SRRM3 regulates a subprogram of highly sensitive microexons important for pancreatic endocrine function.

Simon Bajew

Department of Experimental and Health Sciences
Universitat Pompeu Fabra

DOCTORAL THESIS / 2021

Thesis supervision

Dr. Juan Valcárcel Juárez

Gene Regulation, Stem Cells and Cancer Department,
Centre for Genomic Regulation

Dr. Manuel Irimia Martínez

System Biology, Centre for Genomic Regulation

ABSTRACT

Abstract

Alternative splicing is a post-transcriptional process that allows the generation of multiple transcript and protein isoforms from a single gene by differential processing of exons and introns. This Thesis focuses on microexons, the shortest class of exons, previously characterised in the nervous system of vertebrates as functionally important and regulated by the protein SRRM4. Here, we provide evidence for a program of microexon regulation in endocrine pancreas that is controlled SRRM3, a paralog of SRRM4 sharing an ancestral domain required for microexon inclusion. We find that endocrine pancreas uses a subset of the neural microexon program, forming a nested program of neuroendocrine microexons. We also show that the correct inclusion of these microexons (EndoMICs) is important for the secretory function of pancreatic islets. Finally, we provide insights into the mechanisms by which the nested program of neuroendocrine microexons is differentially regulated between neural and endocrine pancreatic tissues.

Resumen

El empalme alternativo es un proceso postranscripcional que permite la generación de múltiples isoformas de transcripción y proteína a partir de un solo gen mediante el procesamiento diferencial de exones e intrones. Esta tesis se centra en los microexones, la clase más corta de exones, previamente caracterizados en el sistema nervioso de los vertebrados como funcionalmente importantes y regulados por la proteína SRRM4. Aquí, proporcionamos evidencias para un programa de regulación de microexones en el páncreas endocrino que está controlado por SRRM3, un parálogo de SRRM4 que comparte un dominio ancestral requerido para la inclusión de microexones. Encontramos que el páncreas endocrino utiliza un subconjunto del programa de microexones neurales, formando un programa anidado de microexones neuroendocrinos. También mostramos que la correcta inclusión de estos microexones (EndoMICs) es importante para la función secretora de los islotes pancreáticos. Finalmente, proporcionamos información sobre los mecanismos por los cuales el programa anidado de microexones neuroendocrinos se regula de manera diferencial entre los tejidos neurales y pancreáticos endocrinos.

Resum

L'empalmament alternatiu és un procés posttranscripcional que permet generar múltiples isoformes de transcripció i proteïna a partir d'un sol gen mitjançant el processament diferencial d'exons i introns. Aquesta tesi se centra en els microexons, la classe més curta d'exons, prèviament caracteritzats en el sistema nerviós dels vertebrats com a funcionalment importants i regulats per la proteïna SRRM4. Aquí, proporcionem evidències per a un programa de regulació de microexons al pàncrees endocrí que està controlat per SRRM3, un paràleg de SRRM4 que comparteix un domini ancestral requerit per a la inclusió de microexons. Trobem que el pàncrees endocrí utilitza un subconjunt del programa de microexons neurals, formant un programa niat de microexons neuroendocrins. També mostrem que la correcta inclusió d'aquests microexons (EndoMICs) és important per a la funció secretora dels illots pancreàtics. Finalment, proporcionem informació sobre els mecanismes pels quals el programa imbricat de microexons neuroendocrins es regula de manera diferencial entre els teixits neurals i pancreàtics endocrins.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

PREFACE TO INTRODUCTION 3

1.1. GENE REGULATION VIA ALTERNATIVE SPLICING 4

BRIEFLY ON THE EVOLUTION OF CELL TYPES 4
RNA BIOLOGY IN GENE REGULATION AND EXPRESSION IN EUKARYOTES 7
ALTERNATIVE SPLICING DIVERSIFIES TRANSCRIPTOMES AND PROTEOMES WITH FUNCTIONAL CONSEQUENCES 13

1.2. REGULATION AND FUNCTIONS OF MICROEXONS 17

REGULATION OF NEURAL MICROEXONS 17
FUNCTIONS OF NEURAL MICROEXONS 22

1.3. ENDOCRINE FUNCTION OF THE PANCREAS 24

1.4. THESIS OBJECTIVES 29

CHAPTER 2: AN UNCHARACTERISED PROGRAM OF ENDOCRINE PANCREATIC MICROEXONS IS REGULATED BY SRRM3 31

PREFACE TO CHAPTER 2 33

2.1. CHARACTERISATION OF THE PANCREATIC ENDOCRINE MICROEXON PROGRAM 34

2.2. SRRM3-MEDIATED ALTERNATIVE SPLICING OF ENDOMICS AFFECTS INSULIN SECRETORY CAPACITY OF BETA CELLS 39

2.3. SRRM3-MEDIATED ALTERNATIVE SPLICING IN ISLET PHYSIOLOGY 47

SRRM3-/- MICE DISPLAY SYMPTOMS OF HYPERINSULINEMIC HYPOGLYCEMIA 47
IMPLICATION OF MICROEXON DEREGULATION IN TYPE 2 DIABETES AND FURTHER STUDIES 53

2.4. CLOSING REMARKS 55

2.5. METHODS 56

2.6. TABLES 63

CHAPTER 3: REGULATION OF NEUROENDOCRINE MICROEXONS IS ACHIEVED BY DIFFERENTIAL SENSITIVITY TO SRRM3/4 73

PREFACE TO CHAPTER 3 75

3.1. NEURALMICS AND ENDOMICS ARE REGULATED BY DIFFERENT LEVEL OF SRRM4 EXPRESSION 76

3.2. LIBRARY OF MICRO(&)EXONS TO STUDY THE MODEL OF MICROEXON SENSITIVITY TO SRRM4 79

BARCODED MINIGENE LIBRARY TO STUDY PROPERTIES OF MICROEXONS ON A HIGH-THROUGHPUT SCALE 79
STUDYING THE IMPACT OF EXON LENGTHS ON THE REGULATION BY SRRM4 USING THE LIBRARY OF MICRO(&)EXONS 92

3.3. CLOSING REMARKS 101

3.4. METHODS 102

CHAPTER 4: DISCUSSION 109

4.1. ON THE NON-TISSUE SPECIFICITY OF NEURAL-SPECIFIC MICROEXONS 111

BIOLOGICAL FUNCTIONS OF ENDOMICS 113
BIOLOGICAL FUNCTIONS OF NESTED PROGRAMS OF MICROEXONS 116

4.2. ON THE LESSONS FROM THE LIBRARY OF MICRO(&)EXONS 118

4.3. ON THE LIMITATIONS OF THE LIBRARY OF MICRO(&)EXONS 121

EXOGENOUS SPLICING 121
CONTEXT-SPECIFIC EFFECTS OF EXTENSION SEQUENCES 124

4.4. ON THE FUTURE STUDIES ON CIS-ACTING ELEMENTS AND EVOLUTIONARY CONSERVATION UNDERLYING SRRM3/4 SENSITIVITY 125

LENGTH AND INTRONIC MUTANTS 128
EVOLUTIONARY CONSERVATION 131
VALIDATION OF NOVEL CIS-REGULATORY ELEMENTS UNDERLYING MICROEXON REGULATION 132

4.5. CLOSING REMARKS 133

CONCLUSIONS 135

ACKNOWLEDGMENTS 139

BIBLIOGRAPHY 143

CHAPTER 1

INTRODUCTION

1. Chapter 1: Introduction

Preface to Introduction

In this Chapter, I will introduce the main concepts that in my opinion are required for understanding the basis of this thesis. After briefly introducing the evolution of cell type specificity, I will discuss the role and basic mechanisms of RNA splicing in Eukaryotic gene expression and regulation and introduce microexons to lay the foundations of my thesis work. To lay the other foundations of my work, I will then discuss the basics of endocrine biology with pancreas at its centre, connecting it to the previous sections.

1.1. Gene regulation via alternative splicing

a) Briefly on the evolution of cell types

Evolution of life on Earth had many pivotal moments. Two are arguably among the most important — compartmentation within ancient prokaryotic cells that gave rise to eukaryotes and the multicellularity that most eukaryotic organisms adopted that allowed the acquisition of specialized cell functions. The first event led to the origin of eukaryotic cells and their specialised organelles. Although several hypotheses have been proposed in attempt to explain it, with endosymbiosis now being widely accepted (Figure I-1A, left), the implications of that transition were unprecedented (Cavalier-Smith, 2010; Horiike et al., 2001; Lake, 1989; W. F. Martin et al., 2015; Sagan, 1967). Of all organelles in eukaryotic cells, the nucleus encapsulated molecular processes involved in genomic DNA replication, repair, and transcription, segregating protein synthesis to the cytosol (Figure I-1A, middle). This provided not only physical protection of the main genetic material from chemical and biological agents but also allowed cells to evolve molecular mechanisms dedicated to compartment-specific tasks (Hendrickson & Poole, 2018). Importantly, the nucleus established mechanisms of bidirectional trafficking to control both regulatory layers (Tetenbaum-Novatt & Rout, 2010).

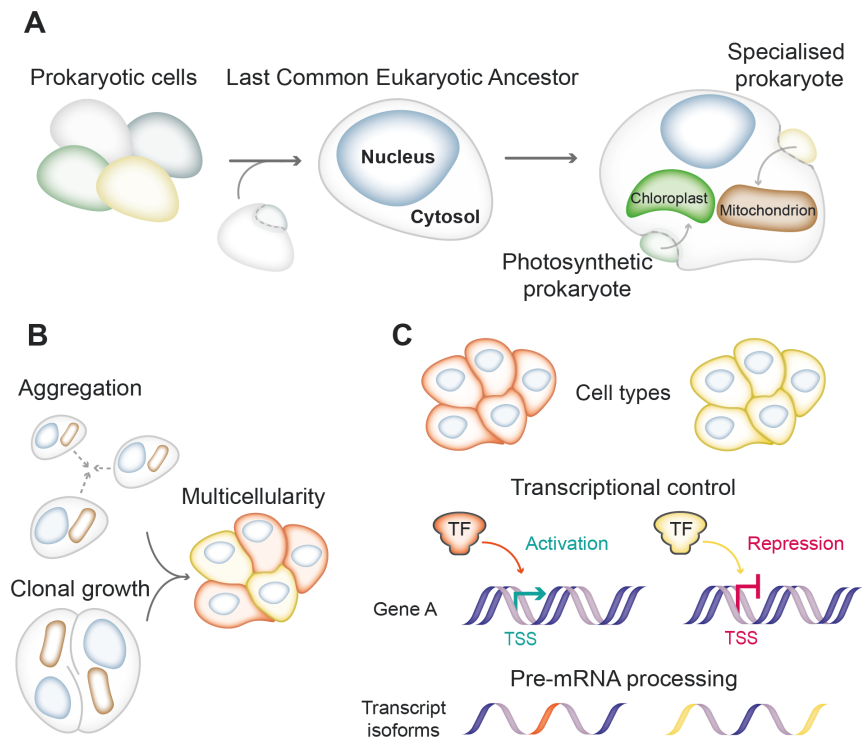

Figure I- 1. Acquisition of nuclei and endosymbiosis led to more complex and compartmented organisation of molecular process within cells. **A)** Schematic of the evolution of eukaryotic cells by endosymbiosis. **B)** Accepted mechanisms of the origin of multicellularity. **C)** Examples of the mechanisms of cell identity specification.

Primitive eukaryotic cells also evolved other organelles, by engulfing either photosynthetic or otherwise specialised prokaryotes (Figure I-1A, right) (Archibald, 2015; Roger et al., 2017; Sagan, 1967). Together with the nucleus, organelles could orchestrate far more complex molecular and cellular processes than prokaryotes and that eventually drove farther specialisation of eukaryotic unicellular organisms. Exposed to environmental, ecological, and genetic pressures, either by

aggregation or clonal growth, unicellular organisms eventually gave rise to multicellular organisms, composed of many cells (Figure I-1B) (Bonner, 1998; King, 2004). One key advantage of multicellularity was the division of tasks within the organism through cellular specialisation (Figure I-1C, top). This functional specialisation can be deployed by several mechanisms. For example, the expression (or not) of cell type-specific transcription factors or repressors, may activate or repress different genetic programs that establish cellular fates (Figure I-1C, middle) (Davis et al., 1987; Lassar et al., 1986). A complementary layer of regulation is also provided by signalling pathways that respond to external and internal cues and propagate responses to the nucleus and other organelles, for example by modulating the activity of transcription regulators (Clevers, 2006). Important for this Thesis, cell function can also be regulated post-transcriptionally at the level of pre-mRNA processing (Figure I-1C, bottom).

First, I need to place pre-mRNA processing in context of the molecular biology of the cell. Although I will focus on one aspect, namely pre-mRNA splicing, it is important to acknowledge that many other RNA processing steps occur in eukaryotic cells, including pre-mRNA capping (addition to N7methyl-Gppp at the 5' end of the transcript), polyadenylation (addition of polyadenosine tail at the 3' end) or RNA editing (chemical modification of bases, for example adenosine to inosine) and that these modifications strongly influence the stability, localisation and fate of the mature mRNAs (Christofi

& Zaravinos, 2019; Elkon et al., 2013; Galloway & Cowling, 2019)

b) RNA biology in gene regulation and expression in Eukaryotes

With the discovery of DNA structure in 1950s (Watson & Crick, 1953), molecular biologists began understanding its role in DNA replication and in storing genetic information. The “central dogma” of molecular biology was proposed to describe a basic flow of genetics information (Figure I-2). At its core, it stated that DNA could be replicated to propagate cells; its content transcribed into a precursor RNA from a basic unit, called a gene; and that the RNA message would be translated into a protein.

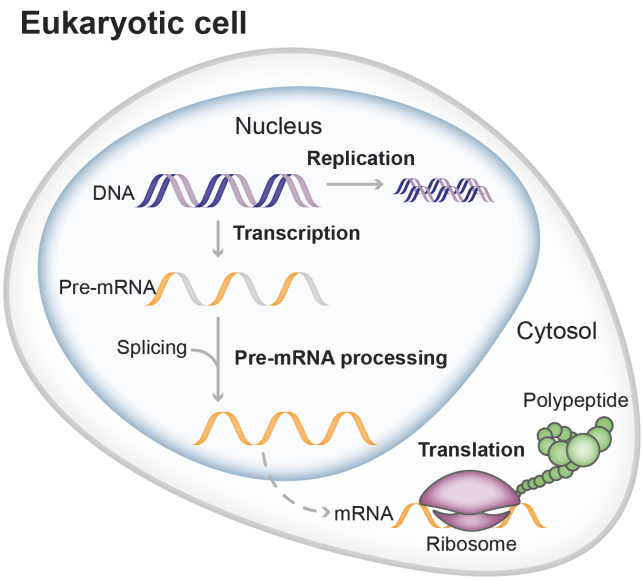

Figure I- 2. "Central dogma in molecular biology" describes basic information flow in living organisms from DNA to protein. Gene regulation and expression was initially proposed as three main processes: DNA replication and transcription to premature messenger RNA (pre-mRNA) occurring in the nucleus, and mature mRNA translation to protein in cytosol. At the level of pre-mRNA, with growing interest and advances in molecular biology, more complex landscape and intertwined relationships were revealed, including but not limited to pre-mRNA capping and RNA splicing.

Although this dogma has been expanded with intertwined processes over the years, one aspect is important for this Thesis in particular.

It was first observed in adenoviruses that genomic DNA and mRNA, when forming a DNA/RNA hybrid, displayed loops of the DNA, implying that internal pieces of the genomic region were not present in the mRNA (Berget et al., 1977; Chow et al., 1977). In the nucleus, "split genes" were shown to have a block-like structure, where sequences that were found in cytoplasmic mRNA, called exons, were interspaced by non-coding sequences, called introns. This discovery implied a necessity of removing introns from nuclear precursor mRNA (pre-mRNA) and the process was called RNA splicing.

It is now well established that RNA splicing requires several sequences, or cis-acting elements, to occur; the 5' splice site (ss) that marks the intron 5' end, the 3'ss at the 3' end of intron, the branch point, a sequence located upstream of 3'ss that included a adenosine involved in the chemical first step of the

splicing reaction and (in higher eukaryotes) a stretch of pyrimidines (Polypyrimidine Tract) between the branchpoint and 3'ss (Figure I-3A). Despite the general lack of conservation of intronic sequences and their length variability, two nucleotides at each of the exon-intron boundaries (GU at 5'ss and AG at 3'ss), as well as a branchpoint adenosine display are required for intron excision across Eukaryotes (Irimia et al., 2007; Irimia & Roy, 2008; Kennedy & Berget, 1997).

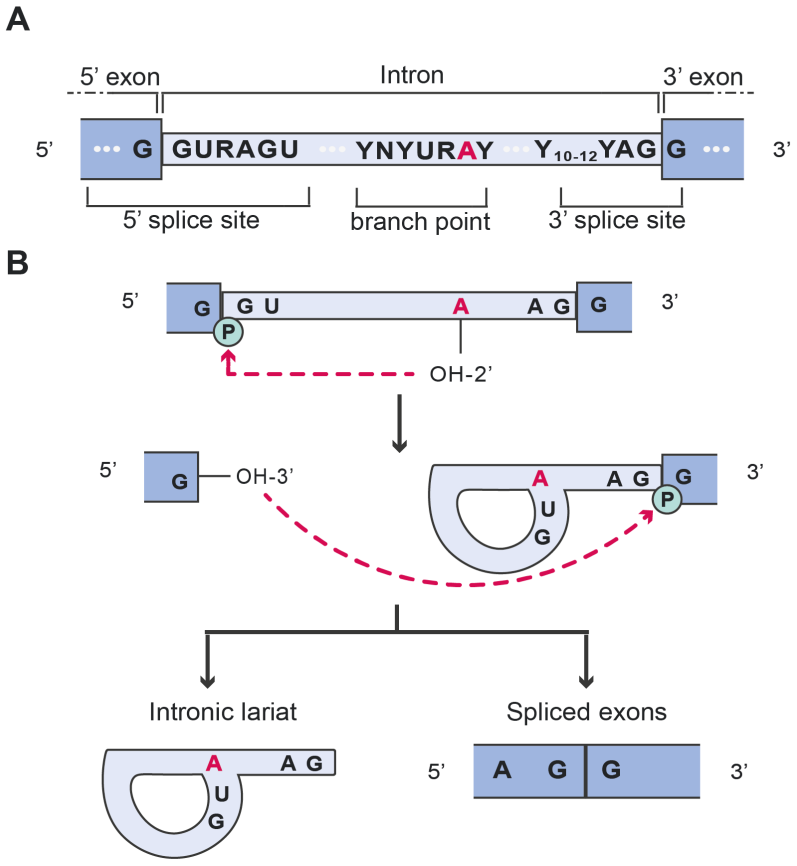

Figure I- 3. Basic pre-mRNA structure and RNA splicing RNA splicing. **A)** Exon-intron boundaries in a gene. Regions and their consensus sequences are highlighted. Adapted from Molecular Biology of The Gene (7th Edition). **B)** Schematic of RNA splicing. Two trans-esterification reactions (1st at the top, 2nd in the middle) lead to excision of intron and ligation of adjacent exons.

From biochemical perspective, RNA splicing is achieved by two subsequent trans-esterification reactions, initiated by a nucleophilic attack of the 2' hydroxyl of the branch point adenosine residue on the phosphodiester bond marking the 5' intron boundary (Figure I-3B, top). During the second reaction, the free 3' hydroxyl group of the 5' exon performs a nucleophilic attack on the phosphoryl group at the 3' intron boundary (Figure I-3B, middle), resulting in the excision of the intron in a lariat configuration and ligation of the exons (Figure I-3B, bottom).

Intron excision is carried out by a macro-molecular complex, called the spliceosome. The major spliceosome that excises most of introns is composed of five small nuclear Ribonucleoprotein complexes (snRNPs) which contain one RNA molecule (U1, U2, U4, U5 and U6 snRNAs) associated with proteins – some common than most snRNPs, some specific of each of them. The dynamic and sequential assembly of these particles at each step of splicing pathway is associated with ATP-dependent structural rearrangements and compositional changes required for recognition of splice sites and trans-esterification reactions (Figure I-4).

During the early spliceosomal formation (Complex E), the 5’ss is recognised by the U1 snRNP through base pairing between the 5' end of its snRNA and the pre-mRNA (typically the last 2 nucleotides of the exon and 6 first nucleotides of the intron) (Du & Rosbash, 2001, 2002; Kaida et al., 2010; Konarska, 1998). U2AF (not shown on Fig. I-4) is a heterodimer of two proteins whose larger (65KDa, U2AF2) subunit binds to polypyrimidine tract and the smaller (35KDa, U2AF1) recognises the 3’ss (Agrawal et al., 2016; Hollins, 2005; Y. Li & Blencowe, 1999; Ruskin et al., 1988; Wang et al., 1995).

The larger subunit of U2AF also helps in the recognition of the branchpoint by the Branchpoint Binding Protein BBP/SF1. U2 snRNP then displaces SF1 and binds to the branch site aided by U2AF. Like the recognition of the 5' ss by U1 snRNP, this involves base pairing interactions, in this case between a region of U2 snRNA known as the branch point recognition sequence and nucleotides flanking the branchpoint adenosine. Assembly of U1 snRNP on the 5' splice site and of U2 snRNP on the branchpoint constitutes Complex A. Splice sites are brought in proximity by the assembly of the tri-snRNP particle (U4/6 – whose snRNAs are extensively base paired with each other – and U5 snRNPs). This association, also known as Complex B, constitutes a pre-catalytic state of the spliceosome.

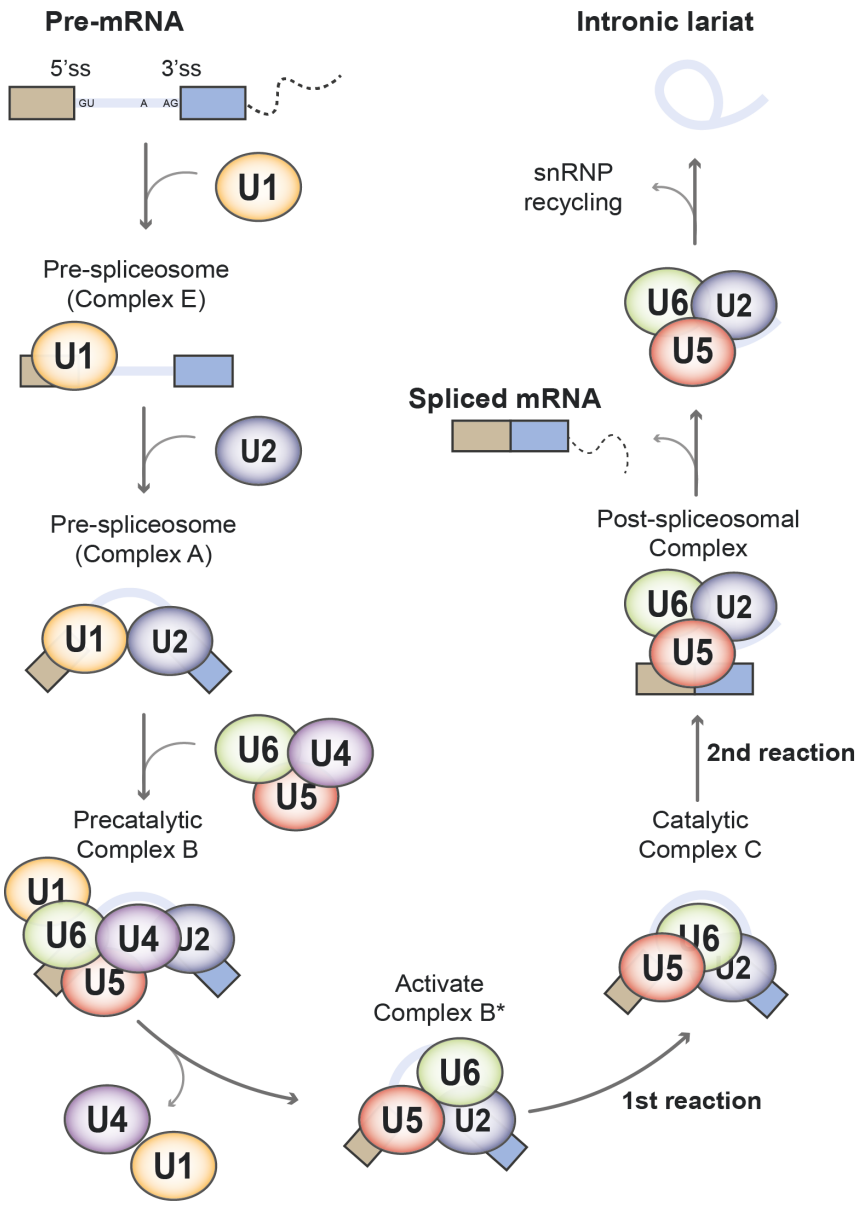

Figure I- 4. RNA splicing is carried out by a macromolecular complex, the spliceosome. The composition of the spliceosome dynamically changes during subsequent formations of catalytically active complexes that facilitate transesterification reactions between the splice sites. Adapted from (Matera & Wang, 2014).

To activate the complex, U6 replaces U1 at the 5' splice site while the release of U4 allows interactions between U2 and U6 that help to shape the catalytic centre of the spliceosome, which is RNA-based and similar in structure to the catalytic core of self-catalytic RNAs (Grabowski & Sharp, 1986; Lamond et al., 1988; van der Veen et al., 1986). The first trans-esterification reaction can now occur between the 5'ss and the branch point and the second catalytic step, involving the 5'ss and 3' ss, results in ligation of exons and release of the intronic lariat structure.

