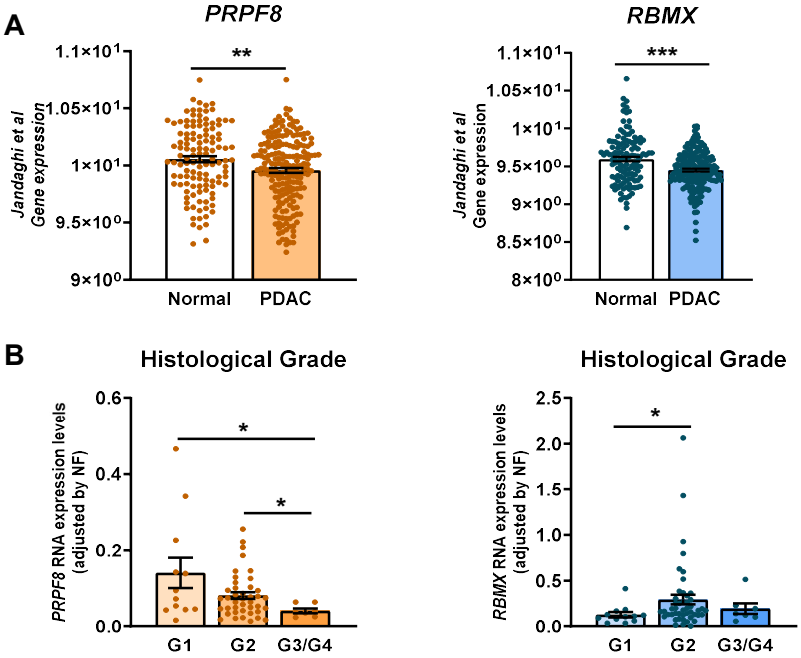


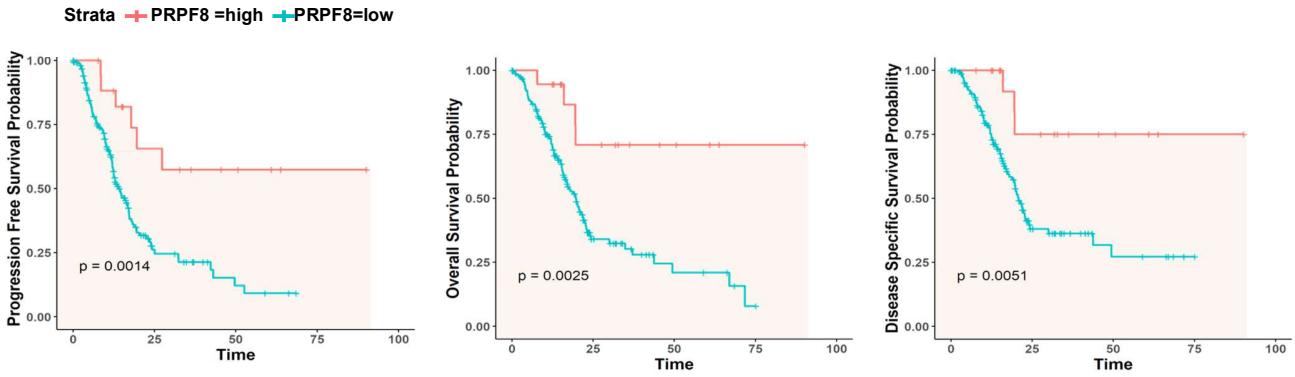
Figure 16. ***PRPF8* and *RBMX* expression in external cohorts.** A) PRPF8 (orange) and RBMX (blue) mRNA levels in an external validation PDAC cohort ("Jandaghi, 2016")[147]. B) Distribution of PRPF8 and RBMX expression levels in the different histological grades of PDAC in the PanCancer cohort [149].

Analyses of patient survival parameters in relation with the expression of the two splicing elements was performed in an RNA-Seq generated from 94 PDAC patient samples described in the present thesis (Chapter III [28]). Of note, high *PRPF8* and *RBMX* expression levels were similarly associated to better patient survival, whereas patients with lower levels showed a lower survival rate, including progression free, overall, and disease specific, survival (**Fig. R17A, B**).

We next sought to examine the possible influence of *PRPF8* and *RBMX* on the splicing process in PDAC. To this end, samples were classified in two groups according

to their low or high *PRPF8* and *RBMX* expression level, and a specific software (SUPPA2) was employed to analyze the number and nature of splicing events in the RNA-Seq. This revealed than only a reduced set of 24 events occurred differentially between low- and high-expressing *PRPF8* samples, while a much larger number, 1324 events, differed in relation to *RBMX* (**Fig. R18A**). Moreover, whereas the profile of splicing events did not reveal major differences depending on *PRPF8* expression, except for a higher 5' alternative splice site, samples with high or low levels of *RBMX* expression displayed strikingly distinct patterns of splicing events, with higher frequency of exon skipping, and 5' and 3' alternative splice site, and lower frequency of alternative first and last exon, as compared to the average of all the calculated events (**Fig. R18B**).

A



B

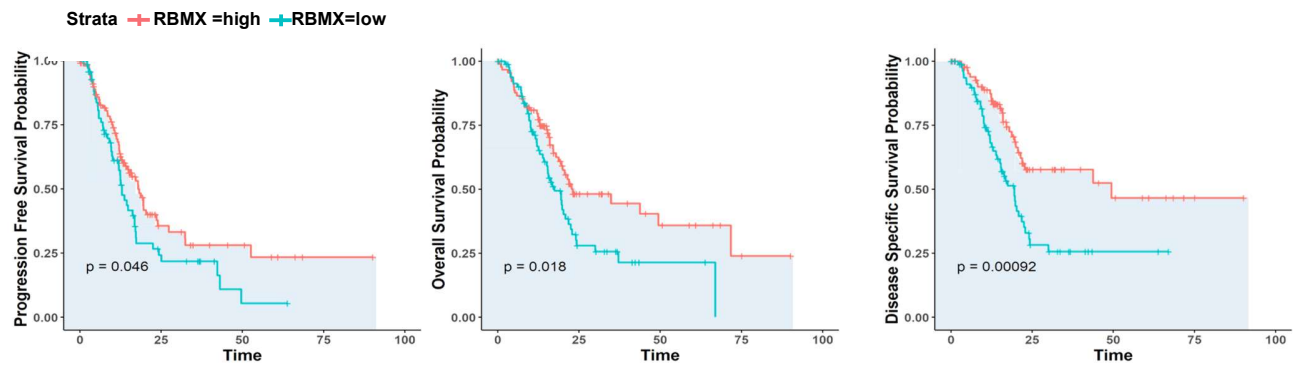


Figure R17. Survival analysis expression levels in PDAC. Kaplan-Meier analyses of progression free survival (left), overall survival (center) and disease specific survival (right) associated with PRPF8 (A) and RBMX (B) expression levels respectively in PanCancer cohort.

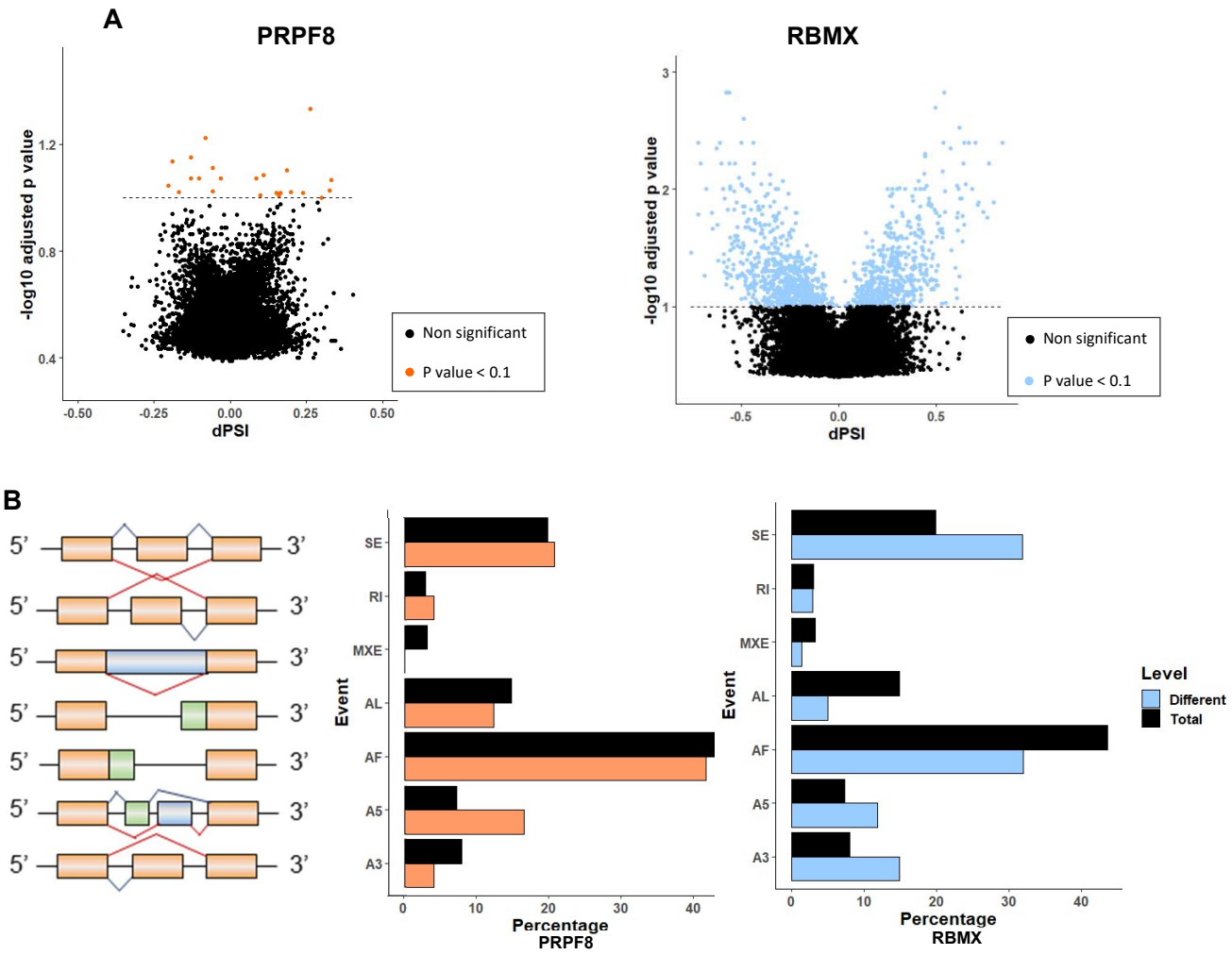
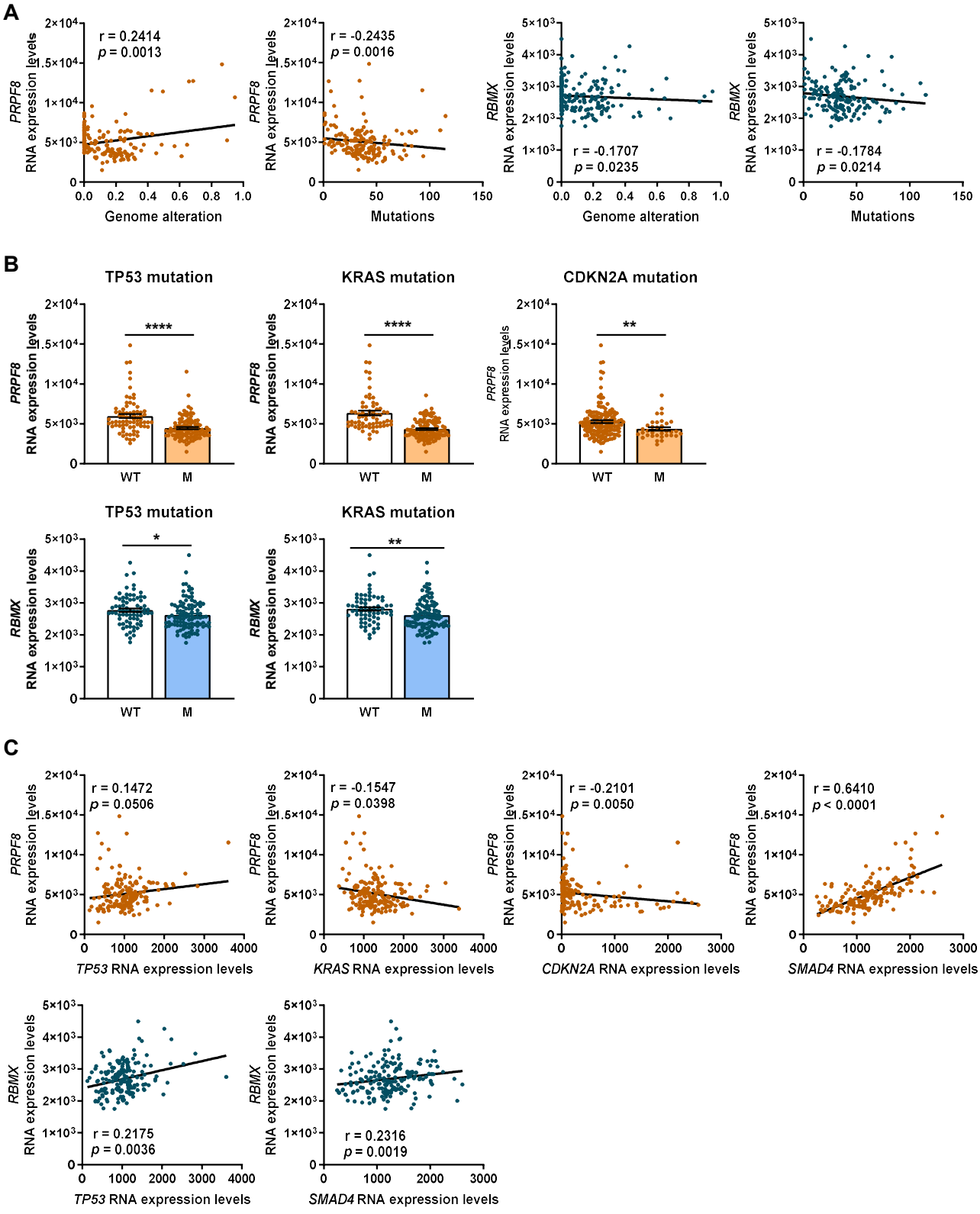


Figure R18. Relationship of PRPF8 (orange) and RBMX (blue) expression levels with splicing event patterns in PDAC. A) Volcano-plot where ΔΨ of total events calculated is plotted against the –log10 p-value of the Fisher's Exact Test to assay differential splicing events between high and low PRPF8 (orange) and RBMX (blue) expression groups of samples, showing their alternative splicing pattern. B) Alternative Splicing events characterization of RNA-Seq samples. Total splicing events detected (black) and significantly different events between PRPF8 (orange) and RBMX (blue) expression groups are classified depending on their type, showing different frequencies (%) between both conditions. SE: Skipping Exon; RI: Retained Intron; MXE: Mutually Exclusive Exons; AF/AL: Alternative First/Last Exons; A5/A3: Alternative 5'/3' Splice Sites.

1.3. Splicing alterations are associated with key PDAC gene mutations

Given the preeminent role in PDAC development and progression of mutations in key genes, namely *KRAS*, *TP53*, *CDKN2A* and *SMAD4*, some of which have already been pathologically linked to altered splicing mechanisms [27], we next evaluated the potential association between *PRPF8* and *RBMX* expression levels and mutations and expression levels of those genes in the PanCancer dataset. Interestingly, this approach revealed that *PRPF8* and *RBMX* expression levels tightly correlate with overall genome alterations and mutations (**Fig. R19A**). More specifically, tumors from patients harbouring *TP53* and *KRAS* mutations displayed lower *PRPF8* and *RBMX* levels, and *CDKN2A* mutations were also related with lower expression of *PRPF8* (**Fig. R19B**). Further analysis indicated that *PRPF8* and *RBMX* expression levels correlated directly with *TP53* and *SMAD4* levels and inversely with *KRAS*, and *CDKN2A* in the case of *PRPF8* **(Fig. R19C**).

Figure R19 (next page). Relationship of *PRPF8* (orange) and *RBMX* (blue) expression levels with expression and mutations of key genes in PDAC. A) Correlations between *PRPF8* and *RBMX* mRNA expression levels and Genome alteration and Mutations in PanCancer cohort. **B**) Correlations between *PRPF8* and *RBMX* mRNA expression levels and TP53, KRAS, and CDKN2A mutations in PanCancer cohort. **C**) Correlations between *PRPF8* and *RBMX* mRNA expression levels and TP53, KRAS, CDKN2A, and SMAD4, mRNA expression levels in PanCancer cohort.


A
PRPF8 RNA expression levels
r = 0.2414
p = 0.0013
Genome alteration
r = -0.2435
p = 0.0016
Mutations
RBMX RNA expression levels
r = -0.1707
p = 0.0235
Genome alteration
r = -0.1784
p = 0.0214
Mutations
B
TP53 mutation
KRAS mutation
CDKN2A mutation
****
****
**
WT
M
*
**
C
r = 0.1472
p = 0.0506
TP53 RNA expression levels
r = -0.1547
p = 0.0398
KRAS RNA expression levels
r = -0.2101
p = 0.0050
CDKN2A RNA expression levels
r = 0.6410
p < 0.0001
SMAD4 RNA expression levels
r = 0.2175
p = 0.0036
r = 0.2316
p = 0.0019

1.4. *PRPF8* and *RBMX* are directly correlated with *in vitro* features

The alterations in expression, splicing and pathological-molecular associations invited to explore in further detail the functional underpinnings of *PRPF8* and *RBMX* in PDAC. To this end, we first evaluated the expression levels of both splicing factors in two widely used PDAC model cell lines, Capan-2 and BxPC-3 cell lines. Since a lower expression of both splicing components was found in tumor tissue compared to non-tumor tissue, we overexpressed *PRPF8* or *RBMX*, using specific expression plasmid, to rescue or mimic their presence in non-tumoral pancreas. Validation of *PRPF8* overexpression confirmed a substantial increase in Capan-2 (over 6-fold), and a more modest but appreciable rise in BxPC-3 (over 70 %) in comparison with empty plasmid (mock) transfected cells (**Fig. R20A**). Similarly, *RBMX* overexpression was confirmed with substantial increases in both Capan-2 (over 10-fold), and BxPC-3 (over 100 %) compared to their respective control (mock transfection; **Fig. R20B**).

In line with our predictions, overexpression of *PRPF8* or *RBMX* decreased cell proliferation, as measured by resazurin assay at 24, 48, and 72 h after transfection. Specifically, a clear, rapid (24 h) and sustained (up to 72 h) decrease was observed in both cell lines after overexpression of *PRPF8*, whereas the effect of *RBMX* upregulation was observable only at 48 h in both cell lines and was long-lasting (72 h) only in BxPC-3 cells (**Fig. R20B**). Interestingly, *PRPF8* and *RBMX* overexpression also impacted on cell migration, which was clearly reduced after 24 h as assessed by a wound-healing assay (**Fig. R20C**). Moreover, *PRPF8* and *RBMX* overexpression similarly blunted colony formation in Capan-2 and BxPC-3 cell lines compared to their respective controls, and markedly reduced the formation of tumorspheres in both cell lines (**Fig. R20D, E**).

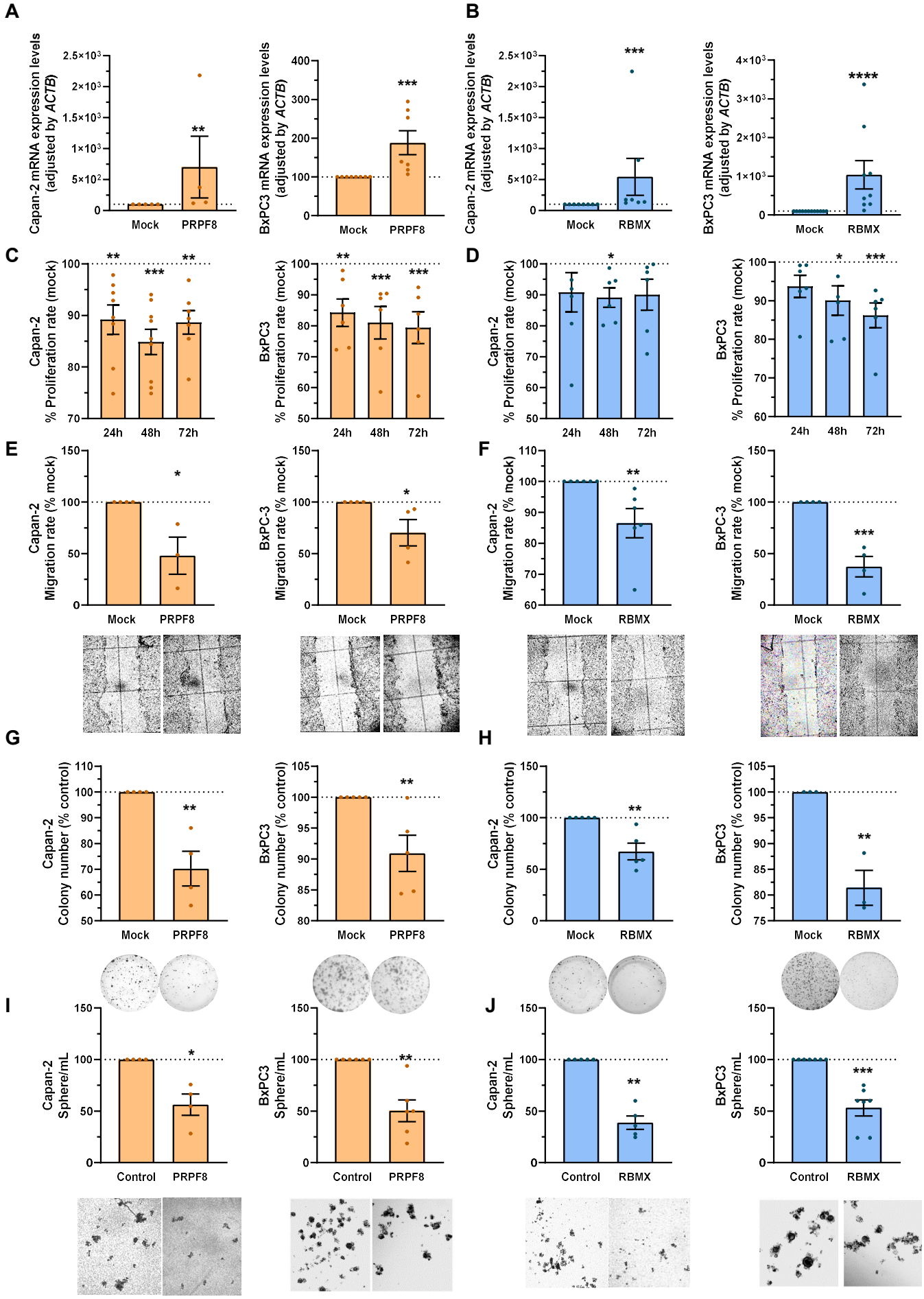

A
Capan-2 mRNA expression levels (adjusted by ACTB)
Mock
PRPF8
BxPC3 mRNA expression levels (adjusted by ACTB)
B
RBMX
C
Capan-2 % Proliferation rate (mock)
BxPC3 % Proliferation rate (mock)
24h
48h
72h
D
E
Capan-2 Migration rate (% mock)
BxPC-3 Migration rate (% mock)
F
G
Capan-2 Colony number (% control)
BxPC3 Colony number (% control)
H
I
Capan-2 Sphere/mL
BxPC3 Sphere/mL
Control
J

Figure R20 (previous page). Effect of PRPF8 and RBMX modulation in PDAC. A, B) qPCR validation of the overexpression of PRPF8 and RBMX respectively in PDAC cell lines. **C,D**) Proliferation rates of Capan-2, and BxPC-3 cell lines after PRPF8 and RBMX overexpression respectively at 24, 48 and 72 hours compared with mock (control; set at 100 %), represented as a dot line. **E,F**) Migration rates of Capan-2, and BxPC-3 cell lines after PRPF8 and RBMX overexpression respectively compared with mock (control; set at 100 %), for 24 hours. Representative images of wound closures. **G,H**) Colony formation capacity of Capan-2, and BxPC-3 cell lines after PRPF8 and RBMX overexpression respectively compared with mock (control; set at 100 %). Representative images of colony formation. **E, J**) Sphere formation capacity of Capan-2, and BxPC-3cell lines after overexpression of PRPF8 and RBMX respectively compared with mock (control; set at 100 %). Representative images of spheres. Data represents mean ± SEM. Asterisks indicates values that significantly differences between groups (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$).

Chapter III

Therapeutic benefit of splicing

Dysregulated splicing factor SF3B1 unveils a dual therapeutic vulnerability to target pancreatic cancer cells and cancer stem cells with an anti-splicing drug

The Splicing Factor 3B Subunit 1 (*SF3B1*) is a spliceosome component essential in pre-RNA processing and the most frequently mutated splicing factor across cancers, particularly in haematological malignancies but also in solid tumors, including PDAC (although at a much lower frequency, 4 % of cases) [110, 204, 209] [reviewed in [52, 210, 211])]. *SF3B1* encodes for a core component of the U2 small nuclear ribonuclear protein (snRNP) and is required for the splicing of most introns, being involved in the recognition of the branch-site, an early stage of spliceosome assembly [52, 96, 210, 211]. Somatic mutations in *SF3B1* in cancer alter the correct recognition of pre-RNA patterns by the spliceosome due to reduced fidelity of branch-point selection and has recently been found to promote tumor glycolysis in PDAC [52, 210-212]. However, the pathological importance of *SF3B1* does not only rely on the well characterized role of *SF3B1* mutations, but growing evidence indicates that alteration of its expression can also have malignant consequences in some cancers, such as prostate cancer [95] and hepatocarcinoma [160]. These studies also underscore the potential of altered SF3B1 as a therapeutic target, as several drugs like Pladienolide-B (a macrocyclic lactone produced by *Streptomyces* sp.) and its derivatives can inhibit SF3B1 function and thereby exert antitumoral effects in several cancers [95, 160].

To date, expression of *SF3B1* and its potential as a therapeutic target have not been explored in detail in PDAC nor in pancreatic cancer stem cells (CSCs), a small population of undifferentiated cells capable of initiating tumor generation, differentiation, and self-renewal, and thus, are key drivers of tumor evolution, metastasis, and relapse [148, 213]. CSCs comprise distinct subsets with inherent characteristics, such as autofluorescence activity or the expression of specific cell surface antigens and receptors (mainly CD133,

EpCAM, CXCR4 or CD44) [159]. Currently, new approaches seek to increase the susceptibility of CSCs to conventional treatments by identifying novel vulnerabilities in these cells. To date, only little evidence suggests splicing dysregulation in PDAC CSCs [214]. Therefore, we aimed to investigate the presence and role of SF3B1 in PDAC as well as its potential value as a therapeutic target.

1. Expression of *SF3B1* in PDAC

Expression levels of *SF3B1* were evaluated by qPCR in RNA isolated from FFPE samples from a cohort of 75 PDAC patients. Main clinical parameters are shown in **Material and Methods Table 1** (see **Appendix 5**). For each patient, tumor tissue was compared with its corresponding NTAT, used as reference. Results revealed that *SF3B1* mRNA expression levels were higher in PDAC tumor tissue compared with NTAT (**Fig. R21A**). Accordingly, IHC staining of 18 randomly selected samples from this same cohort revealed SF3B1 nuclear immunostaining in NTAT (acinar and ductal cells) and cancer cells, where the staining score was higher (**Fig. R21B-C**). Low *SF3B1* expression levels were associated with arterial hypertension (AHT) and type 2 diabetes mellitus (T2DM) in this patient cohort (**Fig. R22**).

To validate our results, *SF3B1* mRNA levels were analyzed in publicly accessible datasets from human samples, the E-MTAB-1791 database (195 PDAC patients and 41 healthy controls) [147], and GSE15471 (36 PDAC samples and corresponding NTAT) [148]. In line with our results, *SF3B1* was overexpressed in both cohorts (**Fig. R21D**: E-MTAB-1791; **Fig. R21E***:* GSE15471). Interestingly, accessible data from the PanCancer study (TCGA) [149] demonstrated that *SF3B1* expression levels were directly associated with neoplasm disease stage, being most expressed in poorly differentiated tumors (**Fig. R21F**). Moreover, *SF3B1* levels were directly associated with lymph node stage, tending

to correlate with metastasis (despite the low number of metastatic patients available; **Fig. R21F**).

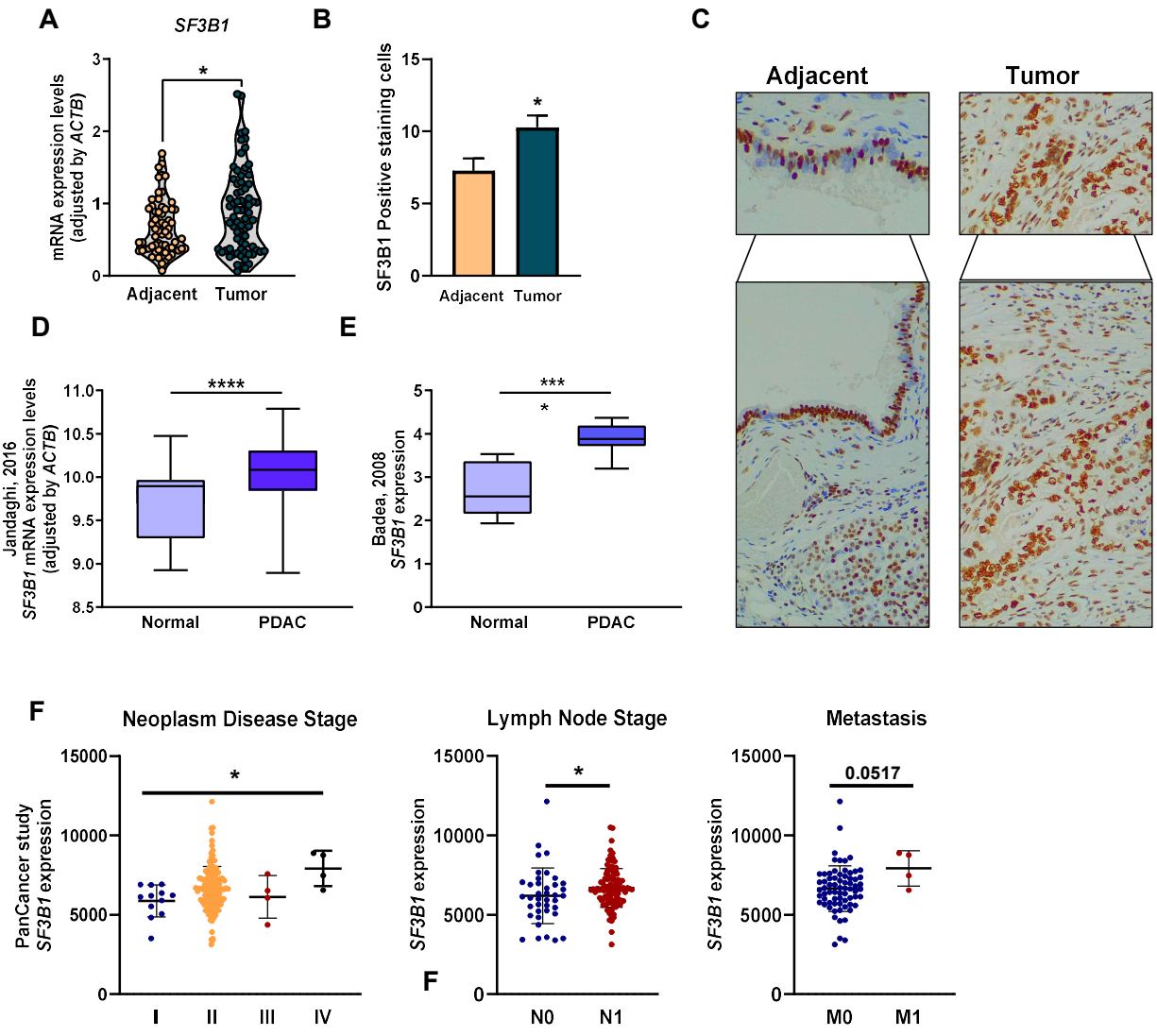


Figure R21. SF3B1 expression in PDAC. A) mRNA levels of SF3B1 adjusted for ACTB gene expression in PDAC FFPE samples compared with non-tumoral adjacent tissue (NTAT). **B**) SF3B1 IHC analysis in PDAC FFPE samples vs. NTAT. **C**) Representative IHC 20X-image; SF3B1 nuclear immunostaining in non-tumoral adjacent tissue is evident in acinar and ductal cells (left panel) and in cancer cells (right panel). **D**) SF3B1 mRNA levels in E-MTAB-1791 [147] comparing PDAC and healthy controls. **E**) SF3B1 mRNA levels in GSE15471 [148] comparing PDAC and NTAT used as a control. **F**) Correlation of SF3B1 mRNA levels with clinical stage, lymph node involvement and distant metastasis (according to WHO) in PanCancer cohort [149]. Data represents mean ± SEM. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

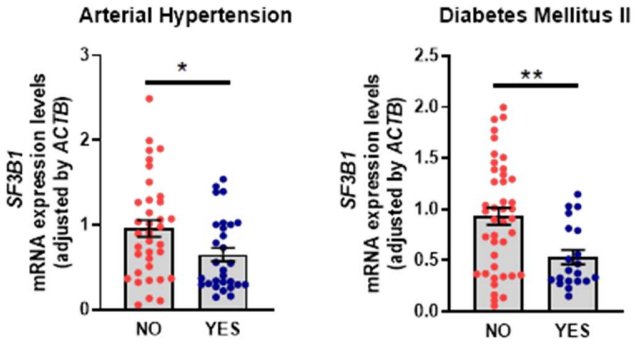


Figure R22. *SF3B1* correlation with clinical parameters. *SF3B1* expression level correlation in patients without/with AHT or T2DM in PDAC FFPE samples compared with non-tumoral adjacent tissue (NTAT).

The potential impact of *SF3B1* expression on alternative splicing in PDAC was assessed with a biocomputational approach that analyzed RNA-Seq data of 94 additional PDAC patient samples (**Appendix 6**), enabling the identification and quantification of splicing events. Samples were first classified into different clusters according to their *SF3B1* expression levels, then the means of the Ψ of each event were compared between groups with high and low expression. This approach detected 482 splicing events that were significantly different according to p value and ΔΨ of the total of 240,941 events detected using SUPPA2 (**Fig. R23A**). Indeed, the general pattern of splicing events differed depending on *SF3B1* expression levels, as these significantly different events displayed a higher frequency of skipped exons, alternative 3' splice sites and alternative 5' splice sites, and lower frequency of alternative first or last exons, compared to the average of all the events calculated (**Fig. R23B**). We used an additional software, rMATs, where we observed a similar pattern of splicing, specifically a higher frequency of alternative 3' splice sites and alternative 5' splice sites (data not shown). These results were supported with a validation cohort where 91,860 events were detected, being 57 of them significantly different (**Fig. R23C**), showing a similar pattern of distinct splicing events depending on *SF3B1* expression levels. Interestingly, exon skipping and

alternative 3' splice site events were over-represented in PDAC samples expressing high *SF3B1* levels, while mutually exclusive exons, alternative first exon and alternative last exon events prevailed in tumors expressing low *SF3B1* levels (**Fig. R24**). Importantly, some of the most pronounced changes were validated in an external PDAC cohort (**Fig. R25**).