A more comprehensive overview of this pathway provided in (Matera & Wang, 2014; Wahl et al., 2009). Astonishingly, it is estimated that the spliceosome may be composed of over 300 different proteins, which raises an important question – what is the molecular advantage of assembling and coordinating such an enormous complex (Irimia & Blencowe, 2012; Papasaikas & Valcárcel, 2016)?

c) Alternative splicing diversifies transcriptomes and proteomes with functional consequences

The current human genome annotation consists of 20,442 protein coding genes and over 230 thousand transcripts (Ensembl, 2021). With an average of 11 transcripts per gene the transcriptomic diversity within the nucleus is enormous. Among the most studied and understood patterns of isoform generation are selection of alternative 5' or 3' splice sites,

inclusion or skipping of alternative exons and retention of introns (Figure I-5A). With the advent of progressively improving and emerging RNA sequencing technologies, identification and quantification of transcript isoforms, and mechanistic studies of RNA binding proteins, it became evident that the transcriptomic output is strongly influenced by the complexity of alternative splicing pathways. Several studies identified cross-regulatory splicing networks that uncovered complex relationships between RNA-binding proteins and exons or introns that they co-regulate (Brooks et al., 2015; Lareau and Brenner, 2015; Papasaikas et al., 2015; Rösel-Hillgärtner et al., 2013; Saha et al., 2017; Saltzman et al., 2011).

The internal interaction networks of the spliceosome intertwined with other mechanisms of gene regulation allow for cell type-specific alternative splicing programs, which are typically under the control of splicing factors displaying restricted expression in specific cell types or tissues (Figure I-5B). Transcriptomic studies in both vertebrate and invertebrate species identified alternative splicing programs that influence gene expression and regulation and that strongly influence cell fate and development, dictate tissue identity, or contribute to disease phenotype (Baralle & Giudice, 2017; Climente-González et al., 2017).

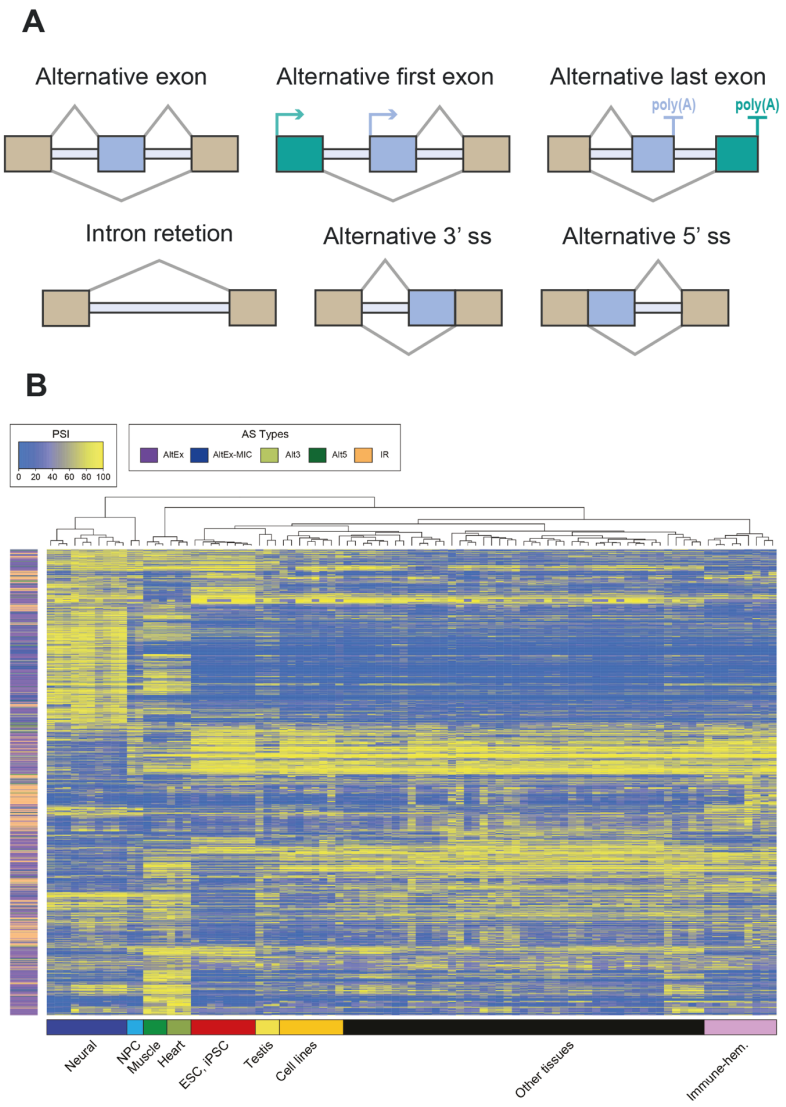

Figure I- 5. Alternative splicing generates multiple isoforms from a single gene, increasing transcriptomic mosaic and proteomic diversity. **A)** Main types of transcript isoforms. **B)** Profiles of alternative splicing events in human transcriptome across selected tissue groups. Adapted from (Tapial et al., 2017a).

For example, in developmental biology, one of the most studied aspects – sex determination and X chromosome dosage compensation in *Drosophila melanogaster* – occurs through a cascade of alternative splicing events in *Sex-lethal* (*Sxl*), *Transformer* (*Tra*) genes, *Doublesex (Dsx*) and *Male-specific-lethal 2 (msl2)* genes, ultimately controlled by specific expression of Sex-lethal protein exclusively in female flies (Förch & Valcárcel, 2003).

Perhaps one of the most compelling studies showcasing the biomedical importance of alternative splicing comes from studies of the genes coding for the protein SMN. *SMN1* gene is mutated in patients with Spinal Muscular Atrophy (Lorson et al., 1999) and a second copy of the gene, *SMN2*, cannot compensate for *SMN1* loss because exon 7 is typically skipped from SMN2 transcripts – but not from SMN1 transcripts – due to a single nucleotide difference at exonic position 6 (Lorson et al., 1999). (Hua et al., 2008; Lorson et al., 1999) identified intronic splicing silencers in flanking introns of exon 7 in *SMN2* gene that represses inclusion of that exon. SMN2 is thus a clinically relevant gene that influences the severity and prognostics of spinal muscle atrophy (SMA) depending on the residual levels of exon 7 inclusion. Using antisense oligonucleotides – short chemically modified RNA molecules – it was possible to promote the inclusion of exon 7. This discovery led to the development and successful implementation of a drug, Nusinersen, that alleviated the

severity of SMA and increased survival of the patients (Darras et al., 2019; Finkel et al., 2017)

1.2. Regulation and functions of microexons

a) Regulation of neural microexons

Microexons were first characterised on the genome-wide scale in the neural context where an enrichment for exons ranging from 3 to 27 nucleotides was observed (Figure I-6A) (Irimia et al., 2014; Y. I. Li et al., 2015). Irimia et al (2014) described several properties of this class of exons. First, and inherently to their length, microexons tend to preserve open reading frame and can encode as little as one amino acid. Microexons were found to overlap or be in proximity to surface-accessible regions of proteins which allows for to fine-tuned regulation of protein-protein interactions by either inclusion or skipping of microexon in a transcript (Figure I-6B) (Irimia et al., 2014). Remarkably, upstream intronic regions of microexons display higher genomic conservation across vertebrates, as compared to constitutive and longer alternative exons. (Figure I-6C). It is in that region where genomic sequences required for splicing, such as splice sites, branch point and polypyrimidine tract, are distinct. Microexons are typically associated with significantly weaker 3'ss and much stronger 5'ss, which together with other intronic regulatory elements implies a different regulatory mechanism than that of longer alternative exons and,

importantly, longer neural exons (Figure I-6D) (Irimia et al., 2014; Raj et al., 2014).

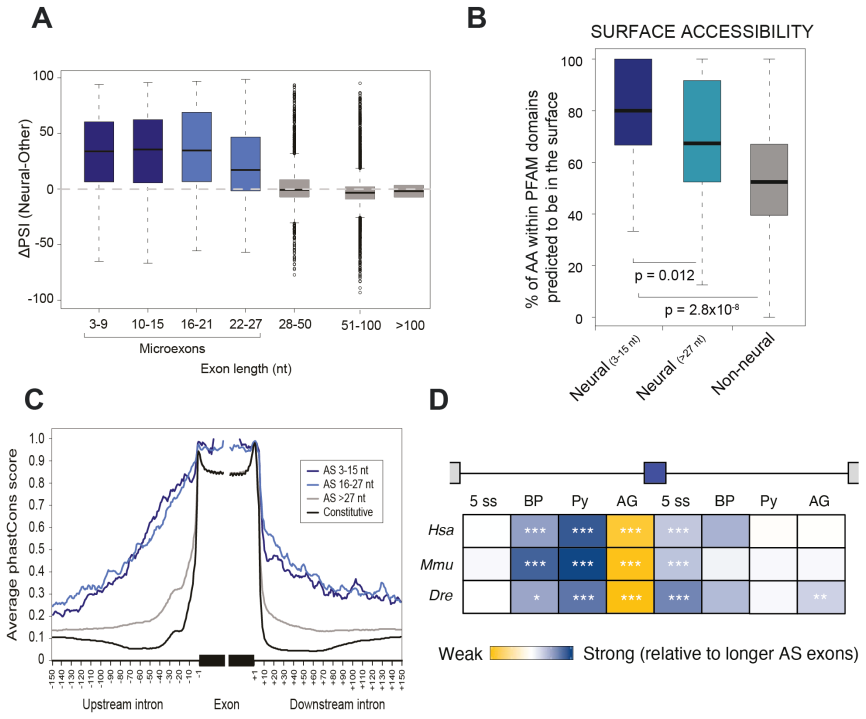

Figure I- 6. Microexons are distinctively short exons with conserved regulatory sequences in upstream introns. **A)** Differences in inclusion levels (ΔPSI) between mean inclusion level in neural samples and other tissues, stratified by the length of exons. **B)** Percent of amino acids predicted to be found in PFAM domains accessible to the surface of the proteins. **C)** Average phastScore conservation score for different classes of exons in 200 nucleotide windows of flanking introns. **D)** Intronic regulatory elements of neural microexons as compared to longer alternative exons. Strengths of splice sites was measured in MaxEnt maximum entropy score, branch point (BP) and Py scores were provided for best predicted sequences. Panel A-C from (Irimia et al., 2014), panel D from (Torres-Méndez et al., 2019).

Owing to their small size, microexons pose unique challenges for spliceosome assembly at splice sites. Inclusion of microexons was shown to largely depend on nSR100/SRRM4, a serine-arginine rich splicing factor (Figure I-7A) (Calarco et al., 2009; Irimia et al., 2014; Raj et al., 2014). The binding of SRRM4 was found in the UGC-enriched region of upstream introns of microexons, of which majority contains the first UGC between nucleotide -50 and 3'ss (Figure I-7B&C).

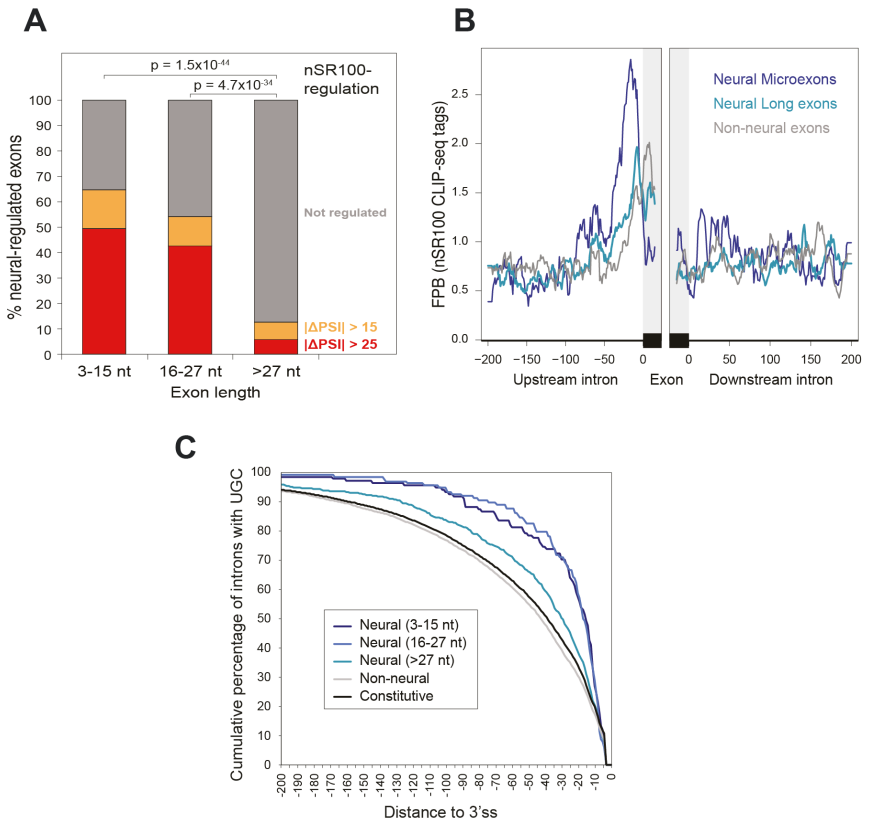

Figure I- 7. Inclusion of neural microexons is mediated by binding of SRRM3 to upstream introns. A) Percent of included microexons under over-expression of SRRM4 in HEK293. **B)** Enrichment of SRRM4 CLIP

(cross-linking and immunoprecipitation) tags for different classes of exons in 200 nucleotide windows of flanking introns, expressed in Fragments Per Billion (FPB). **C)** Cumulative distribution of the first UGC for different classes of exons. Panels A-C from (Irimia et al., 2014).

It was also demonstrated that a paralog of SRRM4, SRRM3, also plays a role in the regulation of neural microexons in, albeit with a functional redundancy between these two proteins (Nakano et al., 2019). This is consistent with a recent evolutionary study that showed that SRRM3 and SRRM4 contain a conserved domain, enhancer of microexons or eMIC, that alone was sufficient to regulate inclusion of microexons in bilaterians(Torres-Méndez et al., 2019).

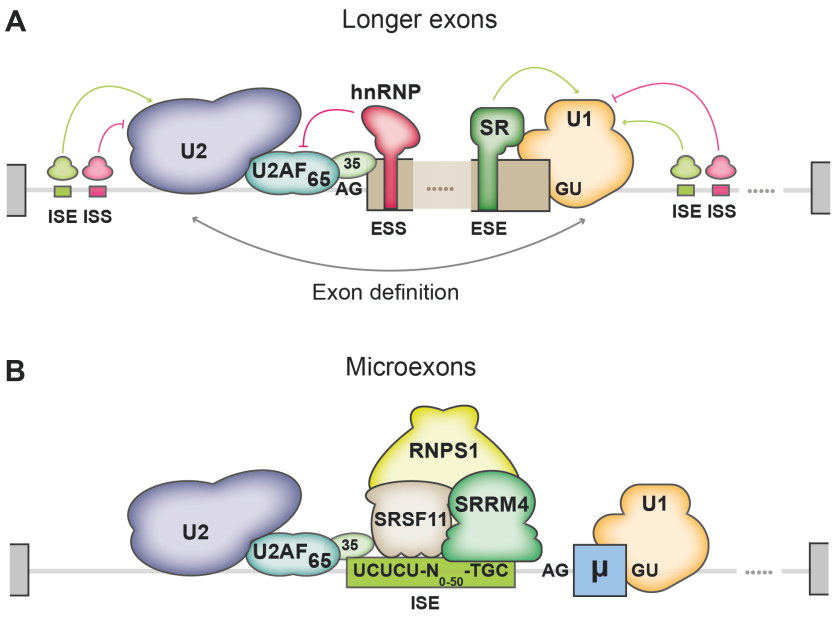

Figure I- 8. SRRM3 and SRRM4 proteins associate with early spliceosomal components and regulate intron definition at upstream region of microexons. **A)** Schematic of the exon-like definition model

assembled at upstream intron of microexons. **B)** Schematic of the exon definition model assembled at flanking introns of longer exons. μ – microexon; ISE – intronic splicing enhancer; ISS – intronic splicing silencer; ESE – exonic splicing enhancer; ESS – exonic splicing silencer. Adapted from (Gonatopoulos-Pournatzis et al., 2018; Scotti & Swanson, 2016).

In vertebrates, where introns tend to be longer that exons, early spliceosomal components, U1 and U2AF, recognise flanking splice sites of exons, constituting exon definition (Berget, 1995; Black, 1995; Grau-Bové et al., 2018; Robberson et al., 1990). It is now well understood that the outcome of alternative splicing of an alternative exon can be further regulated by a plethora of interactions between splicing enhancers or silencer (Figure I-8A) (Ule & Blencowe, 2019). In the context of exon definition, exons longer than 27 nucleotides harbour enhancer or silencer sequences that allow binding of positive or negative regulators that interact with U1 snRNP and U2AF.

However, as microexons are too short to harbour exonic enhancers, they evolved a dependency on 3'ss and intronic cis-regulatory elements to recognise an exon and bring splice sites in proximity for splicing. A search for a mechanistic model of microexon regulation revealed that peptide constructs containing eMIC domain associate with early spliceosomal components involved in 3'ss recognition, U2AF and branchpoint binding protein SF1 (Torres-Méndez et al., 2019). A more systematic study using CRISPR-Cas9 loss of function screens identified nearly 200 genes with potential impact on microexon inclusion. Although those genes represented

various regulatory layers, the strongest positive impact on microexon inclusion was noted for SRRM3, SRSF11 and RNPS1 (Gonatopoulos-Pournatzis et al., 2018). The same study extended the UGC motif found in proximity of upstream 3'ss with binding sites of SRSF11 and RNPS1, providing further evidence of microexon-specific exon definition complex formation (Figure I-8B) (Gonatopoulos-Pournatzis et al., 2018; Irimia et al., 2014; Raj et al., 2014; Torres-Méndez et al., 2019) Importantly, this model suggests that the regulatory mechanism of microexon inclusion may consists of more elements than currently known.

b) Functions of neural microexons

Neural microexons have been studied also due to their relevance in neurobiology, as they are found in genes with essential functions for the nervous system. Inclusion of microexons was shown essential for neuronal development, neurite outgrowth, neurotransmission and signalling (Figure I-9A). For example, regulation of *BAK1* exon 5 alternative splicing patterns in neurons, a 20nt microexon, was shown to modulate neuronal apoptotic sensitivity by promoting the non-sense mediated decay of *BAK1* transcripts upon microexon inclusion (Lin et al., 2020). A deletion of microexon in Eif4g1 translation initiation factor impaired hippocampal synaptic plasticity in mouse, leading to deficits in learning and decreased social behaviour (Gonatopoulos-Pournatzis et al., 2020).

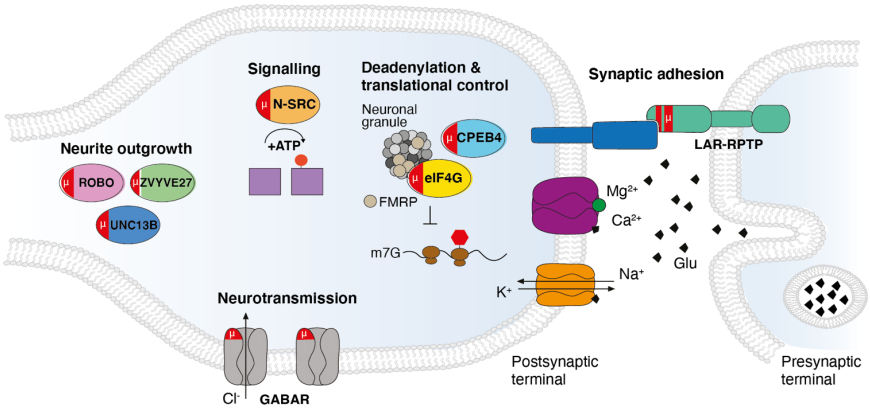

Figure I- 9. Functional impact of Inclusion of neural microexons. Neural microexons are found in genes spanning various regulatory layers required for correct function of the nervous system. Adapted from (Gonatopoulos-Pournatzis & Blencowe, 2020).

More examples of functional impact of microexon inclusion were reviewed in (Gonatopoulos-Pournatzis & Blencowe, 2020). Several neural microexons were also associated with autism spectrum disorder (ASD), where decreased inclusion of microexons forming a protein interaction network correlated with reduced expression of SRRM4 (Irimia et al., 2014). A mouse model of Srrm4 mutant recapitulated behavioural changes associated with ASD and demonstrated further neurological deficits, including deafness and decreased survival at the cellular and morphological levels (Nakano et al., 2012, 2019; Quesnel-Vallières et al., 2015). However, despite their emerging importance in neurobiology and disease, microexons have not yet been investigated outside that context.

As one of the focal points of this thesis is a program of microexons in endocrine pancreas, I need to introduce the functions of alternative splicing in the pancreatic endocrine biology.

1.3. Endocrine function of the pancreas

Pancreas is a heterocrine gland with both endocrine and digestive exocrine function. Pancreatic islets, or islets of Langerhans, are microorgans scattered throughout the gland, and are composed of five endocrine cell types (Figure I-10A). The most abundant are insulin-producing beta cells (around 60%) and glucagon-producing alpha cells (around 30%), although there is some variability between islets within the organ and between individuals (Steiner et al., 2010).

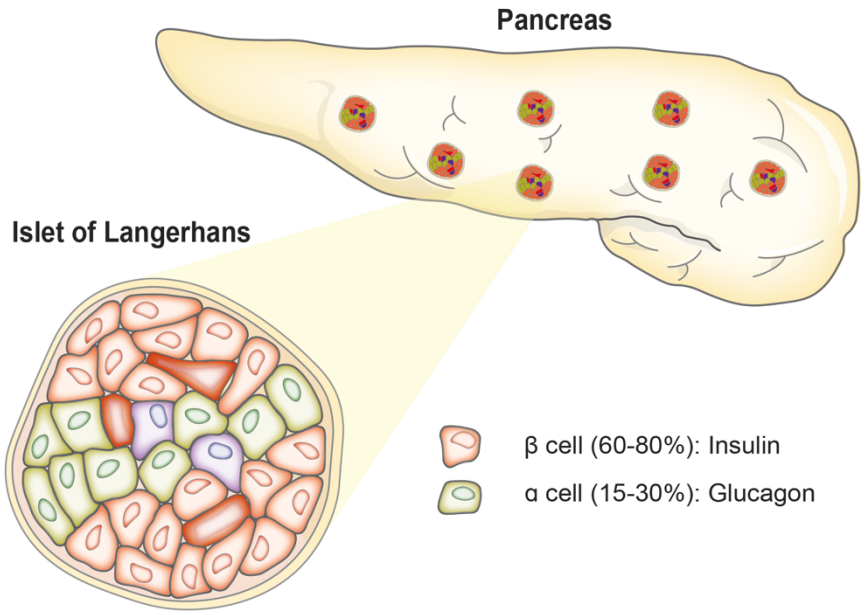

Figure I- 10. Schematic of the composition of the pancreatic islets. Pancreatic islets are composed predominantly of beta and alpha cells that secrete insulin and glucagon, respectively. In response to the high glucose level, insulin lowers the blood glucose level. On the other hand, glucagon stimulates conversion of glycogen stores in liver to glucose in response to the low blood glucose level.

The remaining 10% of the islet cell mass consists of delta, gamma and epsilon cells that secret somatostatin, pancreatic polypeptide, and ghrelin, respectively (Steiner et al., 2010).

The primary function of the pancreatic islets is maintenance of glucose homeostasis through tightly regulated and highly synchronised insulin secretion by beta cells in response to nutrient availability, specifically glucose (Figure I-11). Beta cells can detect changes in circulating glucose concentrations and other nutrients and secrete insulin according to body nutrient state. These is achieved by coupling cell metabolism to electrical activity. Briefly, glucose enters the cell through GLUT1 glucose transported (GLUT2 in rodents) where it is phosphorylated by glucokinase (GK) and enters the glycolytic pathway Figure I-11; 1). Pyruvate generated in this reaction enters the tricarboxylic acid cycle (TCA) in mitochondria (Figure I-11; 2), where the electron transport chain generates ATP from ADP in a calcium dependent manner (A. I. Tarasov et al., 2012). The cytosolic increase in ATP:ADP ration inhibits ATP-dependant potassium channels, which triggers depolarisation of the plasma membrane (Figure I-11; 3-4), (A. Tarasov et al., 2004). Upon sufficient depolarisation of the

membrane, the activation of voltage-dependant calcium channels triggers influx of calcium ions and increase in intracellular calcium concentration (Figure I-11; 5) (Rorsman et al., 2012).

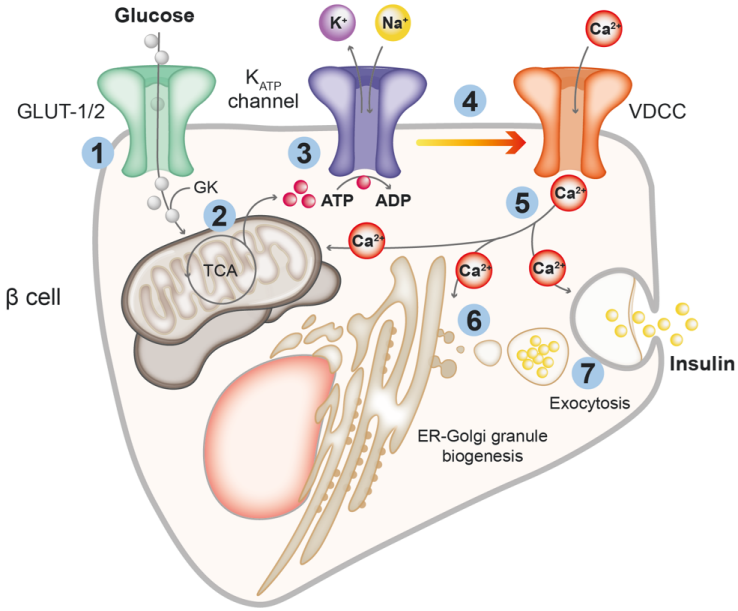

Figure I- 11. Schematic of insulin secretion by beta cells. 1 – Glucose sensing and uptake. 2 – Pyruvate uptake and TCA cycle in mitochondria. 3 – Increase in ATP:ADP ratio. 4 - Plasma membrane depolarization stimulated by the closure of potassium channels. 5 – Voltage-dependent influx of calcium ions. 6 – Insulin processing at ER-Golgi. 7 – Transport of insulin granules to the exocytotic site and insulin release.