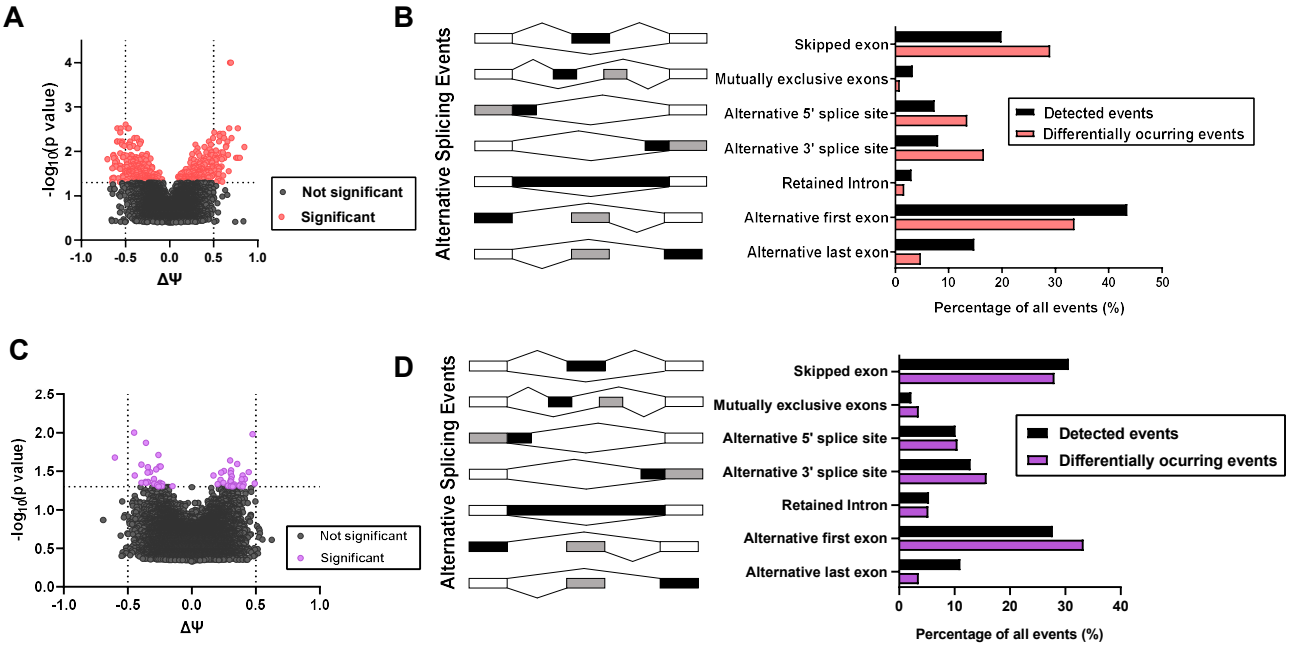


Figure R23. Relationship of *SF3B1* expression levels with splicing event patterns in PDAC. **A**) Volcano-plot where ΔΨ of total events calculated is plotted against the –log10 p-value of the Fisher's Exact Test to assay differential splicing events between high and low *SF3B1* expression groups of samples, showing that *SF3B1* tumor expression may influence alternative splicing pattern. Alternative Splicing event characterization of RNA-Seq samples. **B**) Total splicing events detected (black) and significantly different events between *SF3B1* expression groups (red) are classified depending on their type, showing different frequencies (%) between both conditions. **C,D**)Volcano-plot and differential splicing events between high and low *SF3B1* expression groups of GSE79670 cohort samples.

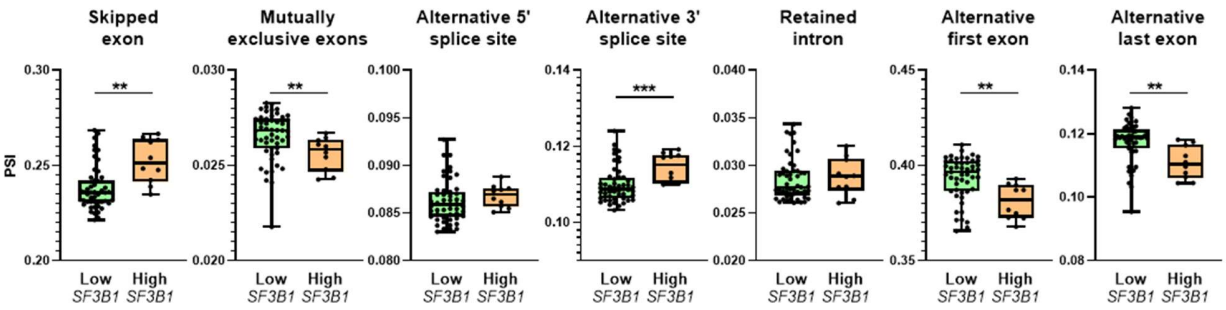
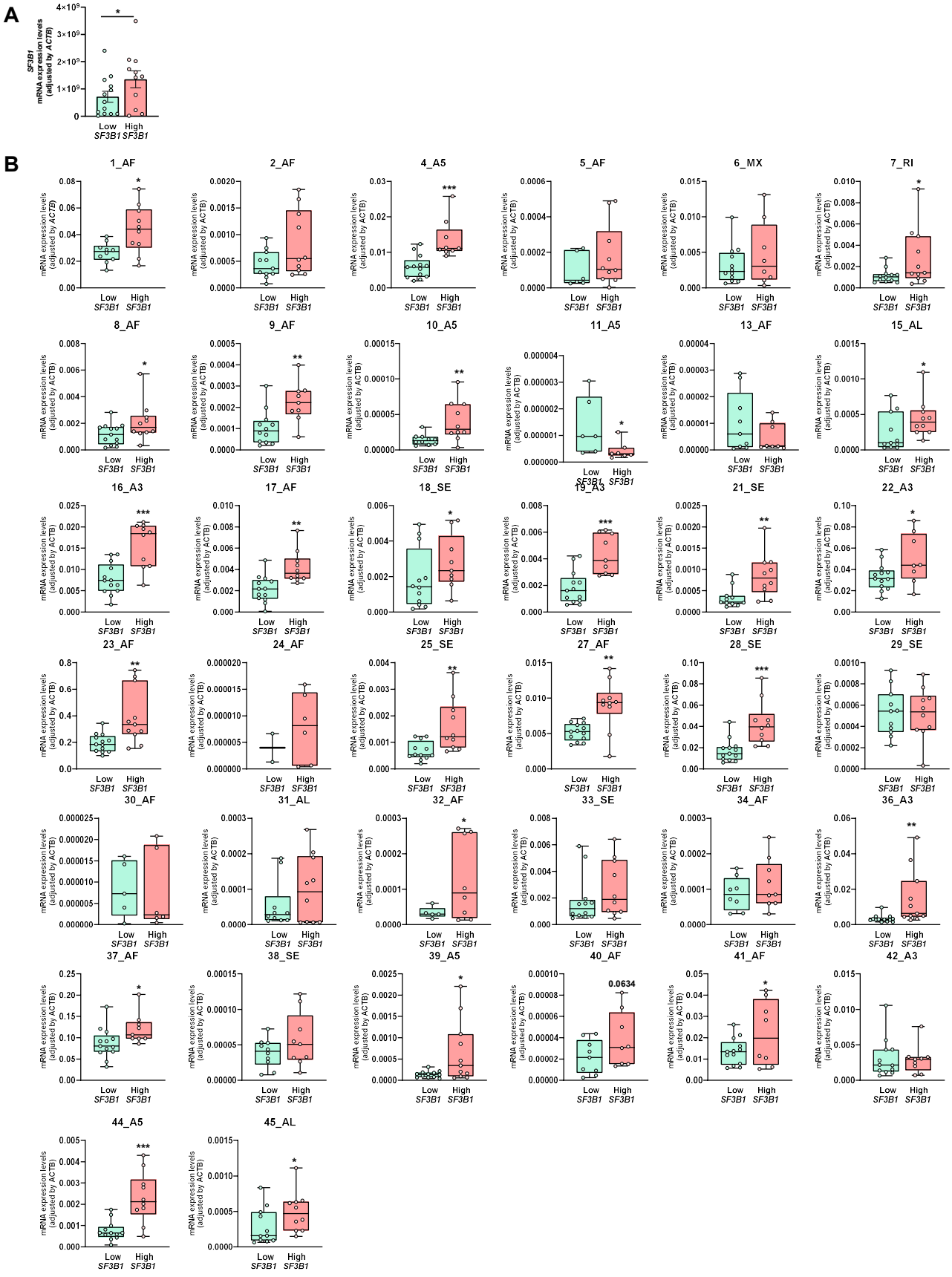


Figure R24. Relationship of *SF3B1* expression levels with splicing event patterns of key genes in PDAC. Levels of expression of significantly different alternative splicing events transcripts between High and Low *SF3B1* expression groups. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

Figure R25 (next page). Relationship validation of *SF3B1* expression levels with splicing event patterns of key genes in PDAC. A) Levels of *SF3B1* expression in PDAC samples from patients of the fresh tissue cohort in low and high *SF3B1* expression groups. **B**) Levels of expression of significantly different alternative splicing events transcripts (**Appendix 7**) between High (n=11) and Low (n=13) *SF3B1* expression groups. Gene expression was normalized to *ACTB* expression. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).


A
SF3B1 mRNA expression levels (adjusted by ACTB)
Low SF3B1
High SF3B1
B
1_AF
2_AF
4_A5
5_AF
6_MX
7_RI
8_AF
9_AF
10_A5
11_A5
13_AF
15_AL
16_A3
17_AF
18_SE
19_A3
21_SE
22_A3
23_AF
24_AF
25_SE
27_AF
28_SE
29_SE
30_AF
31_AL
32_AF
33_SE
34_AF
36_A3
37_AF
38_SE
39_A5
40_AF
41_AF
42_A3
44_A5
45_AL
mRNA expression levels (adjusted by ACTB)
0.0634

KEGG analysis of the genes differentially spliced, depending on *SF3B1* expression, revealed a particularly tight association with the “pancreatic cancer” category (the term with the highest gene ratio i.e., number of hits divided by the total genes of that KEGG term), but also with colorectal cancer and relevant signaling pathways in cancer (**Fig. R26A**). Moreover, analysis of the genes provided by KEGG and Reactome allowed for identification of a number of key signaling pathways, particularly AKT-related (**Fig. R26B**).

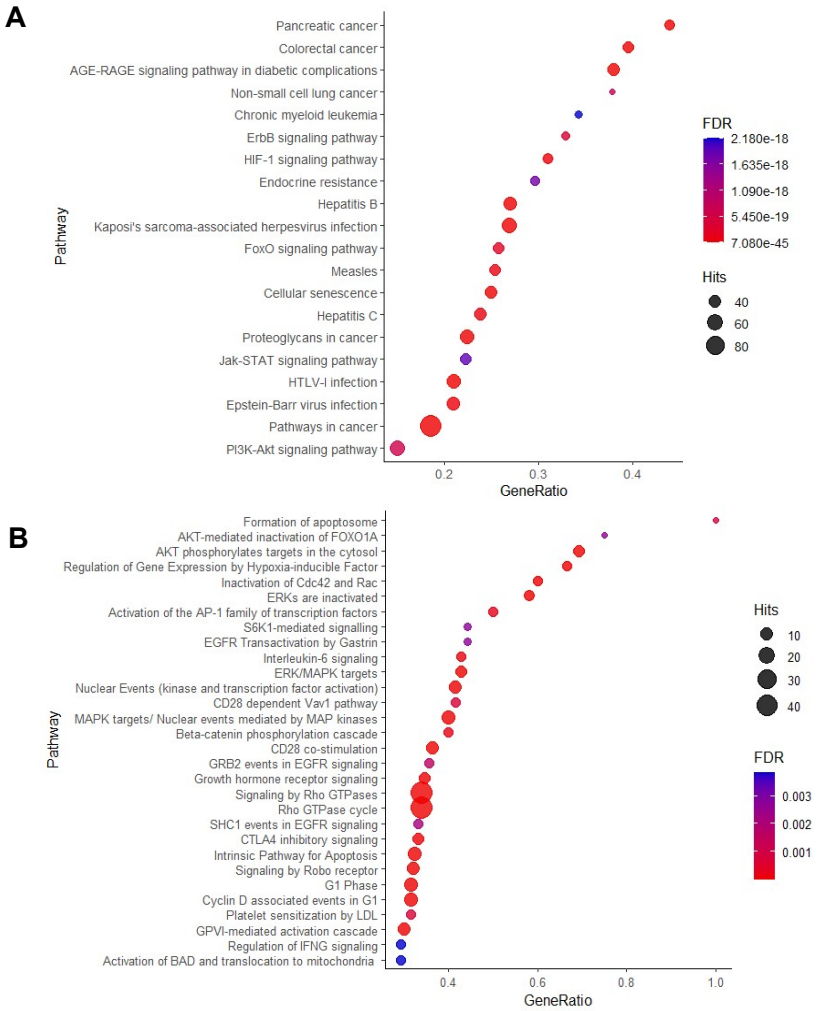


Figure R26. *In silico* relationship between key pathways and differentially spliced genes depending on *SF3B1* expression. A) KEGG analysis of significantly different spliced genes depending on *SF3B1* expression. Ratio of the genes’ hits over the total genes of a pathway (X-axis) is plotted for each pathway (Y-axis). The size of each point denotes the genes hits, and the color represents their significancy. **B**) Reactome analysis of significantly different spliced genes depending on *SF3B1* expression.

Further analysis of our RNA-Seq data and the PanCancer dataset indicated that *SF3B1* expression levels correlated directly with *KRAS*, *BRCA1*, *BRCA2,* and *HNRNPK* and inversely with *CDKN2A* and *TP53* mRNA levels (**Fig. R27A-B**). Conversely, *SF3B1* expression did not seem to be associated with the mutational status of key driver genes (*KRAS, CDKN2A*, *SMAD4*, *TP53, BRCA1, BRCA2,* and *HNRNPK*) in the PanCancer PDAC dataset (**Fig. R27C**).

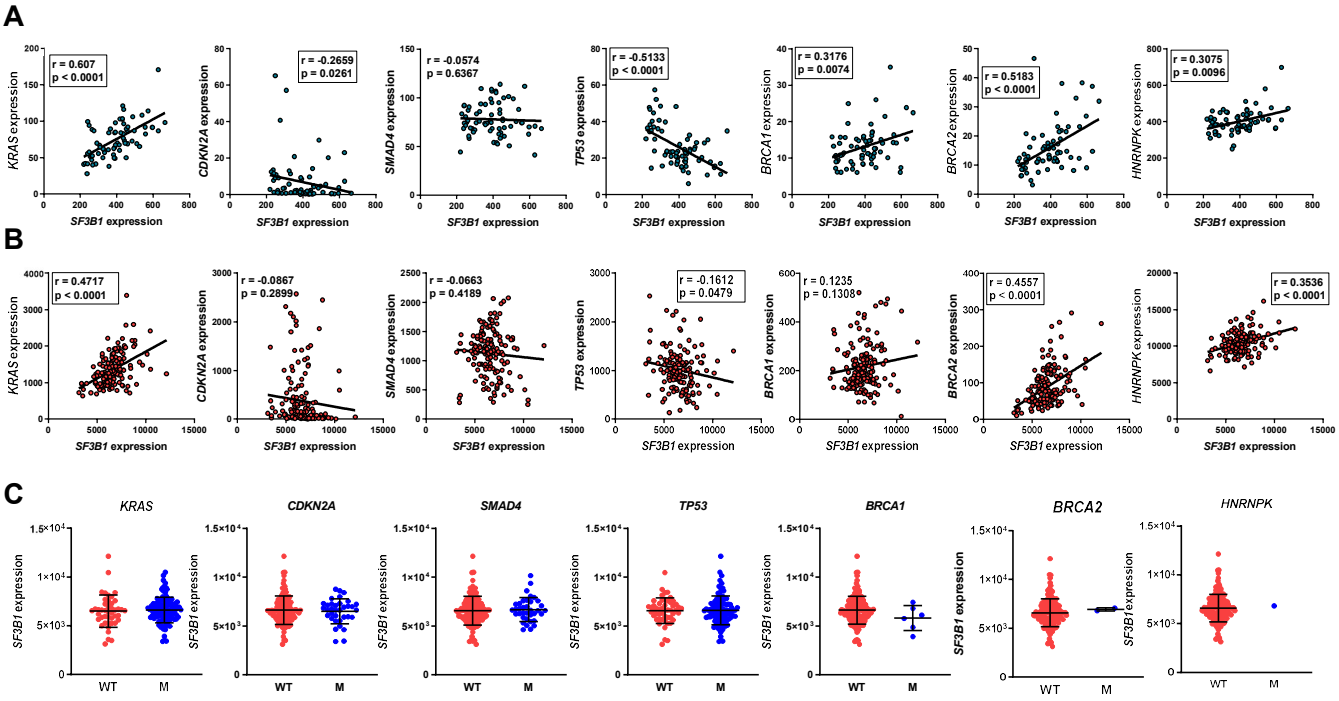


Figure R27. Correlations between *SF3B1* and key genes. Correlations between *SF3B1* and *KRAS, CDKN2A, SMAD4, TP53, BRCA1, BRCA2,* and *HNRNPK* mRNA levels in our RNA-Seq cohort **(A)** and PanCancer cohort **(B)** .**C)** Levels of SF3B1 expression in PDAC samples from patients of the PanCancer cohort with (M) or without (WT) mutations in *KRAS, CDKN2A, SMAD4, TP53, BRCA1, BRCA2* and *HNRNPK.*

2. SF3B1 inhibition alters functional features as well as signaling and splicing mechanisms in PDAC cell lines

To explore the role of SF3B1 in PDAC, we silenced its expression with a specific siRNA or inhibited its function pharmacologically. PDAC cell lines expressed appreciable

mRNA levels of *SF3B1* (**Fig. R28A**) that were efficiently silenced (40-80 %) in all cells tested (**Fig. R28B**).

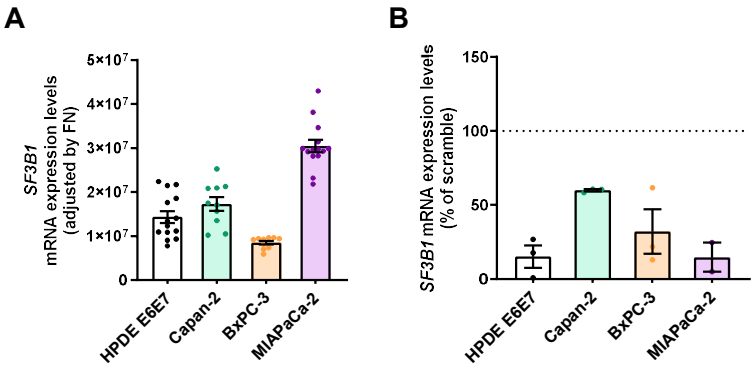


Figure R28. ***SF3B1* expression levels in PDAC cell lines. A)** *SF3B1* basal expression levels in normal pancreatic HPDEE6E7 cells and Capan-2, BxPC-3 and MIAPaCa-2 PDAC cell lines adjusted by a normalization factor calculated from the expression levels of *HPRT1*, *GAPDH* and *ACTB* (n=5-7). **B)** qPCR validation of the silencing efficiency achieved with SF3B1 specific siRNAs in PDAC cell lines. mRNA expression levels were normalized to *ACTB* expression levels. Data are expressed as a percentage of control (Scramble; set at 100%) (n=2-3).

SF3B1 silencing time-dependently decreased cell proliferation in PDAC cell lines: well differentiated Capan-2 (less prominently), moderately differentiated BxPC-3, and poorly differentiated MIAPaCa-2 [155], and particularly in the non-tumoral pancreatic cell line HPDE E6E7 (**Fig. R29A**). We then applied an alternative experimental approach by pharmacologically blunting SF3B1 activity, instead of its expression, using the specific inhibitor Pladienolide-B [101]. Initial screenings in PDAC cell lines using various Pladienolide-B doses led us to select a 1 nM dose for subsequent studies (**Fig. R29C**). Pladienolide-B time-dependently reduced proliferation in all PDAC cell lines (**Fig. R29B**), in a manner that parallel their reported degree of aggressiveness. Interestingly, Pladienolide-B did not alter proliferation of non-tumoral HPDE E6E7 cells, suggesting a tumor cell-specific effect. Intriguingly, both conditions revealed a distinct cell response, where the anti-proliferative effect of either Pladienolide-B or *SFB31* silencing appeared to be associated to changes in intracellular distribution. Specifically, decreases in proliferation were accompanied by an increased proportion of cytoplasmic SF3B1

staining (**Fig. R30**). Comparing the actions of Pladienolide-B and the first line PDAC chemotherapeutic drug Gemcitabine showed that both drugs exerted comparable effects on all PDAC cell lines tested; however, their combination did not produce an additive inhibitory effect (**Fig. R29D**).

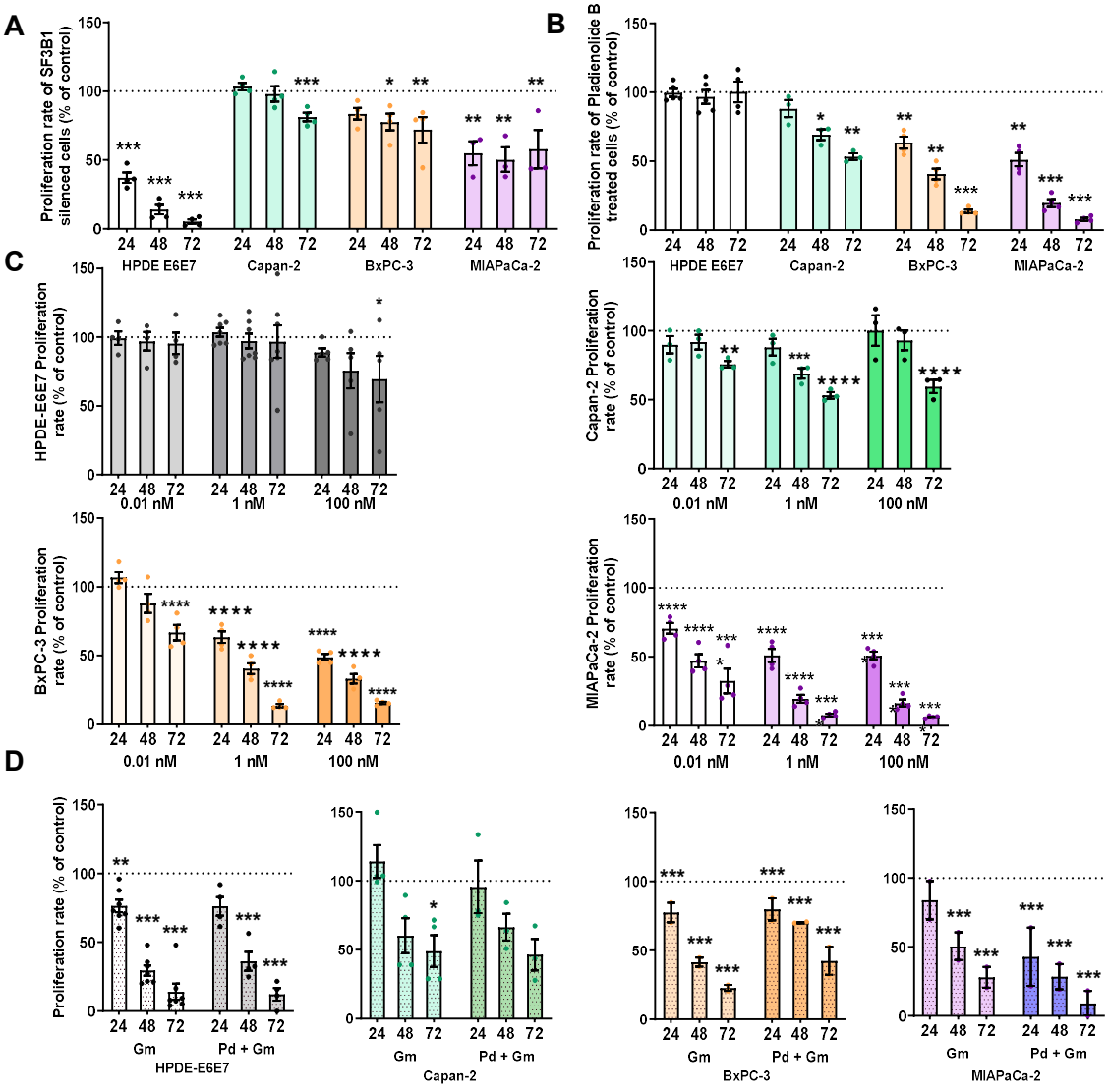
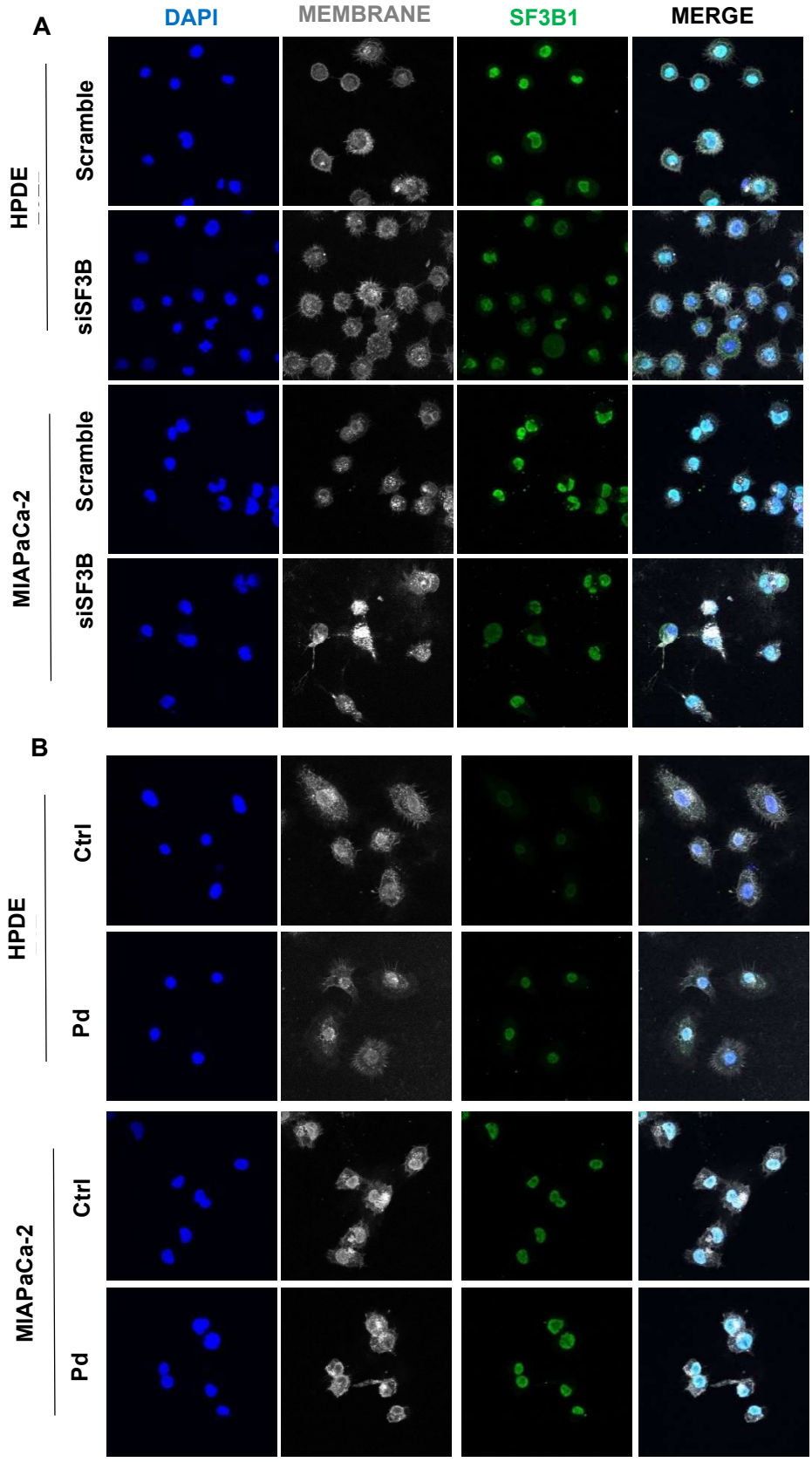


Figure R29. Effects of SF3B1 inhibition on proliferation rates in PDAC cell lines. A) Proliferation rates of HPDE-E6E7, Capan-2, BxPC-3, and MIAPaCa-2 cell lines after SF3B1 silencing compared with scramble control-silenced cells (set at 100 %; dotted line; n=3-4). **B**) Proliferation rates of same cell lines treated with or without (vehicle, set at 100 %; dotted line) splicing (SF3B1) inhibitor Pladienolide-B (n=3-5). **C**) Proliferation rates of HPDEE6E7, Capan-2, BxPC-3 and MIAPaCa-2 cell lines following treatment with different doses of Pladienolide-B compared to vehicle-treated control cells, (set as 100 %, represented as a dotted line; n=3-5).

Asterisks indicate significant differences (*p<0.05; **p<0.01; ***p<0.001).**D**) Gemcitabine (Gm) and Pladienolide-B plus Gemcitabine (Pd+Gm) treated cells compared with vehicle-treated cells (set at 100 %; dotted line; n=3-5).



C

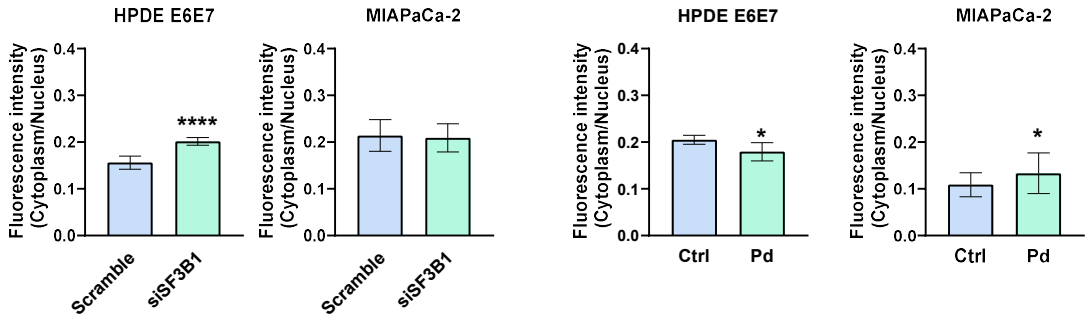


Figure R30. SF3B1 localization in HPDE E6E7 and MIAPaCa-2 cell lines. **A**) Confocal images of Control and Pladienolide-B treated HPDE E6E7 and MIAPaCa-2 cell lines. **B**) Confocal images of Scramble and siSF3B1 in HPDE E6E7 and MIAPaCa-2 cell lines. **C**) Quantification of SF3B1 subcellular distribution in each set of samples. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

Pladienolide-B reduced the migration rate of the three PDAC cell lines assessed in a wound-healing assay, while no such effect was observed in non-tumoral HPDE E6E7 cells (**Fig. R31A**). Interestingly, MIAPaCa-2 cells, regarded as the most aggressive and stem-like [215] of the three PDAC cell lines tested, displayed the most pronounced reductions in migration and proliferation in response to Pladienolide-B. Hence, this cell line was selected to further explore the effects of the drug in subsequent stem-associated assays, using non-tumoral HPDE E6E7 cells in parallel, where appropriate. Pladienolide-B reduced by half the sphere formation (i.e., self-renewal) capacity of MIAPaCa-2 cells compared to vehicle-treated cells (**Fig. R31B**). Likewise, Pladienolide-B inhibited colony formation of MIAPaCa-2 cells with respect to vehicle-treated cells (**Fig. R31C**). Furthermore, Pladienolide-B increased apoptotic in MIAPaCa-2 cells but not in HPDE E6E7 cells (**Fig. R31D**).

To gain mechanistic insights into the observed effects of Pladienolide-B, we explored the activation, expression or splicing of key signaling players/regulatory genes in PDAC cells (**Fig. R32**). Pladienolide-B decreased AKT and increased JNK phosphorylation in MIAPaCa-2 cells (**Fig. R32A**) without altering ERK1/2 phosphorylation (not shown). Intriguingly, Pladienolide-B did not influence the expression of genes relevant to tumor

biology (apoptosis, proliferation, inflammation) in MIAPaCa-2 cells or in HPDE E6E7 cells, including *NFKB1*, *CASP3*, *MKI67*, and *HER2* (**Fig. R32B**).

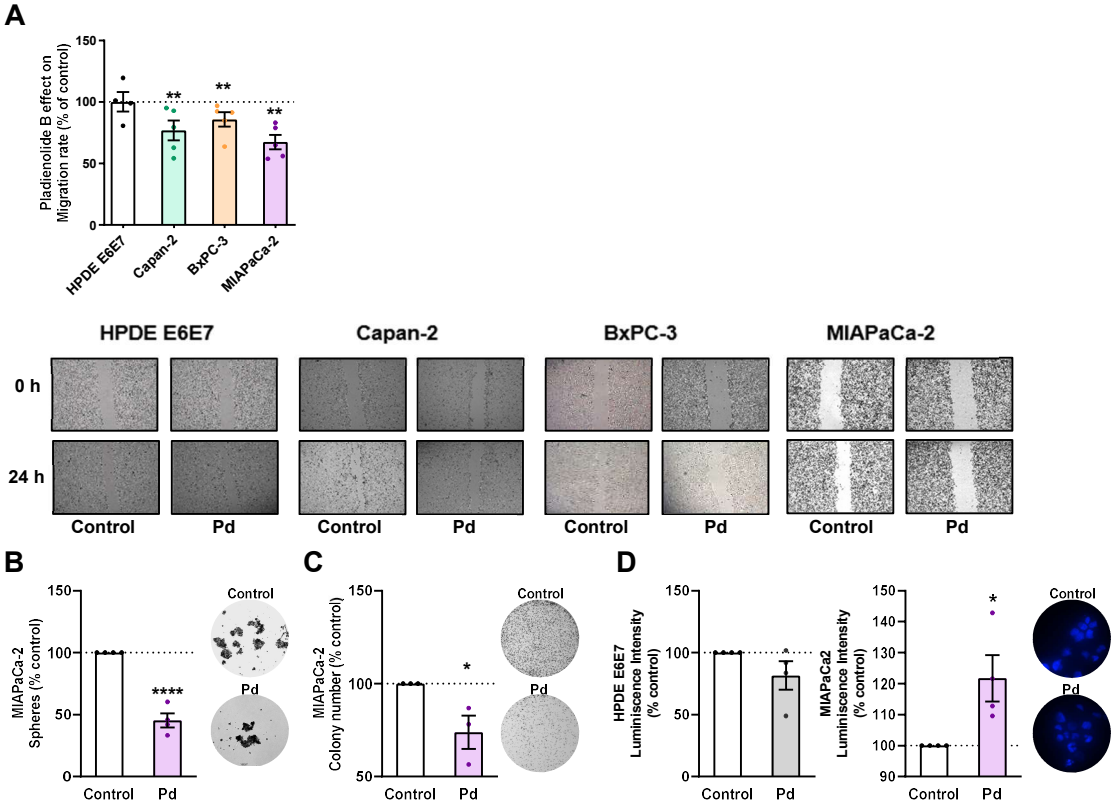


Figure R31. Effect of SF3B1 modulation on PDAC cell lines. A) Migration rates of HPDE E6E7, Capan-2, BxPC-3 and MIAPaCa-2 cell lines treated with or without (vehicle; set at 100 %) Pladienolide-B for 24h. Representative images of wound closures (n=4). **B**) Quantification of sphere formation capacity of MIAPaCa-2 treated with Pladienolide-B or vehicle (control; set at 100 %). Representative images of spheres (n=4). **C**) Colony formation capacity quantification of MIAPaCa-2 treated with Pladienolide-B or with vehicle (control; set as 100 %). Representative images of colony formation (n=3). **D**) Apoptosis quantification using Caspase-3/7 assay in HPDE E6E7 and MIAPaCa-2 treated 24h with Pladienolide-B or vehicle (control; set as 100 %) (n=4). Representative images show MIAPaCa-2 nuclear staining with DAPI. Data represents mean ± SEM. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

Conversely, this drug did modify the expression pattern of splicing-related isoforms of key PDAC-related genes. Thus, Pladienolide-B increased the levels of the pro-apoptotic splice isoform *BCL-XS* but not of the long, anti-apoptotic *BCL-XL* isoform in MIAPaCa-2 cells. Importantly, these effects were not observed in HPDE E6E7 cells (**Fig. R32C**). Furthermore, while Pladienolide-B did not alter total *KRAS* mRNA levels in HPDE

E6E7 or MIAPaCa-2 cells, it augmented the expression of the splice isoform *KRAS4a* only in MIAPaCa-2 cells (**Fig. R32D**).

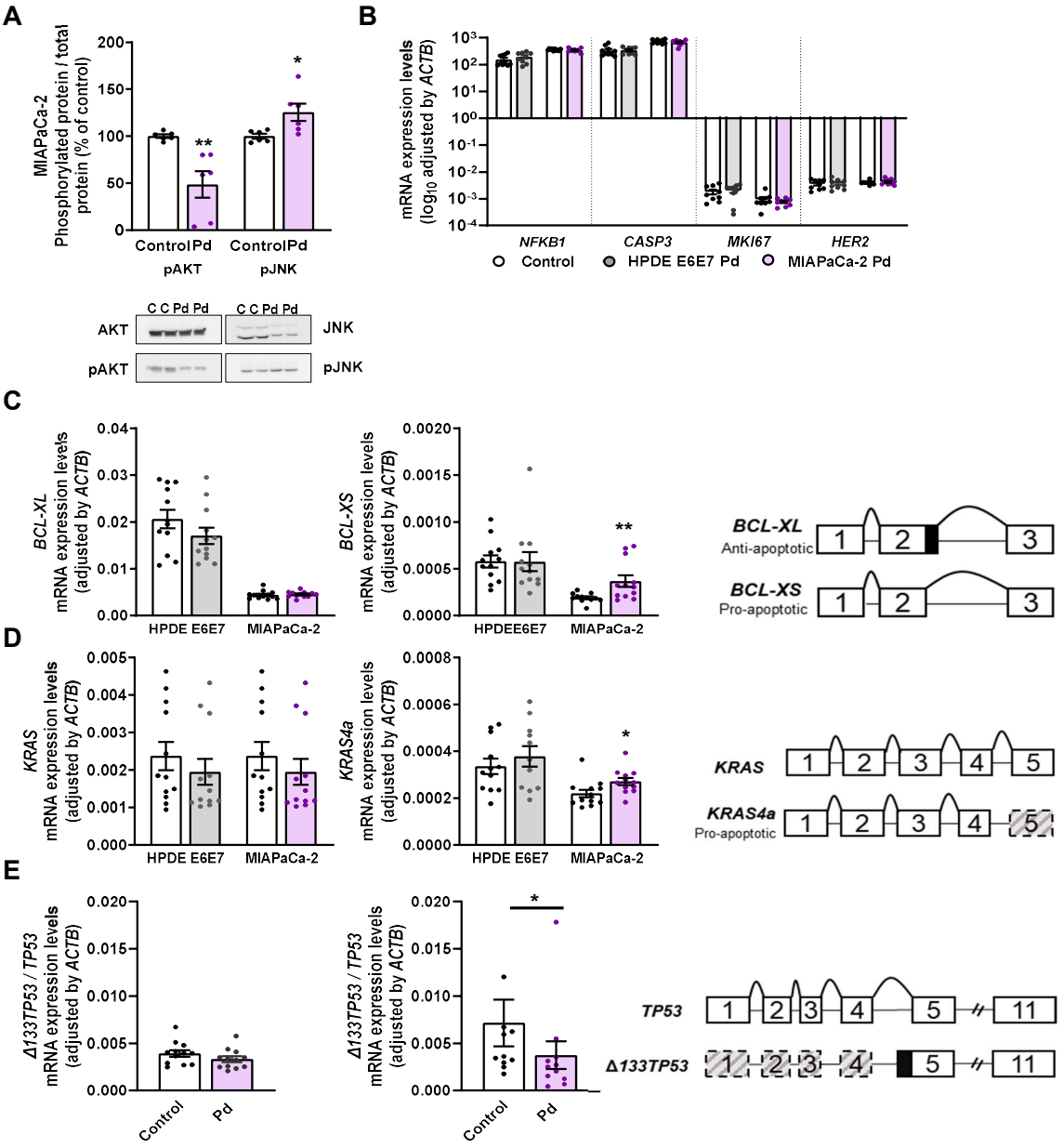


Figure R32. Molecular profile of Pladienolide-B-treated PDAC cell lines. A) Western-blot analysis of p/tAKT p/tJNK in MIAPaCa-2 cell lines after 24h Pladienolide-B treatment. Ponceau-stained membranes served as loading control reference (n=6). **B**) mRNA levels in malignancy-implicated genes. Values represent the $\log_{10}$ of expression compared to vehicle-treated (n=4). **C**) mRNA levels of *BCL-XL* and *BCL-XS* in HPDE E6E7 and MIAPaCa-2 cells treated 24h with or without (vehicle, control) Pladienolide-B (n=4). **D**) mRNA expressions of *KRAS* and *KRAS4a* in HPDE E6E7 and MIAPaCa-2 cells treated 24h with or without (vehicle, control) Pladienolide-B (n=4). **E**) Ratio of *Δ133TP53/TP53* mRNA levels in HPDE E6E7 and MIAPaCa-2 cells treated 24h

with or without (vehicle, control) Pladienolide-B (n=4). Gene expression was normalized to *ACTB* expression. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

Pladienolide-B also modulated *TP53* in MIAPaCa-2, but not in HPDE E6E7 cells, by increasing full-length *TP53* expression while blunting its truncated variant *Δ133TP53*, resulting in a decreased *Δ133TP53*/*TP53* ratio (**Fig. R32E**). Therefore, not only is *SF3B1* overexpressed in PDAC, which may influence the splicing profiles in cancer cells, but the splicing inhibitor Pladienolide-B reduces pivotal functional (proliferation, migration) and stem-associated features (colony- and sphere-formation), likely by altering key signaling and splicing events.

3. Pladienolide-B attenuates PDAC stemness functional properties

To investigate a possible role for SF3B1 in pancreatic CSCs, we tested four previously characterized human PDX-derived cell lines (i.e., A6L, 215, 253, and 354), which contain *bona fide* pancreatic CSCs (**Fig. R33**) [159]. *SF3B1* expression was first evaluated in adherent (ADH) and spheroid (SPH) cell cultures derived from these cell lines, which represent, respectively, cancer- and CSC-enriched cell populations from the corresponding PDXs. While all the tumors analyzed expressed *SF3B1*, levels were lower in spheroid CSC-enriched cultures (both on average and in each line), suggesting that CSCs naturally express less *SF3B1* than their more differentiated counterparts (**Fig. R33A, B**).

To assess the impact of SF3B1 blockade specifically in CSCs, Pladienolide-B effects on PDX-derived cell lines were tested using multiple stem-related assays (**Fig. R33C**). First, the levels of autofluorescence and CD133, established pancreatic CSC markers [159], were evaluated in sphere-derived cells from PDX-derived cell lines. Interestingly, Pladienolide-B transiently (at 48h) reduced autofluorescence in A6L cells while causing

an early and sustained (24-72h) increase in 215 cells, suggesting CSCs enrichment (**Fig. R33D**; not measured in 354 and 253 cells, which lack autofluorescence). Importantly, in all cell lines Pladienolide-B induced an early (except 253 cells) and sustained enrichment in CD133, again suggesting an enrichment in CSC-marker positive cells (**Figure R33E**).

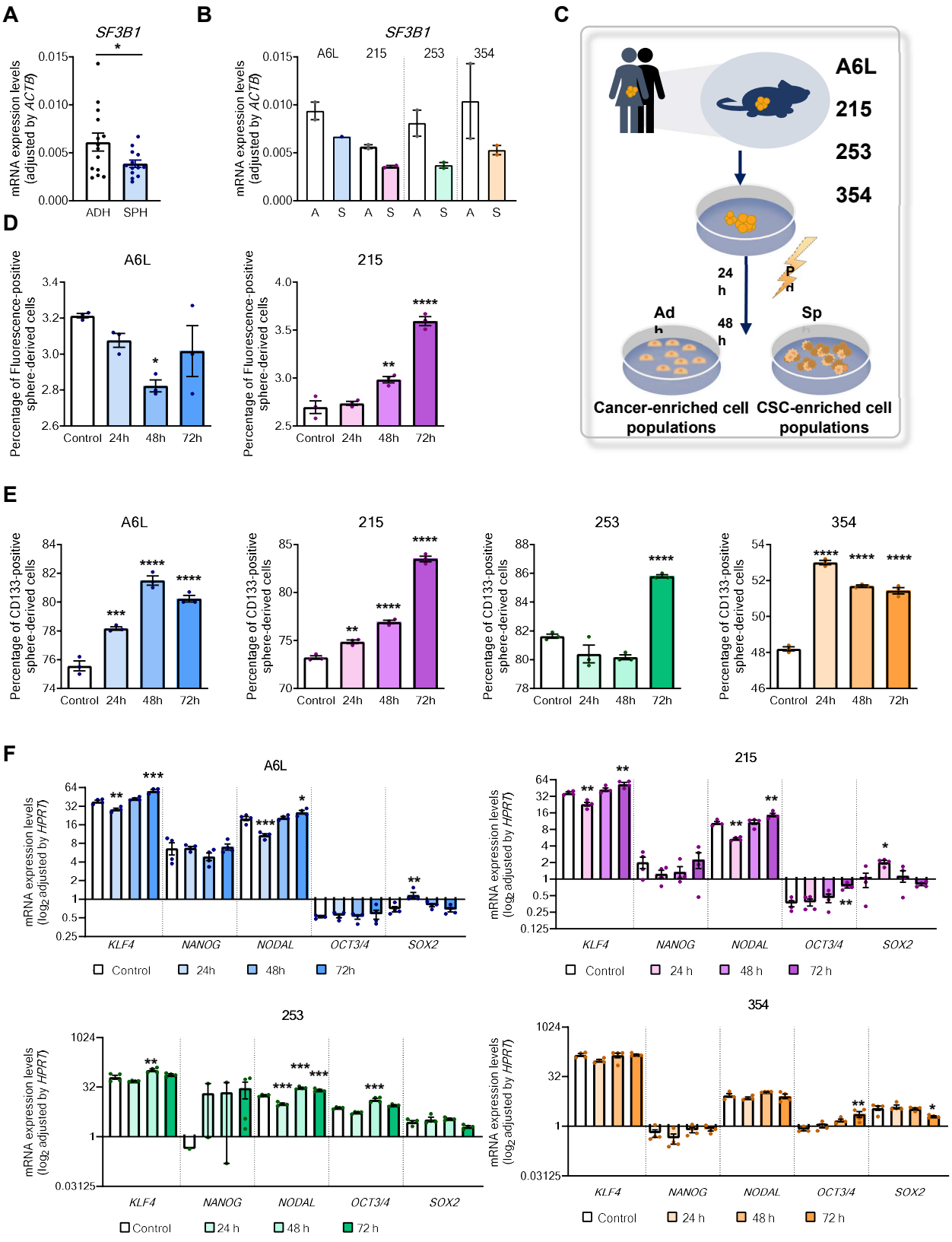

A
SF3B1
mRNA expression levels (adjusted by ACTB)
ADH
SPH
B
SF3B1
A6L
215
253
354
C
A6L
215
253
354
24 h
48 h
Ad h
Sp
Cancer-enriched cell populations
CSC-enriched cell populations
D
A6L
215
Percentage of Fluorescence-positive sphere-derived cells
Control
24h
48h
72h
E
A6L
215
253
354
Percentage of CD133-positive sphere-derived cells
F
A6L
215
253
354
mRNA expression levels (log2 adjusted by HPRT)
KLF4
NANOG
NODAL
OCT3/4
SOX2
Control
24h
48h
72h
24 h
48 h
72 h

Figure R33 (previous page). SF3B1 expression and consequences of its modulation in PDAC CSCs. A, B) SF3B1 mRNA levels (normalized to *ACTB* expression) in adherent (ADH; A)- vs. sphere (SPH; S)-derived PDX PDAC cells. The *SF3B1* mRNA levels are grouped in A or individual in B for the PDX-derived cell set (A6L, 215, 253 and 354) (n=2). **C**) Representative scheme of PDAC PDX-derived cell line generation and treatment with Pladienolide-B. **D, E**) Quantification of flow cytometric analysis of the percentage of autofluorescent (Fluo) and CD133-positive cells in the indicated PDX-derived in vitro cultures treated with or without (vehicle) 1 nM Pladienolide-B (n=3). **F**) Log_2 mRNA expressions levels of genes implicated in stemness normalized to HPRT expression (n=4). Data shown are mean ± SEM. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

We next evaluated the influence of Pladienolide-B on the expression of CSC stem- or pluripotency-related transcription factors in PDX-derived cell lines, which, except for 354 cells, largely showed comparable response patterns (**Fig. R33F**). Specifically, *KLF4* and *NODAL* expression displayed a biphasic response in A6L, 215 and 253 cells, whereas *SOX2* expression increased in A6L and 215 cells at 24h, and decreased at 72h in 354 cells, and *OCT3/4* displayed disparate responses among the cell lines (**Fig. R33F**). These data suggests that Pladienolide-B only marginally influences the transcription of stem-associated genes in PDX-derived PDAC cells.

While the above data could suggest a CSCs enrichment, we examined the functional consequences of Pladienolide-B treatment, by evaluating the capacity of PDX-derived PDAC cell lines to form colonies or spheres. Pladienolide-B reduced the colony-formation capacity of A6L, 215 and 253 cells in a drastic, rapid (24h) and sustained (72h) manner, while 354 cells showed a slightly delayed (72h) response (**Fig. R34A**). Accordingly, Pladienolide-B clearly reduced the capacity of A6L, 215 and 253 PDX-derived PDAC cell lines to form spheres (**Fig. R34B**), mimicking the response of MIAPaCa-2 cells treated with Pladienolide-B (**Fig. R31B**).

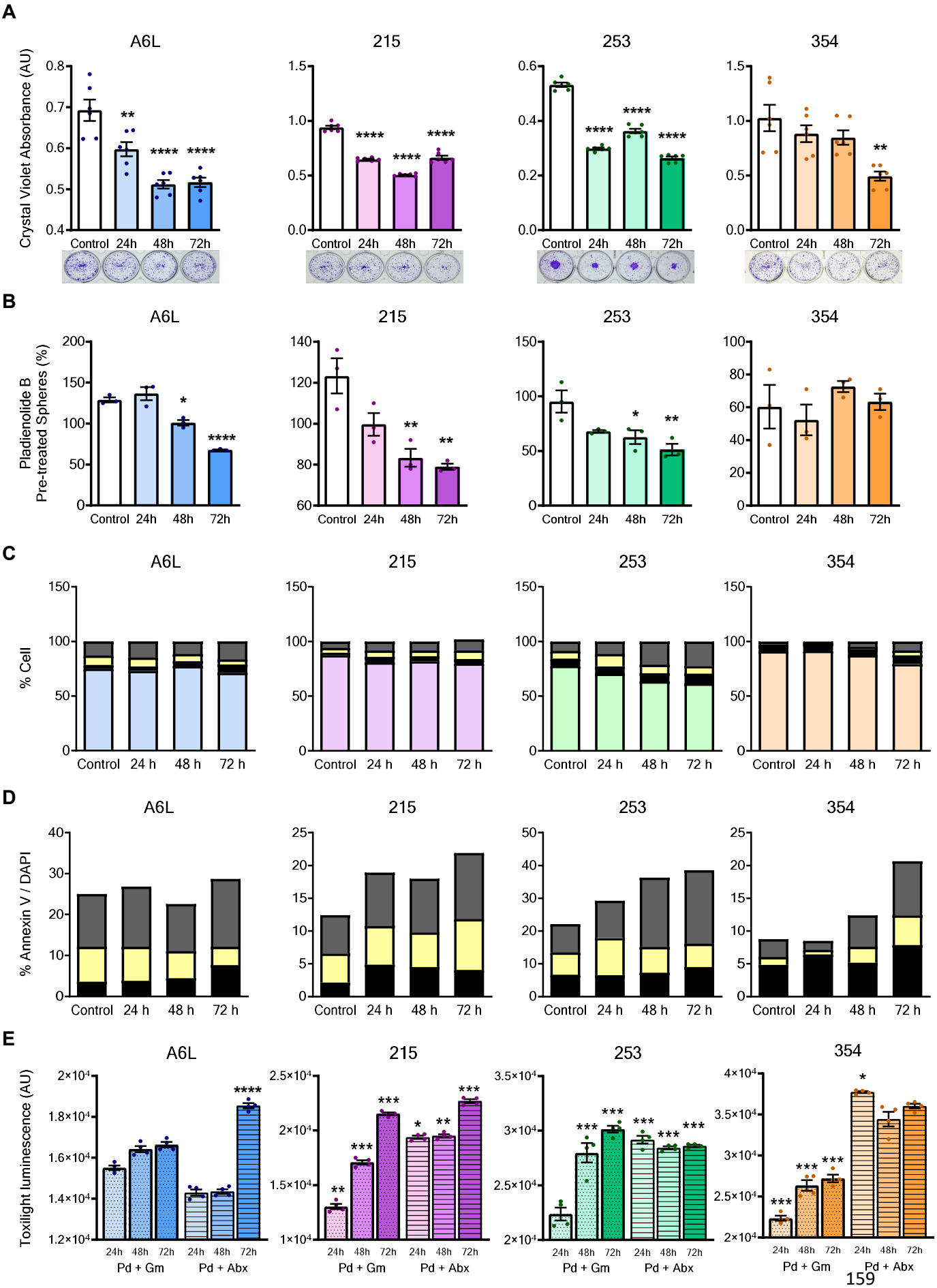

A
A6L
215
253
354
Crystal Violet Absorbance (AU)
Control 24h 48h 72h
B
Pladienolide B
Pre-treated Spheres (%)
C
% Cell
Control 24 h 48 h 72 h
D
% Annexin V / DAPI
E
Toxilight luminescence (AU)
24h 48h 72h 24h 48h 72h
Pd + Gm
Pd + Abx

Figure R34 (previous page). Effect of Pladienolide-B on PDAC CSC functional properties. **A**) Colony formation efficiency represented as crystal violet absorbance (AU; arbitrary units) in PDAC PDX-derived cells after Pladienolide-B treatment compared with vehicle-treated cells. Representative images of colony formation (n=6). **B**) Sphere formation efficiency (number of spheres/mL) in PDAC PDX-derived cells after Pladienolide-B treatment compared to vehicle-treated cells (n=3). **C, D**) Quantification of annexin-V staining in Pladienolide-B-treated cells across PDX-derived in vitro cultures compared with vehicle-treated cells. Top and bottom: live cells (blue); dead cells (black); early apoptosis (yellow) and late apoptosis (grey) (n=1). **E**) Cell death, measured with the bioluminescence Toxilight assay, after treatment with the indicated compounds in combination with Pladienolide-B (n=4). Data represents mean ± SEM. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

The effect of Pladienolide-B on the viability of PDX-derived PDAC cell lines was limited (**Fig. R34C**), with live cells consistently remaining above 60 % in the presence of Pladienolide-B. However, an apparent responsiveness gradient was noticed when apoptosis was assessed, with A6L cells exhibiting higher resistance and 253 cells being more sensitive to Pladienolide-B (**Fig. R34D**). Early and late apoptotic rates revealed a clear time-dependent trend towards increased late apoptosis, particularly in 354 and 253 cells. Thus, the effects observed in sphere and colony formation may result from Pladienolide-B selectively targeting CSCs.

A hallmark of CSCs is their inherent chemoresistance. Thus, we tested the capacity of Pladienolide-B to sensitize pancreatic CSCs to Gemcitabine or Abraxane, two first-line PDAC treatments. A luminescence-based toxicity assay showed that Pladienolide-B increased the cytotoxic capacity of Gemcitabine and Abraxane, with 215 and 253 cell lines showing the highest cell death increase upon addition of Pladienolide-B compared to Gemcitabine or Abraxane alone (**Fig. R34E**). As expected, A6L and 354 cells were more resistant. These results demonstrate that SF3B1 is present in PDAC CSCs and that targeting its function with Pladienolide-B cause alterations that reduce key stemness features, decreasing their ability to form colonies and spheres, and enhancing their susceptibility to Gemcitabine or Abraxane.

4. Pladienolide-B affects PDAC cell and CSC *in vivo* tumor formation

To test whether the inhibitory effects exerted by Pladienolide-B *in vitro* in PDAC cell lines and PDX-derived cell lines could also be observed *in vivo*, we employed two complementary preclinical models. First, PDAC cells were intravenously injected in zebrafish (an adequate system for real-time tracking of CSC-mediated early metastasis and tumor formation) [164]. Specifically, MIAPaCa-2 and A6L cells, stably infected with an mCherry-H2B expressing lentivirus, were treated *in vitro* with Pladienolide-B or vehicle prior to microinjection into circulation [216]. While the inhibitory actions of Pladienolide-B pre-treatment were not observed at 1-dpi, they became evident thereafter. Embryos injected with pre-treated MIAPaCa-2-mCherry-H2B cells showed a marked reduction in cell dissemination at 4-dpi, while those injected with pre-treated A6L-mCherry-H2B PDX-derived cells showed a drastic reduction in tumor cell dissemination and growth at 4-dpi, which was further enhanced at-6 dpi (**Fig R35A, B**; representative images, **R35C**).

Finally, tumor xenografts were generated in nude mice (subcutaneous injections in both flanks, paired tumors) with MIAPaCa-2 cells. After tumors reached approximately 100mm^3 (34-dpi), diluent control (G1) or Pladienolide-B (G2) were injected intratumorally and tumor growth was evaluated over the next 12 days. In line with our *in vitro* and zebrafish results, a single Pladienolide-B dose was sufficient to significantly reduce tumor growth (**Fig. R35A**). Likewise, appreciable differences in tumor weight and size were observed when tumors were resected (**Fig. R35D-F**). Examination of the presence of necrosis in the tissue of xenograft tumors did not reveal any appreciable difference between diluent- and Pladienolide-B-treated tumors. In contrast, Ki67 analysis showed a downward but non-significant trend in Pladienolide-B treated xenografts (**Fig. R36A-C**). Moreover, in line with previous results in cell lines, particularly MIAPaCa-2 cells, we observed similar trends in the alternative splicing of the genes examined (**Fig. R36D-F**).

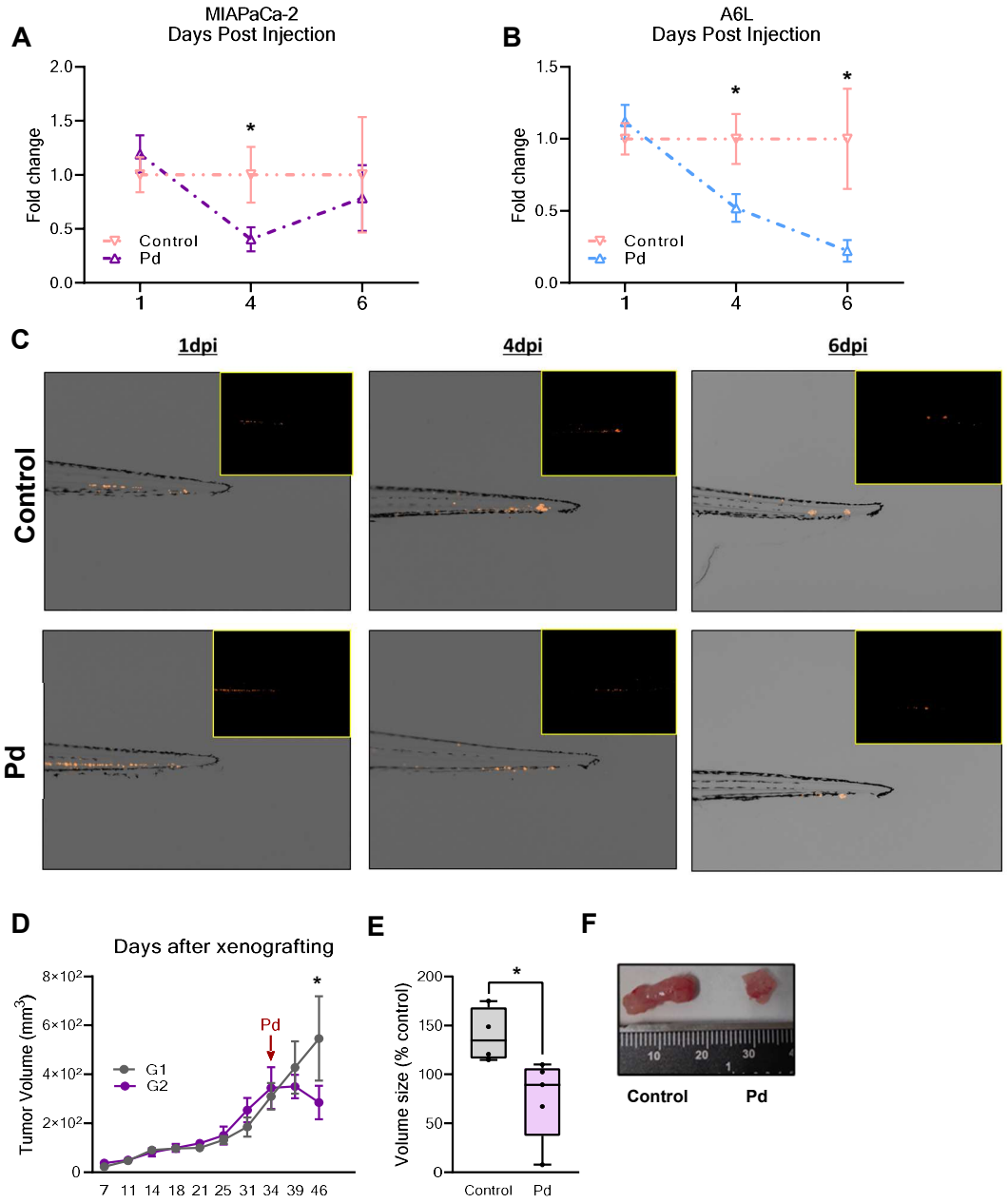


Figure R35. Pladienolide-B reduces malignancy features of PDAC cells and CSCs *in vivo*. **A, B**) Fold-change ± SEM of MIAPaCa-2 and A6L h2b-mCherry cells in zebrafish embryos, calculated by measuring the area and the fluorescence intensity. Cells were injected after 24h of Pladienolide-B (Pd) treatment. Changes in Pd-treated cells were compared to control at indicated days post injection (dpi). **C**) Representative images of Control- and Pd-treated A6L-zebrafish xenografts at indicated dpi. **D**) Tumor volume (mm^3) of MIAPaCa-2-xenografts at indicated days after xenografting. Red arrow indicates Pladienolide-B injection. G1, control-treated mice, G2, Pd-treated mice. **E**) MIAPaCa-2 xenograft tumor volumes, expressed as % ratio, extracted from Control- (n=4) and Pd-treated (n=5) mice at time of euthanasia (d49 after xenografting). **F**) Pictures of paired Control- and Pd-treated tumors. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

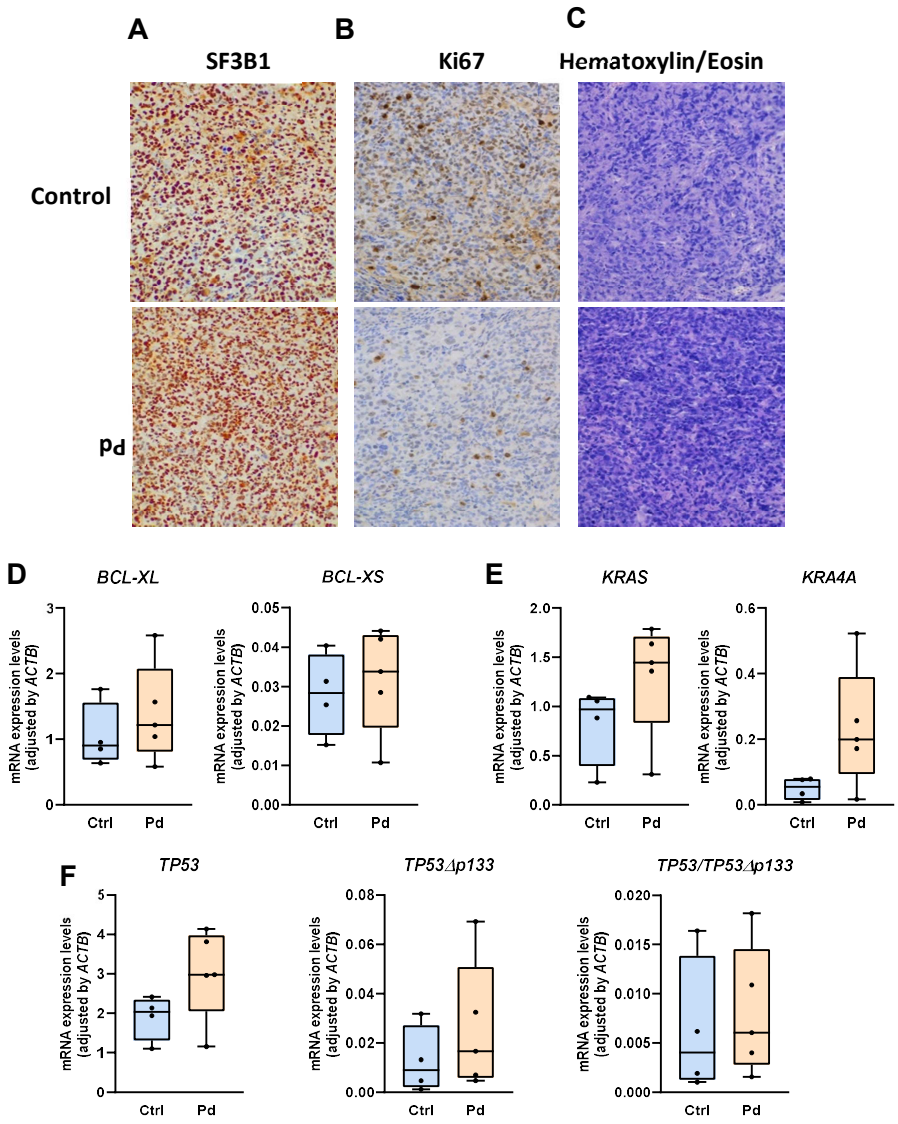


Figure R36. Pladienolide-B effect in mice tumor xenografts. A) SF3B1 IHC analysis in vehicle (Control) and Pladienolide (Pd)-treated mice. Representative IHC 20X-images showing the general SF3B1 nuclear immunostaining in Control- and Pd-treated tumors. **B**) Ki-67 staining in representative tissue sections from tumor xenografts treated with vehicle (Control) and Pd. **C**) Haematoxylin-eosin-stained sections from vehicle (Control) and Pd-treated tumors illustrating the similarly scarce necrosis foci found in these tissues. **D**) mRNA levels of *BCL-XL* and *BCL-XS* in HPDE E6E7 and MIAPaCa-2 cells treated 24h with or without (vehicle, control) Pladienolide-B (n=4). **E**) mRNA expressions of *KRAS* and *KRAS4a* in HPDE E6E7 and MIAPaCa-2 cells treated 24h with or without (vehicle, control) Pladienolide-B (n=4). **F)** Ratio of *Δ133TP53/TP53* mRNA levels in HPDE E6E7 and MIAPaCa-2 cells treated 24h with or without (vehicle, control) Pladienolide-B (n=4). Values represent the log10 of expression compared with vehicle-treated (n=4). Gene expression was normalized to *ACTB* expression. Asterisks indicate significant differences (*p < 0.05; **p < 0.01; ***p < 0.001).

Thus, the use of two different animal models indicates that Pladienolide-B treatment can reduce early metastasis and tumor cell proliferation of PDAC cells as well as retract tumor growth of PDAC xenografts, highlighting the potential of Pladienolide-B to treat PDAC.

Discussion

Cancer represents one of the main problems for human health worldwide, being the second highest cause of mortality in developed countries [1]. Despite the scientific community efforts and resources invested in studying tumor pathologies and cancer to improve our understanding of their molecular nature and to discover novel clinical approaches to combat them, their remarkable heterogeneity and complexity, resulting from both genetic and environmental factors, still hinder the finding of global solutions, and pose a big challenge to the society. Tumor heterogeneity pervades all levels of tumor biology, from molecular to clinical, and impacts all cancers, being also present among patients with the same type of cancer and even within a single tumor of a given patient [217, 218]. To address this challenge, the scientific community adopted the conceptual framework proposed by Hanahan and Weinberg [2-4], based on a group of common hallmarks that characterize and describe all types of cancers. Alteration in the normal gene expression holds a relevant position within these hallmarks, associated to both, genome instability and mutation, and nonmutational epigenetic reprogramming. In this scenario, there is ample evidence that altered patterns of gene expression can be driven by the incorrect functioning of the machinery that regulates the splicing process, which comprises a transversal cancer hallmark, as it is not only shared by all tumoral pathologies where this issue has been investigated, but also impinges upon each and every one of the proposed hallmarks of cancer. Actually, changes in this machinery lead to the aberrant expression of mRNAs and proteins that, beyond cancer, can contribute to the development or progression of diverse pathologies. Specifically, current studies in cancer have revealed that spliceosomic dysregulation may play important roles in tumor progression due to mutations and/or altered expression levels in core components of the splicing machinery [67].

Despite the growing list of alterations in genes and regulatory mechanisms discovered to date and the remarkable advances achieved, these are still insufficient to provide an effective therapeutic strategy to tackle cancer. In this context, the present Thesis is focused on the study of the role of spliceosomic alterations in pancreatic neuroendocrine tumors (PanNETs), pancreatic ductal adenocarcinoma (PDAC) and prostate cancer (PCa). The first, PanNETs, are still poorly known, in part due to their underestimated frequency, but represent a major challenge given their great heterogeneity and complexity. The second, PDAC, is one of the most lethal of all cancers worldwide, having almost the same number of cases than deaths. The third, PCa, is the second most common cancer among men worldwide, after lung cancer, being their fifth leading cause of death by cancer [1].

Although some valuable advances have been attained to establish the contribution of splicing alterations in these cancers, there is still much to be learnt in this field. For all these reasons, the general hypothesis of this Thesis derives from the emerging notion that, during the development and progression of cancer, there are splicing-related molecular dysregulations that substantially influence tumor behavior. In this context, we propose that a careful screening and characterization focused in spliceosomic processes and elements, particularly in the spliceosome molecular machinery, can provide discoveries of key importance to understand the normal physiological regulation of the cells, to elucidate alterations that contribute to tumor development and aggression, and to identify specific components and mechanisms that can serve as novel tools to devise potential diagnostic and prognostic biomarkers and actionable therapeutic targets.

Chapter I

Splicing dysregulation is increasingly regarded as a novel cancer hallmark influencing all key tumor features [108]. In PanNETs, our earlier work unveiled the overexpression of aberrant splicing variants that impart oncogenic properties [77, 78], similar to that found in numerous cancers [79, 80, 86, 219-222]. More recently, we discovered that the splicing machinery is altered in PanNETs, which involves multiple splicing factors and may underlie tumorigenesis (DOI:10.1101/2022.02.09.479525). These findings are in agreement with the idea derived from biocomputational analysis of large data sets, that alteration of the splicing machinery can result in dysregulated splicing in sets of functionally related genes, which may lead to an imbalance in relevant processes in tumors [223, 224] .However, the status and dysregulation of the splicing machinery largely varies for each type of tumor, and therefore, the detailed role and putative oncogenic contribution of individual altered components have to be assessed in their appropriate context. In this study we describe that the splicing factor *CELF4* is altered in PanNETs, where its dysregulation may enhance tumor aggressiveness by acting through the mTOR pathway, which may, in turn, influence PanNET cell response to everolimus.