The role of calcium ions in beta cells are manifold. As mentioned before, calcium ions control ATP synthesis. Calcium ions are also used by endoplasmic reticulum (ER) to correctly fold proteins, assemble the insulin secretory granules, and regulate their migration to the plasma membrane

(Figure I-11; 6) (Lees et al., 2017; Zhang et al., 2020). The remodelling of actin cytoskeleton and thus regulation of insulin exocytosis, has also been shown to depend on calcium ion oscillations (Arous & Halban, 2015). Insulin granules dock to the plasma membrane through the SNARE proteins, including SNAP-25, VAMP-2/8, and syntaxins 1A and 3 (Gaisano, 2017; Leung et al., 2007; Regazzi et al., 1995; Sadoul et al., 1995). Calcium channels in proximity to the exocytic sites deliver calcium ions to Synaptotagmin VII, a calcium-dependent secretory vesicle (Gandasi et al., 2017; Gandasi & Barg, 2014). These associations ensure the fusion of insulin granules to the plasma membrane and the release of insulin (Figure I-11, 7). Although the described mechanisms account for the basic mechanism by which beta cells secrete insulin, other hormones, biomolecules, and metabolic processes contribute to this tightly regulated process (Fu et al., 2013). Different aspects of beta cells biology were reviewed in (Tokarz et al., 2018).

Remarkably, the mechanism of insulin secretion and the electrical activity of the cell closely resemble the release of neurotransmitters by neurons (Arntfield & van der Kooy, 2011). In fact, it has been proposed that beta cells evolved into specialised insulin-secretory cell by co-opting some of the neuronal transcriptional regulatory networks (Arntfield & van der Kooy, 2011). For example, at the level of gene expression, both beta cells and neurons lack the expression of REST which represses the neuronal phenotypes in non-neural cells (Atouf

et al., 1997; D. Martin & Grapin-Botton, 2017; Nica et al., 2013). Moreover, global gene expression patterns and chromatin post-translational modifications, were shown to resemble neuronal patterns more than other cells, including exocrine acinar cells (Nica et al., 2013; van Arensbergen et al., 2010).

Important for this Thesis and consistently with the similarities between beta cells and neurons, recent studies linked the expression of neural-specific splicing factors to the secretory functions of beta cells. Depletion of Nova1, a splicing factor typically associated with neural alternative splicing, was shown to inhibit insulin secretion and sensitise cells to pro-inflammatory cytokines that induce apoptosis by deregulating splicing of transcripts in genes involved insulin receptor signalling and exocytosis, among others (Villate et al., 2014).

Two splicing factors, Rbfox1 and 2, were linked to negative regulation of insulin secretion, as their depletion resulted in increased insulin secretion through stimulation of actin cytoskeleton remodelling (Juan-Mateu et al., 2017).

Similar effect on beta cell survival was observed upon depletion of Nova2, Elavl4 and SRSF6 (Alvelos et al., 2021; Juan-Mateu et al., 2017, 2018). These and other studies demonstrated the importance of alternative splicing in beta cell biology and expanded the regulatory mechanisms underlying beta cell malfunction and diabetes (Colli et al., 2020; Juan-

Mateu et al., 2016; Malakar et al., 2016; Wilhelmi et al., 2021), and reinforced the similarity of beta cells to neurons. However, the identification and functional characterisation of alternative splicing specific or enriched in beta cells has not been studied before.

1.4. Thesis objectives

Main objectives of this Thesis:

Objective 1: Investigate the existence and evolutionary conservation of programs of microexon regulation in non-neural tissues.

Objective 2: Characterise properties, inclusion patterns and biological functions of the program of microexons in endocrine biology of the pancreas.

Objective 3: Investigate cis-acting regulatory elements contributing to the regulation of microexon inclusion in pancreatic beta cells and neurons in the context of regulation by SRRM3/4 proteins.

CHAPTER 2

An uncharacterised program of endocrine pancreatic microexons is regulated by SRRM3

2. Chapter 2: An uncharacterised program of pancreatic endocrine microexons is regulated by SRRM3.

Preface to Chapter 2

Chapter 1 is the corner stone of this Thesis. I will begin with the discovery of endocrine pancreatic microexons and a brief characterisation of this program. Then, I will discuss the regulation by SRRM3 and present analysis and experimental data from beta cell lines and mouse islets of Langerhans arguing for the biological importance of these microexons in the context of pancreatic beta cell biology. Finally, I will introduce the mechanistic implications that we derived from sequence analyses of endocrine pancreatic microexons, which will be further explored in Chapter 3.

2.1. Characterisation of the pancreatic endocrine microexon program.

Our current understanding of microexon regulation and function comes from extensive studies in neural tissues (Gonatopoulos-Pournatzis et al., 2020; Irimia et al., 2014; Nakano et al., 2019; Quesnel-Vallières et al., 2015, 2016; Raj et al., 2014; Torres-Méndez et al., 2019). However, SRRM3/4-regulated programs of microexons had not yet been investigated outside of their neural context.

To systematically investigate the alternative splicing profile of microexons in non-neural tissues, we performed pairwise comparisons of RNA-seq transcriptome data between human cell and tissue type groups from the VastDB database and other sources. In total, we compared 31 groups of cell and tissue types, comprised of 159 samples from VastDB and 18 samples that we collected from publicly available data (see Methods).

After neural tissues, the second most microexon-enriched tissue was endocrine pancreas (Figure 1). We identified a total of 168 exons that were included in endocrine pancreas than at least 5 other tissue groups, excluding neural tissue. More than half of endocrine pancreas-enriched exons were microexons (Figure 1A). As these microexons are not included in exocrine pancreas and are found in genes involved in key endocrine functions, we refer to these microexons as EndoMICs, hereafter.

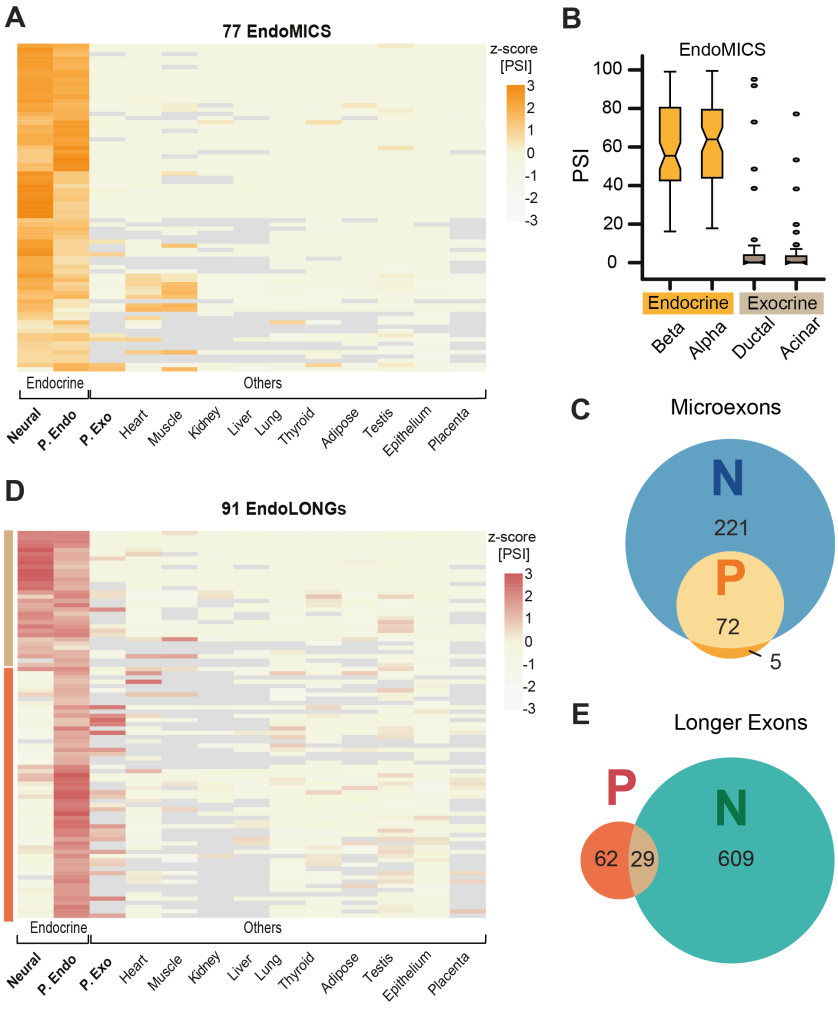

Figure 1. Endocrine pancreas utilises a subset of microexons previously characterised as neural-specific. A) Heatmap of z-scored inclusion level (PSI) for EndoMICs across selected tissues (Endocrine and Others). Neural and endocrine pancreatic programs of tissue-enriched longer exons and microexons were identified through pair-wise transcriptomic comparisons with other tissue groups (Others). Inclusion levels obtained from VastDB for tissue groups with at least three samples per group. **B)** Distribution of inclusion level for EndoMICs in endocrine and exocrine cell subtypes of pancreas. **C)** Overlap between full program

NeuralMICs and EndoMICs. **D)** Heatmap of z-scored inclusion level (PSI) for LongEndoEX across selected tissues (Endocrine and Others). Colour bars correspond to Venn diagram sets in Panel F. **E)** Overlap between NeuralLONGs and EndoLONGs. All data presented for human samples and identified exon programs.

To confirm the enrichment of EndoMICs in endocrine pancreas, we checked their inclusion of two endocrine (alpha, beta) and two exocrine (ductal, acinar) cell populations of the pancreas (Figure 1B). We found that indeed EndoMICs were included more included in alpha and beta cells and scarcely utilised by ductal and acinar cells.

Interestingly, we observed that EndoMICs largely constitute a subset of the larger, previously characterised program of 298 NeuralMICs. This two-tier nested program thus questioned the neural specificity of microexons. This observation will be further expanded in Chapter 3.

We also wondered if the same held true for the longer exons enriched endocrine pancreas, which we named EndoLONGs (Figure 1D-E). These exons did not show the same relative enrichment as EndoMICs, as only 29 exons were part of the neural-enriched 638 longer exons. The remaining 62 exons form a distinct alternative splicing program, with 13 of these exons shared with exocrine pancreas.

NeuralMICs are more conserved than longer neural exons at the genomic level (Irimia et al., 2014). We observed a similarly high genomic conservation for EndoMICs, with 70 and 72 microexons found in the genome of rat and mouse,

respectively. (Figure 2A). In contrast, EndoLONGs in the respective species showed that – as in neural tissues – EndoMICs display higher genomic conservation (Table 4).

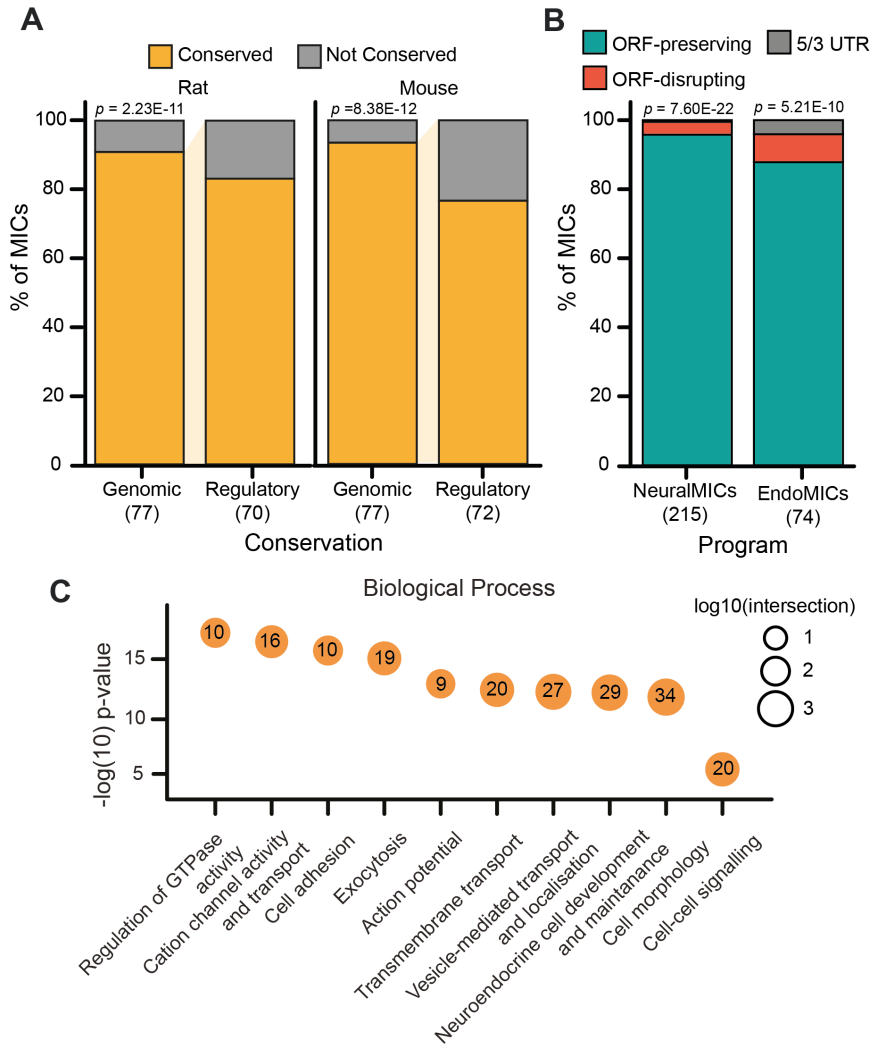

Figure 2. EndoMICs display properties common to NeuralMICs. **A)** Percentage of genomic and regulatory conservation for EndoMICs in rat and mouse. Regulatory conservation was defined under the threshold of the average ΔPSI 15 between endocrine pancreatic samples and other tissues across human, rat and mouse data. P-values were obtained from

Fisher exact test for the number of genomically conserved/non-conserved EndoMICs and EndoLONGs for the respective species. **B)** Percentage of MICS displaying different types of predicted impact on open reading frames or mRNA Untranslated Regions, for human NeuralMICs and EndoMICs. P-values were obtained from Fisher exact test for the number of ORF/Other MICS and LONGs for the respective programs. **C)** Enriched Gene Ontology (GO) categories for EndoMICs-harbouring genes. All data presented is for human exon programs. Numbers correspond to the number of genes harbouring microexons that intersect the GO categories.

Next, to study regulatory conservation, we performed the same tissue enrichment analysis using rat and mouse inclusion data from VastDB. We identified 64 and 58 microexons in rat and murine endocrine pancreatic samples, respectively, with higher inclusion in endocrine pancreas – and similar overlap with NeuralMICs for these species – compared to other tissues.

Microexons of just 3 nucleotides in length can modify a protein sequence by a single amino acid. Microexons are characterised by their lengths displaying a distribution of multiples of 3nt, which allows preservation of open reading (ORF) frame upon their inclusion. Consistent with this, we observed as much ORF-preserving potential between NeuralMICs-only and EndoMICs as compared to NeuralLONGs and EndoLONGS, respectively. exons identified in these tissues.

Finally, we checked potential biological function of EndoMICs-harbouring genes, using all genes with exons with sufficient RNA-seq read coverage in endocrine pancreatic samples as

background for gene ontology analysis. Given the overlap of EndoMICs with the larger neural program, ontology terms for EndoMICs-harbouring genes were enriched in terms related to neurobiology. To comprehensively capture functions with importance for pancreatic endocrine function, we manually curated and generalised ontology terms based on the scientific literature. Our analysis and ontology curation provided an annotation for 57 out of 76 genes that harbour EndoMICs (see Table 5) (Figure 2C). The 10 most enriched terms reflected some of the essential functions that endocrine pancreatic cells share with neurons, such as cation transport across the membrane (basis for the transmission of membrane potential) or vesicle transport (common to the secretion of insulin and neurotransmitters).

2.2. Srrm3-mediated alternative splicing of EndoMICs affects insulin secretory capacity of beta cells

We then sought to understand whether EndoMICs follow the SRRM4-based regulation of NeuralMICs. It has been recently shown that NeuralMICs can be regulated also by a paralog of SRRM4, SRRM3 (Nakano et al., 2019; Torres-Méndez et al., 2019). This redundancy is mediated by an ancestral domain (the enhancer of microexons, eMIC, domain) present in both proteins (Torres-Méndez et al., 2019).

For this reason, we looked at the expression levels of *Srrm3* and *Srrm4* in rat beta cell samples. We chose the rat model due to its high relevance in pancreatic and diabetic research. To our surprise, Srrm4 was not expressed in the endocrine pancreas, unlike Srrm3 (Figure 3A). Interestingly, the levels of Srrm3 were substantially lower than those observed in the neural samples.

To assess the potential role of Srrm3 regulating EndoMICs, we transiently depleted Srrm3 in INS1E rat beta cell line using siRNAs followed by sequencing of polyA-selected RNA (RNAseq) (Figure 3B-D). It is important to note that despite the 47% reduction in *Srrm3* RNA level (Figure 3B) (*Srrm4* undetectable by RT-qPCR, data not shown), we observed pronounced skipping of microexons in the INS1E rat beta cell line (Figure 3D). Out of 171 skipped exons with a minimum of difference in inclusion (ΔPSI) of -15, 116 were microexons. Moreover, 32/55 of the remaining skipped exons had sizes between 28-50nt, which is consistent with previous studies showing that SRRM4 regulates a group of exons between 28 and 50 nucleotides (Irimia et al., 2014). In fact, these exons are also considered as microexons by Li et al (2015). However, while we will continue to show their inclusion profile, for the purpose of this Thesis these exons will not be farther explored in the context of pancreatic endocrine biology.

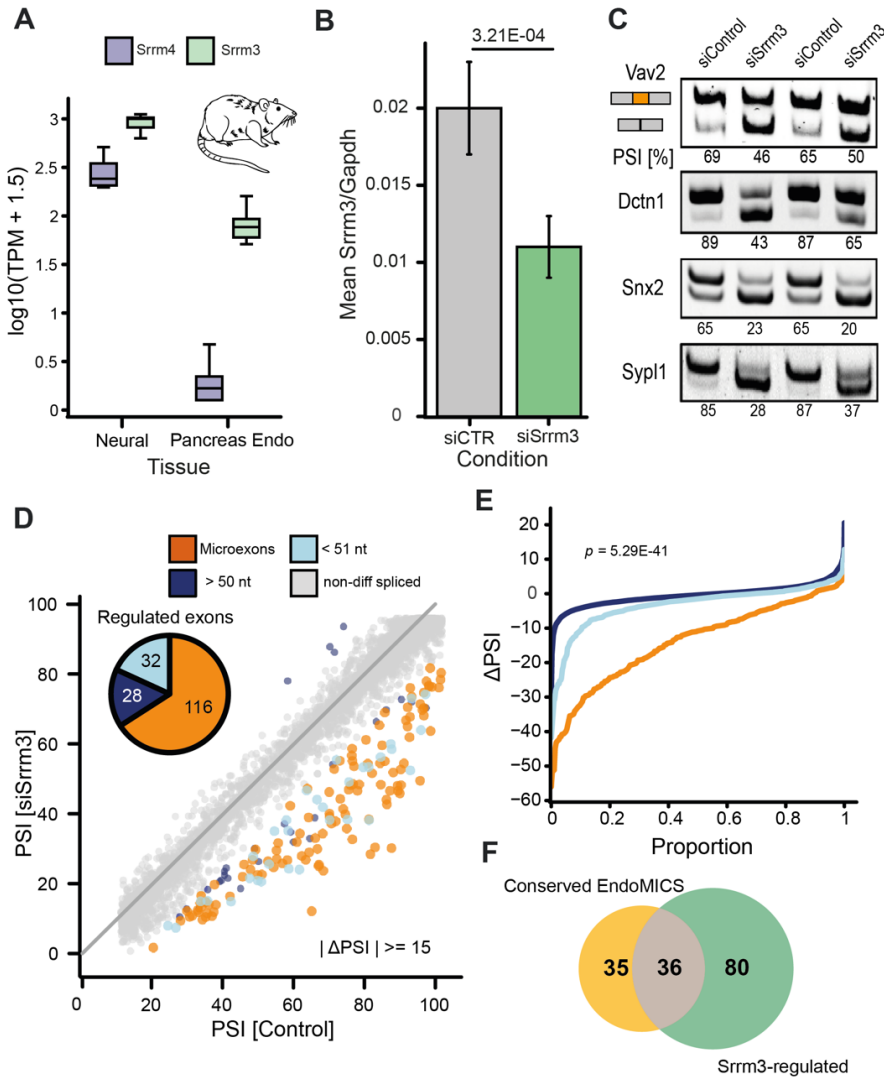

Figure 3. Expression of Srrm3 drives the inclusion EndoMICs. ***A)*** *Expression of Srrm3 and Srrm4 in log10 expression in Neural and PancreasEndo samples.* ***B)*** *qPCR validation of siRNA-mediated Srrm3 knockdown (KD) relative to Gapdh expression in INS1E beta cell line. Error bars correspond to normalised ΔCt values obtained by qPCR (n=5).* ***C)*** *RT-PCR validation showing inclusion/exclusion of selected microexons upon downregulation of Srrm3 in INS1E beta cell line. All data presented for rat samples and cell line.* **D)** Inclusion profile scatter plot for rat longer exons

and microexons in control and siRNA-mediated *Srrm3* KD in INS1E rat beta cell line. Differentially spliced microexons and longer exons are coloured under |ΔPSI| >= 15. Pie chart represents number of affected longer exons and microexons. **E)** Distribution of ΔPSI for all longer exons (blue) and microexons (orange) with coverage. P-value of 1.26E-02 was obtained with two-sided Fisher exact-test between number of respective types of exons at |ΔPSI| >= 15. **F)** Overlap between Srrm3-regulated microexons in INS1E rat beta cells and microexons rat EndoMICs conserved in human. All data presented for rat sequencing data. Experiments in panel B-C performed by Jonàs Juan-Mateu.

We confirmed the specificity of Srrm3-mediated alternative splicing profile in our data by comparing ΔPSI distribution between microexons and longer exons (>50nt) performing two-sided Wilcoxon test that showed a significant difference between distributions (Figure 3E). Finally, we checked the proportion of human EndoMICs conserved in rat in our data. More than half of the microexons with genomic conservation in were skipped in INS1E rat beta cell line (Figure 3F).

As the primary function of beta cells is to synthesise and secrete insulin in response to glucose, we wondered how the Srrm3 downregulation affected secretory capacity of INS1E beta cell line. We performed a glucose-stimulated insulin secretion assay (GSIS, Figure 4A) in INS1E cells incubated in glucose-free medium followed by incubation in Krebs-Ringer solution that contains physiologically relevant concentrations of glucose and salts. As a control, we stimulated cells with Forskolin, which promotes cAMP production in beta cells and stimulates elevated insulin secretion. We observed that, at the

basal level, Srrm3 KD significantly increased the insulin release by 50% compared to non-targeting siRNA (Figure 4A). A similar and significant increase, albeit at higher values of changes in stimulation index, was observed upon Forskolin treatment (Figure 4A), confirming that the downregulation of Srrm3 affects insulin secretion in INS1E beta cell line.

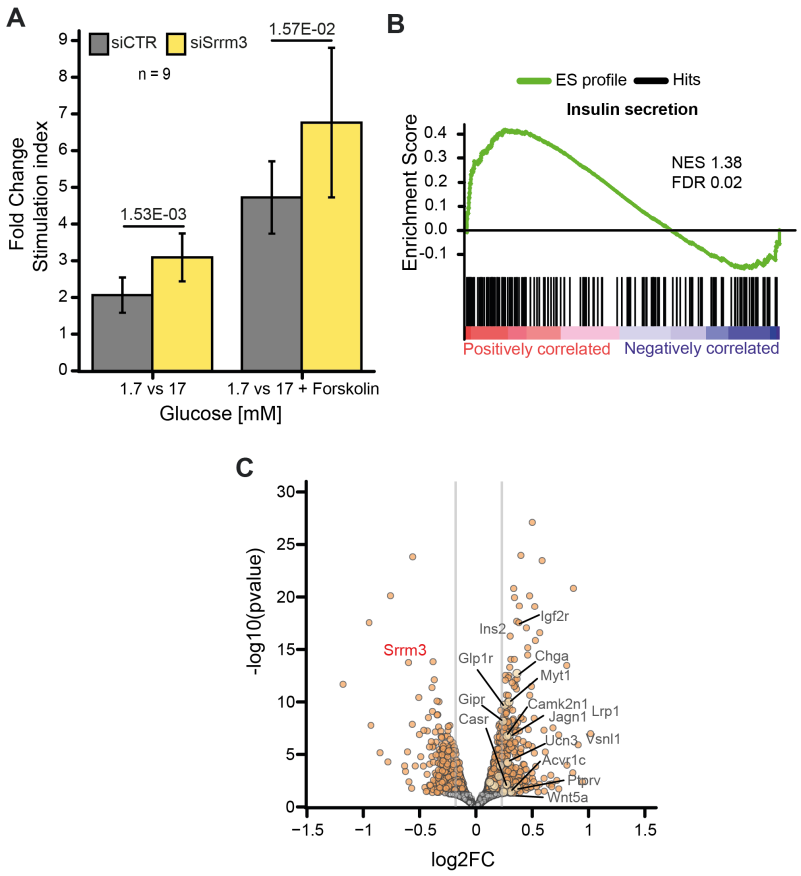

Figure 4. Srrm3 regulates insulin secretory functions of pancreatic beta cells. **A)** Static glucose-stimulated insulin secretion assay in INS1E rat beta cell line upon downregulation of Srrm3 after exposure to low and high glucose concentrations with or without the cAMP activator Forskolin for 30 min, showing increased basal and stimulated secretion upon Srrm3

KD. **B)** Gene set enrichment analysis for Insulin secretion Gene Ontology term in INS1E rat beta cell line upon Srrm3 KD showing significant upregulation of genes belonging to this functional term. **C)** Volcano plot for differentially regulated genes in INS1E rat beta cell line upon Srrm3 KD showing mild gene expression changes. Points were coloured according to log2FC threshold (log2FC ≤ -0.3 for downregulated genes and log2FC ≥ 0.3 for upregulated genes) and significance of FDR ≤ 5%. Genes that contribute the most to the core enrichments in Panel C and passed the differential expression threshold were highlighted. Experiments in panel A performed by Jonàs Juan-Mateu.