CELF4 (CUGBP ELAV-like family member 4) is one of the 6 members of the CELF family of RNA binding proteins associated with regulation of pre-RNA alternative splicing [225]. Earlier studies on *CELF4* expression were conflicting, suggesting either a broad tissue expression or more restricted to nervous tissue [226], while reports on gene mutations and variants in humans and experimental studies on rodents associated this gene to neurological, neurodevelopmental and behavioral defects [227-230]. To date, only a limited number of studies have linked *CELF4* to cancer, particularly in colorectal cancer, where a prognostic role has been proposed through bioinformatic analysis of open databases and

the study of an intronic variant [231, 232] and reviewed in [226]; and in endometrial cancer, where *CELF4* expression seems to be downregulated due to hypermethylation and may also provide prognostic information [233]. Very recently, it has been reported that *CELF4* can be linked to oncogenic splicing alterations in high-grade diffuse glioma, not necessarily through mutational but via transcriptional or epigenetic regulation [234]. Our initial discovery derived from the observation that this splicing factor is overexpressed in tumor tissue compared with non-tumoral adjacent tissue in paired samples, which was well in line with our previous PanNETs us study where the vast majority of splicing machinery components studied were upregulated in PanNET tissue (DOI:10.1101/2022.02.09.479525). The use of surrounding non-tumoral tissue as a reference poses limitations but is commonly accepted as a means to biomarker discovery in NETs, where the access to fully normal tissue of origin is very difficult if not practically impossible. Nevertheless, the relevance of the discovery of this altered marker is reinforced by its quantitative inverse association with the rate of metastasis and abdominal pain, which jointly support the notion that high levels of *CELF4* expression could be explored as a potential tumor biomarker in patients with less metastasis and pain, and hence where tumor initiation might be more difficult to identify. This finding of CELF4 is completely original, as only a marginal finding of unexplored significance has identified a methylation-related expression of this gene in PanNETs [229].

To further understand the possible role of CELF4 in PanNETs and its relation to splicing regulation, we performed a biocomputational analysis of a publicly accessible RNA-Seq (GSE118014). This showed that high or low *CELF4* expression levels are distinctly associated with the expression of a discrete percentage of genes (1.15%) which include a high representation of relevant cancer-related genes. In particular, we observed an inverse correlation with two tumor suppressors, one that is widely known to hold strong links with NETs, *TP53* [235], and a related one, *CDKN2B,* that has also been linked to

PanNETs and advanced neuroendocrine neoplasms [236, 237]. In contrast, *CELF4* expression levels were directly correlated with *TSC1* and *BAD*, two pivotal intermediaries of PI3K/Akt and EGFR/MAPK pathways [7, 12], whose expression appeared to be altered/enriched in relation to CELF expression in the GSE/DAVID analysis. Moreover, when the analysis of *CELF4* was focused on splicing, we observed that high/low *CELF4* expression was associated with a distinct pattern of splicing events, mostly due to a higher usage of exon skipping and alternative last exon, and lower usage of alternative 5' splice site, differences which have been shown to be linked to alterations in the resulting transcript profile and proteome diversity and function [77, 78, 187, 219, 222]. In fact, the most altered splicing events corresponded to genes associated to signaling cascades like MAPK, and regulation of transcription and splicing itself. Thus, taken as a whole, these results demonstrate that altered *CELF4* expression is accompanied by substantial changes in the expression and splicing profiles of functionally and pathologically relevant genes in PanNETs, thereby inviting to explore in more detail their potential relationships.

Accordingly, we next studied the functional consequences of *CELF4* expression modulation using two PanNET model cell lines. This revealed that high *CELF4* expression levels directly increased proliferation of BON-1 and QGP-1 cells, whereas its silencing exerted the opposite effect, decreasing cell proliferation. These results compare favourably with our recent findings in PanNETs studying a related splicing factor, *NOVA1* (DOI:10.1101/2022.02.09.479525), but also in the most aggressive pancreatic adenocarcinoma cell models, where manipulation of *SF3B1* caused these same effects [190]. These parallel observations are in line with recent findings from our group and other labs [160, 191, 238] , and collectively argue in favor of the idea that not only mutations, but transcriptional (and epigenetic) alterations of specific components of the

splicing machinery can entail functionally relevant consequences for key cell functions. Actually, *in vivo* data with xenograft mice provides proof-of-concept that *CELF4* silencing in PanNET BON-1 cells can counteract cell proliferation and blunt tumor growth, paving the way to further explore the therapeutic potential of CELF4 in these rare tumors.

From a mechanistic perspective, the suggestive biocomputational evidence pointing towards a CELF4-dependent alteration in PI3K/Akt/mTOR pathway was confirmed by functional *in vitro* assays, where modulation of CELF4 expression in PanNET cells influenced their response to everolimus, a paradigmatic mTOR pathway inhibitor and first-line drug for the treatment of these tumors [239]. Thus, whereas *CELF4* silencing enhanced the antiproliferative effect of everolimus, its overexpression did not interfere with the inhibitory capacity of the drug. Furthermore, detailed inspection of this signaling cascade with a dedicated phosphoarray illuminated the discrete set of specific components that are particularly influenced by CELF4 expression in each cell line. Interestingly, those precise targets mostly differed between BON-1 and QGP-1 cells, which is not surprising, given the known fundamental differences of these cell models at multiple levels, from genetic to phenotypic and also functional [151, 240], which nevertheless also reflect the remarkable multilayered heterogeneity of PanNETs [5, 241-243]. Notwithstanding, these results reinforce the idea that the presence of high CELF4 levels in PanNETs can influence, likely through mechanisms involving splicing and gene expression regulation, the mTOR pathway, a master signal that impacts cell survival, proliferation, growth, and metabolism, and can also affect angiogenesis and metastasis (49, 50). Indeed, the comprehensive molecular landscape of PanNETs revealed a notable proportion of genes related with this pathway are mutated or altered in these tumors, including from *TSC1* and *TSC2* to *PTEN*, *TS2*, and *PIK3CA* [7, 12, 19, 20, 244]. Future studies should be aimed to further explore the role of CELF4 and the detailed

contribution of individual mTOR pathway components altered after its silencing (e.g., CDK5, ERN1, MAP3K5), which will help to elucidate the oncogenic role of these novel molecules in the PanNET field, as well as to define their potential as actionable therapeutic targets.

Chapter II

Section I: dysregulated splicing factor SRSF2 plays a similar oncogenic role in prostate and pancreatic cancer

The present study provides primary evidence for a previously unknown parallelism between PCa and PDAC, in that both cancers overexpress SRSF2, a critical SR protein for splicing regulation that is involved in multiple functions and impacts on different tumoral features [192]. As a splicing factor, SRSF2 plays key roles to ensure the correct assembly of the spliceosome [245], which may explain why its mutations or altered expression levels can contribute, respectively, to diverse haematological malignancies [197, 198, 246, 247] and solid cancers [199, 201]. While a marked dysregulation of the splicing machinery has recently been described in both PCa and PDAC, to the best of our knowledge, the status and role of SRSF2 has not been explored hitherto in these cancers. The interest in pursuing this goal was reinforced by the finding that their elevated levels in tumor tissue were associated, both in PCa and PDAC, with key clinical characteristic linked to poor prognosis, suggesting a potential oncogenic role.

Evidence in support of an increased *SRSF2* expression in PCa and PDAC tumor tissues is solid, as it derives from both, our in house FFPE sample cohorts and external validation databases. Although altered expression of splicing factors is emerging as a relevant oncogenic player in different cancers [186, 190, 222, 248], the specific direction of the alteration may not be the same in all cases. For example, whereas increased levels of SRSF2 play an oncogenic role in hepatocarcinoma [199, 201], like in PCa and PDAC in the present study, the opposite, a decreased expression, seems to play an antiapoptotic role in renal cancer [199, 201]. The underlying reasons for this apparent dissimilarity are presently unknown, but anyhow the observations point to SRSF2 —and, more globally, splicing factor alterations— as a relevant player in various solid cancers.

The increased levels of SRSF2 may also bear a prognostic potential. Indeed, *SRSF2* expression levels were higher in patients with clinically significant PCa (Gleason score ≥ 7) compared with non-significant PCa, while in the PDAC samples, *SRSF2* expression increased with tumor stage and the size and extent of the main tumor. Likewise, it is interesting to note that a higher expression of *SRSF2* was associated with lymphovascular invasion in both pathologies, which is a clinically relevant feature given its association with metastasis and the fact that has been previously described as a key prognostic factor in PCa [249] and PDAC [250]. In line with these observations, our *in silico* analyses corroborated that *SRSF2* expression was also directly associated with critical molecular markers of aggressiveness in both pathologies. For instance, in PDAC, *SRSF2* expression was directly associated with *TP53*, a universal tumor suppressor mutated in 50–60 % of cancers [251], and whose overexpression has been associated with poorer survival in PCa [252] and PDAC [253]. As well, in PDAC tumor tissue *SRSF2* expression was inversely correlate with *SMAD4*, a tumor suppressor gene whose loss of expression is a negative prognostic indicator and is associated with worse survival in PDAC [254]. Interestingly, *SRSF2* expression was also directly correlated with *SF3B1* expression in PCa and PDAC, a finding that may entail an arising molecular relevance since SF3B1 overexpression has been shown to be causatively associated to increased tumor aggressiveness, and linked to poorer survival in various cancers, as reported by our group in PCa [95], pancreatic cancer [190], hepatocarcinoma [160] and glioblastoma [238], where SF3B1 emerges as a new potential prognostic biomarker and therapeutic target. In this same line, it is worth noting that *SRSF2* overexpression was also coupled to increased expression of other members of its family, SRSF6 in PCa, and *SRSF1* in both cancers. This latter finding compares favourably with the evidence for a role of this factor in both PCa and PDAC. Specifically, in PCa, pioneering work indicated that SRPK1-phosphorylated SRSF1 would mediate the oncogenic selection of splice variants

of VEGF [255]. In PDAC, SRSF1 expression seems to be induced in response to the first line chemotherapeutic drug gemcitabine, thereby altering splicing and function in the MNK/eIF4E pathway and ultimately causing chemoresistance [256]. When viewed together, these results portray increased *SRSF2* expression as a clinically-linked alteration located at the interface of both well-established (e.g. TP53, SMAD4) and emerging, splicing-related (e.g. SF3B1, SRSF1) molecular markers and mechanism involved in tumor development and/or progression in both PCa and PDAC tumors. As a result, it seems reasonable to propose that the implicit prognostic potential of these discoveries deserves further, specific assessment in larger cohorts and, ideally, involving randomized prospective studies.

The second notion derived from the findings on SRSF2 overexpression, and its clinical-molecular liaisons relates to the causative (i.e., oncogenic) potential of this alteration in PCa and PDAC. To explore this issue, we specifically silenced *SRSF2* in representative cell line models for these cancers [155, 257], which, not unexpectedly, overexpressed this splicing factor under basal conditions. Intriguingly, *SRSF2* silencing using a specific siRNA did not cause identical effects in the selected cell lines and the two types of adenocarcinomas. Thus, whereas *SRSF2* silencing markedly reduced proliferation and migration rate in PCa cell lines, it did not significantly alter proliferation rate of PDAC cells, and decreased migration rate significantly only in BxPC-3. These adenocarcinoma type- and cell line-dependent differences, particularly the poor response of MiaPaCa2 PDAC cells, might obviously be due to the distinct intrinsic nature of PCa and PDAC, and the specific phenotypic differences between the two PDAC cell lines used (i.e., mutation profile, as well as invasiveness and aggressiveness, etc. [155]). This notwithstanding, it is worth noting that the four cells lines, and thus both cancer models, responded quite similarly when we measured the effect of *SRSF2* silencing on colony

formation, which depicts the ability of cells to produce progeny "infinitely". This suggests, first, that the cell lines employed are not unresponsive to *SRSF2* silencing but display a differential responsiveness to changes in SRSF2 status, which would involve distinct functional mechanisms. But, most importantly, our results point to the attractive possibility that SRSF2 may serve a relevant role in tumor initiation capacity, which depends on cancer stem cells (CSCs) within the cell population [258], a cell type that has already been shown to interact with splicing factor dysregulation in PDAC by our group [15]. Hence, future efforts should be aimed at ascertaining this possibility.

In conclusion, we discovered that two distinct adenocarcinomas, PCa and PDAC, share a previously unrecognized overexpression of the core splicing factor SRSF2, which is similarly linked to clinical-molecular features suggestive of prognostic potential and oncogenic capacity. In vitro silencing of *SRSF2* in model cell lines differentially influenced functional tumor features indicative of tumor aggressiveness, which warrants further studies to assess its putative capacity as therapeutic actionable target in these cancers. Altogether, these and our previous studies underscore the pertinence of examining alternative splicing, its players, and abnormalities as powerful tools to identify novel biomarkers and therapeutic targets in PCa and PDAC.

Chapter II

Section II: Dysregulation of the splicing machinery as a target for pancreatic ductal adenocarcinoma

There is increasing evidence that in PDAC, like in many other cancers, alternative splicing is severely dysregulated, causing changes that can contribute to its development and progression [96, 101, 112, 190]. Such dysregulations may often derive from alterations in the machinery that controls the splicing process, comprised by the spliceosome core and ancillary splicing factors, which lead to the aberrant expression of RNAs and/or proteins that, in turn, impart oncological features to the transforming cell. Specifically, both biocomputational and experimental studies strongly support that spliceosome-related defects due to altered expression and/or mutations in splicing machinery components may play an important role in PDAC progression [101, 190, 206]. In the present study, analysis of a comprehensive landscape of splicing machinery elements revealed its broad dysregulation and led to exploring the specific role of two splicing factors, whose particular alteration may play a role in PDAC aggressiveness.

In PDAC —unlike other cancers— assessing the molecular differences between tumor tissue and non-tumoral adjacent tissue can provide meaningful, precise information of changes taking place in pancreatic cells during cancer development. In this regard, our results confirm and extend previous data, by demonstrating a profound dysregulation in the expression profile of numerous splicing machinery components in PDAC, which involved more than one-third of the spliceosome components and splicing factors examined, that differed in tumor samples vs. adjacent tissue. The observed changes involved factors from different molecular families, linked to distinct functions on the splicing process (e.g. RNUs, SRSFs, PRPFs, RMBs, etc.), suggesting that the alterations in the splicing machinery in PDAC are not restricted to a limited, particular set

of factors and processing steps, but may widely influence alternative splicing as a whole. These findings provide further support to the rising notion that the splicing machinery is profoundly altered in many diseases, particularly in cancer, although the overarching mechanisms driving this alteration and its overall causal significance remain to be elucidated. To clarify the precise implications of these alterations in PDAC, integrative biocomputational studies on available databases can provide valuable information [110, 223, 259], however, specific experimental studies are mandatory to assess the particular potential contribution of specific dysregulated molecular components.

Our approach examining multiple, representative splicing machinery-related factors enabled to identify concurrent changes in expression levels of several altered molecules, which may hold diagnostic and/or prognostic significance. This idea is based in the combined ability of these factors to discriminate between tumoral and non-tumoral PDAC tissue (as indicated by the significant ROC curves) and their association with critical clinical features, including patient survival. Thus, by systematically exploring components of the splicing machinery in in-house samples and external validation cohorts, our findings unveil a relevant set of factors with potential utility to enhance the arsenal of molecular biomarkers and targets to tackle PDAC. To bring this concept forwards, we selected two factors, PRPF8 and RBMX.

PRPF8 (Pre-MRNA Processing Factor 8, also known as Prp8) is the largest and evolutionarily most conserved protein component of the spliceosome, where it is a component of the snRNP U5 complex [260, 261]. Here, its expression in PDAC tissue was lower than in the adjacent non-tumor tissue, suggesting both, a potential value as a biomarker and a possible pathological role in this cancer. Mutations in *PRPF8* have been implicated in the development of Retinitis Pigmentosa [262], but the role of this factor in cancer is less well understood, with only some studies reporting its ability to reduce cell

growth in colorectal cancer [263] and to modify androgen receptor levels in PCa [188]. Notably, our data shows that *PRPF8* levels sharply decreased in high grade tumors, compared with higher levels in less advanced disease stages. Moreover, reduced *PRPF8* expression levels in PDAC were remarkably associated with decreased survival in all the explored forms: disease specific, progression free and overall survival, which are key unmet problems of PDAC.

RBMX (RNA-binding motif protein, X-linked, also known as HNRPNG) is an essential splicing factor that participates in exon addition or exclusion in the mRNA for many proteins [109]. This factor plays multiple roles in key biological processes, from nervous system development, to transcription control, chromosome biology [264], cell division [265] and DNA stability [266]. In cancer, RBMX also exerts relevant actions, which seem to vary diametrically depending on the type of tumor, behaving as either a tumor promoter or a tumor suppressor. Thus, while its overexpression has been related with hepatocellular carcinoma [267] or T-Cell Lymphomas [268], downregulation is observed in bladder [269], endometrium [270] or neck cancer [271]. In line with the latter, we found that *RBMX* expression is lower in PDAC tumor tissue compared with non-tumor tissue, a reduction that is associated to lower survival rates (progression free, disease specific and overall survival probability) of the patients.

These observations suggest a splicing-related role for PRPF8 and RBMX in PDAC, as it is conceivable that a reduction in the expression of core spliceosomal components and splicing factors may alter spliceosomal catalytic activity and thereby cause functional consequences. However, unequivocal support for this notion requires experimental demonstration. Accordingly, we explored the functional and mechanistic consequences of manipulating the expression of *PRPF8* and *RBMX* by means of appropriate of PDAC cell line models. Indeed, overexpression of *PRPF8* and *RBMX* in two different PDAC cell

lines, mimicking their respective levels in non-tumor reference tissues, showed similar, predicted results. Specifically, overexpression of *PRPF8* and *RBMX* could rescue/normalize cancer cell line profiles by reducing cell proliferation and migration, similar to what have been observed previously in comparable experimental settings (DOI:10.1101/2022.02.09.479525, [91, 160, 190]). Remarkably, the effect of both splicing factors was observed not only in this typical tumor features but also inhibited sphere and colony formation, indicating that their role may extend to the control of self-renewal and stem properties [213]. These findings underscore the powerful functional consequences of alterations in splicing machinery components and argue that its exogenous manipulation could provide means for therapeutic intervention, as we and others have recently proposed in PDAC and other cancers [66, 91, 160, 186, 187].

The relevant role of PRPF8 and RBMX as components of the splicing machinery prompted us to examine the possible implications of their altered expression in RNA splicing in PDAC, by comparing global splicing patterns in tumors with low and high *PRPF8* and *RBMX* expression. Interestingly, this approach revealed clear differences that, in the case of PRPF8, and given its core role, were of an unexpected, limited extent, whereas for RBMX were associated to a high number of significant splicing events. These differences are reflected in the distinct splicing patterns observed, which mainly affect exon skipping and alternative first and last exon, as compared to the average of all the events calculated. These results suggests that, despite its central implication in common gene processing, as an spliceosome U5 component essential for splicing in all tissues to process intron-containing transcripts, PRPF8 may exert its tumor suppressor actions in PDAC by modulating a limited number of gene splicing events, which certainly deserve a close inspection in the future [272]. Conversely, the effects of altered RBMX may involve changes in the splicing of multiple genes of different families and would therefore

implicate different players and mechanisms. In this context, and although the functional pathways conveying the putative tumor suppressor actions of *PRPF8* and *RBMX* are still to be elucidated, it is worth noting that PDAC samples with lower levels of these splicing machinery components display mutational signatures linked to poorer PDAC prognosis, including mutations of key driver genes as *KRAS* and *TP53* [149, 273, 274]. Likewise, transcriptional analyses linked *PRPF8* and *RBMX* expression with that of key PDAC genes, by showing a direct correlation between both genes with *TP53* and *SMAD4*, two key tumoral suppressor genes in PDAC, and an inverse correlation of *PRPF8* expression with *KRAS* and *CDKN2A* mRNA levels [101, 149]. These observations are in line with and provide further support to the recent notion that dysregulations of components of the splicing machinery may exert their actions in connection to altered functioning of well recognized key gene players in PDAC like KRAS and P53 [101, 268].

Taken together, our results demonstrate that the splicing machinery is severely dysregulated in PDAC, where we identified two specific components, PRPF8 and RBMX, that display a downregulated expression closely linked to poorer survival and clinical and molecular markers of bad prognosis. Furthermore, we found that expression of PRPF8 and RBMX is distinctly associated to altered splicing profiles in PDAC, and restoring their expression levels rescued their tumor suppressor ability in vitro in two representative PDAC cell models. We conclude that the splicing machinery is profoundly altered in PDAC, which provides a novel pathway to identify new potential biomarkers and actionable therapeutic targets for this dismal cancer.

Chapter III

In the present study, we not only show that the pivotal splicing machinery component SF3B1 is overexpressed in PDAC, but it can be targeted by Pladienolide-B, which causes antioncogenic effects in both cancer cells and CSCs, paving the way to develop new treatment strategies for this deadly cancer.

SF3B1 dysfunction, through mutation [52, 210, 211] or altered expression [95, 160], is known to increase oncogenic features in various cancers including PDAC [107, 212]. We now show that *SF3B1* is overexpressed in PDAC as compared to its surrounding tissue (our samples) or healthy pancreatic tissue (E-MTAB-1791-cohort) [147]. Importantly, IHC analysis confirmed its overexpression in tumor cell nuclei. Moreover, *SF3B1* levels were associated to relevant clinical parameters, suggesting a potential pathological relevance linked to its dysregulation. These results are in line with our recent studies in prostate cancer [95] and hepatocarcinoma [160], and other studies in chronic lymphocytic leukaemia or endometrial and breast cancers [275-277], collectively reinforcing the growing view that this splicing factor is heavily altered in cancer. Accordingly, we asked whether *SF3B1* expression could be associated with or even contribute to PDAC pathophysiology.

To answer this question, we first biocomputationally explored the link between *SF3B1* expression and its primary regulatory endpoint, alternative splicing, and found that PDAC with high or low *SF3B1* expression displayed strikingly distinct global splicing patterns. Interestingly, high *SF3B1* levels correlated with higher usage of alternative 3' splice sites, resembling common alterations in *SF3B1*-mutated cancers [211], and with elevated exon skipping, which has been linked to PDAC and to *SF3B1*-mutation in myelodysplastic syndromes [278] and *C. elegans* models [279]. Conversely, low *SF3B1* expression was associated with elevated frequency of splicing events not particularly linked to mutant *SF3B1* malignancies [52, 210, 211]. These findings suggest that, as a

key splicing player, *SF3B1* may influence the global splicing pattern in PDAC, thereby potentially having pathological implications [66, 211, 280]. Reactome analysis revealed that the spliced genes associated to *SF3B1* expression were tightly coupled to both PDAC and its central AKT-signaling pathway [281, 282]. Further analysis revealed that *SF3B1* expression levels were linked to key PDAC molecular features, including direct correlations with *KRAS*, *BRCA1*, *BRCA2* and *HNRNPK* expression and inverse correlations with *CDKN2A* and *TP53* expression. This multifaceted association of *SF3B1* expression with global splicing and expression levels of key PDAC genes converges with recent data linking PDAC malignancy to splicing dysregulation [101], strongly suggesting that *SF3B1* overexpression in PDAC, like in prostate cancer [95] and hepatocarcinoma [160], may have pathological consequences.

To interrogate *SF3B1* function in PDAC, we first silenced its expression in normal pancreatic (HPDE E6E7) and PDAC cell lines (Capan-2, BxPC-3, MIAPaCa-2), where a time-dependent decrease in cell proliferation, particularly in HPDE E6E7 cells, was observed. These results agree with findings in mice showing that *Sf3b1* homozygote deletion is embryonic lethal [283], and in cancer cell lines, where *SF3B1* copy number loss represents a vulnerability, suggesting its essential role [284]. Notably, in keeping with this latter study [284] and our previous work [95], pharmacological inhibition of SF3B1 function with Pladienolide-B markedly decreased PDAC cell proliferation without affecting HPDE E6E7 cells, unveiling a difference in cell function when targeting SF3B1 with inhibitors *vs.* modulating its expression genetically [284]. Interestingly, Pladienolide-B's antiproliferative action was comparable to that of Gemcitabine, although their combination did not potentiate each other (at least in established cell lines), suggesting shared mechanism(s) of action(s). Notably, Pladienolide-B not only affected proliferation but also inhibited cell migration and sphere and colony formation, while enhancing cell apoptosis. These findings underscore the promising anticancer capacity of Pladienolide-

B in PDAC cells and expands upon the cancers wherein pharmacologically targeting SF3B1 exerts anticancer actions.

The mechanisms underpinning Pladienolide-B actions in PDAC likely involves alteration of key signals, as suggested by the concomitant reduction of pAKT and increase in pJNK, two critical kinases that regulate vital cellular processes in PDAC and other cancers [281, 282, 285-287]. In PDAC, AKT overexpression is a common feature and is closely linked to cell plasticity [281, 282], which also seems to be linked to JNK, that could act as a tumor suppressor [285-287]. Pladienolide-B inhibition of SF3B1 function not only modulated signaling cascades in tumor cells but also altered splicing of molecules crucial in PDAC, such as *BCL-X*, *KRAS* and *TP53*, favouring the balance of the more pro-apoptotic and/or antioncogenic variants. Specifically, Pladienolide-B increased the *BCL2L1* isoform *BCL-XS*, which binds to and inhibits its antiapoptotic variant *BCL-XL* and *BCL2* itself, thereby promoting the release of proapoptotic BAK [288]. Likewise, Pladienolide-B treatment increased the proapoptotic variant *KRAS4a* [289] without altering full-length *KRAS4*, and reduced the proportion *Δ133TP53/TP53*, likely fostering apoptosis, inasmuch as *Δ133TP53* inhibits p53 [290]. These results provide experimental support that SF3B1 directly impacts relevant splicing phenomena in PDAC, which was prompted by the aforementioned association of *SF3B1* expression levels with distinct splicing event profiles. Hence, Pladienolide-B would act on PDAC cancer cells by altering both, key signaling pathways and splicing mechanisms.

Having established the antioncogenic actions of Pladienolide-B in PDAC cells, we next interrogated its potential effects on CSCs, a unique cell subset increasingly recognized as a relevant player in PDAC maintenance, chemoresistance, disease relapse and metastasis [213]. Although recent evidence suggests a splicing machinery dysregulation in PDAC CSCs [214], *SF3B1*'s role in these cells is still unknown. In our PDX-derived CSC-enriched models, *SF3B1* expression levels were appreciable but

lower than in cancer cell lines. Correspondingly, CSCs presented lower protein levels for related splicing machinery components (SF3B2, SRSF1, hnRNPs) compared to PDAC Panc1 cells, which may be linked to maintenance of the CSC "dedifferentiated" state [214]. Intriguingly, Pladienolide-B appeared to preferentially target cell survival and apoptosis in cancer cells over CSCs in PDX-derived cultures, potentially suggesting CSCs drug resistance. However, further examination revealed that Pladienolide-B was able to affect other crucial CSC features, altering pluripotency-related gene expression (e.g., *KLF4*, *NODAL* and *SOX2*) and decreasing sphere- and colony-formation capacity, which reflect a loss in self-renewal and stem properties. More importantly, Pladienolide-B also reduced CSC chemoresistance, as its combination with chemotherapeutic drugs (e.g., Gemcitabine or Abraxane) markedly increased toxicity.

We were surprised that while Pladienolide-B inhibited CSC functional properties (self-renewal, chemoresistance, tumorigenicity), CSC marker-positive populations increased. We hypothesize that the latter could result from cancer cell plasticity. Indeed, non-CSC hybrid/transient cells can dedifferentiate and convert into CSCs when the CSC compartment is compromised [291, 292]. Since Pladienolide-B enhanced apoptosis in all PDAC PDX-derived cultures concomitant with a decrease in CSC functional phenotypes, we can only assume that CSC-negative cells were attempting to replenish the CSC pool, resulting in increased autofluorescent- and CD133-positive cells. While confirming this hypothesis requires further studies, the fact that Pladienolide-B treatment reduces functional CSC properties is proof enough that the CSC compartment is affected by SF3B1 modulation.

As proof-of-concept that Pladienolide-B's antioncogenic effects *in vitro* are clinically translatable, we tested its actions *in vivo* in two previously validated preclinical models [160, 164]. Indeed, Pladienolide-B pretreatment of MIAPaCa-2 cells or CSCs blunted

their capacity to migrate and proliferate in a zebrafish model, supporting the anti-invasive and anti-metastatic effects of Pladienolide-B. Moreover, Pladienolide-B prevented tumor growth in mice with MIAPaCa-2 tumor xenografts, which did not present systemic or histological problems (metastasis, necrosis), in keeping with our data in prostate cancer [95, 160] and hepatocarcinoma [160], and the antitumoral actions of spliceosome-targeted drugs in PDAC mouse models [101]. Thus, these two distinct models provide suggestive evidence that by inhibiting SF3B1 with Pladienolide-B, the oncogenic properties of both PDAC cells and CSCs are reduced *in vivo*.

In summary, our findings reveal that SF3B1 is overexpressed in human PDAC, where its levels associate with key clinical (lymph node stage), histological (grade), and molecular (e.g., splicing alterations) features. Furthermore, targeting SF3B1 function with Pladienolide-B reduces multiple cancer features in PDAC cells (proliferation, migration, and colony and sphere formation) by altering relevant signaling pathways and splicing events. Importantly, Pladienolide-B treatment reduces CSCs stemness, making CSCs more sensitive to chemotherapy treatment. Finally, this drug's anti-tumoral and anti-CSC effects were also observed in two distinct in vivo preclinical models, xenografted zebrafish and mice. We conclude that SF3B1 overexpression represents a therapeutic vulnerability in PDAC that enables the targeting of splicing with Pladienolide-B not only in cancer cells but also in CSCs, which opens up novel therapeutic avenues for this lethal cancer.

General conclusions

For all the above mentioned, the main conclusions of this Thesis are:

1. The splicing factor *CELF4* is overexpressed in PanNETs, where its levels associate with malignancy features and distinct splicing profiles. Modulation of CELF4 levels predictably influences multiple cancer features *in vitro* in PanNET cell lines, and its silencing inhibits xenograft tumor growth. CELF4 expression impacts pathways and mediators linked to mTOR pathway, which likely explains how it impairs the response of PanNET cells to everolimus treatment.

2. The splicing machinery is severely dysregulated in PDAC, wherein we identified two specific components, PRPF8 and RBMX, that display a downregulated expression, which is closely linked to poorer survival and clinical and molecular markers of bad prognosis. PRPF8 and RBMX expression is distinctly associated to altered splicing profiles, and restoring their expression levels rescued their tumor suppressor ability *in vitro* in PDAC cell models. These factors represent two promising targets that deserve further research as new potential biomarkers and actionable molecular instruments to tackle PDAC.

3. Two distinct adenocarcinomas, PCa and PDAC, share a previously unrecognized overexpression of SRSF2, which is linked to clinical-molecular features suggestive of worse prognostic potential and oncogenic capacity. *SRSF2* silencing differentially influenced functional tumor features indicative of tumor aggressiveness in PCa (more responsive) and PDAC (less responsive). In contrast, *SRSF2* silencing similarly reduced colony formation in cell models of both cancers, suggesting a possible role of this factor in the control of tumor initiation capacity by cancer/cancer stem cells.

4. SF3B1 is overexpressed in human PDAC, wherein its levels associate with key clinical, histological, and molecular features. Furthermore, targeting SF3B1 activity

with Pladienolide-B reduces multiple cancer features in PDAC cells by altering relevant signaling pathways and splicing events. Importantly, Pladienolide-B treatment reduces CSCs stemness, making CSCs more sensitive to chemotherapy treatment.

Global corollary

Taken together, the studies developed in the present Thesis provide novel evidence to advance in the molecular knowledge of the role of the splicing machinery and its alterations in different cancers. In particular, we identify specific splicing factors that are altered in PanNETs, PDAC and PCa, and seem to play a relevant functional role in these tumors, wherein they could serve as useful tools for the development of new biomarkers and could be the target for newly developed splicing-directed drugs, like Pladienolide-B and/or its derivatives. These findings reinforce the pertinence of examining alternative splicing, its players and abnormalities to open up novel avenues for precision medicine in solid cancers.

References

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. **71**(3): p. 209-249.
2. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
3. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
4. Hanahan, D., *Hallmarks of Cancer: New Dimensions.* Cancer Discov, 2022. **12**(1): p. 31-46.
5. Pedraza-Arévalo, S., et al., *Multilayered heterogeneity as an intrinsic hallmark of neuroendocrine tumors.* Rev Endocr Metab Disord, 2018. **19**(2): p. 179-192.
6. Darba, J. and A. Marsa, *Exploring the current status of neuroendocrine tumours: a population-based analysis of epidemiology, management and use of resources.* BMC Cancer, 2019. **19**(1): p. 1226.
7. Scarpa, A., et al., *Whole-genome landscape of pancreatic neuroendocrine tumours.* Nature, 2017. **543**(7643): p. 65-71.
8. Raphael, M.J., et al., *Principles of diagnosis and management of neuroendocrine tumours.* CMAJ, 2017. **189**(10): p. E398-E404.
9. Dasari, A., et al., *Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States.* JAMA Oncol, 2017. **3**(10): p. 1335-1342.
10. Tsai, H.J., et al., *The epidemiology of neuroendocrine tumors in Taiwan: a nation-wide cancer registry-based study.* PLoS One, 2013. **8**(4): p. e62487.
11. Leyden, S., et al., *Unmet needs in the international neuroendocrine tumor (NET) community: Assessment of major gaps from the perspective of patients, patient advocates and NET health care professionals.* Int J Cancer, 2020. **146**(5): p. 1316-1323.
12. Mafficini, A. and A. Scarpa, *Genomic landscape of pancreatic neuroendocrine tumours: the International Cancer Genome Consortium.* J Endocrinol, 2018. **236**(3): p. R161-R167.
13. Mafficini, A. and A. Scarpa, *Genetics and Epigenetics of Gastroenteropancreatic Neuroendocrine Neoplasms.* Endocr Rev, 2019. **40**(2): p. 506-536.
14. Kloppel, G., *Classification and pathology of gastroenteropancreatic neuroendocrine neoplasms.* Endocr Relat Cancer, 2011. **18 Suppl 1**: p. S1-16.
15. Minnetti, M. and A. Grossman, *Somatic and germline mutations in NETs: Implications for their diagnosis and management.* Best Pract Res Clin Endocrinol Metab, 2016. **30**(1): p. 115-27.
16. Jiao, Y., et al., *DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors.* Science, 2011. **331**(6021): p. 1199-203.
17. Guertin, D.A. and D.M. Sabatini, *Defining the role of mTOR in cancer.* Cancer Cell, 2007. **12**(1): p. 9-22.
18. Hua, H., et al., *Targeting mTOR for cancer therapy.* J Hematol Oncol, 2019. **12**(1): p. 71.
19. Zanini, S., et al., *mTOR Pathway in Gastroenteropancreatic Neuroendocrine Tumor (GEP-NETs).* Front Endocrinol (Lausanne), 2020. **11**: p. 562505.
20. Chan, J. and M. Kulke, *Targeting the mTOR signaling pathway in neuroendocrine tumors.* Curr Treat Options Oncol, 2014. **15**(3): p. 365-79.
21. Pea, A., R.H. Hruban, and L.D. Wood, *Genetics of pancreatic neuroendocrine tumors: implications for the clinic.* Expert Rev Gastroenterol Hepatol, 2015. **9**(11): p. 1407-19.
22. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
23. Petersen, G.M., *Familial pancreatic cancer.* Semin Oncol, 2016. **43**(5): p. 548-553.

24. Benzel, J. and V. Fendrich, *Familial Pancreatic Cancer.* Oncol Res Treat, 2018. **41**(10): p. 611-618.
25. Murphy, S.J., et al., *Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor.* Gastroenterology, 2013. **145**(5): p. 1098-1109 e1.
26. Lohr, M., et al., *Frequency of K-ras mutations in pancreatic intraductal neoplasias associated with pancreatic ductal adenocarcinoma and chronic pancreatitis: a meta-analysis.* Neoplasia, 2005. **7**(1): p. 17-23.
27. Biankin, A.V., et al., *Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes.* Nature, 2012. **491**(7424): p. 399-405.
28. Witkiewicz, A.K., et al., *Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets.* Nature Communications, 2015. **6**.
29. Ottenhof, N.A., et al., *Molecular characteristics of pancreatic ductal adenocarcinoma.* Patholog Res Int, 2011. **2011**: p. 620601.
30. Neesse, A., et al., *Stromal biology and therapy in pancreatic cancer: a changing paradigm.* Gut, 2015. **64**(9): p. 1476-84.
31. Tao, L., et al., *Cancer associated fibroblasts: An essential role in the tumor microenvironment.* Oncol Lett, 2017. **14**(3): p. 2611-2620.
32. Cid-Arregui, A. and V. Juarez, *Perspectives in the treatment of pancreatic adenocarcinoma.* World J Gastroenterol, 2015. **21**(31): p. 9297-316.
33. Mottet, N., et al., *EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.* Eur Urol, 2017. **71**(4): p. 618-629.
34. Attard, G., et al., *Prostate cancer.* Lancet, 2016. **387**(10013): p. 70-82.
35. Epstein, J.I., et al., *The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma.* Am J Surg Pathol, 2005. **29**(9): p. 1228-42.
36. Zhang, W., et al., *Exploration on Gleason score variation trend of patients with prostate carcinoma from 1996 to 2019: a retrospective single center study.* Gland Surg, 2021. **10**(2): p. 607-617.
37. Lynch, H.T., et al., *Screening for familial and hereditary prostate cancer.* Int J Cancer, 2016. **138**(11): p. 2579-91.
38. Rebbeck, T.R., *Prostate Cancer Genetics: Variation by Race, Ethnicity, and Geography.* Semin Radiat Oncol, 2017. **27**(1): p. 3-10.
39. Heidenreich, A., et al., *EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013.* Eur Urol, 2014. **65**(1): p. 124-37.
40. Lolli, C., et al., *Systemic Immune-Inflammation Index Predicts the Clinical Outcome in Patients with mCRPC Treated with Abiraterone.* Front Pharmacol, 2016. **7**: p. 376.
41. Kornblihtt, A.R., et al., *Alternative splicing: a pivotal step between eukaryotic transcription and translation.* Nat Rev Mol Cell Biol, 2013. **14**(3): p. 153-65.
42. Baralle, F.E. and J. Giudice, *Alternative splicing as a regulator of development and tissue identity.* Nat Rev Mol Cell Biol, 2017. **18**(7): p. 437-451.
43. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome.* Nat Rev Mol Cell Biol, 2014. **15**(2): p. 108-21.
44. Gallego-Paez, L.M., et al., *Alternative splicing: the pledge, the turn, and the prestige : The key role of alternative splicing in human biological systems.* Hum Genet, 2017. **136**(9): p. 1015-1042.
45. Tyson-Capper, A. and H. Gautrey, *Regulation of Mcl-1 alternative splicing by hnRNP F, H1 and K in breast cancer cells.* RNA Biol, 2018. **15**(12): p. 1448-1457.

46. Ritchie, D.B., M.J. Schellenberg, and A.M. MacMillan, *Spliceosome structure: piece by piece.* Biochim Biophys Acta, 2009. **1789**(9-10): p. 624-33.
47. Turunen, J.J., et al., *The significant other: splicing by the minor spliceosome.* Wiley Interdiscip Rev RNA, 2013. **4**(1): p. 61-76.
48. Sharp, P.A. and C.B. Burge, *Classification of introns: U2-type or U12-type.* Cell, 1997. **91**(7): p. 875-9.
49. Anczukow, O. and A.R. Krainer, *Splicing-factor alterations in cancers.* RNA, 2016. **22**(9): p. 1285-301.
50. Jurica, M.S. and M.J. Moore, *Pre-mRNA splicing: awash in a sea of proteins.* Mol Cell, 2003. **12**(1): p. 5-14.
51. Wang, Y., et al., *Mechanism of alternative splicing and its regulation.* Biomed Rep, 2015. **3**(2): p. 152-158.
52. Dvinge, H., et al., *RNA splicing factors as oncoproteins and tumour suppressors.* Nat Rev Cancer, 2016. **16**(7): p. 413-30.
53. Long, J.C. and J.F. Caceres, *The SR protein family of splicing factors: master regulators of gene expression.* Biochem J, 2009. **417**(1): p. 15-27.
54. Griffiths, D., et al., *Melatonin reduces the production and secretion of prolactin and growth hormone from rat pituitary cells in culture.* Acta Physiol Scand, 1987. **131**(1): p. 43-9.
55. Wong, J.J., et al., *Epigenetic modifications of splicing factor genes in myelodysplastic syndromes and acute myeloid leukemia.* Cancer Sci, 2014. **105**(11): p. 1457-63.
56. Boutz, P.L., et al., *MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development.* Genes Dev, 2007. **21**(1): p. 71-84.
57. Naro, C. and C. Sette, *Phosphorylation-mediated regulation of alternative splicing in cancer.* Int J Cell Biol, 2013. **2013**: p. 151839.
58. Goncalves, V. and P. Jordan, *Posttranscriptional Regulation of Splicing Factor SRSF1 and Its Role in Cancer Cell Biology.* Biomed Res Int, 2015. **2015**: p. 287048.
59. Chen, Y., et al., *Mutually exclusive acetylation and ubiquitylation of the splicing factor SRSF5 control tumor growth.* Nat Commun, 2018. **9**(1): p. 2464.
60. Nilsen, T.W. and B.R. Graveley, *Expansion of the eukaryotic proteome by alternative splicing.* Nature, 2010. **463**(7280): p. 457-63.
61. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.
62. Oltean, S. and D.O. Bates, *Hallmarks of alternative splicing in cancer.* Oncogene, 2014. **33**(46): p. 5311-8.
63. Sveen, A., et al., *Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes.* Oncogene, 2016. **35**(19): p. 2413-27.
64. Kim, J., et al., *Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease.* N Engl J Med, 2019. **381**(17): p. 1644-1652.
65. Wang, E., et al., *Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia.* Cancer Cell, 2019. **35**(3): p. 369-384 e7.
66. Bonnal, S.C., I. Lopez-Oreja, and J. Valcarcel, *Roles and mechanisms of alternative splicing in cancer - implications for care.* Nat Rev Clin Oncol, 2020. **17**(8): p. 457-474.
67. Urbanski, L.M., N. Leclair, and O. Anczukow, *Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics.* Wiley Interdiscip Rev RNA, 2018. **9**(4): p. e1476.
68. Chaudhry, A., et al., *Different splice variants of CD44 are expressed in gastrinomas but not in other subtypes of endocrine pancreatic tumors.* Cancer Res, 1994. **54**(4): p. 981-6.

69. Jakubauskiene, E., et al., *Gastrointestinal tract tumors and cell lines possess differential splicing factor expression and tumor associated mRNA isoform formation profiles.* Cancer Biomark, 2015. **15**(5): p. 575-81.
70. Miyanaga, A., et al., *Diagnostic and prognostic significance of the alternatively spliced ACTN4 variant in high-grade neuroendocrine pulmonary tumours.* Ann Oncol, 2013. **24**(1): p. 84-90.
71. Choi, S., et al., *Function and clinical relevance of RHAMM isoforms in pancreatic tumor progression.* Mol Cancer, 2019. **18**(1): p. 92.
72. Suzuki, H., et al., *Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma.* Nature, 2019. **574**(7780): p. 707-711.
73. Inoue, D., et al., *Spliceosomal disruption of the non-canonical BAF complex in cancer.* Nature, 2019. **574**(7778): p. 432-436.
74. Shimojo, M., et al., *A gapmer antisense oligonucleotide targeting SRRM4 is a novel therapeutic medicine for lung cancer.* Sci Rep, 2019. **9**(1): p. 7618.
75. Loiselle, J.J., J.G. Roy, and L.C. Sutherland, *RBM10 promotes transformation-associated processes in small cell lung cancer and is directly regulated by RBM5.* PLoS One, 2017. **12**(6): p. e0180258.
76. Zheng, M., et al., *ESRP1 regulates alternative splicing of CARM1 to sensitize small cell lung cancer cells to chemotherapy by inhibiting TGF-beta/Smad signaling.* Aging (Albany NY), 2021. **13**(3): p. 3554-3572.
77. Sampedro-Nuñez, M., 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**(6): p. 6593-608.
78. Luque, R.M., 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**(23): p. 19619-33.
79. Durán-Prado, 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**(16): p. 2049-61.
80. Hormaechea-Agulla, D., et al., *The oncogenic role of the In1-ghrelin splicing variant in prostate cancer aggressiveness.* Mol Cancer, 2017. **16**(1): p. 146.
81. Johannessen, L.E., et al., *Upregulation of INS-IGF2 read-through expression and identification of a novel INS-IGF2 splice variant in insulinomas.* Oncol Rep, 2016. **36**(5): p. 2653-2662.
82. Sanchez, C., et al., *Characterization of a novel five-transmembrane domain cholecystokinin-2 receptor splice variant identified in human tumors.* Mol Cell Endocrinol, 2012. **349**(2): p. 170-9.
83. Capdevila, J., et al., *Translational research in neuroendocrine tumors: pitfalls and opportunities.* Oncogene, 2017. **36**(14): p. 1899-1907.
84. Karantanos, T., P.G. Corn, and T.C. Thompson, *Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches.* Oncogene, 2013. **32**(49): p. 5501-11.
85. Lu, C. and J. Luo, *Decoding the androgen receptor splice variants.* Transl Androl Urol, 2013. **2**(3): p. 178-186.
86. Hormaechea-Agulla, D., et al., *The oncogenic role of the spliced somatostatin receptor sst5TMD4 variant in prostate cancer.* FASEB J, 2017. **31**(11): p. 4682-4696.
87. Narla, G., et al., *Targeted inhibition of the KLF6 splice variant, KLF6 SV1, suppresses prostate cancer cell growth and spread.* Cancer Res, 2005. **65**(13): p. 5761-8.

88. Carstens, R.P., et al., *Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer.* Oncogene, 1997. **15**(25): p. 3059-65.
89. Woolard, J., et al., *VEGF165b, an inhibitory vascular endothelial growth factor splice variant: mechanism of action, in vivo effect on angiogenesis and endogenous protein expression.* Cancer Res, 2004. **64**(21): p. 7822-35.
90. Augello, M.A., et al., *Convergence of oncogenic and hormone receptor pathways promotes metastatic phenotypes.* J Clin Invest, 2013. **123**(1): p. 493-508.
91. Jiménez-Vacas, J.M., et al., *Dysregulation of the splicing machinery is directly associated to aggressiveness of prostate cancer.* EBioMedicine, 2020. **51**: p. 102547.
92. Zhang, X., et al., *SRRM4 Expression and the Loss of REST Activity May Promote the Emergence of the Neuroendocrine Phenotype in Castration-Resistant Prostate Cancer.* Clin Cancer Res, 2015. **21**(20): p. 4698-708.
93. Li, Y., et al., *SRRM4 Drives Neuroendocrine Transdifferentiation of Prostate Adenocarcinoma Under Androgen Receptor Pathway Inhibition.* Eur Urol, 2017. **71**(1): p. 68-78.
94. Munkley, J., et al., *Androgen-regulated transcription of ESRP2 drives alternative splicing patterns in prostate cancer.* Elife, 2019. **8**.
95. Jiménez-Vacas, J.M., et al., *Spliceosome component SF3B1 as novel prognostic biomarker and therapeutic target for prostate cancer.* Transl Res, 2019. **212**: p. 89-103.
96. Carrigan, P.E., et al., *Characterization of alternative spliceoforms and the RNA splicing machinery in pancreatic cancer.* Pancreas, 2011. **40**(2): p. 281-8.
97. Wang, J., et al., *Splice variants as novel targets in pancreatic ductal adenocarcinoma.* Sci Rep, 2017. **7**(1): p. 2980.
98. Hayes, G.M., et al., *Targeting the RNA splicing machinery as a novel treatment strategy for pancreatic carcinoma.* Cancer Res, 2006. **66**(7): p. 3819-27.
99. Hayes, G.M., P.E. Carrigan, and L.J. Miller, *Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas.* Cancer Res, 2007. **67**(5): p. 2072-80.
100. Chen, S., et al., *CLK1/SRSF5 pathway induces aberrant exon skipping of METTL14 and Cyclin L2 and promotes growth and metastasis of pancreatic cancer.* J Hematol Oncol, 2021. **14**(1): p. 60.
101. Escobar-Hoyos, L.F., et al., *Altered RNA Splicing by Mutant p53 Activates Oncogenic RAS Signaling in Pancreatic Cancer.* Cancer Cell, 2020. **38**(2): p. 198-211 e8.
102. Huo, Z., et al., *PRPF40A as a potential diagnostic and prognostic marker is upregulated in pancreatic cancer tissues and cell lines: an integrated bioinformatics data analysis.* Onco Targets Ther, 2019. **12**: p. 5037-5051.
103. Ueda, J., et al., *Epithelial splicing regulatory protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases.* Oncogene, 2014. **33**(36): p. 4485-95.
104. Li, M., et al., *miR-193a-5p promotes pancreatic cancer cell metastasis through SRSF6-mediated alternative splicing of OGDHL and ECM1.* Am J Cancer Res, 2020. **10**(1): p. 38-59.
105. Calabretta, S., et al., *Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells.* Oncogene, 2016. **35**(16): p. 2031-9.
106. Jones, S., et al., *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.* Science, 2008. **321**(5897): p. 1801-6.
107. Biankin, A.V., et al., *Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes.* Nature, 2012. **491**(7424): p. 399-405.

108. Lee, S.C. and O. Abdel-Wahab, *Therapeutic targeting of splicing in cancer.* Nat Med, 2016. **22**(9): p. 976-86.
109. Heinrich, B., et al., *Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA.* J Biol Chem, 2009. **284**(21): p. 14303-15.
110. Yang, C., et al., *Genome-wide profiling reveals the landscape of prognostic alternative splicing signatures in pancreatic ductal adenocarcinoma.* Front Oncol, 2019. **9**: p. 511.
111. Rall, C.J. and A.K. Rustgi, *CD44 isoform expression in primary and metastatic pancreatic adenocarcinoma.* Cancer Res, 1995. **55**(9): p. 1831-5.
112. Gansauge, F., et al., *Differential expression of CD44 splice variants in human pancreatic adenocarcinoma and in normal pancreas.* Cancer Res, 1995. **55**(23): p. 5499-503.
113. Gotoda, T., et al., *Expression of CD44 variants and its association with survival in pancreatic cancer.* Jpn J Cancer Res, 1998. **89**(10): p. 1033-40.
114. Zhao, S., et al., *CD44 Expression Level and Isoform Contributes to Pancreatic Cancer Cell Plasticity, Invasiveness, and Response to Therapy.* Clin Cancer Res, 2016. **22**(22): p. 5592-5604.
115. Nie, H., et al., *The short isoform of PRLR suppresses the pentose phosphate pathway and nucleotide synthesis through the NEK9-Hippo axis in pancreatic cancer.* Theranostics, 2021. **11**(8): p. 3898-3915.
116. Kornmann, M., et al., *Expression of the IIIc variant of FGF receptor-1 confers mitogenic responsiveness to heparin and FGF-5 in TAKA-1 pancreatic ductal cells.* Int J Pancreatol, 2001. **29**(2): p. 85-92.
117. Liu, Z., et al., *Human fibroblast growth factor receptor 1-IIIb is a functional fibroblast growth factor receptor expressed in the pancreas and involved in proliferation and movement of pancreatic ductal cells.* Pancreas, 2007. **35**(2): p. 147-57.
118. Cras-Meneur, C. and R. Scharfmann, *FGFR1-IIIb is a putative marker of pancreatic progenitor cells.* Mech Dev, 2002. **116**(1-2): p. 205-8.
119. Liu, Z., et al., *Identification of a fibroblast growth factor receptor 1 splice variant that inhibits pancreatic cancer cell growth.* Cancer Res, 2007. **67**(6): p. 2712-9.
120. Kornmann, M., et al., *IIIc isoform of fibroblast growth factor receptor 1 is overexpressed in human pancreatic cancer and enhances tumorigenicity of hamster ductal cells.* Gastroenterology, 2002. **123**(1): p. 301-13.
121. Friess, H., et al., *Moderate activation of the apoptosis inhibitor bcl-xL worsens the prognosis in pancreatic cancer.* Ann Surg, 1998. **228**(6): p. 780-7.
122. Eskens, F.A., et al., *Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors.* Clin Cancer Res, 2013. **19**(22): p. 6296-304.
123. Seiler, M., et al., *H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers.* Nat Med, 2018. **24**(4): p. 497-504.
124. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nat Rev Drug Discov, 2002. **1**(4): p. 309-15.
125. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review).* Int J Mol Med, 2017. **40**(2): p. 271-280.
126. Fleuren, E.D., et al., *The kinome 'at large' in cancer.* Nat Rev Cancer, 2016. **16**(2): p. 83-98.
127. Deng, Y.N., J.A. Bellanti, and S.G. Zheng, *Essential Kinases and Transcriptional Regulators and Their Roles in Autoimmunity.* Biomolecules, 2019. **9**(4).
128. Martin Moyano, P., V. Nemec, and K. Paruch, *Cdc-Like Kinases (CLKs): Biology, Chemical Probes, and Therapeutic Potential.* Int J Mol Sci, 2020. **21**(20).

129. Ding, S., et al., *Regulation of alternative splicing of tau exon 10 by 9G8 and Dyrk1A.* Neurobiol Aging, 2012. **33**(7): p. 1389-99.
130. Smith, B.E., et al., *Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase.* Nat Commun, 2019. **10**(1): p. 131.
131. Shi, J., et al., *Cyclic AMP-dependent protein kinase regulates the alternative splicing of tau exon 10: a mechanism involved in tau pathology of Alzheimer disease.* J Biol Chem, 2011. **286**(16): p. 14639-48.
132. Kvissel, A.K., et al., *Involvement of the catalytic subunit of protein kinase A and of HA95 in pre-mRNA splicing.* Exp Cell Res, 2007. **313**(13): p. 2795-809.
133. Fackenthal, J.D. and L.A. Godley, *Aberrant RNA splicing and its functional consequences in cancer cells.* Dis Model Mech, 2008. **1**(1): p. 37-42.
134. Tam, B.Y., et al., *The CLK inhibitor SM08502 induces anti-tumor activity and reduces Wnt pathway gene expression in gastrointestinal cancer models.* Cancer Lett, 2020. **473**: p. 186-197.
135. Havens, M.A. and M.L. Hastings, *Splice-switching antisense oligonucleotides as therapeutic drugs.* Nucleic Acids Res, 2016. **44**(14): p. 6549-63.
136. Stein, C.A. and D. Castanotto, *FDA-Approved Oligonucleotide Therapies in 2017.* Mol Ther, 2017. **25**(5): p. 1069-1075.
137. Corey, D.R., *Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy.* Nat Neurosci, 2017. **20**(4): p. 497-499.
138. Shahbazi, R., B. Ozpolat, and K. Ulubayram, *Oligonucleotide-based theranostic nanoparticles in cancer therapy.* Nanomedicine (Lond), 2016. **11**(10): p. 1287-308.
139. Roberts, T.C., R. Langer, and M.J.A. Wood, *Advances in oligonucleotide drug delivery.* Nat Rev Drug Discov, 2020. **19**(10): p. 673-694.
140. Shimojo, M., et al., *The small cell lung cancer-specific isoform of RE1-silencing transcription factor (REST) is regulated by neural-specific Ser/Arg repeat-related protein of 100 kDa (nSR100).* Mol Cancer Res, 2013. **11**(10): p. 1258-68.
141. De Velasco, M.A., et al., *Targeting castration-resistant prostate cancer with androgen receptor antisense oligonucleotide therapy.* JCI Insight, 2019. **4**(17).
142. Ross, S.J., et al., *Targeting KRAS-dependent tumors with AZD4785, a high-affinity therapeutic antisense oligonucleotide inhibitor of KRAS.* Sci Transl Med, 2017. **9**(394).
143. Li, Z., et al., *Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines.* Oncol Rep, 2016. **35**(2): p. 1013-9.
144. Mercatante, D.R., et al., *Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. analysis of apoptosis and cell death.* J Biol Chem, 2001. **276**(19): p. 16411-7.
145. Taylor, J.K., et al., *Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides.* Nat Biotechnol, 1999. **17**(11): p. 1097-100.
146. Mohler, J.L., et al., *Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology.* J Natl Compr Canc Netw, 2019. **17**(5): p. 479-505.
147. Jandaghi, P., et al., *Expression of DRD2 Is Increased in Human Pancreatic Ductal Adenocarcinoma and Inhibitors Slow Tumor Growth in Mice.* Gastroenterology, 2016. **151**(6): p. 1218-1231.
148. Badea, L., et al., *Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia.* Hepatogastroenterology, 2008. **55**(88): p. 2016-27.
149. Bailey, P., et al., *Genomic analyses identify molecular subtypes of pancreatic cancer.* Nature, 2016. **531**(7592): p. 47-52.

150. Kirby, M.K., et al., *RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4.* Mol Oncol, 2016. **10**(8): p. 1169-82.
151. Vandamme, T., et al., *Whole-exome characterization of pancreatic neuroendocrine tumor cell lines BON-1 and QGP-1.* J Mol Endocrinol, 2015. **54**(2): p. 137-47.
152. Kyriazis, A.A., et al., *Morphological, biological, biochemical, and karyotypic characteristics of human pancreatic ductal adenocarcinoma Capan-2 in tissue culture and the nude mouse.* Cancer Res, 1986. **46**(11): p. 5810-5.
153. Tan, M.H., et al., *Characterization of a new primary human pancreatic tumor line.* Cancer Invest, 1986. **4**(1): p. 15-23.
154. Yunis, A.A., G.K. Arimura, and D.J. Russin, *Human pancreatic carcinoma (MIA PaCa-2) in continuous culture: sensitivity to asparaginase.* Int J Cancer, 1977. **19**(1): p. 128-35.
155. Deer, E.L., et al., *Phenotype and genotype of pancreatic cancer cell lines.* Pancreas, 2010. **39**(4): p. 425-35.
156. Kaighn, M.E., et al., *Establishment and characterization of a human prostatic carcinoma cell line (PC-3).* Invest Urol, 1979. **17**(1): p. 16-23.
157. van Steenbrugge, G.J., et al., *The human prostatic carcinoma cell line LNCaP and its derivatives. An overview.* Urol Res, 1989. **17**(2): p. 71-7.
158. Uphoff, C.C. and H.G. Drexler, *Detection of mycoplasma contaminations.* Methods Mol Biol, 2005. **290**: p. 13-23.
159. Miranda-Lorenzo, I., et al., *Intracellular autofluorescence: a biomarker for epithelial cancer stem cells.* Nat Methods, 2014. **11**(11): p. 1161-9.
160. López-Cánovas, J.L., et al., *Splicing factor SF3B1 is overexpressed and implicated in the aggressiveness and survival of hepatocellular carcinoma.* Cancer Lett, 2020. **496**: p. 72-83.
161. Vázquez-Borrego, M.C., et al., *A Somatostatin Receptor Subtype-3 (SST3) Peptide Agonist Shows Antitumor Effects in Experimental Models of Nonfunctioning Pituitary Tumors.* Clin Cancer Res, 2020. **26**(4): p. 957-969.
162. Del Río-Moreno, M., et al., *Dysregulation of the splicing machinery Is associated to the development of nonalcoholic fatty liver disease.* J Clin Endocrinol Metab, 2019. **104**(8): p. 3389-3402.
163. Westerfield, M., *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. 2000, 4th ed. Eugene: University of Oregon Press.
164. Valle, S., et al., *Exploiting oxidative phosphorylation to promote the stem and immunoevasive properties of pancreatic cancer stem cells.* Nat Commun, 2020. **11**(1): p. 5265.
165. Stirling, D. and G. Tomlinson. *Quantifish 1.1. Quantifish: fast, efficient analysis program for the quantification of fluorescence in zebrafish larvae*. University College London, London, UK 2017; Available from: https://zenodo.org/record/1182791.
166. Chan, C.S., et al., *ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup.* Nat Commun, 2018. **9**(1): p. 4158.
167. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.
168. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
169. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
170. Scrucca, L., et al., *mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.* R J, 2016. **8**(1): p. 289-317.

171. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
172. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
173. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.
174. Frankish, A., et al., *GENCODE reference annotation for the human and mouse genomes.* Nucleic Acids Res, 2019. **47**(D1): p. D766-D773.
175. Trincado, J.L., et al., *SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions.* Genome Biol, 2018. **19**(1): p. 40.
176. R Core Team, *R: A language and environment for statistical computing*. 2017, R Foundation for Statistical Computing.
177. Love, M.I., et al., *Tximeta: Reference sequence checksums for provenance identification in RNA-seq.* PLoS Comput Biol, 2020. **16**(2): p. e1007664.
178. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
179. McCarthy, D.J., Y. Chen, and G.K. Smyth, *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.* Nucleic Acids Res, 2012. **40**(10): p. 4288-97.
180. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data.* Genome Biol, 2010. **11**(3): p. R25.
181. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
182. Terfve, C.D., et al., *Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data.* Nat Commun, 2015. **6**: p. 8033.
183. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
184. Kim, H.K., et al., *Alternative splicing isoforms in health and disease.* Pflugers Arch, 2018. **470**(7): p. 995-1016.
185. Evsyukova, I., et al., *Alternative splicing in multiple sclerosis and other autoimmune diseases.* RNA Biol, 2010. **7**(4): p. 462-73.
186. Fuentes-Fayos, A.C., et al., *Splicing machinery dysregulation drives glioblastoma development/aggressiveness: oncogenic role of SRSF3.* Brain, 2020. **In press**.
187. Vázquez-Borrego, M.C., et al., *Splicing machinery is dysregulated in pituitary neuroendocrine tumors and is associated with aggressiveness features.* Cancers (Basel), 2019. **11**(10).
188. Wang, D., et al., *Splicing Factor Prp8 Interacts With NES(AR) and Regulates Androgen Receptor in Prostate Cancer Cells.* Mol Endocrinol, 2015. **29**(12): p. 1731-42.
189. Fucic, A., et al., *Systems Oncology: Bridging Pancreatic and Castrate Resistant Prostate Cancer.* Pathol Oncol Res, 2019. **25**(4): p. 1269-1277.
190. Alors-Perez, E., et al., *Dysregulated splicing factor SF3B1 unveils a dual therapeutic vulnerability to target pancreatic cancer cells and cancer stem cells with an anti-splicing drug.* J Exp Clin Cancer Res, 2021. **40**(1): p. 382.
191. Jimenez-Vacas, J.M., et al., *Spliceosome component SF3B1 as novel prognostic biomarker and therapeutic target for prostate cancer.* Transl Res, 2019. **212**: p. 89-103.

192. Li, K. and Z. Wang, *Splicing factor SRSF2-centric gene regulation.* Int J Biol Sci, 2021. **17**(7): p. 1708-1715.
193. Jeong, S., *SR Proteins: Binders, Regulators, and Connectors of RNA.* Mol Cells, 2017. **40**(1): p. 1-9.
194. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.
195. Sliskovic, I., H. Eich, and M. Muller-McNicoll, *Exploring the multifunctionality of SR proteins.* Biochem Soc Trans, 2022. **50**(1): p. 187-198.
196. Sapra, A.K., et al., *SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo.* Mol Cell, 2009. **34**(2): p. 179-90.
197. Kim, E., et al., *SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition.* Cancer Cell, 2015. **27**(5): p. 617-30.
198. Aujla, A., et al., *SRSF2 mutations in myelodysplasia/myeloproliferative neoplasms.* Biomark Res, 2018. **6**: p. 29.
199. Luo, C., et al., *SRSF2 Regulates Alternative Splicing to Drive Hepatocellular Carcinoma Development.* Cancer Res, 2017. **77**(5): p. 1168-1178.
200. Chen, Z., et al., *CircPLCE1 facilitates the malignant progression of colorectal cancer by repressing the SRSF2-dependent PLCE1 pre-RNA splicing.* J Cell Mol Med, 2021. **25**(15): p. 7244-7256.
201. Kedzierska, H., et al., *Decreased Expression of SRSF2 Splicing Factor Inhibits Apoptotic Pathways in Renal Cancer.* Int J Mol Sci, 2016. **17**(10).
202. Edmond, V., et al., *A new function of the splicing factor SRSF2 in the control of E2F1-mediated cell cycle progression in neuroendocrine lung tumors.* Cell Cycle, 2013. **12**(8): p. 1267-78.
203. Wallace, T.A., et al., *Tumor immunobiological differences in prostate cancer between African-American and European-American men.* Cancer Res, 2008. **68**(3): p. 927-36.
204. Cancer Genome Atlas Research Network, *Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma.* Cancer Cell, 2017. **32**(2): p. 185-203 e13.
205. Park, W., A. Chawla, and E.M. O'Reilly, *Pancreatic Cancer: A Review.* JAMA, 2021. **326**(9): p. 851-862.
206. Shen, Q., et al., *Possible Molecular Markers for the Diagnosis of Pancreatic Ductal Adenocarcinoma.* Med Sci Monit, 2018. **24**: p. 2368-2376.
207. Adesso, L., et al., *Gemcitabine triggers a pro-survival response in pancreatic cancer cells through activation of the MNK2/eIF4E pathway.* Oncogene, 2013. **32**(23): p. 2848-57.
208. Peng, J., et al., *Differential expression of RBM5 and KRAS in pancreatic ductal adenocarcinoma and their association with clinicopathological features.* Oncol Lett, 2013. **5**(3): p. 1000-1004.
209. Kleeff, J., et al., *Pancreatic cancer.* Nat Rev Dis Primers, 2016. **2**: p. 16022.
210. Zhou, Z., et al., *The biological function and clinical significance of SF3B1 mutations in cancer.* Biomark Res, 2020. **8**: p. 38.
211. Obeng, E.A., C. Stewart, and O. Abdel-Wahab, *Altered RNA processing in cancer pathogenesis and therapy.* Cancer Discov, 2019. **9**(11): p. 1493-1510.
212. Yang, J.Y., et al., *SF3B1 mutation in pancreatic cancer contributes to aerobic glycolysis and tumor growth through a PP2A-c-Myc axis.* Mol Oncol, 2021.
213. Hermann, P.C. and B. Sainz, Jr., *Pancreatic cancer stem cells: A state or an entity?* Semin Cancer Biol, 2018. **53**: p. 223-231.
214. Brandi, J., et al., *Proteomic analysis of pancreatic cancer stem cells: Functional role of fatty acid synthesis and mevalonate pathways.* J Proteomics, 2017. **150**: p. 310-322.

215. Bao, B., et al., *Pancreatic cancer stem-like cells display aggressive behavior mediated via activation of FoxQ1.* J Biol Chem, 2014. **289**(21): p. 14520-33.
216. Cabezas-Sáinz, P., et al., *Improving zebrafish embryo xenotransplantation conditions by increasing incubation temperature and establishing a proliferation index with ZFtool.* BMC Cancer, 2018. **18**(1): p. 3.
217. Zhang, J., et al., *Characterization of cancer genomic heterogeneity by next-generation sequencing advances precision medicine in cancer treatment.* Precis Clin Med, 2018. **1**(1): p. 29-48.
218. Turajlic, S., et al., *Resolving genetic heterogeneity in cancer.* Nat Rev Genet, 2019. **20**(7): p. 404-416.
219. Durán-Prado, M., 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**(5): p. 2497-502.
220. Puig-Domingo, M., et al., *The truncated isoform of somatostatin receptor5 (sst5TMD4) is associated with poorly differentiated thyroid cancer.* PLoS One, 2014. **9**(1): p. e85527.
221. Rincón-Fernandez, D., 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**(3): p. 447-457.
222. Ibáñez-Costa, A., et al., *In1-ghrelin splicing variant is overexpressed in pituitary adenomas and increases their aggressive features.* Sci Rep, 2015. **5**: p. 8714.
223. Climente-Gonzalez, H., et al., *The Functional Impact of Alternative Splicing in Cancer.* Cell Rep, 2017. **20**(9): p. 2215-2226.
224. Coltri, P.P., M.G.P. Dos Santos, and G.H.G. da Silva, *Splicing and cancer: Challenges and opportunities.* Wiley Interdiscip Rev RNA, 2019. **10**(3): p. e1527.
225. Nasiri-Aghdam, M., T.C. Garcia-Garduno, and L.F. Jave-Suarez, *CELF Family Proteins in Cancer: Highlights on the RNA-Binding Protein/Noncoding RNA Regulatory Axis.* Int J Mol Sci, 2021. **22**(20).
226. Dasgupta, T. and A.N. Ladd, *The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins.* Wiley Interdiscip Rev RNA, 2012. **3**(1): p. 104-21.
227. Wagnon, J.L., et al., *CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function.* PLoS Genet, 2012. **8**(11): p. e1003067.
228. Yang, Y., et al., *Complex seizure disorder caused by Brunol4 deficiency in mice.* PLoS Genet, 2007. **3**(7): p. e124.
229. Sun, W., et al., *Aberrant sodium channel activity in the complex seizure disorder of Celf4 mutant mice.* J Physiol, 2013. **591**(1): p. 241-55.
230. Karunakaran, D.K., et al., *The expression analysis of Sfrs10 and Celf4 during mouse retinal development.* Gene Expr Patterns, 2013. **13**(8): p. 425-36.
231. Chang, K., C. Yuan, and X. Liu, *A New RBPs-Related Signature Predicts the Prognosis of Colon Adenocarcinoma Patients.* Front Oncol, 2021. **11**: p. 627504.
232. Teerlink, C.C., et al., *An intronic variant in the CELF4 gene is associated with risk for colorectal cancer.* Cancer Epidemiol, 2021. **72**: p. 101941.
233. Huang, R.L., et al., *Integrated Epigenomics Analysis Reveals a DNA Methylation Panel for Endometrial Cancer Detection Using Cervical Scrapings.* Clin Cancer Res, 2017. **23**(1): p. 263-272.
234. Siddaway, R., et al., *Splicing is an alternate oncogenic pathway activation mechanism in glioma.* Nat Commun, 2022. **13**(1): p. 588.

235. Briest, F. and P. Grabowski, *The p53 network as therapeutic target in gastroenteropancreatic neuroendocrine neoplasms.* Cancer Treat Rev, 2015. **41**(5): p. 423-30.
236. Lubomierski, N., et al., *Tumor suppressor genes in the 9p21 gene cluster are selective targets of inactivation in neuroendocrine gastroenteropancreatic tumors.* Cancer Res, 2001. **61**(15): p. 5905-10.
237. van Riet, J., et al., *The genomic landscape of 85 advanced neuroendocrine neoplasms reveals subtype-heterogeneity and potential therapeutic targets.* Nat Commun, 2021. **12**(1): p. 4612.
238. Fuentes-Fayos, A.C., et al., *SF3B1 inhibition disrupts malignancy and prolongs survival in glioblastoma patients through BCL2L1 splicing and mTOR/ss-catenin pathways imbalances.* J Exp Clin Cancer Res, 2022. **41**(1): p. 39.
239. Yao, J.C., et al., *Everolimus for advanced pancreatic neuroendocrine tumors.* N Engl J Med, 2011. **364**(6): p. 514-23.
240. Hofving, T., et al., *The neuroendocrine phenotype, genomic profile and therapeutic sensitivity of GEPNET cell lines.* Endocr Relat Cancer, 2018. **25**(3): p. 367-380.
241. Khanna, L., et al., *Pancreatic Neuroendocrine Neoplasms: 2020 Update on Pathologic and Imaging Findings and Classification.* Radiographics, 2020. **40**(5): p. 1240-1262.
242. Taskin, O.C., et al., *Pancreatic neuroendocrine neoplasms: current state and ongoing controversies on terminology, classification and prognostication.* J Gastrointest Oncol, 2020. **11**(3): p. 548-558.
243. Lakis, V., et al., *DNA methylation patterns identify subgroups of pancreatic neuroendocrine tumors with clinical association.* Commun Biol, 2021. **4**(1): p. 155.
244. Missiaglia, E., et al., *Pancreatic endocrine tumors: expression profiling evidences a role for AKT-mTOR pathway.* J Clin Oncol, 2010. **28**(2): p. 245-55.
245. Bejar, R., *Splicing Factor Mutations in Cancer.* Adv Exp Med Biol, 2016. **907**: p. 215-28.
246. Liang, Y., et al., *SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells.* Leukemia, 2018. **32**(12): p. 2659-2671.
247. Zhang, J., et al., *Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities.* Proc Natl Acad Sci U S A, 2015. **112**(34): p. E4726-34.
248. Lopez-Canovas, J.L., et al., *Splicing factor SF3B1 is overexpressed and implicated in the aggressiveness and survival of hepatocellular carcinoma.* Cancer Lett, 2021. **496**: p. 72-83.
249. Herman, C.M., et al., *Lymphovascular invasion as a predictor of disease progression in prostate cancer.* Am J Surg Pathol, 2000. **24**(6): p. 859-63.
250. Maehara, Y., et al., *Prognosis and surgical treatment of gastric cancer invading the pancreas.* Oncology, 2000. **59**(1): p. 1-6.
251. Baugh, E.H., et al., *Why are there hotspot mutations in the TP53 gene in human cancers?* Cell Death Differ, 2018. **25**(1): p. 154-160.
252. Ariga, H., et al., *Endogenous acyl ghrelin is involved in mediating spontaneous phase III-like contractions of the rat stomach.* Neurogastroenterol Motil, 2007. **19**(8): p. 675-80.
253. Wang, Q., et al., *High expression of RAB27A and TP53 in pancreatic cancer predicts poor survival.* Med Oncol, 2015. **32**(1): p. 372.
254. Shugang, X., et al., *Prognostic Value of SMAD4 in Pancreatic Cancer: A Meta-Analysis.* Transl Oncol, 2016. **9**(1): p. 1-7.
255. Mavrou, A., et al., *Serine-arginine protein kinase 1 (SRPK1) inhibition as a potential novel targeted therapeutic strategy in prostate cancer.* Oncogene, 2015. **34**(33): p. 4311-9.