We wondered how genes related to the regulation of insulin secretion were affected by Srrm3 depletion. To check this, we performed a gene set enrichment analysis using positive and negative regulators of insulin secretion as a query set in ranked list of genes in Srrm3 KD data (Figure 4B) and observed overall upregulation of genes.

To identify relevant targets that could explain the phenotype, we performed differential gene expression analysis upon Srrm3 KD (Figure 4C). Out of 252 significantly regulated genes (FDR 5%, |log2FC| ≥ 0.3), 92 were downregulated and 151 were upregulated. We observed a significant enrichment in upregulated genes. Among the upregulated genes that also contributed the most to the core enrichment observed in GSEA analysis, and passed differential regulation thresholds, we identified rat insulin gene *Ins2*, and genes related to secretory granules, such as *Unc13b*, Igf2r and *Chga*.

To investigate whether SRRM3 depletion has a similar effect in human models, we depleted SRRM3 in EndoC-B1 human beta cell line with siRNA and performed RNAseq.

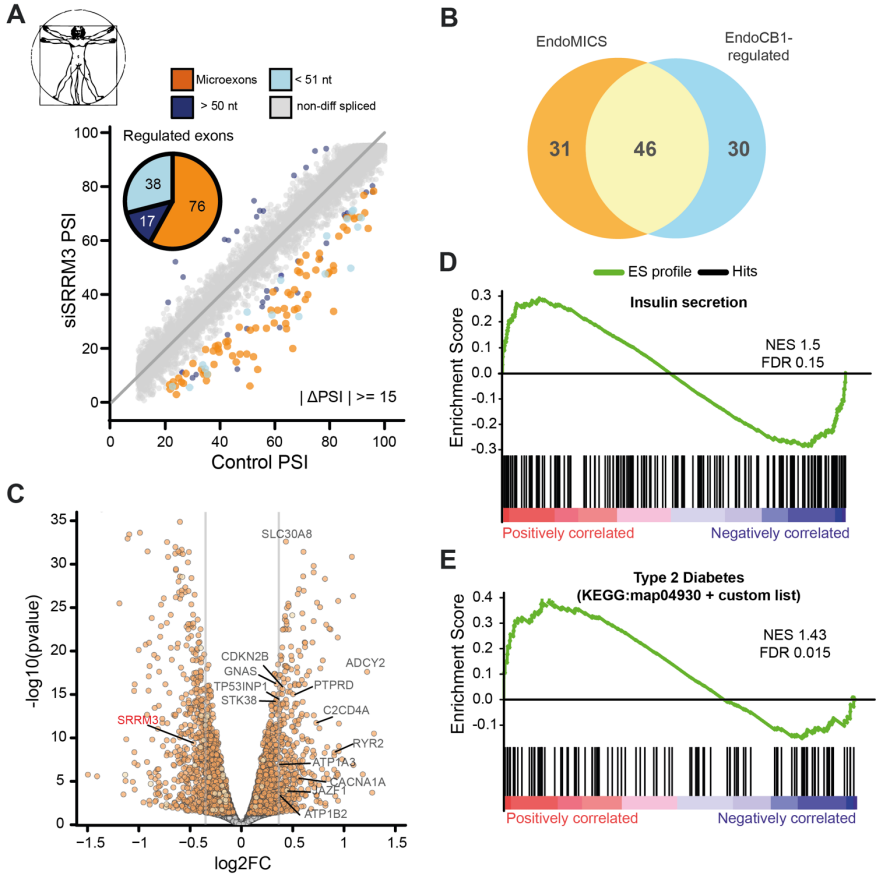

Figure 5. EndoC-B1 human beta cell line recapitulates SRRM3-based regulation of EndoMICs and impact on beta cell function. A) Inclusion profile scatter plot for human longer exons and microexons in control and siRNA-mediated SRRM3 KD in EndoC-B1 human beta cell line. Differentially spliced microexons and longer exons are coloured under |ΔPSI| >= 15. Pie chart represents number of affected longer exons and microexons. **B)** Overlap of human EndoMICs and SRRM3-regulated microexons. **C)** Volcano plot for differentially regulated genes in EndoC-B1 human beta cell line upon SRRM3 KD showing mild gene expression

changes. Points were coloured according to log2FC threshold (log2FC ≤ -0.3 for downregulated genes and log2FC ≥ 0.3 for upregulated genes and significance of FDR ≤ 5%. Genes that contribute the most to the core enrichments in Panel D-E and passed the differential expression threshold were highlighted. **D-E)** Gene set enrichment analysis results (GSEA) for genes with coverage obtained from differential gene expression profile upon downregulation of SRRM3 in human EndoC-B1 beta cell line. GSEA for Insulin secretion (D) and for T2D KEGG Pathway term (E).

Although we achieved lower level of SRRM3 downregulation than in INS1E rat beta cell line (40% depletion, data not shown), we observed a similar pattern of microexons skipping (Figure 5A) and a relatively higher overlap between EndoMICs and SRRM3-regulated microexons in this cell line (Figure 5B). SRRM3 KD showed more effect on gene expression in EndoC-B1 cells (Figure 5C), as 650 genes were down- and 531 genes were up-regulated (FDR ≤ 5%, |log2FC| ≥ 0.3). We did not observe significant upregulation in genes related to insulin secretion as in INS1E Srrm3 KD (Figure 5D). We also wondered how the level of SRRM3 KD that we achieved affected genes associated with Type 2 Diabetes (T2D). We observed a significant upregulation of genes in that set (Figure 5E), including calcium channel-related genes, such as (CACNA1A, *RYR2*), kinases (*PTPRD*, *STK38*) and transcription factors (*JAZ1F*, *PAX4*).

2.3. Srrm3-mediated alternative splicing in islet physiology

a) *Srrm3*$^{-/-}$ mice display symptoms of hyperinsulinemic hypoglycemia

Thus far, we investigated the SRRM3-based regulation of microexons and gene expression in model cell lines. However, immortalised beta cell lines do not fully reflect mature function of the primary beta cells (Scharfmann et al., 2019).

To validate our finding in the physiological context of insulin secretion and islet development, we took advantage of a recently published *Srrm3* knock out mouse model (Nakano et al, 2019). While the authors extensively characterised that model and showed that *Srrm3*$^{-/-}$ mice had reduced growth and impaired motor functions, they solely focused on the neural aspects of microexon biology and therefore did not investigate the potential impact of SRRM3 in pancreatic function.

First, we isolated pancreatic islets from 10 to 12 weeks old mice and analysed alternative splicing and gene expression changes by RNAseq (Figure 6A). We observed a widespread impact on alternative splicing when comparing the transcriptomes from *Srrm3*$^{-/-}$ and wild-type mice, with microexons being preferentially skipped (49 microexons with minimum ΔPSI ≤ -10) (Figure 6B). A subset of 13 microexons was found more included in our data. A third of deregulated microexons overlapped with mouse EndoMICs (Figure 6C). We also found that 80 out of 135 differentially regulated exons

between 28 and 50 nucleotides were more skipped in *Srrm3*$^{-/-}$ mouse islets. Longer exons also displayed deregulation in both directions, with 681 exons skipped and 517 exons more included in *Srrm3*$^{-/-}$ mouse islets.

Differential gene expression analysis identified 406 upregulated and 298 downregulated genes in *Srrm3*$^{-/-}$ mouse islets (FDR ≤ 5%, |log2FC| ≥ 0.3) (Figure 6D). The results obtained from GSEA analyses showed general upregulation of T2D and beta cell developmental gene sets in *Srrm3*$^{-/-}$ mouse islets (Figure 6E-F). Among significantly upregulated genes, we found key transcription factors involved in endocrine commitment and islet cell differentiation such *Pdx1* and *Neurog3*, and genes with known function in T2D or glycemic traits, such as *Ptprd*, *Slc30a8* and *Jazf1*. Importantly, these genes were also found to be upregulated in human beta cell lines upon SRRM3 KD.

While analysing the results for alternative splicing analysis and gene expression, it is important to consider the cumulative effects of constitutive *Srrm3*$^{-/-}$ knockout in a wider context. More specifically, *Srrm3*$^{-/-}$ mice have been shown to display reduced viability and several neurological defects. This in turn may be reflected in deregulation of other biological processes and organ development and could explain widespread changes in alternative splicing observed in *Srrm3*$^{-/-}$ mouse islets as compared to model beta cell lines.

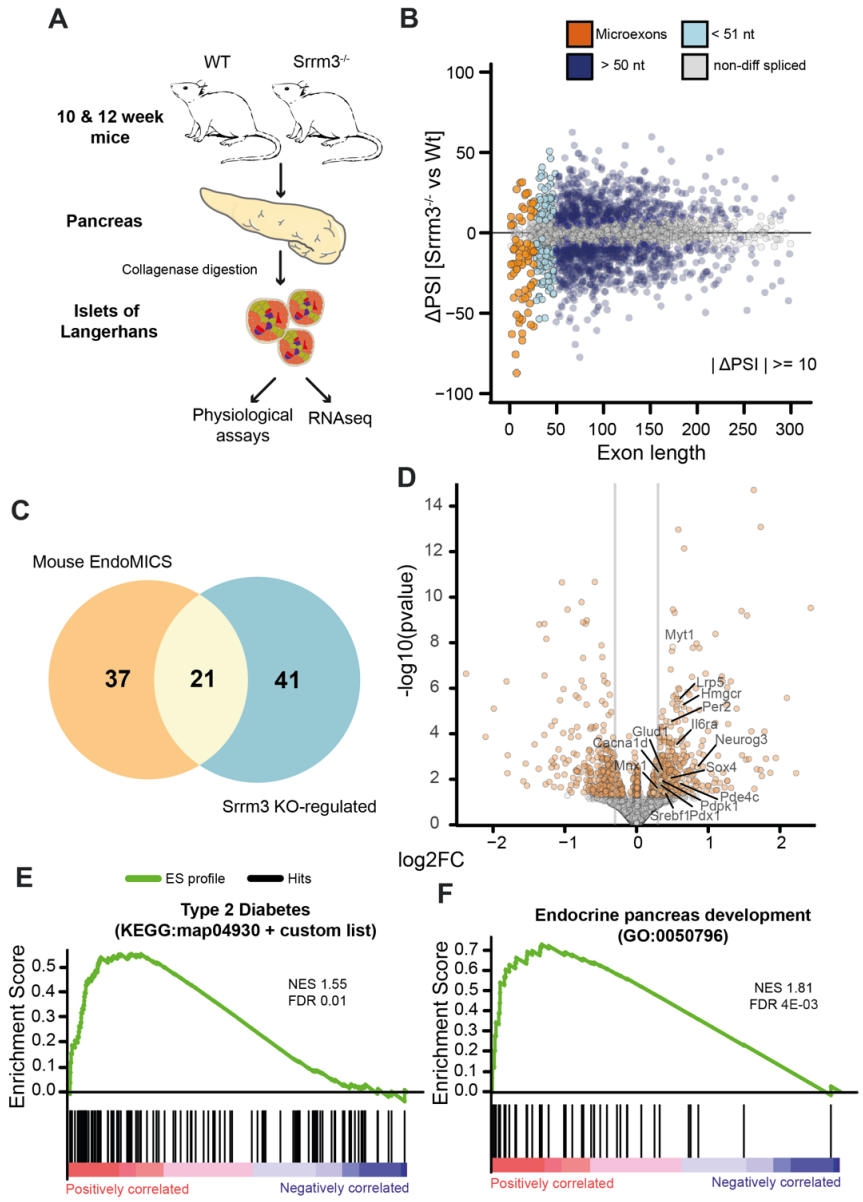

Figure 6. Loss of Srrm3 affects alternative splicing and gene expression related to beta cell biology. **A)** Schematic of data acquisition for RNAseq and experimental validation in Figure 7. **B)** Impact of Srrm3$^{-/-}$ knockout exon inclusion in mouse islets relative to exon length. Exons were coloured by differential profile with a minimum of |ΔPSI| >= 10. Mean dPSI was obtained through paired comparisons between samples. **C)** Overlap

between mouse EndoMICs and microexons regulated in *Srrm3*$^{-/-}$ mouse islets. **D)** Volcano plot for differentially regulated genes between *Srrm3*$^{-/-}$ and wild-type mouse islets. Points were coloured according to log2FC threshold (log2FC ≤ -0.3 for downregulated genes and log2FC ≥ 0.3 for upregulated genes) and significance of FDR ≤ 5%. Genes that contribute the most to the core enrichments in Panel E-F and passed the differential expression threshold were highlighted. **E-F)** Gene set enrichment analysis results (GSEA) for genes with coverage obtained from differential gene expression profile between *Srrm3*$^{-/-}$ and wild-type mouse islets. **E)** GSEA for T2 KEGG Pathway term showing a significant enrichment in upregulated genes. Mouse orthologs were used for this analysis. **F)** GSEA for Endocrine pancreas development Gene Ontology term showing a significant enrichment in upregulated genes. A list of curated genes was additionally provided.

We found several genes involved in organ development enriched in gene ontology analysis (Figure 7A, top). Other functional categories for upregulated genes were involved in lipid metabolic processes that are also involved in insulin secretion and have direct link to T2D. On the other hand, downregulated genes were mainly enriched in mitochondrial genes related to cell metabolism, such as ATP production and cellular respiration (Figure 7A, bottom). Altogether, these results suggested that loss of Srrm3 impacts islet differentiation and identity, and affects pathways related to cellular metabolism, nutrient sensing, and insulin secretion.

Next, to assess the effect of the constitutive Srrm3$^{-/-}$ knockout on islet secretory function, the islets were maintained in culture and the accumulated insulin in media was measured.

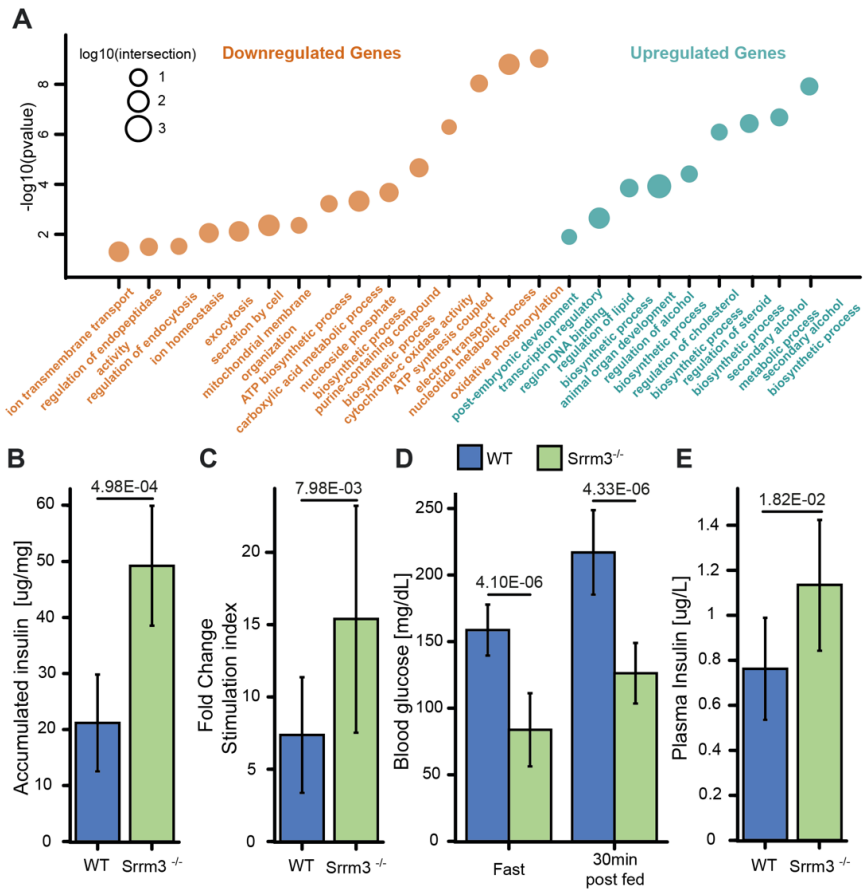

Figure 7. Srrm3$^{-/-}$ mice are characterised by hyperinsulinemic hypoglycemia. **A)** Enriched terms for Biological Process for differentially regulated genes between Srrm3$^{-/-}$ and wild-type mice using all genes with coverage as a background. Results obtained with gProfiler2 for terms with at least 5 genes and terms intersection of 10 genes. Terms were selected with FDR ≤ 5%. **B)** Insulin accumulated in culture media after 18h. **C)** Glucose stimulated insulin secretion fold change following sequential stimulation from 2.8 mM to 20 mM glucose for 30 min. **D)** Blood glucose levels following a fast and fed test. Blood glucose level was measured at the end of a 4h fast and at 30 min post feeding. **E)** Plasma Insulin levels and ratio between plasma insulin and glycemia (data not shown) at 30 min postprandial showing hyperinsulinemia in *Srrm3*$^{-/-}$ mice. Isolation of mouse

islets and experimental validations were performed by Jonàs Juan-Mateu, Amaya Lopez Pascual and Marta Miret-Cuesta.

Islets from *Srrm3*$^{-/-}$ mice showed a 2-fold increase in accumulated secreted insulin in the media relative to the wild-type mice (Figure 7B). To confirm this observation, we performed static GSIS in mouse islets. Consistently with increase in accumulated insulin in media, *Srrm3*$^{-/-}$ mouse islets displayed increased stimulation index following exposure to glucose (Figure 7C). Following these experiments, glucose tolerance was assessed by fast-fed test. In line with previous experiments, *Srrm3*$^{-/-}$ mice displayed hypoglycemia at 4h fast and reduced glycemia at 30 min postprandial (Figure 7D). To test whether hypoglycemia in *Srrm3*$^{-/-}$ mice was caused by excessive insulin secretion from the pancreas, we measured insulin plasma levels at fasting and 30 min postprandial. *Srrm3*$^{-/-}$ mice showed a significant increase in plasma insulin at 30 min postprandial suggesting that hypoglycemia was caused by hyperinsulinemia (Figure 7E). Other experiments in the lab (data not shown) validated these finding by measuring plasma glucagon and ketone bodies that were expected to decrease in response to excessive insulin secretion and inhibited glucose counter-regulatory responses. We confirmed that *Srrm3*$^{-/-}$ mice displayed symptoms of hyperinsulinemic hypoglycemia.

b) Implication of microexon deregulation in Type 2 Diabetes and further studies

Lastly, as we observed an enrichment of T2D-related genes in upregulated genes in EndoC-B1 human beta cell line and *Srrm3*$^{-/-}$ mouse islet data, we wondered whether microexons were also affected in this disease. Due to a limited access to human patient data, and the patient-specific variation in islet composition and thus the quality of publicly available RNAseq, rather than performing an alternative splicing analysis, we checked the general trend of microexon inclusion in two Type 2 Diabetes RNAseq datasets from (Fadista et al., 2014) and (Marselli et al., 2020). Consistent between these two studies we observed significantly more skipping of microexons in human islets.

Finally, we summarised all RNAseq data analysed in this Chapter to provide a data-informed list of putative microexon targets for further studies (Figure 8B). To this end, we looked at EndoMICs regulated by SRRM3/Srrm3 in EndoC-B1 human beta cell line Figure 3), INS1E rat beta cell (Figure 5) line and *Srrm3-/-* mouse data (Figure 6) and two T2D patient datasets data from (Fadista et al., 2014; Marselli et al., 2020).

We selected 28 EndoMICs with information available in at least two of these categories. Among the presented targets, 14 EndoMICs (highlighted in bold) had available information in at least two categories and were found to be skipped in T2D human donor islets.

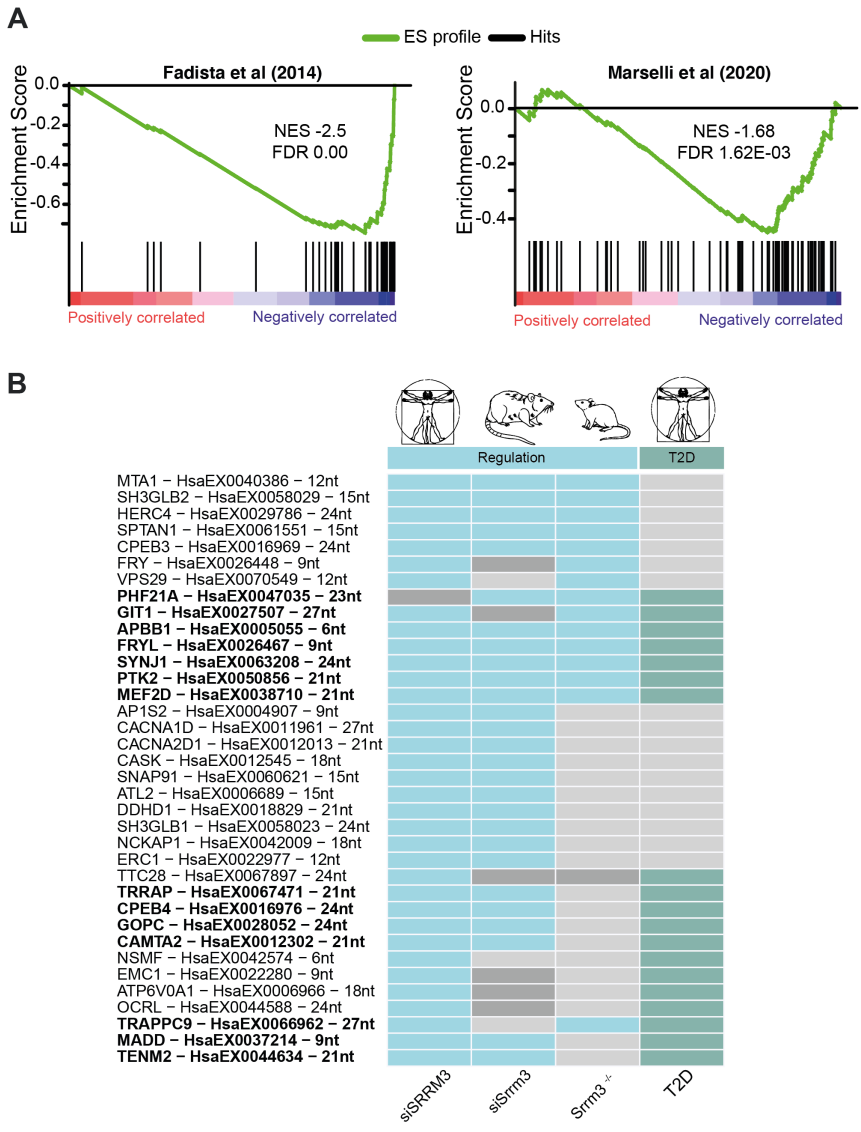

Figure 8. Microexons tend be skipped in Type 2 diabetic patients. A) Alternative Splicing Set Enrichment Analyses for EndoMICs found in two datasets obtained from Type 2 Diabetes human islet donors. **B)** Summary of human EndoMICs for target selection. Category-coloured cells indicate 'Yes', dark grey cells – 'No', grey cells – missing values.

These EndoMICs represented different functional aspects in beta cell biology, such as regulation of gene expression (*MEF2D*, *APBB1, PHF21A, TRRAP, FRYL, CPEB4*), exocytic machinery (*SYNJ1*, *GOPC*), cell signalling (*PTK2*, *GIT1*, TRAPPC9, *MADD*) and calcium homeostasis (*CAMTA2*, *TENM2*).

2.4. Closing remarks

In this Chapter, we validated the program of microexons in endocrine pancreas and their regulation by SRRM3 in rat and human beta cell lines. We showed that depletion of Srrm3 affects secretory capacity of beta cells. Our studies with constitutive *Srrm3*$^{-/-}$ mice confirmed the importance of EndoMICs and Srrm3 regulation in pancreatic islet biology, including secretory functions and beta cell identity. We showed that the loss of Srrm3 leads to hyperinsulinemic hypoglycemia, suggesting that deregulation of EndoMICs contributes to insulin-related pathologies. We also provided a list of EndoMICs for further studies in rat, human and mouse models, with several EndoMICs found in genes related to Type 2 Diabetes.

2.5. Methods

All experimental procedures were performed by Jonàs Juan-Mateu from our group and Amaya Lopez Pascual and Marta Miret-Cuesta from the Irimia group. Individual contributions to the projects were specified in the figure legends.

Identification of EndoMICs programs

VastDB inclusion tables for human (*hg38*) and mouse (*mm10*) were downloaded from VastDB website (link below) (Tapial et al., 2017b). We manually curated available RNA sequencing data in NCBI SRA repository for selected tissues for *Rat norvegicus* (*rno*) and created an inclusion table in accordance with VastDB standards.

Tissue-enriched inclusion of microexons was defined using *Get_Tissue_Specific_AS.pl* script considering the following: i) ΔPSI between a target tissue inclusion average and the averages across other tissues (**--min_dPSI 15**), ii) global ΔPSI, ie. difference between target tissue inclusion average and the average of other tissues **(--min_dPSI_glob 25**), iii) coverage in at least N tissues **(--N_groups 5**) and iv) inclusion value in at least n sample per tissue group (**--min_rep 2**, **--min_rep 1** for *rno*). To capture microexons with biased inclusion in certain tissue groups, we excluded from the tissues with known partial overlap of microexon inclusion from comparisons in those

tissues (for example Muscle and Heart from Neural comparison).

Code availability:

https://github.com/vastgroup/vastdb_framework_code_example/

VastDB:

https://vastdb.crg.eu/wiki/Downloads

Gene ontology analysis

Gene ontology analyses were performed *gprofiler2* R package (Kolberg et al., 2020), with custom annotations and gene backgrounds. Human gene ontology annotation was obtained from Ciampi et al (2021, in submission).

For GO analysis of human EndoMICs, we provided custom gene background of genes harbouring exons and introns passing filtering criteria of *Get_Tissue_Specific_AS.pl* in endocrine pancreatic group. We identified 106 ontology terms for Biological Process passing FDR ≤ 5%. We then manually curated ontology terms and collapsed them into 10 generalised functions important for the endocrine pancreas, taking the FDR value of the largest terms within the new group. GO analyses for rat and mouse RNA sequencing data were performed with the gene orthologs in respective species.