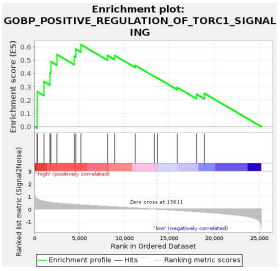
256. Gong, X.G., et al., *Gemcitabine resistance induced by interaction between alternatively spliced segment of tenascin-C and annexin A2 in pancreatic cancer cells.* Biol Pharm Bull, 2010. **33**(8): p. 1261-7.
257. van Bokhoven, A., et al., *Molecular characterization of human prostate carcinoma cell lines.* Prostate, 2003. **57**(3): p. 205-25.
258. Bao, B., et al., *Overview of cancer stem cells (CSCs) and mechanisms of their regulation: implications for cancer therapy.* Curr Protoc Pharmacol, 2013. **Chapter 14**: p. Unit 14 25.
259. Singh, B. and E. Eyras, *The role of alternative splicing in cancer.* Transcription, 2017. **8**(2): p. 91-98.
260. Grainger, R.J. and J.D. Beggs, *Prp8 protein: at the heart of the spliceosome.* RNA, 2005. **11**(5): p. 533-57.
261. Wan, R., et al., *How Is Precursor Messenger RNA Spliced by the Spliceosome?* Annu Rev Biochem, 2020. **89**: p. 333-358.
262. Arzalluz-Luque, A., et al., *Mutant PRPF8 Causes Widespread Splicing Changes in Spliceosome Components in Retinitis Pigmentosa Patient iPSC-Derived RPE Cells.* Front Neurosci, 2021. **15**: p. 636969.
263. Adler, A.S., et al., *An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth.* Genes Dev, 2014. **28**(10): p. 1068-84.
264. Elliott, D.J., et al., *RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology.* Int J Biochem Cell Biol, 2019. **108**: p. 1-6.
265. Matsunaga, S., et al., *RBMX: a regulator for maintenance and centromeric protection of sister chromatid cohesion.* Cell Rep, 2012. **1**(4): p. 299-308.
266. Zheng, T., et al., *RBMX is required for activation of ATR on repetitive DNAs to maintain genome stability.* Cell Death Differ, 2020. **27**(11): p. 3162-3176.
267. Song, Y., et al., *RBMX contributes to hepatocellular carcinoma progression and sorafenib resistance by specifically binding and stabilizing BLACAT1.* Am J Cancer Res, 2020. **10**(11): p. 3644-3665.
268. Schumann, F.L., et al., *RBMX Protein Expression in T-Cell Lymphomas Predicts Chemotherapy Response and Prognosis.* Cancers (Basel), 2021. **13**(19).
269. Yan, Q., et al., *RBMX suppresses tumorigenicity and progression of bladder cancer by interacting with the hnRNP A1 protein to regulate PKM alternative splicing.* Oncogene, 2021. **40**(15): p. 2635-2650.
270. Hirschfeld, M., et al., *HNRNP G and HTRA2-BETA1 regulate estrogen receptor alpha expression with potential impact on endometrial cancer.* BMC Cancer, 2015. **15**: p. 86.
271. Guo, J., et al., *Underexpression of SRSF3 and its target gene RBMX predicts good prognosis in patients with head and neck cancer.* J Oral Sci, 2020. **62**(2): p. 175-179.
272. Galej, W.P., et al., *Crystal structure of Prp8 reveals active site cavity of the spliceosome.* Nature, 2013. **493**(7434): p. 638-43.
273. Cancer Genome Atlas Research Network. Electronic address, a.a.d.h.e. and N. Cancer Genome Atlas Research, *Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma.* Cancer Cell, 2017. **32**(2): p. 185-203 e13.
274. Waters, A.M. and C.J. Der, *KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer.* Cold Spring Harb Perspect Med, 2018. **8**(9).
275. Popli, P., et al., *Splicing factor SF3B1 promotes endometrial cancer progression via regulating KSR2 RNA maturation.* Cell Death Dis, 2020. **11**(10): p. 842.
276. Effenberger, K.A., V.K. Urabe, and M.S. Jurica, *Modulating splicing with small molecular inhibitors of the spliceosome.* Wiley Interdiscip Rev RNA, 2017. **8**(2).

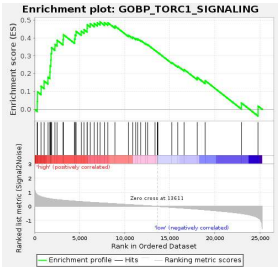
277. Zhang, L., et al., *Knockdown of SF3B1 inhibits cell proliferation, invasion and migration triggering apoptosis in breast cancer via aberrant splicing.* Breast Cancer, 2020. **27**(3): p. 464-476.
278. Liberante, F.G., et al., *Altered splicing and cytoplasmic levels of tRNA synthetases in SF3B1-mutant myelodysplastic syndromes as a therapeutic vulnerability.* Sci Rep, 2019. **9**(1): p. 2678.
279. Serrat, X., et al., *CRISPR editing of sftb-1/SF3B1 in Caenorhabditis elegans allows the identification of synthetic interactions with cancer-related mutations and the chemical inhibition of splicing.* PLoS Genet, 2019. **15**(10): p. e1008464.
280. Wang, E. and I. Aifantis, *RNA Splicing and Cancer.* Trends Cancer, 2020. **6**(8): p. 631-644.
281. Mann, K.M., et al., *KRAS-related proteins in pancreatic cancer.* Pharmacol Ther, 2016. **168**: p. 29-42.
282. Song, M., et al., *AKT as a Therapeutic Target for Cancer.* Cancer Res, 2019. **79**(6): p. 1019-1031.
283. Isono, K., et al., *Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1.* Genes Dev, 2005. **19**(5): p. 536-41.
284. Paolella, B.R., et al., *Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability.* Elife, 2017. **6**.
285. Elghazi, L., et al., *Regulation of pancreas plasticity and malignant transformation by Akt signaling.* Gastroenterology, 2009. **136**(3): p. 1091-103.
286. Xu, R. and J. Hu, *The role of JNK in prostate cancer progression and therapeutic strategies.* Biomed Pharmacother, 2020. **121**: p. 109679.
287. Edling, C.E., et al., *Key role of phosphoinositide 3-kinase class IB in pancreatic cancer.* Clin Cancer Res, 2010. **16**(20): p. 4928-37.
288. Stevens, M. and S. Oltean, *Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing.* Front Genet, 2019. **10**: p. 804.
289. Plowman, S.J., et al., *The K-Ras 4A isoform promotes apoptosis but does not affect either lifespan or spontaneous tumor incidence in aging mice.* Exp Cell Res, 2006. **312**(1): p. 16-26.
290. Bernard, H., et al., *The p53 isoform, Delta133p53alpha, stimulates angiogenesis and tumour progression.* Oncogene, 2013. **32**(17): p. 2150-60.
291. de Sousa e Melo, F., et al., *A distinct role for Lgr5(+) stem cells in primary and metastatic colon cancer.* Nature, 2017. **543**(7647): p. 676-680.
292. Shimokawa, M., et al., *Visualization and targeting of LGR5(+) human colon cancer stem cells.* Nature, 2017. **545**(7653): p. 187-192.

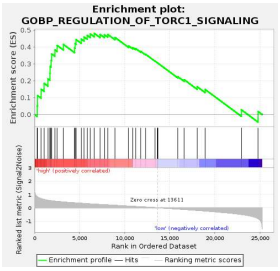
Article

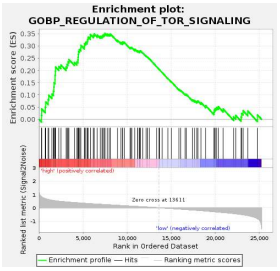
Appendix

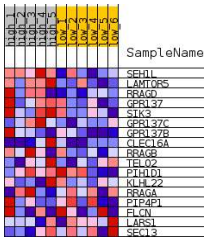
Supplemental Figure 1 (next page). GSEA analysis performed by Gene pattern in Reactome using the TCGA cohort classified by CELF4 expression levels in low and high mRNA expression groups and its corresponded differentially expressed genes in each pathway.

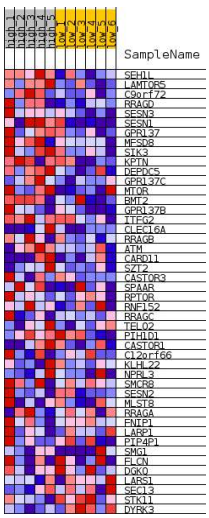

Enrichment plot:
GOBP_POSITIVE_REGULATION_OF_TORC1_SIGNALING

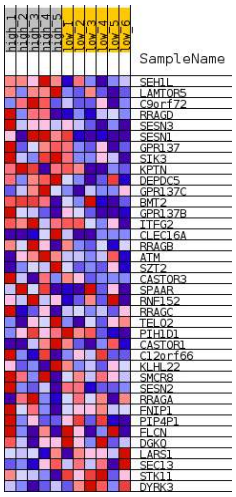

Enrichment plot: GOBP_TORC1_SIGNALING

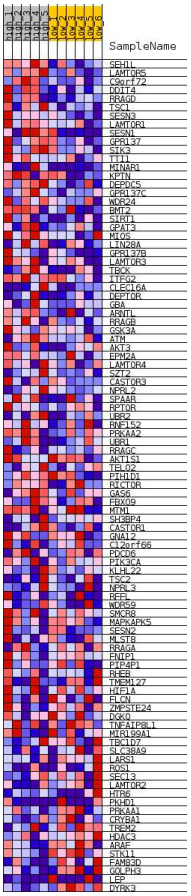

Enrichment plot:
GOBP_REGULATION_OF_TORC1_SIGNALING

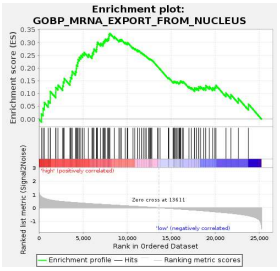

Enrichment plot:
GOBP_REGULATION_OF_TOR_SIGNALING

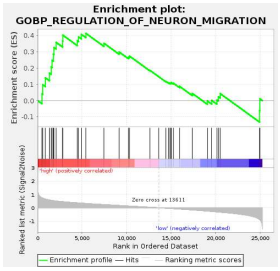

SampleName
SEH1L
LAMTOR5
RRAGD
GPR137
STK3
GPR137C
GPR137B
CLEC16A
RRAGB
TELO2
PTHLD1
KLHL22
RRAGA
PTP4P1
FLCN
LARS1
SEC13

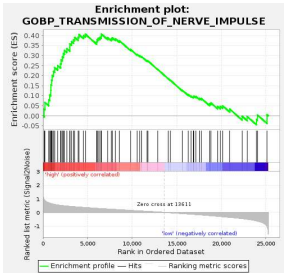

SampleName
SEH1L
LAMTOR5
C9orf72
RRAGD
SESN3
SESN1
GPR137
MFSD8
SIK3
KPTN
DEPDC5
GPR137C
MTOR
BMT2
GPR137B
ITFG2
CLEC16A
RRAGB
ATM
CARD11
SZT2
CASTOR3
SPAAR
RPTOR
RNF152
RRAGC
TELO2
PTHLD1
CASTOR1
C12orf66
KLHL22
NPRL3
SMCR8
SESN2
MLST8
RRAGA
FNIP1
LARP1
PTP4P1
SMG1
FLCN
DGKQ
LARS1
SEC13
STK11
DYRK3

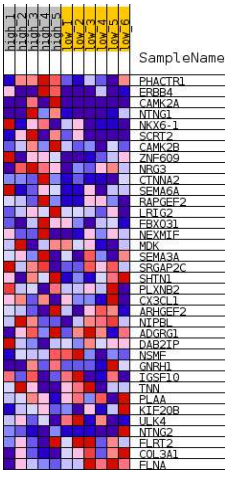

SampleName
SEH1L
LAMTOR5
C9orf72
RRAGD
SESN3
SESN1
GPR137
SIK3
KPTN
DEPDC5
GPR137C
BMT2
GPR137B
ITFG2
CLEC16A
RRAGB
ATM
SZT2
CASTOR3
SPAAR
RNF152
RRAGC
TELO2
PTHLD1
CASTOR1
C12orf66
KLHL22
SMCR8
SESN2
RRAGA
FNIP1
PTP4P1
FLCN
DGKQ
LARS1
SEC13
STK11
DYRK3

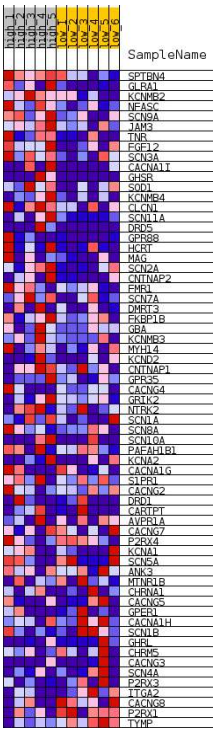

SampleName
SEH1L
LAMTOR5
C9orf72
DDIT4
RRAGD
TSC1
SESN3
LAMTOR1
SESN1
GPR137
SIK3
ITI1
MINAR1
KPTN
DEPDC5
GPR137C
WDR24
BMT2
SIRT1
GPAT3
MIOS
LIN28A
GPR137B
LAMTOR3
TBCK
ITFG2
CLEC16A
DEPTOR
GBA
ARNTL
RRAGB
GSK3A
ATM
AKT3
EPM2A
LAMTOR4
SZT2
CASTOR3
NPRL2
SPAAR
RPTOR
UBR2
RNF152
PRKAA2
UBR1
RRAGC
AKTIS1
TELO2
PTHLD1
RICTOR
GAS6
FBXO9
MTM1
SH3BP4
CASTOR1
GNA12
C12orf66
PDCD6
PIK3CA
KLHL22
TSC2
NPRL3
RFT1
WDR59
SMCR8
MAPKAPK5
SESN2
MLST8
RRAGA
FNIP1
PTP4P1
RHEB
TMEM127
HIF1A
FLCN
ZMPSTE24
DGKQ
TNFAIP8L1
MIR199A1
TRIO17
SLC38A9
LARS1
ROS1
SEC13
LAMTOR2
HTR6
PHLDA1
PRKAA1
CRYBA1
TREM2
HDAC3
ARAF
STK11
FAM83D
GOLPH3
LEP
DYRK3

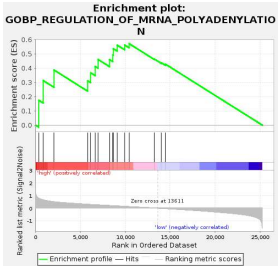

Enrichment plot:
GOBP_MRNA_EXPORT_FROM_NUCLEUS

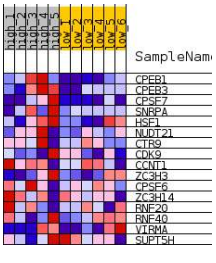

Enrichment plot:
GOBP_REGULATION_OF_NEURON_MIGRATION


Enrichment plot:
GOBP_TRANSMISSION_OF_NERVE_IMPULSE


SampleName
PHACTR1
ERBB4
CAMK2A
NTNG1
NKX6-1
SCRT2
CAMK2B
ZNF609
NRG3
CTNNA2
SEMA6A
RAPGEF2
LRIG2
FBXO31
NEXMIF
MDK
SEMA3A
SRGAP2C
SHTN1
PLXNB2
CX3CL1
ARHGEF2
NTRK1
ADGRG1
DAB2IP
NSMF
GNRH1
IGSF10
TNN
PLAA
KIF20B
ULK4
NTNG2
FLRT2
COL3A1
FLNA


SampleName
SPTBN4
GLRA1
KCNMB2
NFASC
SCN9A
JAM3
TNR
FGF12
SCN3A
CACNA1I
GHSR
SOD1
KCNMB4
CLCN1
SCN11A
DRD5
GPR88
HCRT
MAG
SCN2A
CNTNAP2
FMR1
SCN7A
DMRT3
FKBP1B
GBA
KCNMB3
MYH14
KCND2
CNTNAP1
GPR35
CACNG4
GRIK2
NTRK2
SCN1A
SCN8A
SCN10A
PAFAH1B1
KCNA2
CACNA1G
S1PR1
CACNG2
DRD1
CARTPT
AVPR1A
CACNG7
P2RX4
KCNA1
SCN5A
ANK3
MTNR1B
CHRNA1
CACNG5
GPER1
CACNA1H
SCN1B
GHRL
CHRM5
CACNG3
SCN4A
P2RX3
ITGA2
CACNG8
P2RX1
TYMP


Enrichment plot:
GOBP_REGULATION_OF_MRNA_POLYADENYLATION


SampleName
CPEB1
CPEB3
CPSF7
SNRPA
HSF1
NUDT21
CTR9
CDK9
CCNT1
ZC3H3
CPSF6
ZC3H14
RNF20
RNF40
VIRMA
SUPT5H

APPENDIX 1. Genes differentially expressed accordingly to high and low *CELF4* expression groups of samples.

Ensemble ID	Gene	Fold Change	q-value
ENSG00000096006	CRISP3	-10,3216	8,88E-11
ENSG00000204071	TCEAL6	4,8767	1,88E-09
ENSG00000162782	TDRD5	6,3719	3,89E-09
ENSG00000101489	CELF4	2,2406	5,37E-08
ENSG00000137673	MMP7	-9,4742	5,75E-08
ENSG00000121853	GHSR	9,5267	2,01E-07
ENSG00000177551	NHLH2	7,7089	3,53E-07
ENSG00000125931	CITED1	4,2772	4,16E-07
ENSG00000105509	HAS1	-4,8445	5,47E-07
ENSG00000223770	CACNA2D1-AS1	6,5724	5,47E-07
ENSG00000198739	LRRTM3	6,8189	5,47E-07
ENSG00000196361	ELAVL3	5,3349	1,90E-06
ENSG00000132693	CRP	-9,9749	3,12E-06
ENSG00000164825	DEFB1	-8,7371	4,25E-06
ENSG00000164690	SHH	-6,4918	2,47E-05
ENSG00000007306	CEACAM7	-8,4331	2,65E-05
ENSG00000170827	CELP	7,5848	3,09E-05
ENSG00000162896	PIGR	-8,0802	3,44E-05
ENSG00000172568	FNDC9	7,1742	4,51E-05
ENSG00000099337	KCNK6	-1,9022	5,51E-05
ENSG00000172548	NIPAL4	6,5633	5,51E-05
ENSG00000204642	HLA-F	7,0029	5,75E-05
ENSG00000124216	SNAI1	-2,8360	6,22E-05
ENSG00000183638	RP1L1	4,0980	6,89E-05

ENSG00000131910	NR0B2	5,7413	8,18E-05
ENSG00000102837	OLFM4	-7,3960	0,0001004
ENSG00000128285	MCHR1	3,0197	0,0001275
ENSG00000152969	JAKMIP1	3,3468	0,0001591
ENSG00000181143	MUC16	-7,4561	0,0001608
ENSG00000197251	LINC00336	4,8803	0,0001624
ENSG00000211677	IGLC2	-6,3827	0,000166
ENSG00000230795	HLA-K	7,9412	0,000166
ENSG00000128342	LIF	-4,6062	0,0001786
ENSG00000254585	MAGEL2	5,6652	0,0001857
ENSG00000117983	MUC5B	-6,9898	0,0002114
ENSG00000187479	C11orf96	-2,4220	0,0002316
ENSG00000180861	LINC01559	-7,5254	0,0002321
ENSG00000005108	THSD7A	2,1201	0,0002461
ENSG00000083067	TRPM3	4,5033	0,0002657
ENSG00000164220	F2RL2	-3,8613	0,0003285
ENSG00000260265	LINC02562	8,4724	0,0003285
ENSG00000011677	GABRA3	5,9513	0,000615
ENSG00000154764	WNT7A	-8,1436	0,0006244
ENSG00000248596	AC139491.2	4,8124	0,0006283
ENSG00000135917	SLC19A3	-3,9443	0,0007639
ENSG00000234965	SHISA8	-7,9759	0,0008876
ENSG00000148346	LCN2	-5,7501	0,0008876
ENSG00000253666	AP000424.1	4,7210	0,0008876
ENSG00000120149	MSX2	-2,7502	0,0009477
ENSG00000163630	SYNPR	6,0585	0,0009477
ENSG00000259223	AC009654.1	4,6312	0,0009714
ENSG00000173406	DAB1	3,3396	0,0010834
ENSG00000185686	PRAME	-8,1063	0,0011101

ENSG00000070808	CAMK2A	4,7064	0,0011101
ENSG00000183379	SYNDIG1L	4,8295	0,0011704
ENSG00000145428	RNF175	2,2610	0,0013314
ENSG00000005421	PON1	5,0579	0,0013541
ENSG00000147255	IGSF1	5,5057	0,0013541
ENSG00000112164	GLP1R	4,1285	0,0014186
ENSG00000237686	AL109615.3	3,0881	0,0014228
ENSG00000136244	IL6	-4,8567	0,0015182
ENSG00000256321	AC087235.2	6,1013	0,0015182
ENSG00000165376	CLDN2	-5,8518	0,0015327
ENSG00000171236	LRG1	-3,6490	0,0016169
ENSG00000053524	MCF2L2	2,0379	0,0016169
ENSG00000123342	MMP19	-2,9414	0,0016242
ENSG00000253537	PCDHGA7	-2,1922	0,0016242
ENSG00000102468	HTR2A	5,3062	0,0017708
ENSG00000124227	ANKRD60	7,0290	0,0017708
ENSG00000135144	DTX1	3,4904	0,0018284
ENSG00000137463	MGARP	2,1253	0,001862
ENSG00000146411	SLC2A12	4,1282	0,001862
ENSG00000269256	AC024603.1	5,3623	0,0022042
ENSG00000188580	NKAIN2	4,1342	0,002291
ENSG00000015413	DPEP1	-5,1365	0,0026042
ENSG00000140285	FGF7	-3,5429	0,0026253
ENSG00000261241	LINC02128	8,7911	0,0027713
ENSG00000241158	ADAMTS9-AS1	-3,5667	0,0031034
ENSG00000145888	GLRA1	3,4654	0,0031869
ENSG00000173432	SAA1	-6,6250	0,0032628
ENSG00000108342	CSF3	-4,9764	0,0032974
ENSG00000255406	LINC02730	7,9675	0,0033081

ENSG00000272068	AL365181.2	-5,9851	0,0033084
ENSG00000175352	NRIP3	2,8786	0,003705
ENSG00000155511	GRIA1	4,5065	0,003705
ENSG00000178038	ALS2CL	-2,1606	0,0037952
ENSG00000163519	TRAT1	-2,8819	0,0038295
ENSG00000137766	UNC13C	6,6657	0,0040817
ENSG00000081041	CXCL2	-3,0947	0,0043758
ENSG00000227141	AL160286.1	3,0190	0,0043758
ENSG00000100593	ISM2	3,3190	0,0044932
ENSG00000105388	CEACAM5	-4,9169	0,0045807
ENSG00000001626	CFTR	-4,7880	0,0045807
ENSG00000188257	PLA2G2A	-4,5645	0,0045807
ENSG00000013588	GPRC5A	-3,7894	0,0045807
ENSG00000151490	PTPRO	4,2138	0,0045807
ENSG00000204241	LINC02731	4,4960	0,0045807
ENSG00000158639	PAGE5	7,2916	0,0045807
ENSG00000257048	LINC02417	3,4425	0,0050397
ENSG00000165553	NGB	4,1576	0,0050397
ENSG00000201920	RNA5SP442	4,5468	0,0050397
ENSG00000188263	IL17REL	-5,2530	0,0051949
ENSG00000125735	TNFSF14	-2,9258	0,0051949
ENSG00000249896	LINC02495	3,3519	0,0053562
ENSG00000204060	FOXO6	3,2266	0,0061281
ENSG00000116774	OLFML3	-3,4339	0,0062924
ENSG00000253755	IGHGP	-4,5406	0,0063578
ENSG00000133401	PDZD2	2,8266	0,0064806
ENSG00000164920	OSR2	-3,6517	0,0066019
ENSG00000270605	AL353622.1	2,6358	0,0067092
ENSG00000143196	DPT	-4,2476	0,0071821

ENSG00000214922	HLA-F-AS1	7,3733	0,0072802
ENSG00000163600	ICOS	-3,2719	0,0075298
ENSG00000115009	CCL20	-4,4584	0,007609
ENSG00000185038	MROH2A	-5,5834	0,0076212
ENSG00000255026	AC136475.3	-3,0381	0,0076681
ENSG00000265356	AC004147.4	5,9461	0,0077676
ENSG00000269404	SPIB	-4,4419	0,0078782
ENSG00000162877	PM20D1	3,2498	0,007881
ENSG00000186832	KRT16	-6,5705	0,0081247
ENSG00000182329	KIAA2012	3,1061	0,0084715
ENSG00000116329	OPRD1	5,3673	0,0088877
ENSG00000166825	ANPEP	-2,2298	0,0089131
ENSG00000137648	TMPRSS4	-6,6893	0,0091469
ENSG00000206129	AC006305.1	-5,7204	0,0091469
ENSG00000272405	AL365181.3	-5,2567	0,0091469
ENSG00000148735	PLEKHS1	-5,0904	0,0091469
ENSG00000149968	MMP3	-5,0328	0,0091469
ENSG00000164530	PI16	-5,0235	0,0091469
ENSG00000271584	LINC02550	-4,0063	0,0091469
ENSG00000175592	FOSL1	-2,7862	0,0091469
ENSG00000012171	SEMA3B	-2,3525	0,0091469
ENSG00000109089	CDR2L	-1,6664	0,0091469
ENSG00000261996	AC004706.1	2,3436	0,0091469
ENSG00000231443	AC124944.1	2,8609	0,0091469
ENSG00000257986	LINC02306	5,8520	0,0091469
ENSG00000254607	AP001783.1	6,7696	0,0091469
ENSG00000157005	SST	7,2440	0,0093699
ENSG00000172461	FUT9	7,2467	0,0096157
ENSG00000090539	CHRD	-3,2044	0,0098696

ENSG00000160181	TFF2	-6,9287	0,0100918
ENSG00000188112	C6orf132	-3,2443	0,0101146
ENSG00000163362	INAVA	-5,5583	0,0105607
ENSG00000134398	ERN2	-6,4520	0,0105671
ENSG00000187583	PLEKHN1	-2,9438	0,0105671
ENSG00000147573	TRIM55	-7,3433	0,0108355
ENSG00000008394	MGST1	-3,3576	0,0108355
ENSG00000183775	KCTD16	2,4693	0,0108355
ENSG00000104783	KCNN4	-2,6571	0,0110001
ENSG00000073331	ALPK1	-2,0729	0,0110001
ENSG00000128422	KRT17	-5,2947	0,0111662
ENSG00000081138	CDH7	7,3271	0,011317
ENSG00000007908	SELE	-3,4886	0,0114333
ENSG00000103196	CRISPLD2	-1,9938	0,0114333
ENSG00000082293	COL19A1	4,6400	0,0114333
ENSG00000211664	IGLV2-18	-6,3462	0,0114602
ENSG00000262768	AC100791.2	-5,1664	0,0115185
ENSG00000165124	SVEP1	-3,2649	0,0115811
ENSG00000133067	LGR6	-3,2605	0,0116046
ENSG00000151650	VENTX	-2,2588	0,0116046
ENSG00000230533	AL356234.2	-5,2125	0,0121102
ENSG00000234756	LINC02621	5,6479	0,0121102
ENSG00000120889	TNFRSF10B	-1,5549	0,0122497
ENSG00000196611	MMP1	-5,4710	0,0123954
ENSG00000100196	KDELR3	-2,1423	0,0127228
ENSG00000177338	LINC00469	6,1166	0,0127431
ENSG00000228714	AL691420.1	6,0411	0,0130348
ENSG00000110848	CD69	-2,5892	0,013114
ENSG00000258498	DIO3OS	-4,0568	0,0131642

ENSG00000131037	EPS8L1	-2,6664	0,0132328
ENSG00000174171	AC020659.1	-2,7779	0,0136008
ENSG00000184557	SOCS3	-2,1367	0,0136008
ENSG00000256923	AC084819.1	4,0301	0,0136008
ENSG00000077616	NAALAD2	2,7853	0,0141649
ENSG00000251185	AC025244.1	6,3830	0,0145544
ENSG00000233429	HOTAIRM1	-1,7283	0,0146489
ENSG00000122584	NXPH1	-2,9716	0,0147994
ENSG00000272823	AL445423.1	-3,4589	0,0150555
ENSG00000146352	CLVS2	5,3818	0,0150555
ENSG00000163734	CXCL3	-3,0747	0,015382
ENSG00000142684	ZNF593	1,7894	0,015382
ENSG00000259353	AC090515.4	3,1190	0,015382
ENSG00000162951	LRRTM1	3,1405	0,015382
ENSG00000175206	NPPA	4,3930	0,015382
ENSG00000152208	GRID2	4,6561	0,015382
ENSG00000256310	NDUFA5P6	5,8697	0,015382
ENSG00000183813	CCR4	-3,4988	0,0155776
ENSG00000175946	KLHL38	-2,6479	0,0155776
ENSG00000172985	SH3RF3	-2,5634	0,0155776
ENSG00000244675	AC108676.1	2,4075	0,0155776
ENSG00000261060	AL160286.2	3,7560	0,0159916
ENSG00000170835	CEL	3,4583	0,0160925
ENSG00000205002	AARD	3,7491	0,0160925
ENSG00000148344	PTGES	-4,3812	0,0171231
ENSG00000185499	MUC1	-2,7887	0,0176556
ENSG00000164129	NPY5R	-3,8292	0,0177629
ENSG00000080007	DDX43	-3,3086	0,0177809
ENSG00000166793	YPEL4	3,2225	0,0178484

ENSG00000225431	LINC01671	-5,3920	0,0191081
ENSG00000100427	MLC1	-4,3542	0,0191081
ENSG00000226197	AL583785.1	-4,2384	0,0191081
ENSG00000167105	TMEM92	-3,3782	0,0191081
ENSG00000112964	GHR	1,9911	0,0191081
ENSG00000256574	OR13A1	2,7697	0,0191081
ENSG00000224215	AL606469.1	3,3033	0,0191081
ENSG00000234068	PAGE2	4,6172	0,0191081
ENSG00000117598	PLPPR5	5,3503	0,0191081
ENSG00000256612	CYP2B7P	3,2327	0,0191194
ENSG00000188517	COL25A1	4,0891	0,0191194
ENSG00000125657	TNFSF9	-2,6676	0,0192129
ENSG00000139292	LGR5	3,5322	0,019453
ENSG00000188959	C9orf152	-4,2265	0,0198817
ENSG00000124203	ZNF831	4,0092	0,0198817
ENSG00000117154	IGSF21	2,2377	0,019901
ENSG00000118733	OLFM3	5,2418	0,019901
ENSG00000267327	AC009271.1	-6,2778	0,0199043
ENSG00000101349	PAK5	1,9642	0,0199043
ENSG00000196878	LAMB3	-3,0415	0,0207774
ENSG00000235563	AL445183.2	-4,2911	0,0208472
ENSG00000064787	BCAS1	-4,1164	0,0208472
ENSG00000267651	AC015961.1	1,8927	0,0208472
ENSG00000198074	AKR1B10	-6,2054	0,0209207
ENSG00000186910	SERPINA11	-5,5975	0,0209207
ENSG00000225630	MTND2P28	-4,5015	0,0211547
ENSG00000146216	TTBK1	3,0002	0,0211547
ENSG00000211749	TRBV23-1	-5,4372	0,0215534
ENSG00000226435	ANKRD18DP	-2,9736	0,021979

ENSG00000267560	AC027514.2	1,6541	0,021979
ENSG00000186766	FOXI2	4,5602	0,021979
ENSG00000233608	TWIST2	-2,6528	0,0220369
ENSG00000118156	ZNF541	2,8758	0,0220369
ENSG00000259738	ZNF444P1	4,2895	0,0223471
ENSG00000203414	BTBD7P1	2,8004	0,0224532
ENSG00000159450	TCHH	-3,2206	0,0225462
ENSG00000270112	AC090241.2	4,8487	0,0225462
ENSG00000267209	LINC01897	7,2540	0,0225462
ENSG00000123843	C4BPB	-3,8903	0,0227906
ENSG00000197632	SERPINB2	-6,7272	0,0230235
ENSG00000189334	S100A14	-5,0953	0,0230235
ENSG00000144834	TAGLN3	4,3123	0,0230235
ENSG00000164500	SPATA48	6,2367	0,0230235
ENSG00000261997	AC007336.1	-1,9665	0,023939
ENSG00000163331	DAPL1	4,2211	0,0240037
ENSG00000224259	LINC01133	-4,5002	0,0245394
ENSG00000181092	ADIPOQ	-11,2042	0,025098
ENSG00000171916	LGALS9C	-5,7537	0,0252223
ENSG00000102359	SRPX2	-2,9269	0,0253034
ENSG00000157765	SLC34A2	-4,6218	0,0255086
ENSG00000254675	AP003032.1	6,4363	0,0262218
ENSG00000267127	AC090360.1	3,3886	0,0267254
ENSG00000125730	C3	-3,7892	0,0270886
ENSG00000167771	RCOR2	2,2475	0,0270886
ENSG00000075891	PAX2	-5,8161	0,0272996
ENSG00000236939	BAALC-AS2	3,4921	0,0272996
ENSG00000239542	RN7SL399P	3,5397	0,0272996
ENSG00000182586	LINC00334	2,9238	0,0273337

ENSG00000250771	AC106865.1	2,9983	0,0273337
ENSG00000150672	DLG2	1,9009	0,0276341
ENSG00000135116	HRK	3,4669	0,0276341
ENSG00000166866	MYO1A	-3,4679	0,0279718
ENSG00000233058	LINC00884	2,4023	0,0279718
ENSG00000155886	SLC24A2	3,2473	0,0279718
ENSG00000184956	MUC6	-5,7608	0,0280342
ENSG00000134258	VTCN1	-4,2433	0,0284655
ENSG00000248211	TRPC7-AS1	-2,1649	0,0284655
ENSG00000120729	MYOT	-5,2116	0,0289679
ENSG00000165606	DRGX	4,4693	0,0294139
ENSG00000188910	GJB3	-4,7654	0,0298219
ENSG00000172156	CCL11	-3,5683	0,0298665
ENSG00000266970	AC061992.2	-2,9482	0,0302304
ENSG00000107742	SPOCK2	1,8107	0,0302304
ENSG00000183091	NEB	2,7687	0,0312408
ENSG00000206073	SERPINB4	-6,2137	0,0313957
ENSG00000117525	F3	-1,8681	0,0314411
ENSG00000087245	MMP2	-2,8882	0,0317849
ENSG00000224739	AC016735.1	-5,4013	0,032192
ENSG00000076716	GPC4	-2,6367	0,0323708
ENSG00000163359	COL6A3	-2,2225	0,0323708
ENSG00000187134	AKR1C1	-2,1017	0,0323708
ENSG00000196187	TMEM63A	-1,6637	0,0323708
ENSG00000116035	VAX2	2,9924	0,0323708
ENSG00000152910	CNTNAP4	3,0980	0,0323708
ENSG00000186369	LINC00643	3,4421	0,0323708
ENSG00000121351	IAPP	5,6930	0,0323708
ENSG00000237574	LINC01856	3,1121	0,0323722

ENSG00000162383	SLC1A7	-2,7026	0,0324407
ENSG00000238062	SPATA3-AS1	-4,3113	0,0326235
ENSG00000176170	SPHK1	-1,5551	0,0326235
ENSG00000239819	IGKV1D-8	-5,2480	0,0328717
ENSG00000255162	AP004833.2	5,6364	0,0330445
ENSG00000198547	C20orf203	2,5926	0,0331555
ENSG00000236244	SLC35F3-AS1	-3,1290	0,0334916
ENSG00000170961	HAS2	-3,0715	0,0334916
ENSG00000124302	CHST8	4,2425	0,0334916
ENSG00000203995	ZYG11A	2,8075	0,0336536
ENSG00000228705	LINC00659	-6,0804	0,0337103
ENSG00000247774	PCED1B-AS1	-1,6750	0,0337738
ENSG00000149451	ADAM33	-3,5303	0,0342411
ENSG00000213355	CNN2P8	6,0699	0,0344118
ENSG00000185915	KLHL34	2,6034	0,0347614
ENSG00000164107	HAND2	-3,0672	0,0349936
ENSG00000087495	PHACTR3	3,7037	0,035476
ENSG00000052344	PRSS8	-5,2382	0,035879
ENSG00000137273	FOXF2	-2,2269	0,0359537
ENSG00000128283	CDC42EP1	-2,1942	0,0360392
ENSG00000197549	PRAMENP	5,9439	0,0363346
ENSG00000072182	ASIC4	3,5574	0,0370785
ENSG00000184811	TRARG1	-6,6364	0,0378412
ENSG00000175793	SFN	-3,4950	0,0384947
ENSG00000182326	C1S	-2,2177	0,0384947
ENSG00000011332	DPF1	1,8623	0,0384947
ENSG00000225637	AP001046.1	3,3804	0,0384947
ENSG00000233355	CHRM3-AS2	-2,7381	0,0395774
ENSG00000177234	LINC01561	-4,3050	0,039627

ENSG00000186191	BPIFB4	-4,4741	0,0400867
ENSG00000267142	AC092296.3	2,8306	0,0405271
ENSG00000249309	AC020703.1	3,5583	0,0405271
ENSG00000169330	MINAR1	3,3046	0,0405787
ENSG00000124107	SLPI	-4,4836	0,0414238
ENSG00000135960	EDAR	-3,8877	0,0417621
ENSG00000134830	C5AR2	-2,8839	0,0417621
ENSG00000148600	CDHR1	2,7494	0,0417621
ENSG00000230018	PPIAP39	-2,2699	0,0420515
ENSG00000137869	CYP19A1	-2,6093	0,0426914
ENSG00000173805	HAP1	3,4151	0,0426914
ENSG00000008226	DLEC1	4,1252	0,0426914
ENSG00000229442	THEMIS3P	4,8562	0,0426914
ENSG00000197406	DIO3	-3,9649	0,042746
ENSG00000070748	CHAT	-4,9611	0,0429599
ENSG00000138311	ZNF365	2,0713	0,0429599
ENSG00000167676	PLIN4	-4,0209	0,0434564
ENSG00000272463	AL357054.4	2,5016	0,0434564
ENSG00000232524	AC073323.1	4,3756	0,0434564
ENSG00000248174	LINC02268	-2,8550	0,0435045
ENSG00000016402	IL20RA	1,7837	0,0440944
ENSG00000253767	PCDHGA8	-3,1738	0,0448404
ENSG00000227910	AC092634.3	-1,6302	0,0448404
ENSG00000271662	AC233280.2	2,6480	0,0448404
ENSG00000267251	AC139100.1	2,7176	0,0448404
ENSG00000269779	AC010329.2	4,4234	0,0448404
ENSG00000197705	KLHL14	3,3688	0,045156
ENSG00000241635	UGT1A1	-5,9417	0,0452754
ENSG00000114638	UPK1B	-4,5999	0,0452754

ENSG00000132718	SYT11	2,2952	0,0452754
ENSG00000073067	CYP2W1	-4,0540	0,0453206
ENSG00000130635	COL5A1	-2,5905	0,0456173
ENSG00000077274	CAPN6	-3,1992	0,0457099
ENSG00000111199	TRPV4	-2,2597	0,046018
ENSG00000257906	LINC02156	2,5147	0,0464148
ENSG00000124813	RUNX2	-1,7932	0,0467959
ENSG00000175287	PHYHD1	-1,6477	0,0467959
ENSG00000215023	AC131097.1	3,3968	0,0467959
ENSG00000211898	IGHD	-3,4718	0,0469657
ENSG00000110195	FOLR1	-4,8217	0,0472067
ENSG00000215784	FAM72D	-2,8330	0,0475827
ENSG00000105991	HOXA1	-2,6617	0,0486616
ENSG00000122861	PLAU	-1,7035	0,0487105

APPENDIX 2. Specific data for differential variants splicing events between high and low *CELF4* expression groups of samples.

Ensemble ID	SUPPA name	dPSI	p-value
ENSG00000105516	A5:chr19:49136912-49137606:49136912-49137765:-	0,4154	0,0085
ENSG00000168653	A5:chr1:39492074-39494395:39492070-39494395:+	0,1607	0,0125
ENSG00000268963	AF:chrHG1497_PATCH:153566964:153567010-153567806:153567358:153567428-153567806:+	-0,2030	0,0155
ENSG00000268963	A5:chrHG1497_PATCH:153569094-153569170:153568409-153569170:+	-0,2000	0,0155
ENSG00000268963	AF:chrHG1497_PATCH:153566847:153567010-153567806:153567358:153567428-153567806:+	-0,1952	0,0155
ENSG00000268963	AF:chrHG1497_PATCH:153566845:153567010-153567806:153567358:153567428-153567806:+	-0,1814	0,0155
ENSG00000261361	A5:chrHG79_PATCH:136232717-136233246:136232603-136233246:+	0,2766	0,0187
ENSG00000261361	A5:chrHG79_PATCH:136232769-136233246:136232603-136233246:+	0,3293	0,0187
ENSG00000214967	SE:chr16:16485953-16486450:16486519-16487313:+	0,2099	0,0190
ENSG00000197746	A3:chr10:73581770-73585594:73581764-73585594:-	0,1164	0,0235
ENSG00000261493	AF:chrHG79_PATCH:136290546-136291197:136291435:136290546-136293155:136293579:-	0,3025	0,0240
ENSG00000154274	AL:chr4:37590521-37591710:37593448:37590521-37624447:37625117:+	0,3228	0,0250
ENSG00000112936	AL:chr5:40980011-40980089:40980452:40980011-40981494:40981868:+	-0,5089	0,0280
ENSG00000103260	AF:chr16:765641:765984-766933:766574:766715-766933:+	0,2761	0,0285
ENSG00000103260	AF:chr16:765641:765984-766933:766427:766715-766933:+	0,4048	0,0285
ENSG00000101856	SE:chrX:118370654-118374272:118374427-118377114:+	0,1463	0,0290
ENSG00000141750	A5:chr17:37373426-37374120:37373426-37374322:-	-0,4733	0,0305
ENSG00000106541	AF:chr7:16841427-16844559:16844704:16841427-16872880:16872932:-	-0,5035	0,0315
ENSG00000262206	AF:chrHG747_PATCH:66098024:66098077-66110538:66099059:66099174-66110538:+	-0,1974	0,0320

ENSG00000169903	SE:chr3:149192838-149193610:149193699-149205406:+	-0,1384	0,0320
ENSG00000227460	SE:chrHSCHR6_MHC_SSTO:33548924-33549618:33549659-33550054:+	0,3673	0,0325
ENSG00000196275	SE:chr7:74216354-74217639:74217822-74219118:-	-0,3230	0,0330
ENSG00000153879	AF:chr19:33864236:33864801-33870050:33865218:33865519-33870050:+	0,1911	0,0330
ENSG00000196275	AL:chr7:74210490:74212604-74216326:74213351:74213856-74216326:-	0,2365	0,0330
ENSG00000164176	SE:chr5:83476339-83525673:83525702-83549902:-	-0,2038	0,0335
ENSG00000198035	SE:chr10:48221539-48229091:48229191-48231779:+	0,2622	0,0335
ENSG00000198035	SE:chr10:48231814-48235122:48235173-48235819:+	0,2622	0,0335
ENSG00000233024	SE:chr16:18462243-18466580:18466711-18468689:-	-0,1561	0,0345
ENSG00000233024	SE:chr16:18462243-18466435:18466711-18468689:-	0,2301	0,0345
ENSG00000167476	AF:chr19:2255343-2256382:2256416:2255343-2269333:2269758:-	-0,1613	0,0360
ENSG00000132698	SE:chr1:156031234-156035702:156035897-156038061:+	0,1311	0,0360
ENSG00000198689	A5:chrX:135077144-135080267:135077048-135080267:+	-0,1719	0,0365
ENSG00000146556	A5:chr2:114354159-114354248:114354155-114354248:+	-0,1432	0,0365
ENSG00000146556	RI:chr2:114353270:114353406-114353645:114353780:+	0,1432	0,0365
ENSG00000164638	A3:chr7:5331452-5334491:5331452-5334496:+	-0,2077	0,0366
ENSG00000164638	A3:chr7:5330868-5331324:5330868-5331361:+	-0,1995	0,0366
ENSG00000164638	A3:chr7:5322713-5327437:5322713-5327440:+	0,1989	0,0366
ENSG00000180964	SE:chrX:102508945-102509533:102509605-102510006:-	0,1136	0,0370
ENSG00000241360	AF:chr22:38054734:38055363-38061562:38060918:38061110-38061562:+	0,1244	0,0390
ENSG00000010818	SE:chr6:143097328-143104658:143104752-143158072:-	-0,2954	0,0415
ENSG00000228299	A3:chrHSCHR6_MHC_DBB:31231330-31231577:31231313-31231577:-	-0,2042	0,0420
ENSG00000217801	AF:chr1:998459:998581-1001210:999875:999969-1001210:+	0,1791	0,0420
ENSG00000228299	RI:chrHSCHR6_MHC_DBB:31229479:31229511-31229951:31230070:-	0,2338	0,0420
ENSG00000186716	AL:chr22:23655208-23656155:23656186:23655208-23656739:23656882:+	-0,4077	0,0425

ENSG00000186716	SE:chr22:23655208-23656155:23656260-23656739:+	-0,3274	0,0425
ENSG00000133135	AF:chrX:105937068:105937638-106016143:105969894:105970627-106016143:+	-0,1568	0,0430
ENSG00000182054	AF:chr15:90634876-90643808:90643927:90634876-90645508:90645736:-	0,1373	0,0440
ENSG00000126562	SE:chr17:40939560-40939796:40939917-40940148:+	0,3177	0,0440
ENSG00000215481	A5:chr22:25041715-25042970:25041712-25042970:+	0,3407	0,0455
ENSG00000158109	SE:chr1:3542079-3542277:3542453-3544064:+	0,1169	0,0460
ENSG00000112245	AF:chr6:64282475:64282624-64286341:64283143:64283518-64286341:+	-0,3787	0,0465
ENSG00000130332	SE:chr19:2324191-2325979:2326138-2328386:-	0,1654	0,0465
ENSG00000112245	SE:chr6:64288933-64289162:64289236-64289962:+	0,1938	0,0465
ENSG00000130332	AF:chr19:2324195-2324741:2324905:2324195-2328386:2328585:-	0,2075	0,0465
ENSG00000169116	SE:chr4:75858580-75937635:75938360-75959093:+	0,0850	0,0480
ENSG00000229194	A5:chrHSCHR6_MHC_MCF:29757725-29758190:29757545-29758190:+	-0,5959	0,0485
ENSG00000065613	SE:chr10:105768114-105770574:105770666-105777918:+	0,1899	0,0485
ENSG00000104112	SE:chr15:51974034-51974714:51974766-51975276:+	0,1346	0,0490
ENSG00000183889	A5:chr16:16429994-16434163:16429849-16434163:+	-0,3965	0,0495
ENSG00000223532	A3:chrHSCHR6_MHC_COX:31312945-31313369:31312927-31313369:-	-0,1915	0,0495
ENSG00000183889	MX:chr16:16427741-16429718:16429849-16434163:16427741-16429909:16429994-16434163:+	0,2789	0,0495
ENSG00000261125	AL:chrHG185_PATCH:39873994:39874278-39882111:39877065:39881406-39882111:-	-0,3188	0,0500
ENSG00000261125	A5:chrHG185_PATCH:39888646-39888947:39888646-39888971:-	-0,2349	0,0500
ENSG00000261125	SE:chrHG185_PATCH:39883396-39884014:39884088-39884453:-	-0,2285	0,0500
ENSG00000125870	A5:chr20:16710809-16712313:16710709-16712313:+	0,1115	0,0504
ENSG00000223606	A5:chrHSCHR6_MHC_DBB:29758059-29758524:29757879-29758524:+	-0,6849	0,0509
ENSG00000099290	SE:chr10:51885209-51885817:51885939-51887343:+	-0,5536	0,0509
ENSG00000147255	SE:chrX:130420022-130420411:130420437-130420579:-	-0,5401	0,0509

ENSG00000230510	AL:chr19:46984053:46984697-47028920:47021936:47023229-47028920:-	-0,3945	0,0519
ENSG00000234539	A3:chrHSCHR6_MHC_COX:32032799-32032939:32032798-32032939:-	-0,3090	0,0519
ENSG00000187094	AF:chr3:42305124-42305315:42305439:42305124-42306189:42306395:-	-0,2324	0,0519
ENSG00000234539	AL:chrHSCHR6_MHC_COX:32014365:32014441-32032715:32031499:32032203-32032715:-	0,3090	0,0519
ENSG00000167842	AL:chr17:5390004-5390321:5390479:5390004-5392143:5394057:+	-0,1733	0,0524
ENSG00000131779	AF:chr1:145516252:145516456-145517273:145516560:145516657-145517273:+	0,1321	0,0524
ENSG00000134313	SE:chr2:8877129-8888017:8888130-8890242:-	0,3711	0,0524
ENSG00000170390	SE:chr4:151119255-151120180:151120230-151124947:+	-0,1657	0,0529
ENSG00000170390	A3:chr4:151125041-151141855:151125041-151141858:+	-0,1526	0,0529
ENSG00000163453	AF:chr4:57907099-57931655:57931769:57907099-57976043:57976551:-	-0,0439	0,0529
ENSG00000132485	SE:chr1:71530820-71531361:71531435-71532459:-	0,1124	0,0529
ENSG00000132485	A5:chr1:71530820-71531361:71530820-71532459:-	0,1257	0,0529
ENSG00000170390	SE:chr4:151170836-151174626:151174708-151177172:+	0,1526	0,0529
ENSG00000176358	A5:chr17:47921519-47925257:47921519-47925275:-	0,5130	0,0534
ENSG00000171863	A3:chr2:3623038-3623182:3623038-3623191:+	-0,1821	0,0539
ENSG00000171863	RI:chr2:3622948:3623038-3623191:3623274:+	-0,1696	0,0539
ENSG00000165688	SE:chr9:139307344-139309005:139309099-139310743:+	0,1085	0,0539
ENSG00000114942	A3:chr2:207024408-207024637:207024408-207024683:+	0,1715	0,0539
ENSG00000107159	SE:chr9:35676203-35676294:35676386-35677787:+	0,1816	0,0544
ENSG00000146731	AF:chr7:56127558:56128109-56128500:56128304:56128371-56128500:+	-0,1488	0,0549
ENSG00000146731	SE:chr7:56120178-56122062:56122196-56123317:+	0,1242	0,0549
ENSG00000101079	SE:chr20:35317187-35335375:35335410-35350082:-	0,1781	0,0549
ENSG00000237727	A5:chrHSCHR6_MHC_COX:31570545-31570712:31570484-31570712:+	-0,2461	0,0554
ENSG00000272886	SE:chrHG957_PATCH:53326843-53338193:53338306-53346257:-	-0,2424	0,0554
ENSG00000160963	A3:chr7:101063380-101090965:101063380-101090971:+	0,3630	0,0554

ENSG00000215077	RI:chrHSCHR6_MHC_QBL:32873080:32873454-32873960:32874088:+	-0,2446	0,0559
ENSG00000234798	RI:chrHSCHR6_MHC_SSTO:31929439:31929653-31929738:31930776:-	-0,2517	0,0564
ENSG00000167523	AL:chr16:89724829-89726948:89727891:89724829-89735694:89736039:+	-0,4256	0,0569
ENSG00000175416	SE:chr5:175819946-175823480:175823533-175824608:-	-0,1371	0,0569
ENSG00000184194	SE:chrX:53078666-53100170:53100292-53105707:+	0,1574	0,0569
ENSG00000167523	AF:chr16:89724220:89724274-89724656:89724279:89724356-89724656:+	0,4718	0,0569
ENSG00000167523	A5:chr16:89724274-89724656:89724243-89724656:+	0,7343	0,0569
ENSG00000125735	A5:chr19:6667460-6669862:6667460-6669970:-	-0,3541	0,0574
ENSG00000132970	MX:chr13:27246126-27250686:27250861-27255191:27246126-27254172:27254338-27255191:+	-0,2064	0,0574
ENSG00000223439	SE:chrHSCHR6_MHC_MANN:29972131-29972274:29972321-29972489:+	-0,2063	0,0574
ENSG00000131795	A3:chr1:145508075-145508207:145508075-145508210:+	-0,2232	0,0579
ENSG00000204348	A3:chr6:31939502-31939706:31939358-31939706:-	-0,2635	0,0584
ENSG00000101335	SE:chr20:35173471-35176435:35176596-35177480:+	0,0834	0,0589
ENSG00000167711	AF:chr17:1646130:1646202-1648286:1646320:1646370-1648286:+	0,1748	0,0589
ENSG00000197766	A5:chr19:859765-860617:859744-860617:+	0,1135	0,0594
ENSG00000263077	AF:chrHG183_PATCH:62582202-62583106:62583316:62582202-62583436:62583649:-	-0,4033	0,0599
ENSG00000263077	AF:chrHG183_PATCH:62582202-62582307:62582764:62582202-62583436:62583649:-	-0,3671	0,0599
ENSG00000263077	RI:chrHG183_PATCH:62577909:62578133-62579370:62579429:-	0,4033	0,0599
ENSG00000079385	A3:chr19:43023387-43025419:43023253-43025419:-	0,4204	0,0599
ENSG00000188153	SE:chrX:107938151-107938497:107938669-107939527:+	-0,3232	0,0604
ENSG00000197859	AF:chr9:136397286:136397692-136401685:136399431:136400039-136401685:+	-0,1631	0,0609
ENSG00000197859	AF:chr9:136397286:136397692-136401685:136399975:136400039-136401685:+	-0,1222	0,0609
ENSG00000161057	SE:chr7:102995396-102996141:102996240-103002404:+	0,1539	0,0609

ENSG00000099960	A3:chr22:21386144-21386820:21386141-21386820:-	0,2141	0,0609
ENSG00000198618	RI:chr21:20230097:20230329-20230438:20230594:+	-0,1904	0,0614
ENSG00000165175	AF:chrX:38660685:38660740-38663936:38662986:38663666-38663936:+	-0,1411	0,0614
ENSG00000101079	SE:chr20:35282104-35283204:35283242-35284763:-	0,1077	0,0614
ENSG00000165175	SE:chrX:38660740-38663563:38663666-38663936:+	0,1401	0,0614
ENSG00000156521	A5:chr10:71903728-71905177:71903728-71906289:-	0,1649	0,0614
ENSG00000214253	AF:chr7:100884187-100884407:100884458:100884187-100888241:100888356:-	-0,4519	0,0617
ENSG00000214253	AF:chr7:100884187-100884407:100884458:100884187-100895176:100895597:-	-0,4302	0,0617
ENSG00000105866	SE:chr7:21468410-21468907:21470461-21516697:+	-0,2118	0,0619
ENSG00000107404	A5:chr1:1274033-1274667:1274033-1274742:-	-0,1137	0,0619
ENSG00000151131	SE:chr12:105382031-105385490:105385626-105388256:+	0,1544	0,0619
ENSG00000132704	SE:chr1:157716689-157718343:157718408-157718665:-	0,6513	0,0619
ENSG00000153086	SE:chr2:135616927-135619539:135619588-135620965:+	0,1937	0,0622
ENSG00000153086	SE:chr2:135596309-135601935:135602041-135602803:+	0,2434	0,0622
ENSG00000100523	SE:chr14:53560162-53569749:53569769-53570401:-	-0,2703	0,0629
ENSG00000135127	SE:chr12:120527875-120528715:120528823-120530804:+	-0,1367	0,0629
ENSG00000135127	SE:chr12:120528823-120529051:120529205-120530804:+	0,2419	0,0629
ENSG00000100523	SE:chr14:53513667-53518562:53518645-53521156:-	0,2438	0,0629
ENSG00000223510	A5:chr17:14139747-14139889:14139747-14140013:-	0,4848	0,0634
ENSG00000124702	SE:chr6:42985433-42985591:42985706-42985883:+	0,1247	0,0637
ENSG00000124702	SE:chr6:42985084-42985257:42985433-42985591:+	0,1292	0,0637
ENSG00000117305	A3:chr1:24134813-24143165:24134705-24143165:-	-0,2249	0,0644
ENSG00000163623	AF:chr4:85416997-85417646:85418103:85416997-85418712:85419603:-	0,1169	0,0644
ENSG00000257961	AF:chrHSCHR17_1:43483448-43502917:43502999:43483448-43506977:43507633:-	0,2697	0,0644
ENSG00000117425	SE:chr1:45286376-45287514:45287576-45288274:-	0,6091	0,0644
ENSG00000265880	SE:chrHG962_PATCH:70874350-70875773:70875912-70879384:+	-0,4011	0,0649

ENSG00000265880	SE:chrHG962_PATCH:70883894-70884516:70884638-70889798:+	-0,3706	0,0649
ENSG00000228049	SE:chr7:102306610-102307537:102307711-102309362:-	-0,2277	0,0649
ENSG00000160563	SE:chr9:134759486-134769272:134769379-134814768:-	0,1052	0,0649
ENSG00000265880	SE:chrHG962_PATCH:70863813-70871837:70871896-70873448:+	0,3706	0,0649
ENSG00000156096	SE:chr4:70359559-70360859:70361128-70391405:-	0,4322	0,0652
ENSG00000156096	AF:chr4:70359559-70360859:70361626:70359559-70391405:70391732:-	0,4333	0,0652
ENSG00000162444	SE:chr1:10057389-10064151:10064365-10067628:+	-0,1502	0,0654
ENSG00000205609	A5:chr16:28414969-28415104:28414969-28415108:-	-0,1779	0,0659
ENSG00000136630	AF:chr1:221051699:221053791-221054536:221054411:221054441-221054536:+	-0,2202	0,0664
ENSG00000160014	AF:chr19:47104331:47104433-47109084:47104624:47104694-47109084:+	-0,5544	0,0666
ENSG00000160014	AF:chr19:47104331:47104433-47109084:47108450:47108562-47109084:+	-0,4778	0,0666
ENSG00000160014	AF:chr19:47104331:47104433-47109084:47104621:47104694-47109084:+	-0,4131	0,0666
ENSG00000144736	A3:chr3:72893604-72897349:72893574-72897349:-	-0,1989	0,0679
ENSG00000142892	SE:chr1:77629627-77632404:77632515-77634945:-	-0,1161	0,0684
ENSG00000173281	AF:chr8:8999178-9008073:9008206:8999178-9008857:9009084:-	0,1572	0,0684
ENSG00000102144	AF:chrX:77359671:77359902-77365364:77361859:77362168-77365364:+	-0,1492	0,0687
ENSG00000102144	AF:chrX:77320713:77320769-77365364:77359671:77359902-77365364:+	0,1405	0,0687
ENSG00000182993	AF:chr12:14956600:14956685-15056941:15038650:15038740-15056941:+	-0,2945	0,0689
ENSG00000234846	A3:chrHSCHR6_MHC_SSTO:31821858-31822285:31821117-31822285:-	-0,2689	0,0689
ENSG00000183160	SE:chr12:108986173-108987940:108988321-108991746:-	-0,2622	0,0689
ENSG00000234846	SE:chrHSCHR6_MHC_SSTO:31821117-31821666:31821858-31822285:-	-0,2352	0,0689
ENSG00000182993	AL:chr12:14956685-14975846:14977349:14956685-15056941:15057544:+	0,2045	0,0689

ENSG00000234846	A3:chrHSCHR6_MHC_SSTO:31819931-31820106:31819845-31820106:-	0,2515	0,0689
ENSG00000137285	AF:chr6:3226903-3227721:3227969:3226903-3231791:3231964:-	-0,1807	0,0694
ENSG00000117407	AF:chr1:44399484:44399557-44401262:44400059:44400208-44401262:+	-0,5290	0,0699
ENSG00000197734	SE:chr14:78227459-78234795:78234866-78235741:+	-0,3139	0,0699
ENSG00000171566	AF:chr4:155470087-155470130:155470138:155470087-155471452:155471587:-	0,2290	0,0699
ENSG00000171566	AF:chr4:155470087-155470130:155470138:155470087-155471452:155471552:-	0,3447	0,0699
ENSG00000171566	AF:chr4:155470087-155470130:155470138:155470087-155471452:155471535:-	0,3590	0,0699
ENSG00000171566	AF:chr4:155470087-155470130:155470138:155470087-155471452:155471549:-	0,3750	0,0699
ENSG00000117407	SE:chr1:44399557-44399784:44399932-44401262:+	0,4699	0,0699
ENSG00000108187	A3:chr10:70066663-70092541:70066660-70092541:-	-0,3203	0,0704
ENSG00000134884	AF:chr13:107212005-107214180:107214299:107212005-107219921:107219971:-	-0,2024	0,0704
ENSG00000122566	SE:chr7:26237352-26237451:26237486-26240192:-	-0,0495	0,0704
ENSG00000122566	A3:chr7:26231958-26232115:26230748-26232115:-	0,0704	0,0704
ENSG00000155115	AF:chr6:111279763:111280029-111280375:111280131:111280245-111280375:+	0,0931	0,0704
ENSG00000229833	A3:chr19:7694746-7695459:7694746-7695709:+	0,2613	0,0704
ENSG00000110002	SE:chr11:123986189-123987273:123987387-123988204:+	-0,1548	0,0709
ENSG00000151572	SE:chr12:101473053-101477456:101477596-101480438:+	0,4970	0,0709
ENSG00000204071	A3:chrX:101396327-101396686:101396322-101396686:-	-0,2090	0,0714
ENSG00000204071	AF:chrX:101396758-101397325:101397477:101396758-101397674:101397942:-	-0,2090	0,0714
ENSG00000103175	A3:chr16:84360561-84362928:84360561-84362932:+	0,2178	0,0714
ENSG00000153283	SE:chr3:111286494-111296349:111296396-111297874:+	-0,5934	0,0719
ENSG00000187955	AL:chr8:121381724-121382575:121383156:121381724-121383391:121384266:+	-0,4498	0,0719
ENSG00000187955	AL:chr8:121381724-121382976:121383320:121381724-121383391:121384266:+	-0,4118	0,0719

ENSG00000248712	RI:chr11:119063861:119063953-119064476:119064595:-	0,1597	0,0719
ENSG00000184574	AF:chr12:6730630-6740751:6740815:6730630-6745073:6745613:-	0,4248	0,0719
ENSG00000156096	AF:chr4:70359559-70360859:70361579:70359559-70391405:70391732:-	0,4297	0,0719
ENSG00000234465	AF:chr19:44084696:44084739-44085364:44084790:44084821-44085364:+	-0,3724	0,0729
ENSG00000129824	AF:chrY:2709527:2709668-2710206:2709961:2710014-2710206:+	-0,1613	0,0729
ENSG00000101474	SE:chr20:24944658-24945011:24945034-24949528:-	-0,1016	0,0729
ENSG00000101474	SE:chr20:24944658-24949528:24949720-24950162:-	0,0688	0,0729
ENSG00000148358	SE:chr9:132869818-132872649:132872705-132887194:+	0,1679	0,0734
ENSG00000117069	A3:chr1:77510298-77515943:77510298-77515947:+	0,2140	0,0734
ENSG00000168824	A5:chr4:4388900-4389331:4388370-4389331:+	0,3467	0,0734
ENSG00000168824	RI:chr4:4418962:4419654-4420511:4420785:+	0,4159	0,0734
ENSG00000183185	A3:chr3:97731411-97736499:97731265-97736499:-	-0,5418	0,0739
ENSG00000146477	A3:chr6:160858243-160863790:160858243-160863793:+	-0,4212	0,0739
ENSG00000152684	SE:chr5:52084248-52095719:52096954-52097243:+	-0,1046	0,0739
ENSG00000225264	AL:chr7:29638411:29638517-29724389:29690659:29690955-29724389:-	-0,1648	0,0744
ENSG00000152467	SE:chr19:58545529-58546808:58546870-58547338:+	-0,1349	0,0744
ENSG00000112306	A5:chr6:133136363-133137600:133136227-133137600:+	-0,0232	0,0744
ENSG00000106049	SE:chr7:27672064-27689092:27689252-27702317:-	0,0868	0,0744
ENSG00000164070	AF:chr4:128703295:128703812-128715232:128704704:128704835-128715232:+	0,1393	0,0744
ENSG00000157303	SE:chr9:95840275-95841753:95841884-95846819:+	0,4054	0,0744
ENSG00000223887	SE:chrHSCHR6_MHC_DBB:30516224-30519340:30519520-30519846:+	-0,3207	0,0749
ENSG00000134884	SE:chr13:107209479-107209907:107209964-107211780:-	0,0899	0,0749
ENSG00000111615	SE:chr12:75895780-75897684:75897854-75898978:-	0,1106	0,0749
ENSG00000184986	AF:chr14:105992940:105992990-105995059:105993442:105993775-105995059:+	0,1846	0,0749
ENSG00000184986	AF:chr14:105992940:105992990-105995059:105994961:105994998-105995059:+	0,2000	0,0749

ENSG00000223887	MX:chrHSCHR6_MHC_DBB:30515462-30516162:30516224-30519846:30515462-30519340:30519520-30519846:+	0,2780	0,0749
ENSG00000223887	SE:chrHSCHR6_MHC_DBB:30515462-30516162:30516224-30519846:+	0,3413	0,0749
ENSG00000137819	A5:chr15:69629844-69652305:69629840-69652305:+	-0,5608	0,0759
ENSG00000088325	SE:chr20:30366787-30368742:30368849-30370052:+	-0,2883	0,0759
ENSG00000197283	A3:chr6:33412394-33414352:33412394-33414358:+	-0,2243	0,0759
ENSG00000197283	SE:chr6:33409536-33410230:33410271-33410666:+	-0,2243	0,0759
ENSG00000138463	SE:chr3:122514382-122525704:122525797-122545647:+	-0,1467	0,0759
ENSG00000181444	A3:chr7:149463328-149466179:149461767-149466179:-	0,0887	0,0759
ENSG00000144744	SE:chr3:69127069-69129263:69129304-69129485:-	0,1178	0,0759
ENSG00000100890	SE:chr14:35592116-35595941:35595988-35596685:+	0,1449	0,0759
ENSG00000234539	A3:chrHSCHR6_MHC_COX:32033365-32033558:32033342-32033558:-	0,1501	0,0759
ENSG00000144744	SE:chr3:69120768-69124581:69124661-69126949:-	0,1826	0,0759
ENSG00000197283	AL:chr6:33415710-33416566:33416830:33415710-33419537:33421466:+	0,3268	0,0759
ENSG00000137819	AF:chr15:69591286:69591395-69629680:69606742:69607133-69629680:+	0,5608	0,0759
ENSG00000135778	AL:chr1:233105864-233112048:233112205:233105864-233113909:233119628:+	0,4058	0,0764
ENSG00000135778	AL:chr1:233105864-233112048:233112205:233105864-233113909:233114468:+	0,5808	0,0764
ENSG00000164615	SE:chr5:134074482-134076753:134077213-134079677:+	0,0629	0,0769
ENSG00000258145	A5:chrHSCHR6_MHC_MANN:32768243-32768284:32768243-32768323:-	0,1142	0,0769
ENSG00000160014	AF:chr19:47104331:47104433-47109084:47104493:47104694-47109084:+	-0,2146	0,0774
ENSG00000197734	A5:chr14:78234866-78235741:78234808-78235741:+	0,3156	0,0774
ENSG00000142484	RI:chr17:4685798:4685934-4686149:4686332:+	-0,1553	0,0784
ENSG00000096696	A5:chr6:7581802-7582875:7580005-7582875:+	-0,0960	0,0784
ENSG00000138660	SE:chr4:113182018-113184144:113184242-113186165:+	0,1041	0,0784
ENSG00000174576	SE:chr11:66190412-66190594:66190703-66191049:+	-0,2200	0,0789

ENSG00000240253	SE:chr2:131452907-131453040:131453164-131453915:+	0,2130	0,0789
ENSG00000214253	AF:chr7:100884187-100884407:100884458:100884187-100888179:100888329:-	-0,4919	0,0791
ENSG00000109270	SE:chr4:100813191-100815112:100815157-100815492:-	-0,0695	0,0794
ENSG00000109270	A3:chr4:100806832-100808453:100806811-100808453:-	0,0821	0,0794
ENSG00000180573	A5:chr6:26124977-26138282:26124861-26138282:+	0,1124	0,0794
ENSG00000152932	AF:chr5:57878048:57878241-57913470:57878867:57879059-57913470:+	0,1140	0,0794
ENSG00000110002	A3:chr11:123987387-123988074:123987387-123988204:+	0,1415	0,0794
ENSG00000064489	SE:chr19:19256831-19257082:19257193-19257364:-	0,3104	0,0794
ENSG00000198053	AF:chr20:1875154:1875163-1876085:1875429:1875442-1876085:+	-0,1755	0,0799
ENSG00000198053	A5:chr20:1915412-1917966:1915400-1917966:+	-0,1474	0,0799
ENSG00000099290	SE:chr10:51885939-51886920:51886982-51887343:+	0,1850	0,0802
ENSG00000258986	AL:chr14:104941015:104941062-105061502:105059330:105059995-105061502:-	-0,0595	0,0804
ENSG00000182916	A3:chrX:102585232-102585851:102585232-102585905:+	0,1766	0,0804
ENSG00000169967	AF:chr2:128100770-128145028:128145508:128100770-128145824:128146041:-	-0,1159	0,0809
ENSG00000145293	A5:chr4:83352066-83369073:83352037-83369073:+	0,0790	0,0809
ENSG00000145293	SE:chr4:83369174-83372196:83372398-83375875:+	0,1063	0,0809
ENSG00000117862	RI:chr1:52485803:52486077-52486598:52486684:-	0,1468	0,0809
ENSG00000182180	AL:chr10:75008696:75009076-75011521:75010502:75010749-75011521:-	-0,2309	0,0813
ENSG00000182180	AF:chr10:75011781-75012017:75012162:75011781-75012290:75012392:-	-0,2076	0,0813
ENSG00000182180	A5:chr10:75011781-75012228:75011781-75012290:-	-0,1615	0,0813
ENSG00000148344	SE:chr9:132511016-132511800:132511921-132515166:-	-0,3420	0,0814
ENSG00000172062	SE:chr5:70237335-70238185:70238385-70238545:+	-0,1068	0,0819
ENSG00000172062	A3:chr5:70242003-70247768:70242003-70248266:+	0,1270	0,0819
ENSG00000204463	AF:chr6:31619553-31620024:31620170:31619553-31620201:31620428:-	-0,1652	0,0824
ENSG00000170185	SE:chr4:144124719-144125570:144125685-144127186:+	-0,1370	0,0824