RNA sequencing

RNA quality was checked using Bioanalyzer (Agilent) and strand-specific Illumina libraries were prepared and sequenced at CRG Genomics Unit. Samples were sequenced on HiSeq2500 and 125 nts paired-end reads were generated.

Alternative splicing analysis

Pair-end FASTQ reads were aligned with *vast-tools* software (v2.5.1) (Irimia et al., 2014; Tapial et al., 2017b) and the following database versions: i) human **vastdb.hs2.23.06.20** based on **GRCh38.p13 Ensembl** genome assembly ii) rat **vastdb.rno.23.06.20** based on **Rno_6.0 Ensembl** genome assembly, iii) mouse **vastdb.mm2.23.06.20** based on **GRCm38.p5 Ensembl** genome assembly.

Briefly, VAST-TOOLS fragments the reads into 50nt read groups of which one random sub-read is used for quantification to avoid counting bias. 50nt reads are then aligned to a reference genome and mapped to splice junction database. Alternative exons were defined by 10 < average PSI < 90 in a group, non-overlapping PSI distribution between two sample groups **(--min_range 5**) Default threshold for differentially spliced exons was |ΔPSI| ≥ 15 (**--min_dPSI 15**), unless otherwise indicated, and the coverage of minimum LOW of at least 15 reads.For more details on methods, read threshold

and commands, refer to vast-tools website (https://github.com/vastgroup/vast-tools/).

Gene expression analysis

Gene counts were obtained with *STAR* software (v.2.7.1a) (Dobin et al., 2013). Genome indices built for human (*hg38*), mouse (*mm10*) and rat (*rn6*) using read_length – 1 per author's recommendation. Pair-end FASTQ files were then aligned using the following parameters: **--twopassMode Basic --quantMode GeneCounts**.

Gene counts were processed using DESeq2 R package (v1.32.0) (Love et al., 2014). We filtered out low gene counts by keeping genes with a row sum of at least 10. Gene counts were normalised to respective library sizes. The following models were fitted: i). for human SRRM3 KD - **~Condition (siRNA or control) + Replicate**, ii) for rat Srrm3 KD: **~Condition + Replicate,** and iii) for mouse knockout $Srrm3^{-/-}$ **~Sex+Age+Genotype**. Log2 fold changes were estimated with apeglm shrinking of effect size from *apeglm* R package using (Zhu et al., 2019).

Gene set enrichment analysis

GSEA analyses were performed with *GSEAPreranked* module of *GSEA* software (v4.1.0) (Mootha et al., 2003; Subramanian et al., 2005), and the following parameters: i) Number of

permutations **10000**, ii) Collapse/Remap to gene symbols: **No_collapse**, and iii) Min.size **5**.

We performed Alternative Splicing Set Enrichment Analyses (ASSEA) with the same software and parameter, providing a set of exons and ranked list of exons with coverage and their ΔPSI values for a given dataset.

Culture of rat INS-1E and human EndoC-βH1 beta cell lines

Rat insulin-producing INS-1E cells, kindly provided by Dr. C. Wollheim (University of Geneva, Geneva, Switzerland), were cultured in RPMI 1640 GlutaMAX-I medium (Invitrogen) as described previously (Ortis F et al. Mol Endocrinol. 2006). Human insulin-producing EndoC-βH1 cells, purchased from Univercell Biosolutions, were grown on plates coated with Matrigel/fibronectin (100 and 2 µg/ml, respectively, Sigma), and cultured in DMEM as described previously (Ravassard et al., 2011).

Gene silencing

The small interfering RNAs (siRNA) targeting the rat and human *SRRM3* gene used in this study were obtain from ON-TARGETplus siRNA Dharmacon, (Horizondiscovery). Allstars Negative Control siRNA (Qiagen, Venlo, Netherlands) was used as a negative control (siCTL). Transient transfection was performed using 30 nmol/L siRNA and Lipofectamine RNAiMAX (Invitrogen, Carlsbad, CA).

Mouse model

Srrm3 gene-trapped mice were kindly provided by Prof. Botond Banfi (University of Iowa, USA)(Nakano et al., 2019). Mouse genotyping was conducted on tail biopsies as described (Nakano et al. 2019) using published primers. In all analysis, wild-type littermates were used as controls. All protocols were carried out in accordance with the European Community Council Directive 2010/63/EU and approved by the local Ethics Committee for Animal Experiments (Comitè Ètic d'Experimentació Animal-Parc de Recerca Biomèdica de Barcelona, CEEA-PRBB).

Glucose tolerance test and measurement of circulating hormone levels

Adult mice (9-14 week old) were first fasted for 4h and then fed for 30 min. Blood glucose levels were measured from tail blood using a standard glucometer (Glucomen Aero 2K) at 4h fasting and 30 min postprandial. Plasma insulin was measured from tail blood at 30 min postprandial using a mouse insulin ELISA kit (Mercodia, Uppsala, Sweden).

Mouse islet isolation

Mouse pancreas from adult mice (9-14 week old) were perfused with a solution containing 1 mg/mL collagenase-P (Roche) and digested for 12-15 min at 37°C. Subsequently, pancreatic islets were isolated using a density gradient

purification with HISTOPAQUE-1077, washed and hand-picked (Villarreal et al., 2019). Isolated islets were cultured for 24h in RPMI 1640 medium supplemented with 10% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, 10 mM, HEPES, 1 mM sodium pyruvate and 0.05 mM 2-mercaptoethanol.

Insulin secretion

INS-1E cells were pre-incubated for 1 h in glucose-free RPMI 1640 Gluta MAX-I medium (Life Technologies, Inc.) followed by incubation with Krebs-Ringer solution for 30 min. Cells were then exposed to 1.7, 17, or 17 mM glucose plus 20 µm Forskolin, or to 35 mM KCl during 30 min. EndoC-βH1 cells were preincubated with culture medium containing 2.8 mmol/L glucose for 18 h. Cells were incubated in Krebs-Ringer buffer for 1 h and sequentially stimulated with 1 mmol/L glucose, 20 mmol/L glucose, or 20 mmol/L glucose and 10 µmol/L Forskolin for 40 min, as described elsewhere (Andersson et al., 2015). Isolated mouse islets were preincubated in Krebs-Ringer containing 2.8 mmol/L glucose for 2h, and then stimulated sequentially with 2.8 mmol/L glucose and 20 mmol/L glucose.

Insulin release and insulin content were measured using the rat, mouse, or human insulin ELISA kit (Mercodia, Uppsala, Sweden) in cell-free supernatants and acid/ethanol-extracted cell lysates, respectively. Results were normalised by total protein content, determined by the Bradford dye method.

2.6. Tables

Table 1. Human pancreatic endocrine samples.

STUDY	SRA	TYPE	SOURCE	NREADS (M)
PRJNA306754	SRR3048070	Beta	Ackermann et al, 2015	52,194,897
PRJNA306754	SRR3048071	Beta	Ackermann et al, 2015	16,263,361
PRJNA306754	SRR3048072	Beta	Ackermann et al, 2015	13,068,171
PRJNA306754	SRR3048073	Beta	Ackermann et al, 2015	16,263,361
PRJNA306754	SRR3048074	Beta	Ackermann et al, 2015	74,954,527
PRJNA306754	SRR3048075	Beta	Ackermann et al, 2015	78,508,860
PRJNA306754	SRR3048076	Beta	Ackermann et al, 2015	73,291,067
PRJNA306754	SRR3048077	Beta	Ackermann et al, 2015	15,555,955
PRJNA306754	SRR3048078	Beta	Ackermann et al, 2015	12,328,421
PRJNA306754	SRR3048079	Beta	Ackermann et al, 2015	12,328,421
PRJNA306754	SRR3048080	Beta	Ackermann et al, 2015	13,010,125
PRJNA306754	SRR3048081	Beta	Ackermann et al, 2015	13,010,125
PRJNA306754	SRR3048082	Beta	Ackermann et al, 2015	10,223,206

PRJNA306754	SRR3048083	Beta	Ackermann et al, 2015	114,783,044
PRJNA306754	SRR3048055	Alpha	Ackermann et al, 2015	35,792,489
PRJNA306754	SRR3048056	Alpha	Ackermann et al, 2015	15,681,374
PRJNA306754	SRR3048057	Alpha	Ackermann et al, 2015	12,495,498
PRJNA306754	SRR3048058	Alpha	Ackermann et al, 2015	12,495,498
PRJNA306754	SRR3048059	Alpha	Ackermann et al, 2015	97,704,374
PRJNA306754	SRR3048060	Alpha	Ackermann et al, 2015	72,405,803
PRJNA306754	SRR3048061	Alpha	Ackermann et al, 2015	12,246,620
PRJNA306754	SRR3048062	Alpha	Ackermann et al, 2015	15,249,800
PRJNA306754	SRR3048063	Alpha	Ackermann et al, 2015	12,246,620
PRJNA306754	SRR3048064	Alpha	Ackermann et al, 2015	14,992,946
PRJNA306754	SRR3048065	Alpha	Ackermann et al, 2015	12,027,850
PRJNA306754	SRR3048066	Alpha	Ackermann et al, 2015	12,027,850
PRJNA306754	SRR3048067	Alpha	Ackermann et al, 2015	16,001,371
PRJNA306754	SRR3048068	Alpha	Ackermann et al, 2015	12,858,633
PRJNA306754	SRR3048069	Alpha	Ackermann et al, 2015	12,858,633

PPRJNA587101	SRR10384238	Alpha	Alvarez-Dominguez JR et al, 2020	33,445,063
PRJNA587101	SRR10384239	Alpha	Alvarez-Dominguez JR et al, 2020	32,615,784
PRJNA587101	SRR10384240	Beta	Alvarez-Dominguez JR et al, 2020	34,221,736
PRJNA587101	SRR10384241	Beta	Alvarez-Dominguez JR et al, 2020	34,221,736

Table 2. Summary table of EndoMICs and EndoLONGs with genomic conservation between human-rat and human-mouse.

Human - Rat		
	Conserved	Non-conserved
EndoMICs	70	7
EndoLONGs	39	52
Human - Mouse		
	Conserved	Non-conserved
EndoMICs	72	5
EndoLONGs	42	49

Table 3. Summary table of the predicted impact of EndoMICs and EndoLONGs on the open reading frame

Endocrine Pancreas		
	CDS	Other
EndoMICs	63	14
EndoLONGs	31	60
Neural		
	CDS	Other
EndoMICs	253	39
EndoLONGs	355	282

Table 4. Human EndoMICs genomic and regulatory conservation table.

HUMAN	**RAT**	**MOUSE**	**ΔPSI_HS2**	**ΔPSI_RNO**	**ΔPSI_MM2**
HsaEX0004128	RnoEX0009339	MmuEX0004901	76.53	77.94	73.91
HsaEX0004907	RnoEX0010152	MmuEX0005412	90.37	82.46	63.6
HsaEX0005055	RnoEX0010357	MmuEX0005519	81.49	60.09	73.33
HsaEX0005085	RnoEX0010382	MmuEX0005537	89.2	NA	NA
HsaEX0006681	RnoEX0012769	MmuEX0006695	46.96	NA	11.24
HsaEX0006689	RnoEX0012777	MmuEX0006704	66.79	55.24	27.05
HsaEX0006966	RnoEX0013261	MmuEX0006887	77.48	83.49	50.75
HsaEX0011961	RnoEX0015956	MmuEX0008762	35.38	53.05	50.66
HsaEX0012013	RnoEX0016085	MmuEX0008808	40.21	32.98	20.42
HsaEX0012052	RnoEX0016166	MmuEX0008835	68.62	75.09	NA
HsaEX0012060	RnoEX0016178	MmuEX6099084	54.27	48.1	50.7
HsaEX0012302	RnoEX0016587	MmuEX0009028	87.31	51.43	58.33
HsaEX0036507	RnoEX0016949	MmuEX0026936	87.87	-1.09	5.54

HsaEX0013877	RnoEX0018699	MmuEX0010235	54.43	50.49	55.82
HsaEX0014432	RnoEX0019523	MmuEX0010675	37.41	2.16	6.15
HsaEX0035450	RnoEX0020329	MmuEX0026173	25.4	10.37	10.34
HsaEX0016967	RnoEX0024034	MmuEX0012537	55.08	45.66	NA
HsaEX0016969	RnoEX0024042	MmuEX0012540	29.66	55.27	41.36
HsaEX0016976	RnoEX0024055	MmuEX0012543	25.46	42.6	35.4
HsaEX0018829	RnoEX0026692	MmuEX0014010	27.22	25.91	16.91
HsaEX0020438	RnoEX0029035	MmuEX0015285	32.38	2.34	NA
HsaEX0020578	RnoEX0029241	MmuEX0015403	79.09	NA	NA
HsaEX0020591	RnoEX0029275	MmuEX0015418	32.28	38.58	NA
HsaEX0021080	RnoEX0030071	MmuEX0015805	57.35	65.7	1.14
HsaEX0022280	RnoEX0031881	MmuEX0008514	87.29	83.36	78.76
HsaEX0022977	RnoEX0032836	MmuEX0017219	38.37	25.57	25.02
HsaEX0023051	RnoEX0032971	MmuEX0017296	29.59	-1.86	19.7
HsaEX0023360	RnoEX0033412	NA	66.16	9.66	NA
HsaEX0026448	RnoEX0036957	MmuEX0019666	41.95	75.34	86.55
HsaEX0026467	RnoEX0036977	MmuEX0019673	56.46	74.46	85.11
HsaEX0027507	RnoEX0038505	MmuEX0020446	65.57	50.09	72.57
HsaEX0028052	RnoEX0039278	MmuEX0021486	36.94	25.72	43.45
HsaEX0028492	RnoEX0039735	MmuEX0021779	63.36	11.73	6.84
HsaEX0029786	RnoEX0041675	MmuEX0022805	25.75	47.36	64.98
HsaEX0035103	RnoEX0047965	MmuEX0025977	44.8	NA	NA
HsaEX0012545	RnoEX0048142	MmuEX0009244	30.0	32.04	20.17
HsaEX0037214	RnoEX0051839	MmuEX0027514	37.65	18.19	12.63
HsaEX0038710	RnoEX0053753	MmuEX0028570	77.62	68.08	59.55
HsaEX0040386	RnoEX0056153	MmuEX0029871	49.97	65.98	61.92
HsaEX0040951	RnoEX0056975	MmuEX0030346	84.55	NA	NA

HsaEX0042009	RnoEX0058510	MmuEX0031113	77.6	78.88	64.97
HsaEX0042667	RnoEX0059513	MmuEX0031572	47.79	50.05	92.62
HsaEX0043771	RnoEX0060933	MmuEX0032311	81.72	48.97	13.71
HsaEX0042574	RnoEX0061203	MmuEX0031485	64.24	21.91	49.06
HsaEX0044588	RnoEX0062060	MmuEX0032953	39.31	35.85	52.53
HsaEX0045196	RnoEX0062995	MmuEX0033441	57.19	1.46	18.3
HsaEX0047035	RnoEX0065375	MmuEX0034764	76.74	76.05	83.64
HsaEX1002327	RnoEX0066326	MmuEX0035345	39.17	5.73	11.24
HsaEX0049351	RnoEX0068537	MmuEX0036514	71.0	2.76	11.45
HsaEX0050856	RnoEX0071038	MmuEX0037737	48.53	63.9	57.61
HsaEX0050855	RnoEX0071039	MmuEX0037746	35.87	-0.44	18.53
HsaEX0051110	RnoEX0071375	MmuEX0037892	42.28	NA	27.07
HsaEX0055468	RnoEX0076838	MmuEX0040382	31.04	60.95	NA
HsaEX0057701	RnoEX0080130	MmuEX0041968	74.72	87.65	-3.87
HsaEX0058023	RnoEX0080499	MmuEX0042195	50.81	25.09	27.04
HsaEX0058029	RnoEX0080506	MmuEX0042199	27.04	65.08	78.16
HsaEX0059697	RnoEX0082681	MmuEX0043333	28.25	8.51	0.02
HsaEX0060621	RnoEX0084229	MmuEX0044070	79.28	29.08	20.82
HsaEX0061551	RnoEX0085728	MmuEX0044723	65.26	58.12	44.9
HsaEX0063208	RnoEX0088100	MmuEX0045982	49.49	43.86	76.86
HsaEX0063243	RnoEX0088147	MmuEX0046015	32.22	18.12	31.12
HsaEX0063881	RnoEX0088919	NA	54.62	30.2	NA
HsaEX0064198	RnoEX0089348	MmuEX0046693	33.77	NA	NA
HsaEX0044634	RnoEX0089695	MmuEX0032988	45.27	35.71	NA
HsaEX0066962	RnoEX0093434	MmuEX0048796	37.95	36.09	14.62
HsaEX0067471	RnoEX0094213	MmuEX0049240	56.29	31.44	67.71
HsaEX0008744	RnoEX0094491	MmuEX0008444	35.84	NA	NA

HsaEX0067897	RnoEX0094713	MmuEX0049506	55.7	1.19	20.09
HsaEX0070549	RnoEX0098264	MmuEX0051510	54.47	46.06	51.6
HsaEX0070972	RnoEX0099033	MmuEX0051821	63.49	NA	NA
HsaEX0006858	NA	NA	26.12	NA	NA
HsaEX0015478	NA	MmuEX0011499	36.69	NA	1.1
HsaEX1043449	NA	NA	77.89	NA	NA
HsaEX0069169	NA	NA	25.32	NA	NA
HsaEX0005072	NA	MmuEX1004073	55.45	NA	55.62
HsaEX0004874	NA	MmuEX0005395	57.28	NA	9.7
HsaEX0043854	NA	MmuEX0032375	62.66	NA	NA

Table 5. Custom gene ontology annotation for human EndoMICs.

TERMS	**GENERALISED TERM**	**P-VALUE**	**GENES**
GO:0008016, GO:0060047 GO:0086091, GO:0061337 GO:0086010, GO:0086003 GO:0086001, GO:0099171 GO:0099623, GO:0060306 GO:0099622	Action potential	8.71E-04	ENSG00000069849,ENSG00000065609 ENSG00000067191,ENSG00000088538 ENSG00000145362,ENSG00000153956 ENSG00000157388,ENSG00000165995 ENSG00000108262
GO:2001257, GO:0006941 GO:0034220, GO:0006812	Cation channel activity and transport	2.88E-02	ENSG00000065609,ENSG00000067191 ENSG00000145362,ENSG00000153956 ENSG00000165995,ENSG00000179915 ENSG00000134769,ENSG00000157388 ENSG00000033627,ENSG00000047932

GO:0043269, GO:0032412 GO:0034765, GO:0030003 GO:0010959, GO:0043270 GO:0006811, GO:0051928 GO:1901019, GO:0032414 GO:1904064, GO:0034767 GO:1901021, GO:2001259 GO:1904427, GO:0070588 GO:0006816, GO:0070509 GO:1901016			ENSG00000069849,ENSG00000167703 ENSG00000005379,ENSG00000078668 ENSG00000125991,ENSG00000147044
GO:0007155, GO:0034113 GO:0007157	Cell adhesion	1.38E-02	ENSG00000163531,ENSG00000091129 ENSG00000133026,ENSG00000145934 ENSG00000147044,ENSG00000196776 ENSG00000198910,ENSG00000064309 ENSG00000179915,ENSG00000153707
GO:0007399, GO:0048699 GO:0022008, GO:0048666 GO:0031175, GO:0051960 GO:0060284, GO:0050767 GO:0045664, GO:0051962 GO:0010720, GO:0050769 GO:0045666	Neuroendocrine cell identity	4.16E-04	ENSG00000061676,ENSG00000163531 ENSG00000197694,ENSG00000107864 ENSG00000134769,ENSG00000169398 ENSG00000065609,ENSG00000088538 ENSG00000091129,ENSG00000116604 ENSG00000122126,ENSG00000133026 ENSG00000145362,ENSG00000145934 ENSG00000148737,ENSG00000159082 ENSG00000167632,ENSG00000198513 ENSG00000198910,ENSG00000064309 ENSG00000073910,ENSG00000075539 ENSG00000103494,ENSG00000153707 ENSG00000163697,ENSG00000165802 ENSG00000166313,ENSG00000179915 ENSG00000079691

GO:0000902, GO:0051130 GO:0048858, GO:0032990 GO:0120035, GO:0031344 GO:0120039, GO:0000904 GO:0048667, GO:0120031 GO:0097485, GO:0031346 GO:0120032, GO:0060491 GO:0010927, GO:0050808	Cell morphology	2.87E-04	ENSG00000061676,ENSG00000163531 ENSG00000197694,ENSG00000079691 ENSG00000169398,ENSG00000065609 ENSG00000088538,ENSG00000091129 ENSG00000133026,ENSG00000198513 ENSG00000198910,ENSG00000073910 ENSG00000075539,ENSG00000100266 ENSG00000153707,ENSG00000163697 ENSG00000165802,ENSG00000166313 ENSG00000179915,ENSG00000074054 ENSG00000097033,ENSG00000100523 ENSG00000107864,ENSG00000134769 ENSG00000106976,ENSG00000145934 ENSG00000159082,ENSG00000196776 ENSG00000111490,ENSG00000122126 ENSG00000078668,ENSG00000103494 ENSG00000145362,ENSG00000177380
GO:0099537, GO:0098916 GO:0099177, GO:0050804 GO:0086019	Cell-cell signalling	7.06E-07	ENSG00000108262,ENSG00000008282 ENSG00000107864,ENSG00000134769 ENSG00000005379,ENSG00000065609 ENSG00000067191,ENSG00000106976 ENSG00000147044,ENSG00000157388 ENSG00000159082,ENSG00000165995 ENSG00000177380,ENSG00000078668 ENSG00000100266,ENSG00000153707 ENSG00000165802,ENSG00000179915 ENSG00000145362,ENSG00000153956
GO:0032940, GO:0006887 GO:1903530, GO:0051046 GO:0099643, GO:0007269 GO:1903532, GO:0051047 GO:0016079, GO:0017156 GO:0017157, GO:1903305	Exocytosis	7.19E-03	ENSG00000108262,ENSG00000163531 ENSG00000197694,ENSG00000033627 ENSG00000074054,ENSG00000005379 ENSG00000065609,ENSG00000133026 ENSG00000147044,ENSG00000148737 ENSG00000157388,ENSG00000159082 ENSG00000165995,ENSG00000167632 ENSG00000177380,ENSG00000196776 ENSG00000166313,ENSG00000166747 ENSG00000179915

GO:2000300, GO:0046928 GO:0045921, GO:1903307			
GO:0043087, GO:0031267	Regulation of GTPase activity	0.050363456110191541	ENSG00000108262,ENSG00000137449 ENSG00000169398,ENSG00000065609 ENSG00000111490,ENSG00000122126 ENSG00000128512,ENSG00000067208 ENSG00000082805,ENSG00000110514
GO:0051050, GO:0022898 GO:0034762, GO:0034764	Transmembrane transport	5.04E-02	ENSG00000069849,ENSG00000074054 ENSG00000097033,ENSG00000065609 ENSG00000106976,ENSG00000133026 ENSG00000145362,ENSG00000147044 ENSG00000148737,ENSG00000153956 ENSG00000157388,ENSG00000159082 ENSG00000165995,ENSG00000167632 ENSG00000196776,ENSG00000166313 ENSG00000166747,ENSG00000047932 ENSG00000067191,ENSG00000179915
GO:0051050, GO:0022898 GO:0034762 GO:0034764	Vesicle-mediated transport and localisation	4.45E-04	ENSG00000108262,ENSG00000047932 ENSG00000074054,ENSG00000065609 ENSG00000106976,ENSG00000122126 ENSG00000147044,ENSG00000157388 ENSG00000159082,ENSG00000165995 ENSG00000196776,ENSG00000100266 ENSG00000166747,ENSG00000067191 ENSG00000167632,ENSG00000177380 ENSG00000179915,ENSG00000097033 ENSG00000148737,ENSG00000008282 ENSG00000134769,ENSG00000078668 ENSG00000069849,ENSG00000145362 ENSG00000153956,ENSG00000197694 ENSG00000125991

CHAPTER 3

Regulation of neuroendocrine microexons is achieved by differential sensitivity to SRRM3/4

3. Chapter 3: Nested regulation of neuroendocrine microexons is achieved by differential sensitivity to SRRM3/4

Preface to Chapter 3

In this Chapter, I will build upon the observation that EndoMICs are almost fully contained in a larger program of neural microexons. First, I will provide evidence that EndoMICs and NeuralMICs (that are not found in endocrine pancreas) display differential sensitivity to SRRM4. I will then present our approach to study this model on a high-throughput scale and discuss the general experimental and analytical pipelines. I will also highlight some important issues that we identified and discuss our preliminary findings on the impact of exon lengths on the sensitivity to SRRM4.

3.1. NeuralMICs and EndoMICs are regulated by different level of SRRM4 expression

In the previous Chapter, we identified a program of EndoMICs in endocrine pancreas and showed that it forms a nested program contained within a larger program of neural microexons. We also showed that EndoMICs are regulated by SRRM3 rather than SRRM4. However, one of the outstanding questions that we did not answer was the mechanism by which endocrine pancreas utilises only a subset of microexons from the full neural-enriched program. Neural-exclusive microexons, i.e. not shared with endocrine pancreas, are refer to as NeuralMICs.

Preliminary data demonstrated that different levels of over-expression of different eMIC domain from several species generated by (Torres-Méndez et al., 2019) displayed different microexon-enhancing activities: b (i.e. eMIC activity). y looking at the inclusion profile of EndoMICs and NeuralMICs (defined as those neural-enriched microexons that are not EndoMICs), we observed that EndoMICs required lower eMIC activity than NeuralMICs (data not shown). This suggested an interesting regulatory model based on differential sensitivity of microexons to SRRM3/4, where the most sensitive microexons respond to lower levels of expression / activity of these regulators. Interestingly, as EndoMICs seemed to be more sensitive to SRRM4 constructs than NeuralMICs, it would explain how the nested program was regulated.