ENSG00000063245	SE:chr19:56200360-56200663:56200737-56201233:+	0,0658	0,0824
ENSG00000063245	A3:chr19:56204162-56204317:56204162-56204320:+	0,0663	0,0824
ENSG00000260952	SE:chrHG75_PATCH:34814155-34814305:34814336-34820183:+	-0,0800	0,0829
ENSG00000124721	SE:chr6:38834453-38834545:38834652-38835838:+	0,0057	0,0829
ENSG00000099290	AF:chr10:51829270:51829471-51838435:51837886:51837912-51838435:+	0,1456	0,0829
ENSG00000125144	A3:chr16:56701292-56701878:56701289-56701878:-	0,2866	0,0829
ENSG00000134258	SE:chr1:117695991-117699196:117699543-117712729:-	0,6147	0,0829
ENSG00000214253	AF:chr7:100884187-100884407:100884458:100884187-100887838:100888012:-	-0,5792	0,0830
ENSG00000105371	A3:chr19:10398082-10398212:10398082-10398289:+	-0,3874	0,0834
ENSG00000204642	SE:chr6:29693340-29693788:29693820-29694660:+	-0,2351	0,0834
ENSG00000204642	A3:chr6:29693083-29693224:29693083-29693229:+	-0,1846	0,0834
ENSG00000095970	A3:chr6:41126844-41127530:41126804-41127530:-	-0,2118	0,0839
ENSG00000226232	A3:chr16:70016448-70016561:70016397-70016561:-	-0,0895	0,0839
ENSG00000226232	AL:chr16:70010291:70010635-70011980:70011509:70011906-70011980:-	-0,0844	0,0839
ENSG00000226232	AF:chr16:70020412-70023859:70023987:70020412-70029972:70030091:-	0,0895	0,0839
ENSG00000122035	AF:chr13:27844464:27845205-27845621:27845314:27845341-27845621:+	0,1466	0,0839
ENSG00000095970	SE:chr6:41126518-41126611:41126804-41127530:-	0,1834	0,0839
ENSG00000181392	A5:chr19:36496339-36499119:36496339-36499456:-	0,3803	0,0839
ENSG00000197442	SE:chr6:136935424-136943986:136944119-136958463:-	-0,1182	0,0844
ENSG00000183160	A5:chr12:108986173-108987940:108986173-108988113:-	-0,2660	0,0847
ENSG00000177045	A3:chr19:46270413-46271300:46269369-46271300:-	-0,2965	0,0849
ENSG00000177045	SE:chr19:46269369-46269608:46270413-46271300:-	-0,1587	0,0849
ENSG00000110107	SE:chr11:60665423-60665573:60665743-60666013:-	0,1282	0,0849
ENSG00000107140	SE:chr9:35606980-35607324:35607406-35607579:+	0,0718	0,0854
ENSG00000107140	A5:chr9:35607406-35607579:35606980-35607579:+	0,0952	0,0854
ENSG00000147044	SE:chrX:41413168-41414853:41414888-41416285:-	-0,5175	0,0857

ENSG00000147044	MX:chrX:41413168-41414853:41414888-41419032:41413168-41416285:41416353-41419032:-	-0,3505	0,0857
ENSG00000105583	A3:chr19:12779431-12779942:12779425-12779942:-	0,0334	0,0859
ENSG00000122641	AF:chr7:41730140-41730787:41732162:41730140-41739585:41740207:-	0,3735	0,0859
ENSG00000080618	SE:chr13:46638876-46641442:46641552-46648008:-	0,4437	0,0859
ENSG00000172922	AL:chr11:65482367:65482788-65487152:65486007:65486240-65487152:-	-0,1824	0,0862
ENSG00000172922	AL:chr11:65482367:65482788-65487152:65486238:65486627-65487152:-	-0,1534	0,0862
ENSG00000178922	SE:chr1:43917990-43918574:43918675-43919084:-	0,1373	0,0864
ENSG00000144229	SE:chr2:138208593-138237045:138237053-138320791:+	-0,5228	0,0869
ENSG00000119523	SE:chr9:101981118-101983261:101983419-101983829:-	-0,0940	0,0869
ENSG00000119523	AF:chr9:101981118-101983261:101983329:101981118-101983829:101984238:-	-0,0683	0,0869
ENSG00000118418	A5:chr6:79911443-79911993:79911443-79912034:-	-0,0498	0,0869
ENSG00000115828	SE:chr2:37571994-37579932:37580078-37586723:+	0,0909	0,0869
ENSG00000011485	SE:chr19:46879831-46886668:46886733-46887037:+	0,1320	0,0869
ENSG00000183145	AF:chr21:38378502:38378533-38380457:38378863:38379176-38380457:+	-0,2968	0,0874
ENSG00000183145	SE:chr21:38380523-38385851:38385918-38390174:+	0,3283	0,0874
ENSG00000110786	A3:chr11:18764976-18765553:18764880-18765553:-	0,4887	0,0874
ENSG00000171189	SE:chr21:30963545-30968846:30968890-30971150:-	0,5215	0,0874
ENSG00000097007	A3:chr9:133589842-133729451:133589842-133729454:+	-0,3292	0,0879
ENSG00000149591	AF:chr11:117070971:117071058-117073718:117072466:117072657-117073718:+	-0,1987	0,0879
ENSG00000149591	AF:chr11:117070110:117070545-117073718:117072466:117072657-117073718:+	-0,1424	0,0879
ENSG00000172115	A5:chr7:25163746-25164615:25163746-25164819:-	0,0877	0,0879
ENSG00000236882	SE:chr5:95188418-95188496:95188630-95192582:+	0,1482	0,0879
ENSG00000172115	A3:chr7:25164463-25164819:25164434-25164819:-	0,2000	0,0879
ENSG00000141560	A5:chr17:80678841-80680680:80678717-80680680:+	0,2698	0,0879
ENSG00000180113	SE:chr6:46665824-46668736:46668799-46669595:+	0,3383	0,0879

ENSG00000139329	A3:chr12:91502777-91505175:91502464-91505175:-	-0,1223	0,0884
ENSG00000139329	A3:chr12:91502464-91505175:91502060-91505175:-	0,0775	0,0884
ENSG00000132475	AF:chr17:73775265-73775623:73775684:73775265-73775738:73775859:-	-0,3202	0,0889
ENSG00000132475	AF:chr17:73775265-73775623:73775684:73775265-73775738:73775839:-	-0,1978	0,0889
ENSG00000132475	AF:chr17:73775265-73775623:73775684:73775265-73775738:73775860:-	-0,0897	0,0889
ENSG00000138085	A5:chr2:27435475-27436045:27435335-27436045:+	0,0422	0,0889
ENSG00000132475	A5:chr17:73775265-73775728:73775265-73775738:-	0,0847	0,0889
ENSG00000099940	AF:chr22:21213271:21213635-21224625:21213771:21213928-21224625:+	0,1057	0,0889
ENSG00000157593	SE:chr6:44223381-44224079:44224233-44224422:-	0,1414	0,0889
ENSG00000132475	AF:chr17:73775265-73775738:73775859:73775265-73781457:73781567:-	0,1591	0,0889
ENSG00000132475	AF:chr17:73775265-73775728:73775861:73775265-73781457:73781567:-	0,1971	0,0889
ENSG00000189423	SE:chr17:20320004-20321496:20321704-20323692:+	0,2736	0,0889
ENSG00000157593	MX:chr6:44223381-44224079:44224233-44225137:44223381-44224422:44224615-44225137:-	0,3107	0,0889
ENSG00000168811	SE:chr3:159710912-159711238:159711279-159711355:+	0,5248	0,0889
ENSG00000265685	AF:chrHG871_PATCH:5077546:5077808-5138602:5136610:5136720-5138602:+	-0,6112	0,0892
ENSG00000265685	AF:chrHG871_PATCH:5077546:5077808-5138602:5135985:5136720-5138602:+	-0,2281	0,0892
ENSG00000153923	SE:chr1:87111837-87112846:87112905-87113000:+	-0,4667	0,0894
ENSG00000164761	SE:chr8:119941168-119942727:119943007-119945170:-	-0,2667	0,0894
ENSG00000184886	AF:chr17:34890847:34890881-34892943:34891403:34891449-34892943:+	-0,1537	0,0894
ENSG00000162664	A3:chr1:90470803-90472904:90470803-90473171:+	0,1128	0,0894
ENSG00000145365	AF:chr4:113199590-113202837:113202929:113199590-113206796:113207059:-	0,1632	0,0894
ENSG00000184886	A5:chr17:34891449-34892943:34891440-34892943:+	0,1870	0,0894
ENSG00000050555	A3:chr9:133957548-133960911:133957548-133960995:+	0,2916	0,0894

ENSG00000117560	SE:chr1:172628689-172629235:172629280-172633474:+	0,5805	0,0894
ENSG00000013619	SE:chrX:149639762-149641952:149642074-149671544:+	-0,2248	0,0899
ENSG00000197694	SE:chr9:131371563-131371930:131371944-131373993:+	-0,1459	0,0899
ENSG00000254647	A5:chr11:2182218-2182372:2182218-2182398:-	0,1417	0,0899
ENSG00000144837	A5:chr3:119327794-119328315:119327746-119328315:+	0,2006	0,0899
ENSG00000175894	AF:chr21:45987889-45993657:45993693:45987889-46131348:46131495:-	-0,1953	0,0904
ENSG00000174938	AF:chr16:29908442-29909174:29909249:29908442-29910262:29910430:-	-0,4009	0,0909
ENSG00000153446	SE:chr16:5105351-5105688:5105783-5106054:-	-0,3728	0,0909
ENSG00000153446	A3:chr16:5094837-5097879:5094479-5097879:-	-0,2262	0,0909
ENSG00000076248	SE:chr12:109537089-109539707:109539804-109540644:+	0,0908	0,0909
ENSG00000167562	SE:chr19:53073583-53075490:53075621-53077330:+	0,1820	0,0909
ENSG00000177628	SE:chr1:155208441-155209677:155209868-155210421:-	0,2257	0,0909
ENSG00000161583	A3:chr17:34807804-34807948:34807733-34807948:-	0,2294	0,0909
ENSG00000149089	SE:chr11:34916657-34918308:34918402-34937775:-	-0,0824	0,0914
ENSG00000151729	A3:chr4:186066404-186066509:186066404-186066913:+	0,0502	0,0914
ENSG00000149089	SE:chr11:34912099-34916557:34916657-34937775:-	0,0884	0,0914
ENSG00000213079	SE:chr6:155055064-155063090:155063152-155095123:+	-0,1228	0,0919
ENSG00000213079	A3:chr6:155054637-155054929:155054637-155054981:+	0,1436	0,0919
ENSG00000261361	A5:chrHG79_PATCH:136232769-136233246:136232717-136233246:+	0,0884	0,0924
ENSG00000102024	AF:chrX:114795509:114795587-114844555:114827819:114827944-114844555:+	0,4096	0,0924
ENSG00000215513	A3:chr22:20397288-20397878:20397240-20397878:-	-0,2974	0,0929
ENSG00000167088	A3:chr18:19202762-19203709:19202762-19203836:+	-0,1080	0,0929
ENSG00000188612	SE:chr17:73164496-73170847:73170918-73177152:-	-0,0562	0,0929
ENSG00000188612	AF:chr17:73177283-73178477:73178658:73177283-73178909:73179078:-	0,0447	0,0929
ENSG00000188612	AF:chr17:73177283-73178477:73178658:73177283-73178909:73179051:-	0,0632	0,0929
ENSG00000143612	SE:chr1:154187050-154192312:154192413-154192818:-	0,0761	0,0929

ENSG00000143612	SE:chr1:154185100-154186369:154186422-154186933:-	0,0842	0,0929
ENSG00000241399	SE:chr2:160634475-160636516:160636689-160637393:-	0,2226	0,0929
ENSG00000162642	AL:chr1:85715639:85718385-85724207:85723126:85723154-85724207:-	0,3009	0,0929
ENSG00000146938	SE:chrX:5947473-5950732:5950791-6069036:-	0,3051	0,0929
ENSG00000146938	AF:chrX:6069812-6145730:6145888:6069812-6146582:6146904:-	0,3133	0,0929
ENSG00000101638	A3:chr18:44268882-44272132:44268830-44272132:-	-0,2198	0,0934
ENSG00000205078	A3:chr16:77246543-77246744:77246543-77246754:+	0,0829	0,0934
ENSG00000238227	AF:chr9:139008870-139009712:139010120:139008870-139010331:139010731:-	-0,2550	0,0935
ENSG00000238227	AF:chr9:139008870-139009918:139010284:139008870-139010331:139010731:-	-0,2192	0,0935
ENSG00000238227	AF:chr9:139008870-139009918:139010284:139008870-139010331:139010709:-	-0,1991	0,0935
ENSG00000238227	AF:chr9:139008870-139009712:139010120:139008870-139010331:139010709:-	-0,1932	0,0935
ENSG00000172260	AF:chr1:72400994-72566428:72566613:72400994-72748002:72748417:-	-0,2573	0,0937
ENSG00000172260	SE:chr1:72076829-72163691:72163822-72241855:-	0,2935	0,0937
ENSG00000146070	AF:chr6:46690662-46702917:46703079:46690662-46703287:46703430:-	-0,3032	0,0939
ENSG00000143443	AL:chr1:151021328-151022057:151022189:151021328-151022914:151023896:+	-0,0940	0,0939
ENSG00000112308	AF:chr6:24716552-24720088:24720226:24716552-24720446:24720620:-	-0,2541	0,0944
ENSG00000168827	SE:chr3:158402457-158407952:158408112-158408928:+	-0,2203	0,0944
ENSG00000111412	SE:chr12:117158252-117160872:117161028-117175595:-	0,1443	0,0949
ENSG00000111412	AF:chr12:117155698-117157568:117157932:117155698-117175595:117175641:-	0,2634	0,0949
ENSG00000188820	SE:chr6:116782592-116783035:116783617-116784446:+	-0,1842	0,0954
ENSG00000188820	AF:chr6:116782533:116782592-116784446:116783630:116783683-116784446:+	0,3076	0,0954
ENSG00000178233	AL:chr6:44241243-44243140:44247181:44241243-44274680:44275243:+	-0,1913	0,0959

ENSG00000170584	AF:chr5:162884616-162886001:162886496:162884616-162886868:162887061:-	-0,1862	0,0959
ENSG00000170584	AF:chr5:162884616-162886333:162886842:162884616-162886868:162887061:-	-0,1735	0,0959
ENSG00000183495	A3:chr12:132446499-132464239:132446499-132464242:+	-0,1696	0,0959
ENSG00000183495	SE:chr12:132464338-132466034:132466141-132466638:+	-0,1614	0,0959
ENSG00000170584	A3:chr5:162884086-162884568:162884011-162884568:-	-0,1326	0,0959
ENSG00000171540	AL:chr5:76924538:76926619-76932646:76931756:76932537-76932646:-	0,0000	0,0959
ENSG00000110700	A3:chr11:17099024-17099166:17099020-17099166:-	0,0714	0,0959
ENSG00000215845	AF:chr1:161007865-161008341:161008519:161007865-161008670:161008767:-	0,1076	0,0959
ENSG00000215845	A3:chr1:161008463-161008670:161007865-161008670:-	0,1371	0,0959
ENSG00000183495	SE:chr12:132504763-132505624:132505866-132508322:+	0,1442	0,0959
ENSG00000215845	SE:chr1:161007865-161008341:161008463-161008670:-	0,1499	0,0959
ENSG00000101134	A5:chr20:53092551-53171472:53092443-53171472:+	-0,2429	0,0964
ENSG00000168065	SE:chr11:64329907-64331780:64331900-64332694:+	-0,3563	0,0969
ENSG00000005075	A5:chr7:102114131-102114778:102114131-102114808:-	0,0451	0,0969
ENSG00000165494	SE:chr11:82868673-82872369:82872494-82874721:+	0,1460	0,0969
ENSG00000137574	SE:chr8:56686278-56695307:56695371-56698278:+	0,1500	0,0969
ENSG00000117525	SE:chr1:94996152-94997877:94998036-94998646:-	0,1515	0,0969
ENSG00000168490	SE:chr8:22085899-22086301:22086422-22089309:-	0,2672	0,0969
ENSG00000168995	SE:chr19:51646059-51647663:51647941-51648169:+	0,3893	0,0969
ENSG00000142875	SE:chr1:84640739-84641482:84641490-84644860:+	-0,4134	0,0979
ENSG00000171160	SE:chr10:99376168-99376435:99376544-99376965:-	0,1545	0,0979
ENSG00000107130	AF:chr9:132934857:132935006-132963237:132962872:132962890-132963237:+	0,1893	0,0979
ENSG00000171476	A5:chr4:57522178-57522506:57522178-57522609:-	0,1999	0,0979
ENSG00000171476	A5:chr4:57522178-57522466:57522178-57522609:-	0,2335	0,0979
ENSG00000234495	AL:chrHSCHR6_MHC_APD:28879088:28879500-28887794:28882558:28883072-28887794:-	0,2840	0,0979
ENSG00000171476	A5:chr4:57522178-57522510:57522178-57522609:-	0,3137	0,0979

ENSG00000171476	AF:chr4:57522178-57522466:57522673:57522178-57523941:57524072:-	0,4526	0,0979
ENSG00000171476	AF:chr4:57522178-57522466:57522673:57522178-57547421:57548065:-	0,4704	0,0979
ENSG00000229833	SE:chr19:7694746-7695459:7695545-7695709:+	0,0665	0,0982
ENSG00000150459	A3:chr13:21721066-21721306:21721066-21721325:+	-0,0386	0,0984
ENSG00000150459	SE:chr13:21715134-21720944:21721066-21721325:+	0,0517	0,0984
ENSG00000128242	AF:chr22:30953388-30956728:30956763:30953388-30970453:30970565:-	-0,3944	0,0989
ENSG00000133477	AF:chr22:40390953:40391535-40415172:40405692:40406128-40415172:+	-0,1218	0,0989
ENSG00000169710	A3:chr17:80038466-80038568:80038411-80038568:-	-0,0618	0,0989
ENSG00000259785	SE:chr15:82596742-82597088:82597212-82598071:+	0,0000	0,0989
ENSG00000038219	SE:chr4:13571752-13574326:13574466-13578462:-	0,1333	0,0989
ENSG00000117461	SE:chr1:46512297-46521467:46521643-46527601:-	0,1989	0,0989
ENSG00000137731	AF:chr11:117693432-117695369:117695459:117693432-117698710:117698807:-	0,2270	0,0989
ENSG00000117461	A5:chr1:46597628-46598371:46597628-46598456:-	0,4056	0,0989
ENSG00000228049	A3:chr7:102307711-102309362:102306610-102309362:-	-0,0744	0,0994
ENSG00000125931	AF:chrX:71522784-71525675:71525764:71522784-71526327:71526837:-	-0,5359	0,0999
ENSG00000176845	AF:chr17:81037567:81037861-81042814:81038075:81038105-81042814:+	-0,1040	0,0999
ENSG00000166922	AL:chr15:32976870-32983911:32984505:32976870-32988715:32989090:+	0,1865	0,0999
ENSG00000166922	AL:chr15:32976870-32983911:32984505:32976870-32988715:32989289:+	0,2760	0,0999
ENSG00000079263	A5:chr2:231109795-231110578:231109786-231110578:+	0,4186	0,0999
ENSG00000079263	SE:chr2:231110655-231112631:231112780-231113600:+	0,4559	0,0999

APPENDIX 3. Significative phosphosites protein between QGP-1 *CELF4* silencing cells and its scramble.

Phosphorilated Protein	q-value
TSC2_T1462	0,0005
RPS6KA1_T573	0,0010
RPS6KB2_S411	0,0010
RPS6KB2_T389	0,0012
BAD_S91	0,0021
RPS6KA1_T359	0,0042
PRKCA_T638	0,0109
AKT1_T474	0,0130
RPS6KB2_T229	0,0142
PTEN_S370	0,0163
BAD_S112	0,0198
PPP2CA_Y307	0,0228
AKT_T308	0,0341
AKT1_Y72	0,0436
RPS6KB2_S423	0,0513

APPENDIX 4. Significative phosphosites protein between BON-1 *CELF4* silencing cells and its scramble.

Phosphorilated Protein	q-value
BAD_S134	0,0000
AKT1_T474	0,0000
PFKFB2_S483	0,0000
TSC2_T1462	0,0000
PDPK1_S241	0,0000
EIF2S1_S51	0,0003
AKT1S1_T246	0,0007
BAD_S155	0,0009
EIF4EBP1_T36	0,0021
BAD_S136	0,0026
PRKCA_T638	0,0034
RPS6KB2_S418	0,0056
AKT1_S124	0,0092
PRKAB1_S182	0,0121
AKT2_S474	0,0123
RPS6KB2_T421	0,0158
AKT_T308	0,0169
EIF4EBP1_T70	0,0175
EIF4G1_S1108	0,0178
BAD_S91	0,0225
GSK3A_Y216	0,0245
MTOR_S2448	0,0315
PRKCA_Y657	0,0316
AKT1_T450	0,0399
PTEN_S380	0,0445

APPENDIX 5. Additional clinical characteristics of the FFPE PDAC cohort patients.

Characteristic		Samples n=75 (100%)
Smoker	Yes	28 (37.3)
	Former smoker	11 (14.7)
Drinker		50 (66.7)
Body Mass Index	< 20	10 (13.04)
	20 – 25	45 (60.87)
	25 – 30	16 (21.74)
	> 30	3 (4.35)
Diabetes	Type 1 Diabetes Mellitus	3 (4.2)
	Type 2 Diabetes Mellitus	23 (30.6)
Arterial Hypertension		44 (58.6)
Hypercholesterolemia		44 (58)
Prior Acute Pancreatitis		7 (9.3)
Family Background	Pancreatic cancer	1 (1.3)
	Other cancers	2 (2.7)
Perineural Invasion		56 (74.7)
Linfovascular Invasion	Small blood vessel	22 (29.3)
	Big blood vessel	20 (26.7)
Grading	Grade X	2 (2.7)
	Grade 1	15 (20)
	Grade 2	43 (57.3)
	Grade 3	8 (10.7)
	Grade 4	1 (1.3)

APPENDIX 6. Clinicopathological characteristics of 94 pancreatic adenocarcinoma patients profiled by RNA-Seq

Characteristic	Samples n=94 (100%)
Age	Median 67 (range 37-86)
Sex	
Female	46 (48.9)
Male	48 (51.1)
T stage	
T1	1 (1.1)
T2	2 (2.1)
T3	91 (96.8)
T4	0 (0)
N stage	
N0	12 (12.8)
N1	82 (87.2)
M stage	
M0	91 (96.8)
M1	3 (3.2)
Grading	
G1	4 (4.3)
G2	60 (63.8)
G3	30 (31.9)
Vascular invasion	
Yes	10 (10.6)
No	84 (89.4)
Perineural invasion	
Yes	89 (94.7)
No	5 (5.3)
Fat invasion	
Yes	85 (90.4)
No	9 (9.6)

APPENDIX 7. Specific data for human transcript variants primers used in the validation of SF3B1 expression levels with splicing event patterns of key genes.

	Ensembl id	Gene name	SUPPA_name	dPSI	p-Value	Forward	Reverse
1_AF	ENSG00001531 87	*HNRNPU*	ENSG0000015318 7,20;AF:chr1:2448 62730-244863617:24486 3653:244862730-244863674:24486 4091:-	0,48 437 2	0,02 248	AGCTAGGAG AGGAGAACG GG	CGAGCTCATCTT CCCCTTCC
2_AF	ENSG00001720 71	*EIF2AK3*	ENSG0000017207 1,14;AF:chr2:8861 3853-88624746:886249 29:88613853-88626967:886274 64:-	0,48 487 1	0,01 199	GCTCCCACC TCAGCGAC	GTCTCATCGTCT GGTTCCGG
3_A3	ENSG00001640 39	*BDH2*	ENSG0000016403 9,15;A3:chr4:1030 92718-103095203:10309 2696-103095203:-	0,48 878 4	0,02 098	ACCCGGTGC CTCTTGTTTT A	AACTTCATTGGCA AACTGATCAA
4_A5	ENSG00002240 32	*NA*	ENSG0000022403 2,7;A5:chr5:11216 1295-112161703:11216 0991-112161703:+	0,49 279 1	0,03 397	CATCGACTA TGCCAGGGA GT	GCAGGGCAAGCA TAAAGTCA
5_AF	ENSG00001025 54	*KLF5*	ENSG0000010255 4,14;AF:chr13:730 54976:73055158-73061861:730590 05:73059588-73061861:+	0,49 342 0	0,00 500	CGCTTGGCC TATAACTTG GT	TGGAGAGACTGG GATTGCTT
6_MX	ENSG00001403 91	*TSPAN3*	ENSG0000014039 1,15;MX:chr15:770 54279-77055789:770558 63-77070892:770542 79-77056064:770562 55-77070892:-	0,49 374 0	0,02 697	CCGTGCTGG TCTTTCTCAA C	TCCCAAAACCAC TACAACAACT

7_RI	ENSG00001033 63	*ELOB*	ENSG0000010336 3,15;RI:chr16:2771 414:2771604-2771994:2772102:-	0,50 664 7	0,00 375	GAGACCCTG GCTGAGAAC TT	GTGTCTCTCCCA GTCCTTCC
8_AF	ENSG00001717 29	*TMEM51*	ENSG0000017172 9,14;AF:chr1:1515 2532:15152784-15214895:151537 33:15153954-15214895:+	0,51 832 0	0,02 697	GAACTTCAA AGGGCTGGA CG	GCTCCCATTCCC TCTCTGAG
9_AF	ENSG00001270 22	*CANX*	ENSG0000012702 2,15;AF:chr5:1796 98417:179698603-179705679:17969 8988:179699102-179705679:+	0,51 907 1	0,01 582	TCGGACTCC TACCCCTTTT G	GTTCCTGGAGCC GAGACTT
10_A5	ENSG00001063 99	*RPA3*	ENSG0000010639 9,11;A5:chr7:7639 144-7640320:7639144-7640723:-	0,52 166 7	0,01 249	TTTCAGAGA CAGCGCGAT TG	ATTTCTCGGCAC CAATCAGC
11_A5	ENSG00002037 60	*CENPW*	ENSG0000020376 0,8;A5:chr6:12634 0444-126346205:12634 0399-126346205:+	0,52 650 9	0,00 949	GAGATTCCA TCCCTTCTC GG	GCCAGTACATGC TCCTTGTT
12_SE	ENSG00001408 48	*CPNE2*	ENSG0000014084 8,17;SE:chr16:570 92790-57093950:570940 91-57110708:+	0,52 839 5	0,04 945	TGAAGGCGT GGTTTTATG GC	GTAGGTTCTGGC CACTCACT
13_AF	ENSG00001647 13	*BRI3*	ENSG0000016471 3,10;AF:chr7:9825 2379:98252528-98282351:982816 86:98281937-98282351:+	0,53 092 0	0,01 998	AGTTGCCGC GTTCTCTCT	TATCCCCAGGTA CCCTCTCC
14_AF	ENSG00001601 79	*ABCG1*	ENSG0000016017 9,18;AF:chr21:422 16077:42216188-42225671:422191 18:42219304-42225671:+	0,53 390 5	0,01 499	TTCACCTTC CGCCATGAT CA	ATTCAGCAGGTC CGTCTCAG
15_AL	ENSG00001977 56	*RPL37A*	ENSG0000019775 6,10;AL:chr2:2165 00031-216501341:21650	0,54 160 4	0,01 399	ATGAAGAGA CGAGCTGTG GG	AGCGTTGCATTT GGTCCATT

			1462:216500031-216529270:216529454:+				
16_A3	ENSG00002459 10	*NA*	ENSG00000245910,8;A3:chr8:66922725-66925437:66922392-66925437:-	0,542885	0,02248	CGAAGAGCCGTTAGTCATGC	ATACATGCCGCGTGATCCTA
17_AF	ENSG00001235 62	*MORF4L2*	ENSG00000123562,17;AF:chrX:103685260-103686631:103686705:103685260-103687989:103688064:-	0,547206	0,01399	TGCTTGCTTGGAGATCAGGA	CAGGGAAGGTTCTGCAATCA
18_SE	ENSG00001356 78	*CPM*	ENSG00000135678,12;SE:chr12:68871956-68885792:68885885-68932678:-	0,551704	0,04196	GCTGCGCTGGATTTCAACTA	CCCACAACAAGAACCCACAG
19_A3	ENSG00001705 84	*NUDCD2*	ENSG00000170584,11;A3:chr5:163457080-163457562:163457005-163457562:-	0,555673 1	0,02398	AGCTGATGAGGGAACATGGA	GAAGTCCAACAATTTGCTGCA
20_A5	ENSG00001417 50	*STAC2*	ENSG00000141750,7;A5:chr17:39217173-39217867:39217173-39218069:-	0,555728 2	0,03247	CTCCCCATGCCCAGTCC	CATGTTCCTGGAAGCTGTGC
21_SE	ENSG00001368 95	*GARNL3*	ENSG00000136895,19;SE:chr9:127385145-127387193:127387331-127388904:+	0,555790 2	0,02797	CTGCAGCTGTGAATGAGGTC	GTCTGAGGAGAATTTCGGGC
22_A3	ENSG00001627 04	*ARPC5*	ENSG00000162704,16;A3:chr1:183627594-183630461:183623471-183630461:-	0,560535	0,00599	AGGGTCCATTGTTCGTGTCT	TCCCGAGGCAGATAATCCAC
23_AF	ENSG00001982 42	*RPL23A*	ENSG00000198242,14;AF:chr17:28719985:28720030-28720707:28720299:28720328-28720707:+	0,568048	0,00450	ACCCTTTTCACAAGATGGCG	GTGACGTGCGGATCTTCTTC

24_AF	ENSG00000075426	*FOSL2*	ENSG00000075426,12;AF:chr2:28392802:28392836-28404107:28395511:28395776-28404107:+	0,566878	0,02331	CAGGGCTGGAGAATAAAGAGT	AGGGTATGGGTTGGACATGG
25_SE	ENSG00002298233	*PET100*	ENSG00000229833,10;SE:chr19:7630846-7631232:7631321-7631473:+	0,570591	0,00400	TAAGGAACCGAGAGCAGAGG	TGGGATCCGCTTCACTCTTC
26_AF	ENSG00002400665	*PSMB9*	ENSG00000240065,8;AF:chr6:32844136:32844274-32856138:32854148:32854289-32856138:+	0,570656	0,04570	GCTTCTCTGCTCTCCCGTTA	ATCAGAACCCATCACAACGC
27_AF	ENSG00000011566	*MAP4K3*	ENSG00000011566,15;AF:chr2:39378123-39379752:39379809:39378123-39436892:39437301:-	0,577151	0,01399	CAGGAGGACTTCGAGCTGAT	CATTCCGTGCCTTGTAGACG
28_SE	ENSG00001162209	*TMEM59*	ENSG00000116209,12;SE:chr1:54047372-54048650:54048682-54053000:-	0,583466	0,01049	AGAAGAGGAGTTGTACGCATG	TCCATCATCCACAAACTGACA
29_SE	ENSG00001563681	*ANKRD9*	ENSG00000156381,9;SE:chr14:102508549-102508641:102508846-102509529:-	0,588917	0,02198	ACCCTCCTATCTCCTCCCAG	GTTCCGGGGATGTGGGTATA
30_AF	ENSG00001532550	*RBMS1*	ENSG00000153250,20;AF:chr2:160367391-160416138:160416178:160367391-160493289:160493807:-	0,591632	0,01499	AGCTTCATGGGCAAAGTGTG	AGGTTCGTTTTGCTGAGCTG
31_AL	ENSG00001057550	*ZNF85*	ENSG00000105750,15;AL:chr19:20935047-20942801:20943977:20935047-20945499:20945709:+	0,595843	0,04795	TCTGGAGCAAGGGAAAGAGG	GGAGAATTGCTTGAACCCAGG

32_AF	ENSG00000113387	*SUB1*	ENSG00000113387,12;AF:chr5:32531633:32531916-32588512:32585559:32585625-32588512:+	0,604633	0,01049	ACAGTCCTCTTAGTGCACCA	CTCCACCATCTCTGAGCTGT
33_SE	ENSG00000170906	*NDUFA3*	ENSG00000170906,16;SE:chr19:54106011-54106353:54106446-54106811:+	0,605968	0,00400	CCGTCATGATCAACAAGGCC	ACAATCCTGGCTAACACGGT
34_AF	ENSG00000091490	*SEL1L3*	ENSG00000091490,11;AF:chr4:25847864-25860601:25860631:25847864-25863434:25863537:-	0,608459	0,03497	CGCAAAAGGGAATGTTGTACC	GCTCAGCTTTGGGTATCACTG
35_AL	ENSG00000127022	*CANX*	ENSG00000127022,15;AL:chr5:179726759-179728591:179728907:179726759-179729360:179729649:+	0,619790	0,01424	AGAAGATGGTGGCACTGTCA	TCTATTCCGGAGCTCACGTG
36_A3	ENSG00000230989	*HSBP1*	ENSG00000230989,7;A3:chr16:83809425-83811421:83809425-83817786:+	0,638175	0,00799	AAACAAGATACCTGCCACGC	TCGATCTTCTCTTGGCTTGGA
37_AF	ENSG00000241343	*RPL36A*	ENSG00000241343,10;AF:chrX:101391016:101391046-101391459:101391202:101391235-101391459:+	0,643030	0,00749	GCAAGCATGGTTAACGTCCC	AATCCTTGCCCTTCTTGTACTG
38_SE	ENSG00000241343	*RPL36A*	ENSG00000241343,10;SE:chrX:101391822-101392050:101392095-101395335:+	0,655676	0,00599	GAAAGCGGCGTTATGACAGG	TAGGAAAGCAGGGCACATTC
39_A5	ENSG00000182899	*RPL35A*	ENSG00000182899,17;A5:chr3:197951328-197954003:197951311-197954003:+	0,664481	0,01199	GTAAGTTTATGACACTGCAACACA	CATGGGCCCGAGTTACTTTT

40_AF	ENSG00001600014	*CALM3*	ENSG00000160014,17;AF:chr19:46601074:46601176-46605827:46601276:46601437-46605827:+	0,676117	0,02797	GCGGCGAGGGAAAGTAGT	GAATCTCTCGTCCCCACCC
41_AF	ENSG00001667441	*NNMT*	ENSG00000166741,7;AF:chr11:114295825:114296710-114297951:114297363:114297398-114297951:+	0,676463	0,00300	CCTAGACGGTGTGAAGGGAG	GTCAGTGACGACGATCTCCT
42_A3	ENSG00002238665	*HLA-DPB1*	ENSG00000223865,11;A3:chr6:33080534-33080672:33080534-33080689:+	0,684339	0,00000	TTCTCTCTCTGCGTGGTGAG	TTCTCGAGTTCTGTGGTCCT
43_AF	ENSG00001850888	*RPS27L*	ENSG00000185088,14;AF:chr15:63156521-63157400:63157477:63156521-63157916:63158021:-	0,697218	0,00000	AAGCAAACCTAAGGCACAGC	TGGACTTTGTACTAGGCGTTTC
44_A5	ENSG00001126995	*COX7A2*	ENSG00000112695,11;A5:chr6:75237988-75240277:75237988-75240301:-	0,702383	0,00500	AAAGGGTGGGGTAGCTGATG	CTTCAAACATTCCAAAGGCCTT
45_AL	ENSG00001977756	*RPL37A*	ENSG00000197756,10;AL:chr2:216500031-216501341:216504086:216500031-216529270:216529454:+	0,755854	0,01399	ATGAAGAGACGAGCTGTGGG	AGCGTTGCATTTGGTCCATT
46_SE	ENSG00001977756	*RPL37A*	ENSG00000197756,10;SE:chr2:216499398-216499832:216499866-216499949:+	0,773962	0,01399	ACGTACCAAGAAAGTCGGGA	TCCCTTTCTCCTTTGCCACA