To investigate the regulation of this nested program of microexons, we took advantage of the available doxycycline-inducible human SRRM4 cell line generated by (Torres-Méndez et al., 2019) to titrate its expression in cells not expressing it endogenously (HEK293) and assess the differential response of Neural and EndoMICs (see Methods).

We therefore performed a titration of doxycycline, to progressively increase the level of expression of human SRRM4 and performed RNA deep sequencing analysis. Consistent with the previous findings using only eMIC domains from different species, we observed that indeed EndoMICs required lower levels of SRRM4 to become included compared to NeuralMICs (Figure 1A). We wondered whether these differences could be attributed to different strength and/or composition of the *cis*-acting elements and splice sites in the flanking introns. However, we did not observe any pronounced differences between these two groups of microexons in terms of these genomic features (Figure 1B).

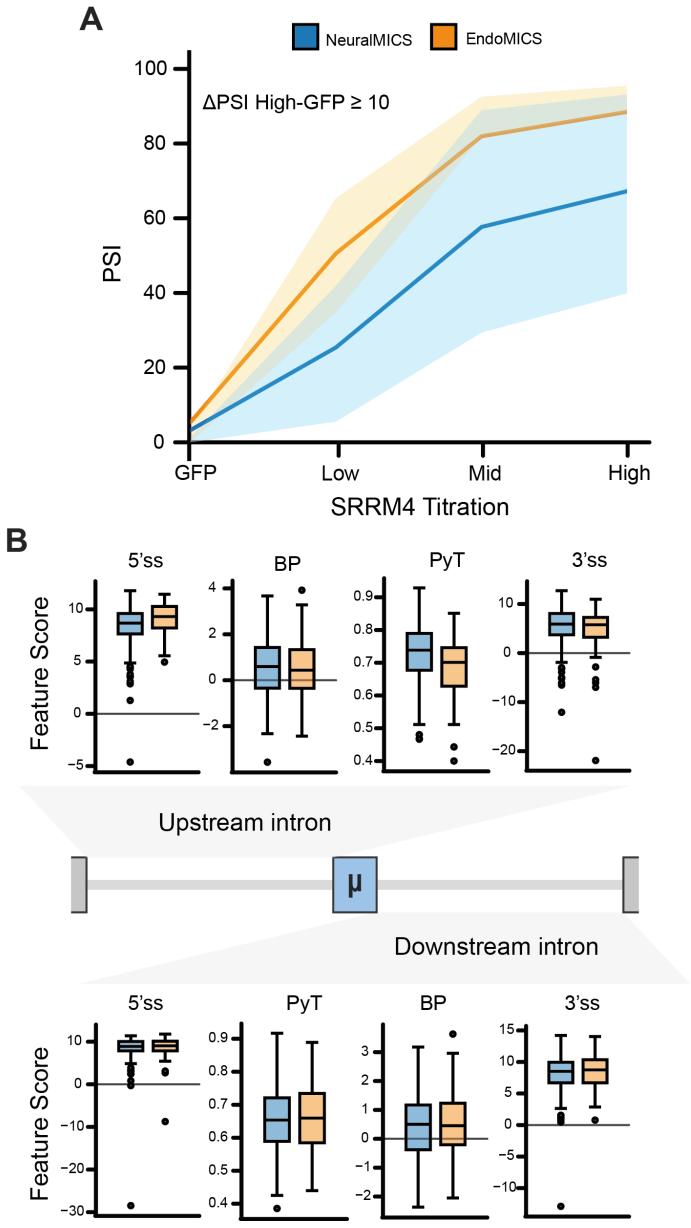

Figure 1. Nested program of NeuralMICs and EndoMICs as a model for differential sensitivity to SRRM3/4. **A)** Differential sensitivity of endogenous NeuralMICs and EndoMICs with ΔPSI ≥ 10 (High-GFP) upon over-expression of SRRM4 in Flp-In T-Rex 293 cell line. The lines

correspond to the median inclusion values per program; shades represent the interquartile range. **B)** Selected genomic features for human Neural and EndoMICs 5'ss – 5' splice site (Maximum Entropy), 3'ss – 3' splice site (Maximum Entropy), PyT – polypyrimidine tract (median pyrimidine content), BP – branch point (median branch point score), μ – microexon. Genomic features were obtained using Matt software.

3.2. Library of Micro(&)Exons to study the model of microexon sensitivity to SRRM3/4

a) Barcoded minigene libraries to study properties of microexons on a high-throughput scale

The most distinct feature of microexons is their length. Previous work established that their short length establishes special cis- and trans-acting requirements for their recognition (Gonatopoulos-Pournatzis et al., 2018; Quesnel-Vallières et al., 2015; Torres-Méndez et al., 2019). As the genomic features in intronic sequences of NeuralMICs and EndoMICs do not alone explain the differential sensitivity to SRRM3/4, we wondered how their lengths, in the context of other features, may affect their regulation.

To answer this question, we designed a high-throughput library (Library of Micro(&)Exons) using minigene constructs (Figure 2A). The basic unit contained a barcode sequence at the 5' of the transcribed region – a unique sequence of 34 nucleotides (see Methods)-, the upstream and downstream exons of a microexon present in the *DNM1* gene with their flanking

intronic sequences (5' end of the downstream intron for the 5' exon and 3' end of upstream intron for the 3' exon) intron of 3'exon) and a minimal unit of microexon variants required for the inclusion of the microexon, covering microexon sequences, 93 nucleotides of the upstream intron and 25 nucleotides of the downstream intron, respectively, in order to preserve the branch point sequence, polypyrimidine tract and 3' splice site AG and the 5’ splice site.

The use of barcodes was designed to allow us to assign each microexon sequence in the library to a specific barcode mark and therefore be able to assign each spliced transcript to a specific pre-mRNA transcript variant. Thus, we should be able to assign skipping reads to specific variants and therefore facilitate the assessment of exon inclusion/skipping ratios. In fact, a similar approach has been previously used to investigate sequence determinants of the 5’ splice site recognition and deciphering general rules that regulate alternative splicing (Mikl et al., 2019; Rosenberg et al., 2015; Wong et al., 2018a).

For the library we selected 241 events, of which 213 microexons belonged to either the NeuralMICs or EndoMICs programs. As control, we also included 28 constitutive exons (42 nts long) and 39 cryptic exons. Among the selection criteria for all exons were the presence of TGC or TCTC motifs in the upstream intron (corresponding to SRRM3/4 and SRSF11 binding, respectively) and their regulation in response to

SRRM3/4. Length variants were generated for a given exon from the middle point of its sequence, in stepwise 3 nucleotide sequence extensions or deletions, the variants ranging in length from 6-42 nucleotides (Figure 1B). For instance, a 15-nt microexon was extended with three different donor sequences to create variants of 18, 21, 24, ..., 42 nts. In addition, deletion variants corresponded to 12, 9 and 6 nts. The pool of designed variants was then cloned into the barcoded *DNM1* minigenes replacing the *DNM1* microexon, to assemble the library input (Figure 1C, left). The library input was subsequently transfected into HEK293 cells (see Methods) and SRRM4 was titrated in three expression tiers. Isolated RNAs from the transfected cells, analysed by RT-PCR using expression vector-specific oligonucleotide primers and deep sequencing of the amplification products, constitute the output of the experiment (Low, Mid and High).

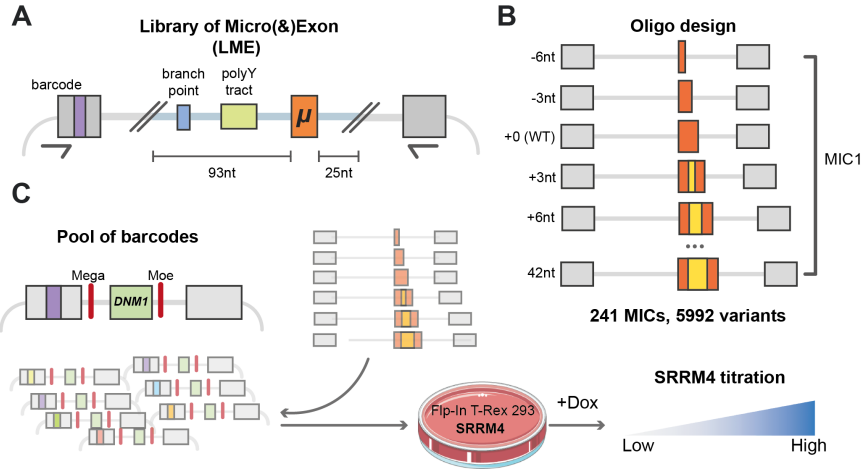

Figure 2. Design, generation, and experimental procedure for the analysis of the high-throughput Library of Micro(&)Exons (LME). **A)** Schematic illustrating the general structure of the barcoded LME constructs. On a backbone of flanking exons and intronic sequences, a library of microexons and their corresponding flanking intronic sequences (93 nts of the 3' end of the upstream intron and 25 nts of the 5' end of the downstream intron) was generated. Barcode sequences at the 5' end of the transcribed region uniquely identify each library variant. **B)** Schematic of the LME library design showing variant generation from a wild-type exon sequence (WT). Sequences were either removed or added from the middle point of a given exon in 3 nt steps. **C)** Experimental procedure prior to RNA sequencing. The constructs were prepared in two-step cloning strategy: i) barcodes were cloned in the *DNM1* minigene, and ii) oligonucleotides designed for variants bearing the cloning sites Mega and Moe were cloned in the DNM1 minigene, replacing 93 nts of downstream intron – DNM1 exon – 25 nts of upstream intron. Fln-In T-Rex 293 cells were then transfected with the pool of plasmids bearing all variants and SRRM4 was titrated using different concentrations of doxycycline. LME was designed by Manuel Irimia. Library cloning and experiments were performed by Sophie Bonnal.

We sequenced the library input in triplicate using paired-end sequencing with an average of ~4.3 million reads per replicate. In total, we identified over 93 thousand barcode-variant associations (BVAs) for 5635 out of 5992 variants (see Methods) (Figure 3A). Our barcoding and cloning strategy should generate at random a unique association between a barcode and a variant. However, we observed that while 87% of the associations supported by at least 5 reads were unique, the remaining 13% of the BVAs were ambiguous (Figure 3B); i.e. a given barcode was associated with more than a variant. We attributed this to template switching of extending DNA

polymerase during the PCR-amplification of the library input, a known factor that introduces variation in high-throughput data (Figure 3C) (Kebschull & Zador, 2015). In this model, barcode ambiguity emerges from Taq polymerase switching template during amplification of region of the minigene vector between the barcode (upstream exon and the remaining part of the intron) that is shared between all library constructs (see Methods) and the insert variant. For the presented Library, the effect of template switching on BVAs was reduced, albeit not eliminated, by increasing the extension time for amplification of the input library and decreasing the number of cycles during the PCR amplification (data not shown).

Next, we resolved the ambiguous assignments to obtain a set of trusted BVAs. Briefly, we considered BVAs supported by at least 5 reads. For a barcode with only two associated variants, we decided to keep the most supported variant provided that the support of the second variant represented less than 10% of the number of reads supporting the predominant variant. For a barcode with more than two mis-assignments (i.e. at least three associated variants), we then identified the second most supported variant. We discarded that barcode if the second variant was supported by more than 2 reads; otherwise, we kept the most supported associated variant. Our final set of trusted BVAs contained associations supported by at least 5 reads and present in at least 2 replicates, plus a set of BVAs found in any single replicate that were supported by at least 10 reads. We removed variants that were supported by

1 barcode only. In total, we obtained information for 5381 variants (out of 5992 variants included in the library design -89.8%-) assigned to over 74 thousand barcodes.

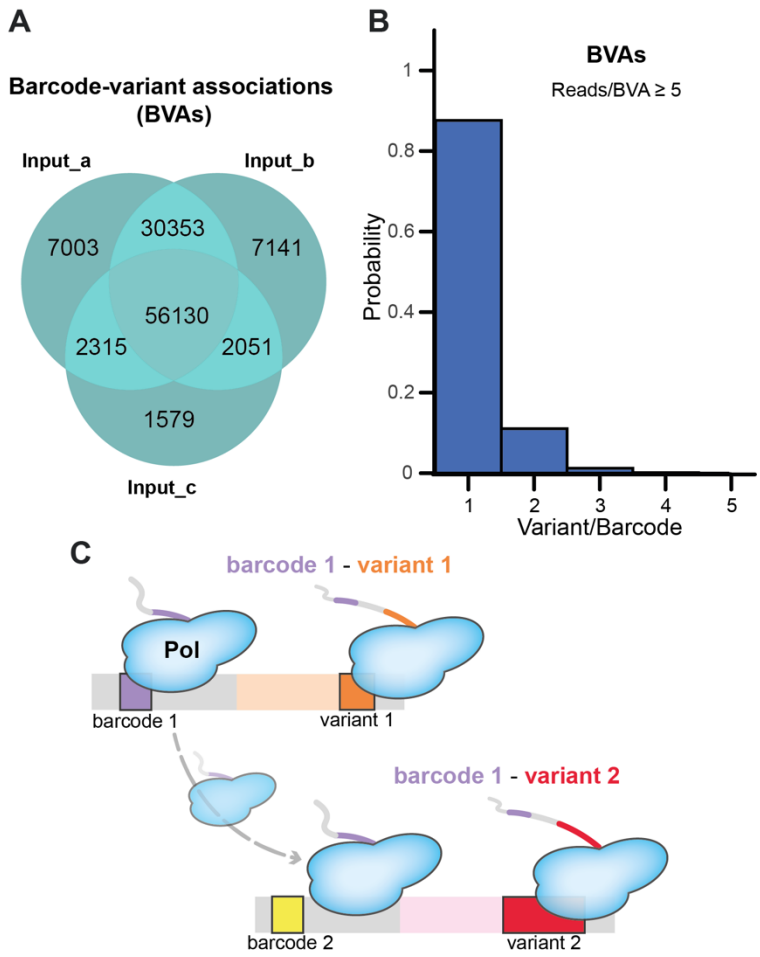

Figure 3. PCR-amplification of the LME input library introduces variant mis-assignments. **A)** Overlap of barcode-variant associations (BVA) between input sequence replicates. **B)** Distribution of the variant number per barcode in input sequencing data for barcode-variant associations, supported by at least 5 reads, showing that while the majority (89.8%) of assignments are unique, a fraction of barcodes was assigned more than 1

variant. **C)** Schematic illustrating the template switching mechanism during PCR amplification of the input library. Pol – Taq DNA polymerase.

Previous work in many systems has demonstrated that alternative splicing can be influenced by a multitude of sequence elements and locations in the pre-mRNA, reviewed by (Wang & Burge, 2008; Witten & Ule, 2011). We therefore next wondered whether barcode sequences could affect the inclusion of a variant associated with it. To test this, we sequenced the input library of barcoded *DNM1* minigenes prior to their replacement with the pool of microexon variants. The *DNM1* input library was transfected in Flp-In T-Rex 293 cell line and SRRM4 expression was induced with doxycycline (High concentration, see Methods). We quantified the inclusion of *DNM1* microexon in control cells (GFP) and in conditions of SRRM4 over-expression (Figure 4A). While the median inclusion of *DNM1* microexon increased upon over-expression of SRRM4, we observed a distribution of inclusion values in both conditions that did not depend on the number of the reads. To check whether the inclusion values were affected at random, we selected barcodes associated with *DNM1* microexon from the 1st and the 4th quartile of the distribution in control (GFP) cells (Figure 4B). We then compared this distribution with the distribution of inclusion of these associations under conditions of SRRM4 over-expression. We reasoned that if barcodes had a strong effect on the inclusion levels of the microexon, barcodes would alter inclusion similarly in both conditions, and consequently they would be

more often observed in the same quartiles of the two distributions.

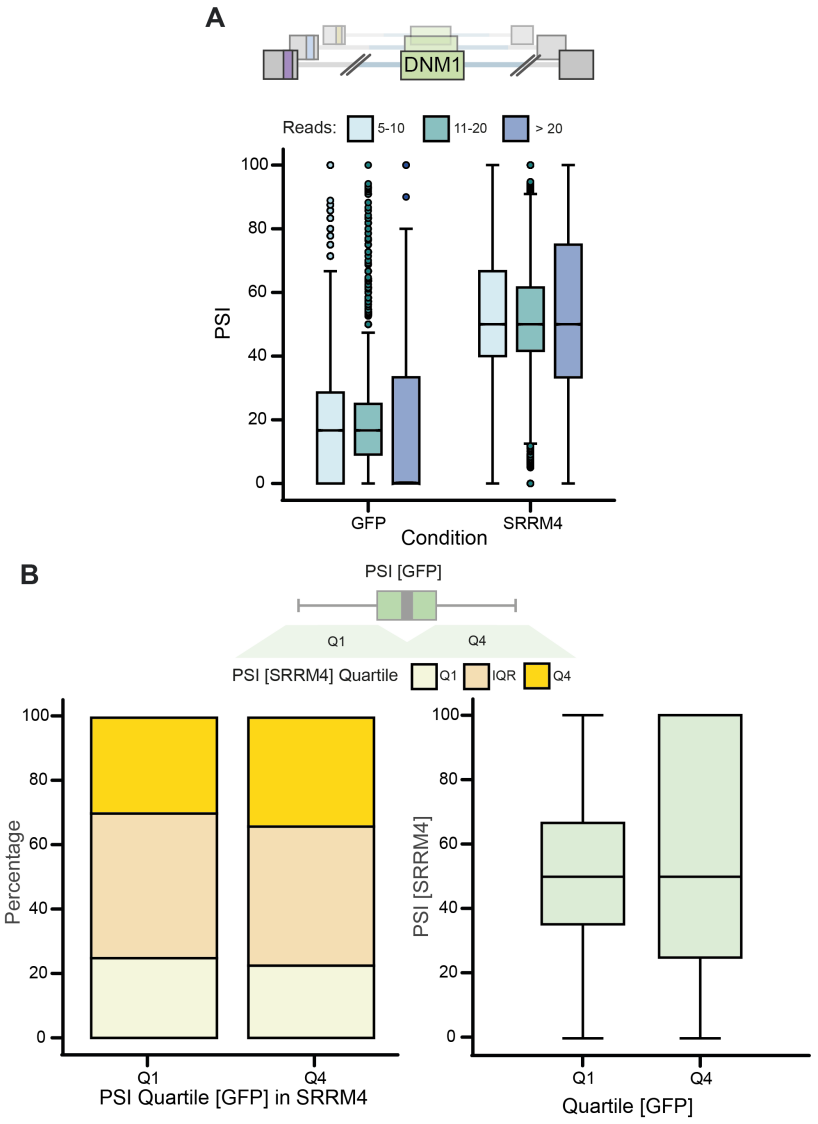

Figure 4. Barcodes seem to stochastically influence inclusion level of their associated exon. A) Distribution of inclusion levels for the pool of barcoded *DNM1* minigene constructs (first step of the library cloning) when transfected into control (GFP) and High expression of SRRM4 over-

expression Flp-In T-Rex 293 cells. Colored bars indicate results stratified by the number of reads for barcode-DNM1 associations supported by at least 5 reads. **B)** Left: Quartile distribution of inclusion levels in High expression of SRRM4 cells for barcodes found in quartile Q1 and quartile Q4 in GFP control cells. Right: PSI distribution corresponding to Left panel.

We observed, however, that they distributed across the full range of inclusion values. This, and the ambiguous BVAs from the library input had important implications for the quantification of the inclusion level of the variants. Using the set of trusted BVAs, we then quantified the inclusion level of the variants in the LME upon over-expression of SRRM4. However, considering the ambiguous BVAs identified in the library input and the stochastic effect of the barcodes on exon inclusion, we needed to resolve two issues. First, the quantification of the skipped variants is only as reliable as the BVAs. The mis-assigned variants would still be quantified for a given barcode's skipping reads, leading to a quantification bias. To resolve this, we quantified the inclusion of mis-assigned variants and calculated their proportion to the sum of inclusion reads of the correctly assigned variant. We then used this metric as a proxy for the correction of skipping reads. Specifically, we multiplied the number of skipped reads for the trusted BVA by a fraction of that proportion (see Methods).

Second, to correct for the stochasticity of inclusion influenced by the presence of barcodes, we calculated the median inclusion per variant in each condition and removed the outliers

based on the difference between calculated inclusion level and its median (see Methods).

After these two steps of filtering, we calculated mean inclusion values per variant and selected variants present in at least three conditions. This provided data for 5212 variants (87% of the total number of variants in the original library design). For this Chapter, we focused on NeuralMICs and EndoMICs identified in Chapter 2, and a set of constitutive and cryptic exons.

To assess the impact of the length extension or deletion, we first checked how reliably the behaviour of the wild type clones in the library reflected the levels of microexon inclusion of the endogenous genes. For this we examined the levels of inclusion of the NeuralMICs and EndoMICs wild-type constructs by deep sequencing of RT-PCR products and compared them to the levels of inclusion in endogenous genes, as determined by RNAseq. While the trend of the response to SRRM4 was preserved, we observed generally lower inclusion levels in the context of the library relative to the endogenous inclusion profile (Figure 5A). As further elaborated in the Discussion section, these discrepancies could be related to dependencies on sequences not included in the minigenes, the chimeric nature of library-derived transcripts, different levels of expression of library-derived and endogenous pre-mRNAs, absence of chromatin organization in plasmids, etc. In any case, to be able to move forward with our analyses, we

selected the most reliable (i.e. resembling endogenous inclusion level) wild-type constructs by calculating the absolute difference in ΔPSI between the inclusion at High SRRM4 levels and inclusion in GFP control cells for both datasets. We discarded wild-type constructs with an absolute difference higher than 50. We then compared the sensitivity to SRRM4 between Neural and EndoMICs wild-type construct and recapitulated the profile observed for these microexons at the endogenous level (Figure 5B).

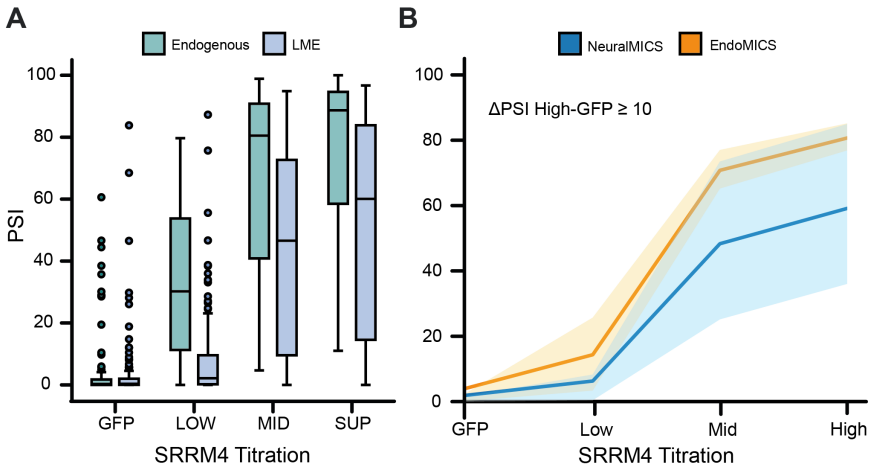

***Figure 5. The behaviour of wild-type exon transcripts generated from the LME library does not always reflect the endogenous level of inclusion upon SRRM4 overexpression*. A)** Comparison of inclusion levels for NeuralMICs and EndoMICs in endogenous and LME-context by RNAseq. **B)** Differential sensitivity for NeuralMICs and EndoMICs LME wild-type constructs upon overexpression of SRRM4 in Flp-In T-Rex 293 cells for a subset of constructs selected based on the similarity to the response to SRRM4 of inclusion in endogenous genes. The lines correspond to median inclusion values per program; shades represent the interquartile range.

Next, we investigated the effect of extension or deletion of sequences on the inclusion levels of NeuralMICs and EndoMICs with reliable behaviour of the wild-type constructs and reliable effects of variants generated with at least two different sequences (for length extension variants) (Figure 6). Although the extension sequences were designed not to contain splicing enhancer and silencer motifs (see Methods), it is possible that they display regulatory effects in the context of their insertion sites (e.g. through sequence motifs at the boundary of their insertion sites). Hence the filtering for consistent behaviour of variants generated with at least two different sequences.

We compared the effects of the sequences in two dimensions: 1) looking at the effect of the length extension in each condition and 2) the response to SRRM4 titration in the range of length extensions (3-9, 12-18 and 21-33 nucleotides). We observed that, consistently, Extension Sequences 0 and 1 promoted more inclusion compared to the wild-type constructs. This effect was particularly evident for the variants extended by 12-18 and 21-33 nucleotides with Extension Sequence 0 in control (GFP) and Low concentration of SRRM4 conditions. The response to SRRM4 for those variants, both for Sequence 0 and 1, plateaued in Mid concentration of SRRM4. Variants generated by addition of 3-9 nucleotides of Sequence 0 and 1 responded gradually to the titration of SRRM4, albeit with lower starting inclusion level in GFP control.

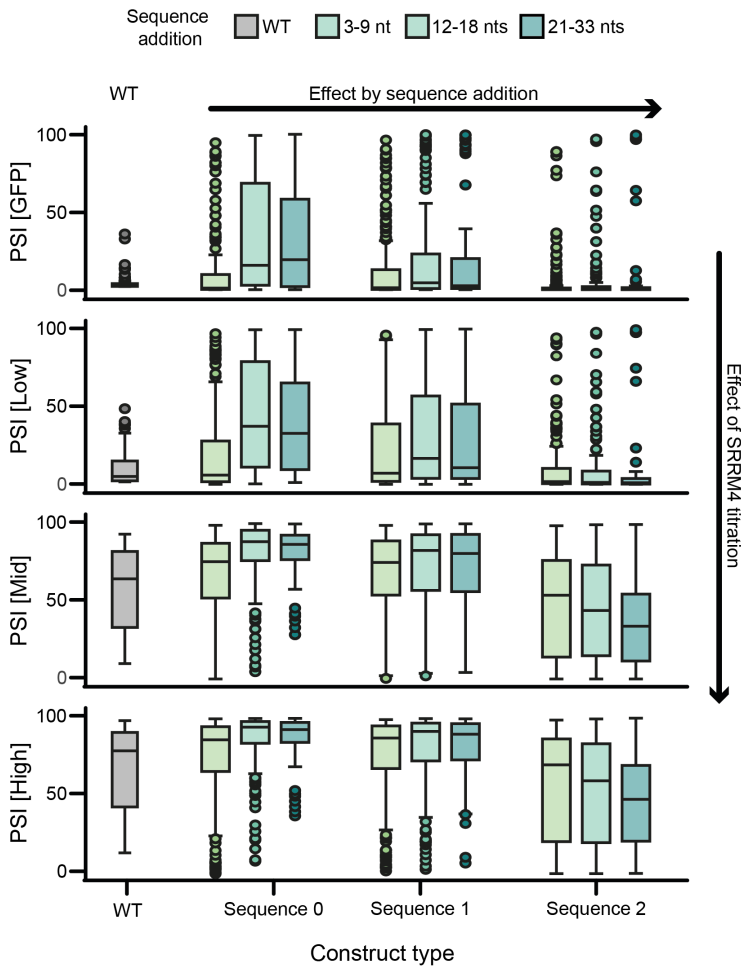

Figure 6. Sequences designed for length extensions introduce context-specific effects. General effect on the inclusion of length variants for Neural and EndoMICs per sequence used to extend exon length variants. Inclusion of wild-type constructs is indicated as a reference. Length extension was stratified by step size in three groups: 3-9 nts, 12-18 nts and 21-33 nts. Data shown correspond to the variants generated from the reliable (similar response to SRRM4 than endogenous) wild-type constructs.

In contrast, variants with Extension Sequence 2 displayed much lower inclusion level in the first two conditions and slower response to SRRM4 and, in general, lower inclusion compared to variants with Sequences 0 and 1. These results suggest that our "neutral" sequences are in fact far from neutral, with Sequence 0 likely harbouring splicing enhancer activity and Sequence 2 silencer activity (see Discussion).

b) Studying the impact of exon length on the regulation by SRRM4 using the Library of Micro(&)Exons

Thus far, we resolved reliable barcode-variant associations, quantified the inclusion level of the variants generated for this library, assessed the reliability of the data and the effect of the Extension Sequences on the variant inclusion for NeuralMICs and EndoMICs. As we identified similarities and differences between the Extension Sequences, we next assessed the inclusion profile for each of them separately to investigate the impact of microexon length perturbation on the sensitivity to SRRM4.

Impact of length addition on variants for NeuralMICs and EndoMICs

To investigate this, we calculated the change in inclusion between the variant and its wild-type construct (ΔPSI Var – WT). Like for the general effect of Extension Sequence on inclusion, we look at our data in two dimensions – *row-wise* for the changes between length addition classes and *column-wise* to assess the effect on the sensitivity to SRRM4 within each length class (Figure 7A).

We observed that variants of EndoMICs showed higher increases in their inclusion by length addition in control (GFP) cells and Low concentration of SRRM4 cells relative to their wild-type constructs and compared to NeuralMICs. Their response to SRRM4 also increased but plateaued in Mid and High concentration of SRRM4, while the variants for NeuralMICs required higher level of SRRM4 (Figure 7B). Predominantly, Extension Sequence 0 had a positive effect on the variant inclusion, but a small subset of variants was, unexpectedly, affected negatively by length extension with this Sequence (red cells in the heatmap). Similar, albeit lower, positive effect on the sensitivity to SRRM4 in response to titration and length extension was achieved with the Extension Sequence 1.

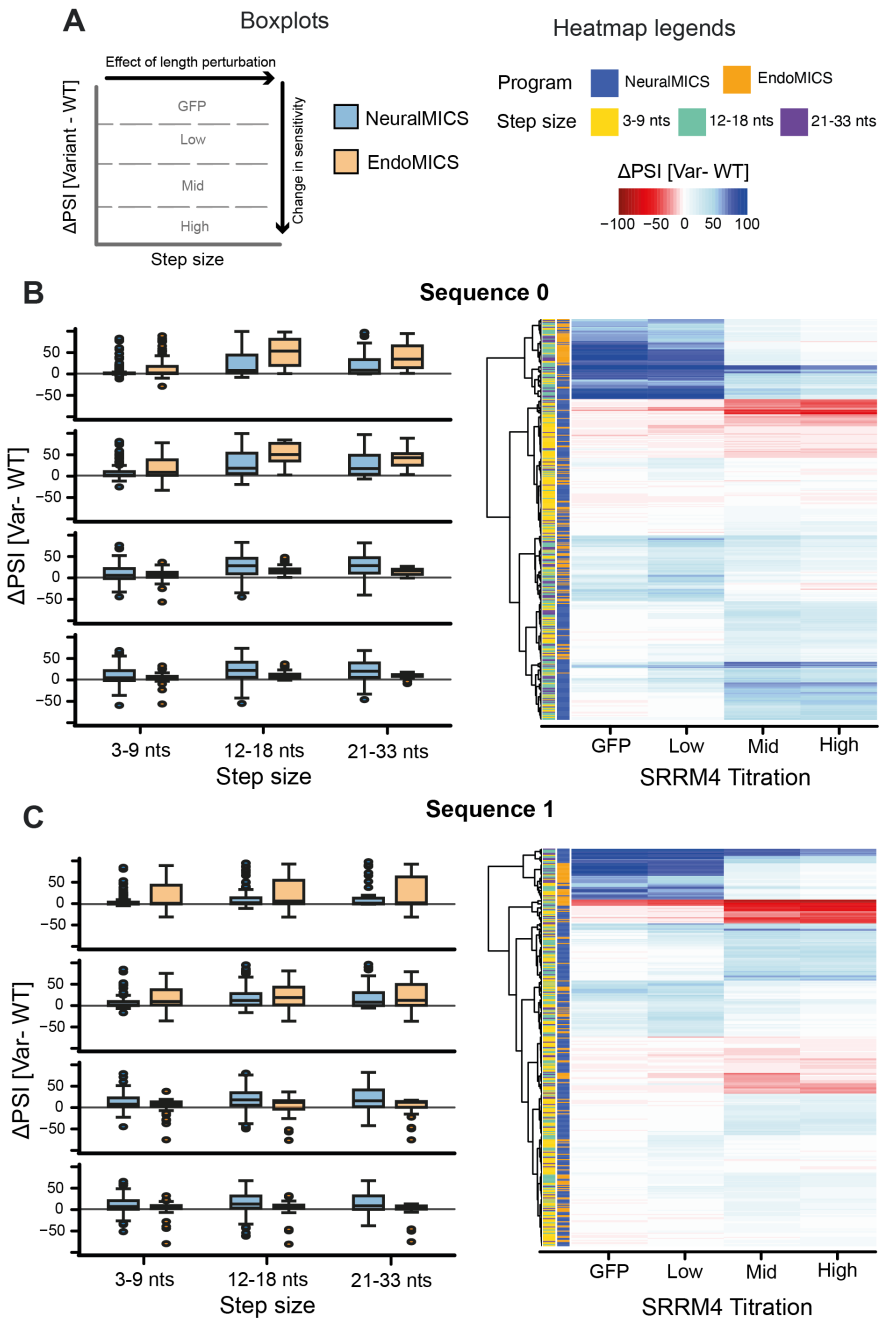

Figure 7. Length extension of microexons with sequences 0 and 1 impacts variant inclusion in similar ways. A) Schematic of the interpretation of Panel B and C results with corresponding legends. **B-C)**

Comparison of variation in inclusion profiles for length extension variants between Neural and EndoMICs for Extension Sequences 0 and 1. Heatmaps in the right panels correspond to variation in inclusion boxplots shown in the left panels. **B)** Variation in inclusion profile for Extension Sequence 0. **C)** Variation in inclusion profile for Extension Sequence 1.

As we observed before that Extension Sequence 2 was characterised by much lower inclusion profiles for all three sequence extension bins, compared to Sequences 0 and 1, we were curious to see how the variants of NeuralMICs and EndoMICs responded to SRRM4 upon length extension with that Sequence (Figure 8).

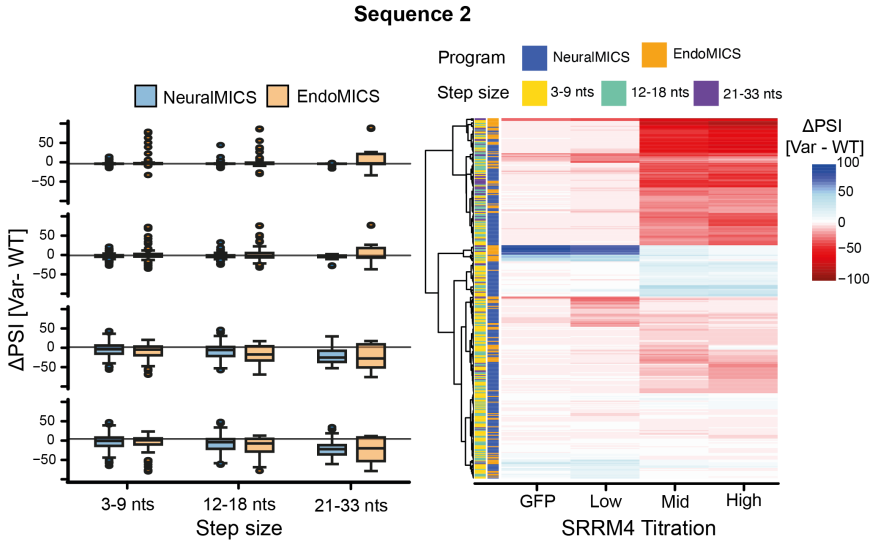

Figure 8. Length extension with Extension Sequence 2 leads to skipping relative to wild-type constructs. *Left:* Variation in inclusion profile for Extension Sequence 2 variants for NeuralMICs and EndoMICs. *Right*: Heatmap of variation in inclusion level corresponding to the *Left* panel.

We observed almost no differences between the microexon programs, particularly in the control and Low expression of SRRM4. With the increasing SRRM4 levels, however, the effect of the lengths extension with Sequence 2 led to gradually increasing skipping relative to the wild-type constructs, both for NeuralMICs and EndoMICs (Figure 8, left). Indeed, the general effect of the Sequence 2 was negative and only a subset of variants responded to SRRM4 titration with low level of sensitivity (Figure 8, right).

Impact of length addition on variants for cryptic microexons

To independently confirm that our Extension Sequences influence the inclusion of variant constructs, we looked at the length extension variants generated for cryptic microexons in our library. We selected microexons based on the presence of a TGC motif in the neighbourhood of the exon in the upstream intron and ΔPSI ≤ 10 (High-GFP) upon over-expression of SRRM4 at the endogenous level of inclusion (Figure 9A). We observed that in the exogenous context of the library, two of these microexons displayed higher inclusion level, suggesting that their regulation may be altered by the architecture of the minigene.

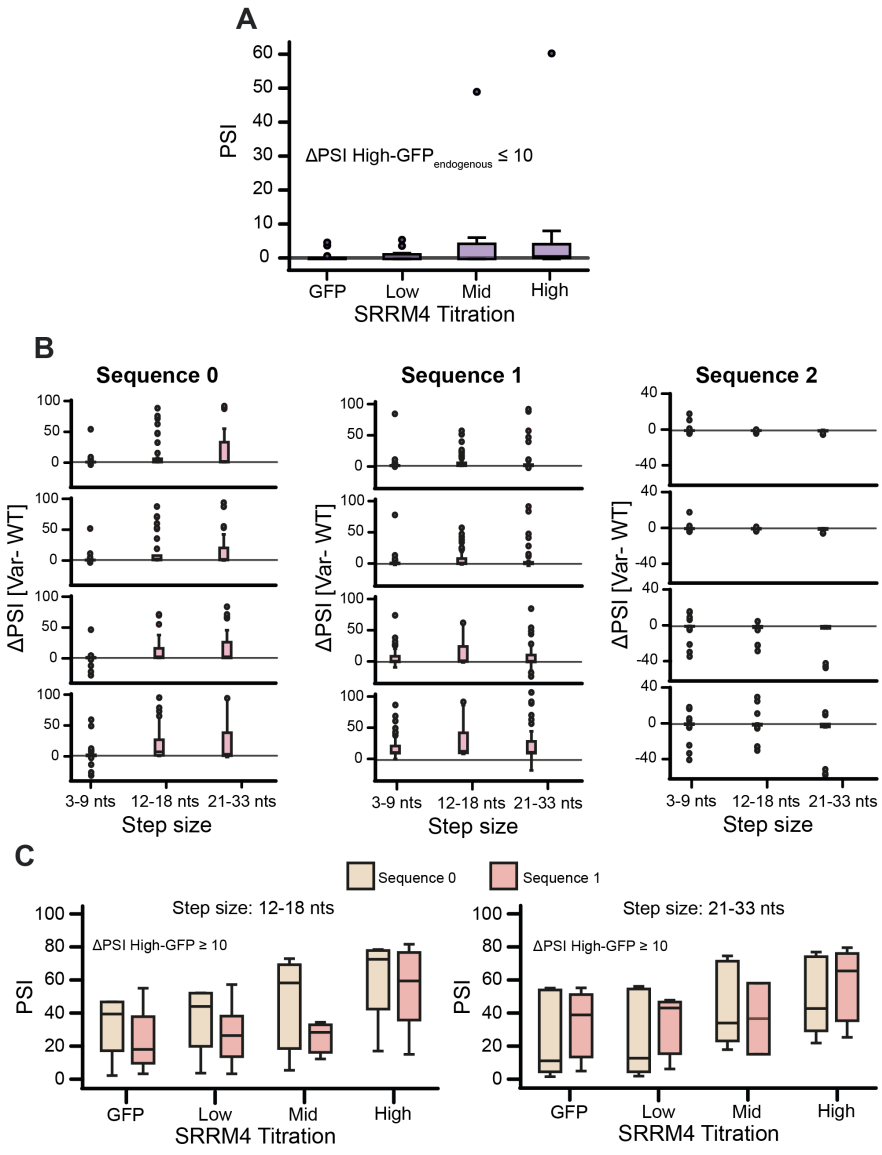

Figure 9. Length extension of cryptic microexons increases sensitivity to SRRM4 **A)** Inclusion level for 16 constructs of TGC-harbouring cryptic exons for microexons with ΔPSI ≤ 10 (High-GFP) upon over-expression of SRRM4 in Flp-In T-Rex 293 cell line. **B)** Comparison of inclusion profiles for length extension variants for cryptic microexons per Extension Sequence used. **C)** Inclusion profile for SRRM4-sensitive length

variants of cryptic exons with ΔPSI ≤ 10 (High-GFP) upon over-expression of SRRM4 in Flp-In T-Rex 293 cell line. Variants in 12-18 nts and 21-33 nts groups are shown.

Consistent with the observations made for NeuralMICs and EndoMICs variants, we observed that Sequences 0 and 1 had a positive effect on the inclusion of the variants, while Sequence 2 displayed little to negative impact (it should be kept in mind that the basal inclusion is already ~0 for these cryptic exons). Although the length extension was sufficient for cryptic microexons to promote their inclusion in all four conditions, we wondered if some variants acquired responsiveness to SRRM4. To test this, we selected variants with ΔPSI ≥ 10 (High-GFP) upon over-expression of SRRM4. 54 variants in two stepwise increase classes (12-18 nts and 21-33 nts) for Extension Sequences 0 and 1, corresponding to 4 unique microexons, increased their response to SRRM4, with Sequence 0 promoting higher inclusion level than Sequence 1 for 12-18 nts class.

Importantly, these results suggest that exon length may be one of the determinants of inclusion for some cryptic microexons that harbour a TGC motif in their upstream intron – a regulatory motif required for regulation of microexons by SRRM3/4.

Impact of length deletion

Next, we investigated how the decrease of exon length may affect microexons and constitutive exons. As 16 out of 28

constitutive exons harboured a TGC motif in their upstream intron, we wondered whether shortened variants could display sensitivity to SRRM4.

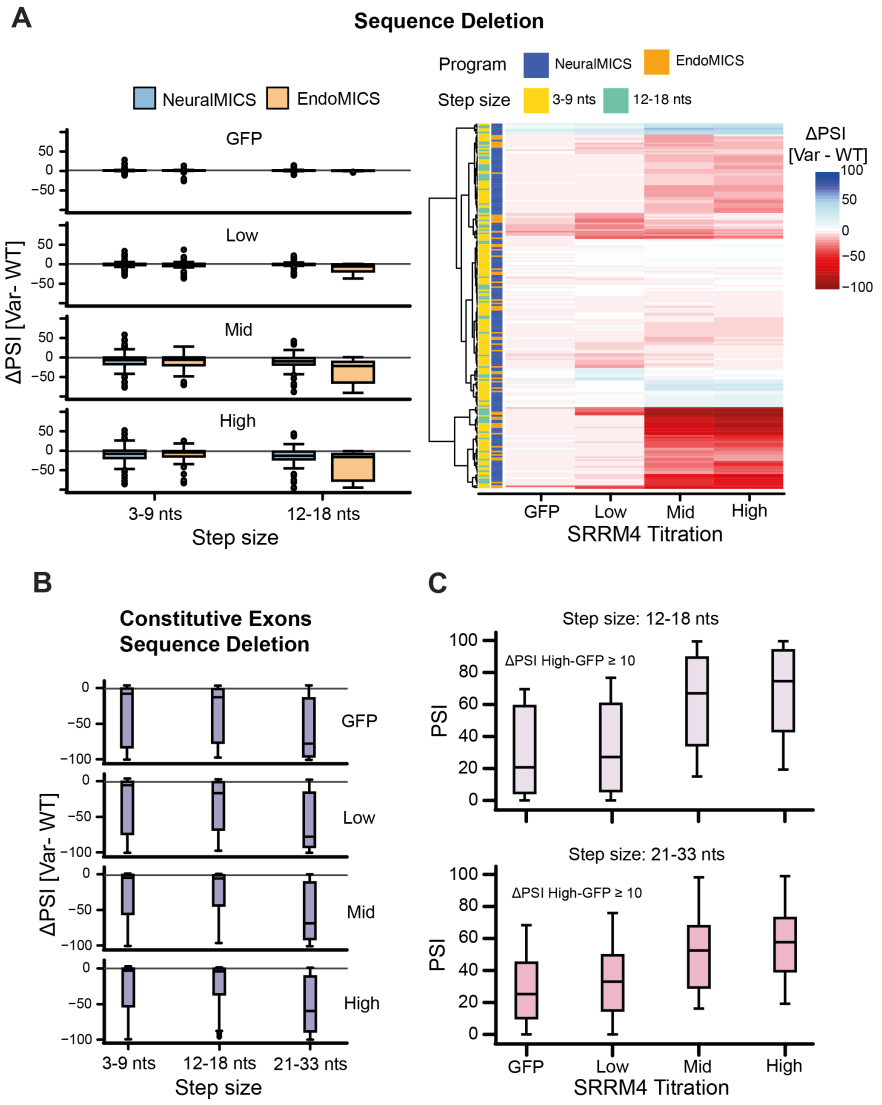

Figure 10. Sequence deletion decreases inclusion of EndoMICs and sensitises shortened constitutive exons to SRRM4. **A**) *Left:* Inclusion profile for sequence deletion variants for NeuralMICs and EndoMICs. *Right*:

Heatmap of inclusion level corresponding to the *Left* panel. **B)** Inclusion profile for sequence deletion from constitutive exons. **C)** Inclusion profile for SRRM4-sensitive length variants of constitutive exons with ΔPSI ≥ 10 (High-GFP) upon over-expression of SRRM4 in Flp-In T-Rex 293 cell line. Variants in 12-18 nts and 21-33 nts groups are shown.

We observed a predominantly negative impact of length decrease on inclusion level, both for NeuralMICs and EndoMICs (Figure 10A) and constitutive exons (Figure 10B) relative to their wild-type constructs. For NeuralMICs and EndoMICs the negative effects were stronger in Mid and High concentration of SRRM4, however comparable between these two groups. Only a small subset of these microexons displayed mild increase in sensitivity to SRRM4 (Figure 10A, right).

All the variants for constitutive exons displayed reduction in inclusion level relative to their wild-type constructs, regardless of the tier of SRRM4 expression. However, we identified 34 variants for 12-18 nts and 21-33 nts, from 11 unique exons, that acquired gradual response to SRRM4. This singularly suggested that exons that are normally constitutive and thus do not depend on SRRM4 for their full inclusion may acquire sensitivity to SRRM4 depending on their endogenous *cis*-regulatory architecture.

3.3. Closing remarks

In this Chapter, we prototyped a high-throughput barcoded library to study microexons and their properties in the context of sensitivity to SRRM3/4. While the barcoded libraries have been used before to study the mechanisms of alternative splicing, our Library of Micro(&)Exons is the first attempt to investigate microexons on a high-throughput scale.

As our first case study, we selected the length of microexons and characterised the effects of three Extension Sequences on the inclusion of NeuralMICs and EndoMICs variants. We also provided insights into the determinants of inclusion for cryptic microexons and the constitutively regulated exons.

Our results indicated that the inclusion and sensitivity to SRRM4 of length extension variants depended on the sequence context introduced by all three Extension Sequences. We also showed that some cryptic microexons upon length extension can be regulated by SRRM4. Interestingly, we observed that the shortening of constitutive exons can lead to SRRM4-regulated inclusion.

3.4. Methods

Bioinformatic analyses were performed using Python 3.6 scripting language and custom scripts developed for this Thesis. All experimental procedures were performed by Sophie Bonnal from our group.

DNM1 minigene cloning

The minigene encompassing 'exon (116 nts) – intron (433 nts)– microexon (12 nts) – intron (647 nts) – exon (171 nts) – first 25 nts of the downstream intron' was constructed for the DNM1 human microexon (HsaEX0020439).

The corresponding genomic region was cloned using EcoRV and NotI restriction sites under a Cytomegalovirus promoter. Furthermore, PT1 and PT2 (Sakamoto et al, 1992) sequences were added at the 5' and 3' ends of the minigene, respectively, for detection by reverse transcription PCR after transfection.

Library design and cloning

The length of each microexon was modified by either removing sequences (3 nts at each deletion step until the final length reached 6 nts) or adding sequences (3 nts at each step of addition from 3 independent Extension Sequences 0, 1 and 2, until the of 42 nts was reached). Each length modification was performed in the middle of the exons.

A total of 5992 variants (average of 24 variants per event), including wild-type sequences, surrounded by 93 nts of their 3'ss and 25 nts of their 5'ss were ordered to Twist Bioscience as ssDNA. Additionally, 20 nts long sequences, named Mega and Moe, were added at the 5' and 3' end of each variant, respectively, for cloning into the DNM1 minigene.

The plasmid bearing the minigene was amplified by PCR around the world using Taq Precision Plus enzyme (Agilent Technologies) with primers bearing Mega and Moe and designed to remove the DNM1 microexon and its surrounding 93 and 25 nts. The ssDNA library was amplified with Taq Precision Plus enzyme (Agilent Technologies) in a 10 cycles PCR. The variants were cloned using the Gibson Assembly cloning strategy with a mix of enzymes provided by the CRG Protein Technologies Unit. The plasmids were transformed in Stellar cells (Takara Bio).

Barcoding strategy

For quantification of the patterns of alternative splicing, barcodes were added between PT1 and the first exon in the DNM1 wild type minigene, prior to the cloning of the variants. For that purpose, the plasmid bearing the wild type minigene was amplified by PCR around the world using Taq Precision Plus enzyme (Agilent Technologies) with primers in PT1 and first exon.

The barcodes (34 nts):

NNNNAGCTNNNNTCAGNNNNTAGCNNNCAGTNNN

were ordered to IDT as (hand mix) ssDNA and amplified by PCR in 10 cycles PCR using Taq Precision Plus enzyme (Agilent Technologies). The barcodes were cloned using the Gibson Assembly cloning strategy with a mix of enzymes provided by the CRG Protein Technologies Unit. The plasmids were transformed in Stellar cells (Takara Bio).

Libraries transfection

Tetracycline inducible Flp-In T-REx 293 cells bearing either GFP or SRRM4 and generated in (Torres-Méndez et al., 2019) were used.

Each condition was done in 6 replicates. Each replicate consisted in a transfection of 6 wells from a 6 wells plate. For each well, 400 000 cells were seeded in 1 ml medium and 80 ng of plasmids bearing the (barcoded) variants were transfected using Lipofectamine 2000 (Invitrogen), following the manufacturer's instructions. At the time of transfection, doxycycline at various concentrations was added to induce the expression of the protein of interest.

Non-induced GFP line and SRRM4 expressing lines treated with 4.8, 19.2 and 50 ng/ml doxycycline were used. The cells

from 6 wells of each condition were pooled 24h after transfection.

RNA isolation

RNA was isolated using Illustra RNAspin Mini RNA Isolation Kit (GE Healthcare).

Amplification of the inputs and outputs of the libraries for RNA deep-sequencing

Library input: To match the barcodes with the variants in each plasmid, we amplified the pool of plasmids using primers bearing sequences from Illumina (Read 1 and Read 2) and complementary to PT1 and down to Moe. Each input was amplified in 3 independent PCRs (7 cycles each using Taq Precision Plus (Agilent Technologies)).

Library output: The pattern of alternative splicing was assessed by RT (with oligo-d(T)/random hexamer) with AMV Reverse Transcriptase (Promega) and 10 cycles PCR with Taq Precision Plus (Agilent Technologies) using 6 primers (3 forward and 3 reverse) bearing sequences from Illumina for sequencing Read 1 – (N/NB/NBK) – PT1 and Read 2 – (N/NB/NBK) – part of last exon. PCR elongation times of 5 minutes were used in all cases.

The products of amplification were purified on AMPure XP beads (Beckman Coulter). Amplicons were checked using

Bioanalyzer (Agilent) and the final steps amplification (consisting in the addition of the final adaptors) were performed at CRG Genomics Unit. Samples were sequenced on HiSeq2500 and 125 nts paired-end reads were generated.

RNA deep-sequencing

RNA quality was checked using Bioanalyzer (Agilent) and strand-specific Illumina libraries were prepared and sequenced at CRG Genomics Unit. Samples were sequenced on HiSeq2500 and 125 nts paired-end reads were generated.

Identification of barcode-variant associations in library input sequencing data

To identify barcodes in raw FASTQ forward reads, we searched sequences for 10 nts of PT1 primer, 34 nts barcode and 10 nts downstream sequence, allowing at most 3 substitutions in the flanking sequence around the barcodes.

'.{13}(CTTGCTCAAC){e<=4}(.{26})(GAATGTCTAC){e<=4}.{33}').

The sequences passing this criterion were selected for further parsing of their reverse mates. To identify variants in reverse mates, FASTQ files were first reverse complemented a subjected to search with Python regex specified below allowing for identification of 5' splice sites.

'*.{58})GT(.{10}).{13}TCGTAGCACG'*

Using unique 5'splice – 3' splice site dictionary for the variants in our library, we identified a corresponding 3' splice site and matched the sequence in between the splice sites to variant sequences. We accepted variants with full splice sites and variant sequence matches only.

Barcode-Variant Associations filtering

Associations supported by at least 5 reads were selected for processing. Barcodes associated to one variant were kept. For a barcode with two associated variants, the most supported variant was selected if the support of the second variant represented less than 10% of the number of reads supporting the predominant variant. For a barcode with more than two mis-assignments the second most supported variant was identified. The association was discarded if the second variant was supported by more than 2 reads; otherwise, the most supported associated variant was selected. The final set of BVAs consisted of associations present in at least 2 replicates, plus a set of BVAs found in any single replicate that were supported by at least 10 reads. Variants supported by 1 barcode were removed.

Inclusion level quantification and PSI correction

Barcodes in forward FASTQ reads were identified as previously described. Sequences were filtered using the set of trusted BVAs. Reverse mates first reverse complemented and

scanned with Python regex specified below to identify inclusion and skipping variants.

'TGAGCGTGTTGGG(.)GCCAGCGAGACCG'*

Reads were counted per each BVA, and inclusion value was calculated:

$$PSI = \frac{Inclusion\ reads * 100}{Inclusion\ reads + Exclusion\ reads}$$

PSI values were corrected to account mis-assigned BVA that contributed to skipping read counts. The read count for these BVAs as a proportion of total number of reads for trusted BVA was calculated and used as a proxy to correct PSI values, as follow:

$$Exclusion\ reads_{corrected} = Exclusion\ reads * \frac{100 - proportion}{100}$$

PSI values were then recalculated using corrected number of exclusion reads.

Removal of barcode outliers

For a given variant in each condition, median inclusion value and absolute difference between median PSI and actual per-barcode PSI were calculated. Removal of barcodes was scaled to the inclusion values as follow: i) for variants with 10 < median PSI > 90, difference of < 10; ii) for variants with 30 < median PSI > 70; iii) else variants were kept if difference was smaller than 20.

CHAPTER 4

DISCUSSION

4. Chapter 4: Discussion

4.1. On the non-tissue specificity of neural-specific microexons

Prior to the beginning of the projects that shaped this Thesis, the research on microexons provided extensive evidence for their regulatory mechanisms, evolutionary conservation, and biological functions (Gonatopoulos-Pournatzis et al., 2018, 2020; Gonatopoulos-Pournatzis & Blencowe, 2020; Irimia et al., 2014; Quesnel-Vallières et al., 2015, 2016; Torres-Méndez et al., 2019). However, none of these studies had investigated the regulation of microexons in non-neural tissues.

In this Thesis, we provide the first evidence of SRRM3-regulated microexons in beta cells of endocrine pancreas (EndoMICs). Moreover, we show that EndoMICs represent a subset of the microexons previously characterised as neural-specific. Interestingly, we also show that beta cells express much lower levels of SRRM3 than neurons and barely express SRRM4. These observations suggest to us that the relative lower expression of SRRM3/4 might be at the basis of the regulation of the nested program of microexons whereby, in an activity-dependent manner, distinct subsets of microexons are included at different levels of SRRM3/4 activity due to their differential sensitivity to these regulators. The model states that the most sensitive microexons in endocrine pancreas would respond to lower expression of SRRM3. On the other hand, NeuralMICs would require higher levels/activity of SRRM3/4

for inclusion. This model challenged the *on-off* switch model for microexon regulation in neural tissues and suggested more complex and diverse regulatory mechanisms.

In summary, in this Thesis we characterised the program of EndoMICs in endocrine pancreas and investigated the mechanisms by which the program of nested microexons is regulated (Figure D-1).

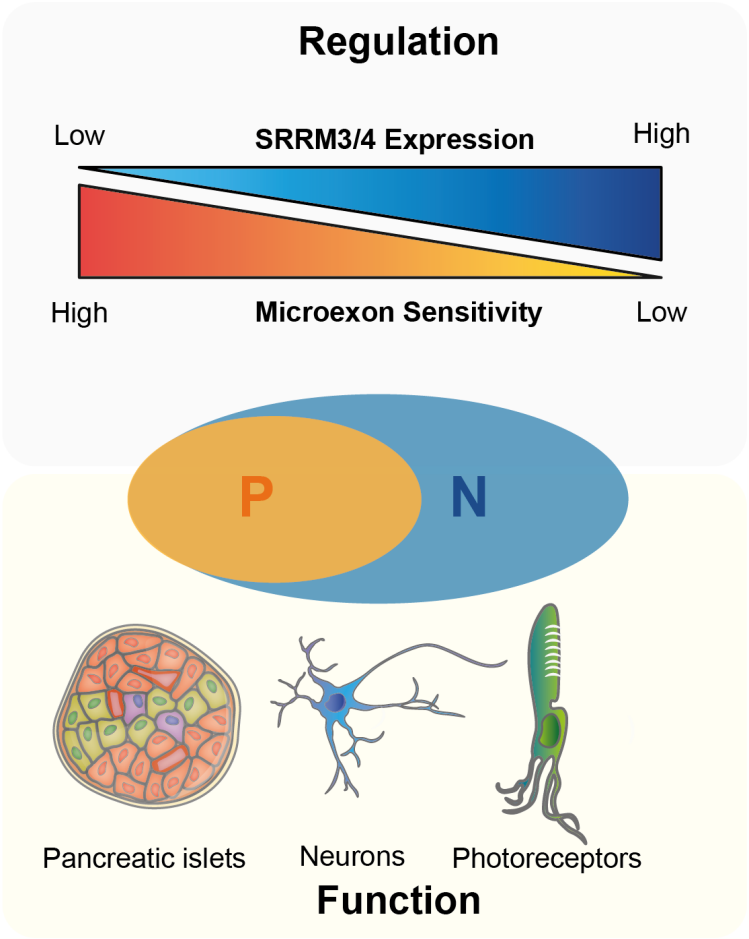

Figure D- 1. Summary of the model of differential sensitivity of microexons to SRRM3/4. The model describes the inverse relation between microexon sensitivity and the expression/activity of SRRM3/4 (top panel). This relationship in turn dictates microexon usage in three neuroendocrine tissues, where it leads to emergence of combinations of functions specific for a given program (bottom panel).

First, let us focus on the biological implications of the nested program of microexons, which precedes our mechanistic studies with our library of mutants, discussed in the following section.

a) Biological functions of EndoMICs

In Chapter 2, our results using cell lines (INS1E rat and EndoC-B1 human beta cell lines) were instrumental to confirm that EndoMICs are regulated by SRRM3. Using a rat beta cell line, we showed that knock down of SRRM3 caused an increase in insulin secretion, confirming that EndoMICs play important roles in regulating the secretory capacity of beta cells.

EndoMICs are in fact harboured in genes with functions regulating insulin secretion at almost all stages of the process. Among these 76 genes, we find calcium channel-related genes (*CACNA1D*, *CACNA2D1*, *CACNB1*, *CACNB2*), mitochondrial integrity and activity genes (*SH3GLB1*, *SH3GLB2*, *ATP6V0A1*, *ATP1B3*), and transcription factors (*MEF2D*, *TCF7L2*), all of which are required for beta cell homeostasis. We also found that differential regulation of gene expression, caused likely by indirect effects of splicing deregulation (including mRNA

degradation by introduction of premature stop codons and Nonsense-Mediated-Decay -NMD-), could also contribute to explain the phenotype that we observed. For example, we observed upregulation of the insulin-coding gene Ins2 as well as its receptor *Igf2r* in rat cell line. Considering the diversity of the effects of depletion of SRRM3 in the transcriptome of these cell lines, it may be difficult to break down the phenotype into discrete cause-and-effect pieces rather than because of the collective impact of a multitude of transcriptome changes.

However, the study of constitutive *Srrm3*$^{-/-}$ knock out mouse provided important phenotypic cues. The loss of Srrm3 profoundly deregulated islet homeostasis, resulting in hyperinsulinemic hypoglycemia, a condition where low blood glucose level is caused by excessive secretion of insulin. This condition is also evident by the decrease in ketone bodies that we observed, which represent one of the main energy sources for the brain. Experiments performed in our group also suggested that the loss of Srrm3 in mice results in increased numbers of alpha and bihormonal cells (data not shown), strongly suggesting the contribution of EndoMICs to the overall identity of islets.

Following these results, it would be interesting and instructive to compare the islet-specific loss of Srrm3 with the results obtained with the full knock out *Srrm3*$^{-/-}$ model we have used used in this Thesis (Nakano et al., 2019; Quesnel-Vallières et al., 2015). As these were shown to display neurological defects consistent with Srrm3/4 loss of function, effects of

neuroendocrine signalling on beta cells cannot be ruled out and therefore further work with islet-specific Srrm3 knock out would be necessary to provide more precise insights into the functions of EndoMICs and the phenotypic consequences of the loss of Srrm3 that do not depend on the gross effects of its knock out and of defects in nervous system. It would be therefore of considerable interest to study whether a nervous-system-specific knock out also has effects on endocrine tissues like pancreas.

In Chapter 2 we proposed 28 EndoMICs that in our opinion are good candidates for further functional studies in the context of beta cell biology, based on the data that we generated. Particularly relevant may be 6 EndoMICs that we identified as deregulated in RNAseq datasets from Type 2 Diabetes human islet donors and regulated in human and rat beta cell lines, as well as in the constitutive *Srrm3*$^{-/-}$ mouse model.

For example, isoform studies on an EndoMIC present in the gene *MEF2D*, encoding a transcription factor important in pancreatic cell differentiation, could be instructive in understanding how SRRM3-regulated microexons can contribute to beta cell development and identity through their effects on MEF2D-dependant transcriptional networks. It has been previously shown that expression of MEF2D is also developmentally regulated by class II histone deacetylases in beta cells (Lenoir et al., 2011). *MEF2D* EndoMIC would therefore provide an interesting link between microexon regulation, transcription, and post-translational modifications

relevant for epigenetic regulation in the context of beta cell biology. It would be also in line with our results in mouse suggesting that beta cell identity is partially lost upon the loss of Srrm3.

The remaining 22 targets may offer different insights into beta cell biology depending on the assay availability and experimental feasibility in each cell line or mouse model.

b) Biological functions of nested program of microexons

Considering the shared functions of EndoMICS in beta cells and neuron, it would be interesting to investigate the subset of EndoMICs with different sensitivities to SRRM3/4 and their biological functions in respective cell types. However, it should be noted that previous attempts to probe the effects of single microexons using CRISPR-Cas9 knock outs did not yield substantial phenotypic defects in zebrafish. This was described in the PhD Thesis by Laura Lopez Blanch "*Deciphering the role of alternatively spliced microexons during vertebrate nervous system development and function"* (2020). For this reason, separating the functional roles of EndoMICs in neuron and beta cells may prove more challenging.

Nonetheless, an important example of the biological functions of the nested program has been recently provided. Consistent with the functional redundancy of SRRM3 and SRRM4 in neural tissue described by (Nakano et al., 2019), recent work in our group (Ciampi et al, 2021; bioRxiv) demonstrated that

while both proteins are expressed during the differentiation of photoreceptors, only the expression of SRRM3 is maintained in mature cells. In fact, higher expression of SRRM3 in photoreceptors compared to other neural tissues fits into our model, as SRRM3-regulated microexons in photoreceptors displayed the lowest sensitivity (data not shown) and therefore require higher levels/activity of SRRM3/4 for inclusion. These results confirmed that differential sensitivity of microexons in the nested program is important as a regulatory mechanism of the biological functions of cells that use this mechanism.

It is also important to recall that the LME studies were conducted in HEK293 cell line. Although important aspects of the tiered regulation were recapitulated in this system, we could not address the contribution of tissue-specific gene expression and regulation on activity of SRRM4. Although we previously discussed the similarities between neurons and beta cells, both at the functional and molecular levels (Chapter 1, Section 1.3; Chapter 2), some transcriptional networks, post-translational modifications and/or signalling pathways may contribute to the overall sensitivity of microexons to SRRM3/4 depending on a specific biological context. Previous studies focused on alternative splicing and insulin secretion identified other splicing factors that contribute to this phenotype (Juan-Mateu et al., 2017, 2018).

It will be therefore instrumental to understand the upstream regulatory mechanisms of SRRM3/4 expression in neurons and beta cells, as well as to investigate other splicing factors

that may regulate microexons (Gonatopoulos-Pournatzis et al., 2018). In line with the first point, the SRRM3 locus has been shown to harbour a glucose-dependent enhancer hub in its vicinity, as well as several single nucleotide substitutions associated with Type 2 Diabetes (Miguel-Escalada et al., 2019). It remains to be investigated to what extent these associations control the expression of SRRM3 and how it may affect beta cell function and deregulation of microexons in diabetes and other insulin-related pathologies.

4.2. On the lessons from the Library of Micro(&)Exons

In Chapter 3 we prototyped a high-throughput barcoded library to study microexon length. It was the first time that this approach was applied to investigating the regulatory mechanisms of microexons.

Using barcodes, we were able to quantify inclusion levels of the wild-type constructs and assess the impact of length perturbation on variant inclusion. We were able to recapitulate the endogenous inclusion of most microexons in our system, proving the usefulness of our method for this and further studies that we have designed (see below).

We showed that length extension using two Extension Sequences generally increased inclusion of the variants. This effect was slightly clearer for EndoMICS than for NeuralMICs

in GFP control and Low expression of SRRM4 conditions, but not in Mid and High expression levels of SRRM4, where the effect was difficult to discern as it did not depend on the length increase and the variants did not display an overall increase in sensitivity to SRRM4.

Analysis of the effects of the third Extension Sequence and variants generated with it provided striking results. While several variants generated with first two Extension Sequences displayed progressive decrease in inclusion in response to SRRM4 expression, almost all variants of the third sequence displayed enhanced skipping, suggesting that the sequence contained a powerful splicing silencer element, despite our best efforts to use sequences devoid of splicing regulatory elements (Ke et al., 2011, 2018). We confirmed with Prof. Larry Chasin, Columbia University (New York), arguably one of the most knowledgeable experts in comprehensive analysis of splicing regulatory elements, that the sequences we used did not contain any predictable splicing regulatory element. The fact that they did underscores the limited knowledge that we have of the sequence determinants involved in the splicing process and its regulation.

In any case, our results led us to the conclusion that Extension Sequences might impact variants in a context-specific manner (discussed in the next section). Collectively, we concluded that the effects of length on the sensitivity to SRRM4 for variants of NeuralMICs and EndoMICs in our library are influenced by their sequence context and by contribution from *cis*-acting

regulatory elements that were not investigated in this library (including for example neighbouring intronic sequences that have been implicated in microexon regulation like UGC motifs (Gonatopoulos-Pournatzis et al., 2018; Quesnel-Vallières et al., 2015).

Furthermore, our results assessing the effects of length extension on cryptic microexons provided some interesting insights. Not only we observed increase in response to SRRM4 upon length extension with two Extension Sequences – and confirmed the repressive effect of the third sequence independently of the context of variants for NeuralMICs and EndoMICs – but also identified several cryptic microexons with increased sensitivity as the length of the microexon was increased. It would be interesting to investigate other cryptic microexons that harbour UGC motifs to answer questions about the restrains that dictate their low inclusion.

Perhaps unexpectedly, we observed response to SRRM4 in constitutive exons of 42 nts, raising an important question about the role of SRRM4 in splicing of these exons. It may be the case that the contribution of SRRM3/4 occurs through a passive association to spliceosome, however it remains an interesting aspect of the constitutive splicing regulation. The outstanding questions include: why the library clones do not include constitutive exons at 100%?, or why an exon that has evolved to not require SRRM proteins for inclusion becomes responsive under certain conditions? These questions could

be further investigated, for example in SRRM3/4 knock out cell lines.

These important and interesting issues raise questions about the interpretation of results obtained using our library which are relevant to the design of further studies. For this reason, it is important that I discuss the limitations of our approach.

4.3. On the limitations of the Library of Micro(&)Exons

In our study, we only tested the variants in the genomic context of the flanking introns and exons of the *DNM1* microexon. However, while investigating the 5´splice site usage using similar barcoded library high-throughput approach, (Wong et al., 2018b) observed considerable differences depending on the gene context in which they placed a given 5´ss mutation. It is therefore important to consider this the future studies of microexons using our library. The context in which we assembled the library may explain several observations that we made so far. Let us consider the following aspects.

a) Exogenous splicing

The generally lower inclusion level that we observed for Neural and EndoMICs in the context of LME compared to endogenous inclusion may be due to the *exogenous* nature of the library. The first aspect that we considered was titration effects of components of the splicing machinery due to the higher levels

of minigene-derived transcripts over individual endogenous microexon-containing transcripts, due to the strong promoter driving expression of library transcripts and the amount of transfected library input. However, we did not observe considerable differences in the inclusion or variants (nor their endogenous counterparts) when comparing 10 ng with 80 ng (for the presented results) of the library input transfection in GFP and High expression of SRRM4 (data not shown).

A second aspect we considered is whether differences may arise from features of splicing regulation and its dynamics that occur in the genomic context of a given microexon vs the non-chromatinised nature of plasmids of our library. For example, it has been shown that spliceosomal components interact with elongating polymerase II, which adds another layer to the regulation of the splicing outcome (Das et al., 2007; Spiluttini et al., 2010). The elongation rate and transcriptional pausing were shown to influence splice site availability and consequent commitment of the spliceosome, reviewed in (Naftelberg et al., 2015). For example, (Maslon et al., 2019) investigated a slow elongating form of RNA polymerase II in a neuronal differentiation model and demonstrated that, at least in the context of that model, alternative splicing of longer genes involved in synapse signalling required more kinetic coupling between transcription and splicing than pluripotent cells. Another study showed that ultraviolet irradiation reduced protein associations to the carboxy-terminal domain of the RNA polymerase II through hyperphosphorylation of that

region that in turn inhibited elongation rate and splicing efficiency (Muñoz et al., 2009). While the mechanism of co-transcriptional coupling between transcription and splicing is indisputable, a recent study highlighted the differences in its impact on splicing between genes and cell lines (Bedi et al., 2021). In the context of our library, the extent to which co-transcriptional coupling occurs could be one of the determinants contributing to the differences between the exon inclusion in library-derived transcripts and endogenously spliced exons.

Another important aspect related to exogenous splicing is the interplay between minigene architecture and regulatory elements required for a given microexon and its variants. The impact of this was particularly evident for several wild-type microexon constructs. For example, two constructs of EndoMICs, *MADD* and *ERGIC3*, showed a 30% inclusion in GFP control in the context of LME, while their endogenous inclusion in the control was close to 0. A minigene of *SULT1B1* cryptic microexon displayed a remarkable inclusion of over 50% in Mid- and High concentration of SRRM4, in contrast to its endogenous counterpart, whose inclusion did not exceed 10%. In our library we kept only 93 nts and 25 nts of endogenous upstream and downstream introns, respectively. Perhaps, in the case of these microexons, while these sequences were still sufficient to preserve splicing regulation, more distal elements were required for full splicing outcome. Collectively, these examples reinforced the concept that

additional regulatory elements not included in our library play a role in microexon regulation.

b) Extension Sequences

The sequences used to extend the length of exons were carefully designed to be devoid of known predictable splicing regulatory elements, such as splicing enhancers and silencers. However, these sequences did exert specific effects in the context of a given variant. It would be therefore interesting to scan the sequences at the boundaries between wild-type exonic sequences and the newly added sequence elements to identify potential newly formed sequence combinations creating regulatory motifs.

Indeed, several studies have highlighted the variety of sequences that can impact splicing regulation. A study of the evolution of *FAS* exon 6 using high-throughput mutagenesis showed that 2/3 of all possible single nucleotide substitutions alter inclusion of that exon (Baeza-Centurion et al., 2019). It was also observed that the effects of a mutation depend on the context of other sequence variants relative to the starting genotype. Another saturation mutagenesis study directly investigated the effects of mutations on exon definition model and recruitment of the core spliceosomal components (Ke et al., 2018). They showed that even 1 nucleotide substitution could affect splicing outcomes by decreasing the affinity of protein binding to regulatory elements or through effects the secondary RNA structures.

In our library of nearly 6 thousand variants generated in the context of 3 different microexons and with a maximum of 11 stepwise additions of 3 nts (depending on the starting length of a given exon), the probability of a sequence-dependent effect on splicing may be difficult to estimate. However, it is possible to identify sequences and quantify their effect on splicing using machine learning algorithms.

These considerations would therefore apply to any high-throughput studies on splicing regulation and are important for the interpretation of our results and for the design of future studies using these approaches.

4.4. On the future studies on *cis*-acting elements and evolutionary conservation underlying SRRM3/4 sensitivity

Based upon the results with our prototyped library and caveats emerging from their analysis, we have generated two types of the second-generation library. The first type was dedicated to length and *cis*-acting elements, to study the contribution of sequence motifs and their location within intronic sequences to the sensitivity to SRRM4. The second type was focused on the evolutionary conservation of intronic sequences, in the hope to investigate the proposed mechanism across vertebrates.

First, we selected microexons based on the following criteria: i) the wild-type constructs recapitulated the inclusion levels of endogenous microexons upon over-expression of SRRM4; ii) the wild-type constructs displayed different sensitivities to SRRM4, for both Neural and EndoMICs subsets; and iii) the upstream introns contained different genomic features expected to influence microexon recognition, for example varying number of TGC motifs and strengths of splice sites

We also included responsive cryptic microexons and constitutive microexons as controls. In total, we selected 41 exons for further studies. The design of the library is summarised below. Importantly, we generated a dox-inducible SRRM3 cell line to recapitulate the regulation of EndoMICs in endocrine pancreas. While the effects of over-expression of SRRM4 and SRRM3 are similar in non-neuroendocrine cells (preliminary data, not shown), investigating the effects of both regulators in parallel may reveal differences and similarities between these factors and effectively provide more resolution to the model of differential sensitivity.

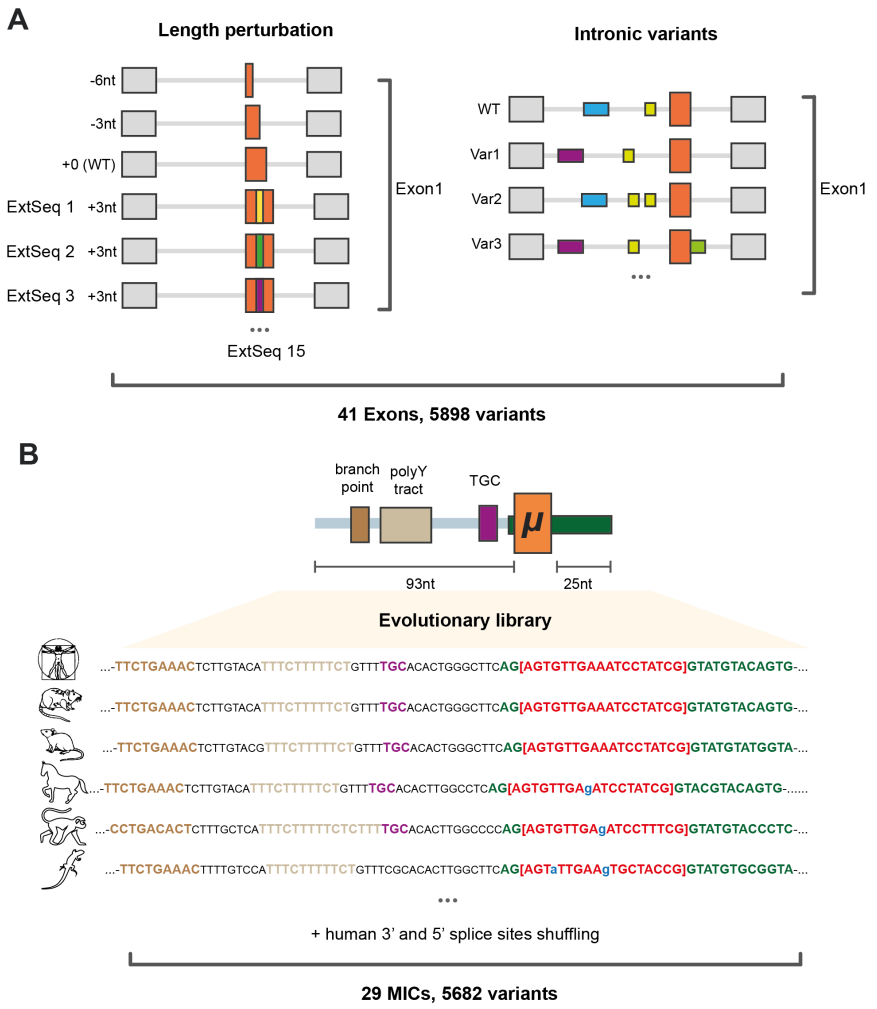

***Figure D- 2. Design of the second-generation high-throughput libraries to study sensitivity of microexons to SRRM3/4*. A)** Length variants were generated by sequence deletion and sequence extension by additional 13 Extension Sequences to resolve context-specific effects on variant inclusion. For each of the 41 exons selected, mutations of selected intronic cis-acting elements implied in the regulation of microexons were introduced, removed, and/or mutated, where possible. **B)** Evolutionary library was designed for 29 microexons and included the homologous microexon and flanking sequences for 24 vertebrates. Sequences shown

correspond to examples of homologous sequences for a microexon found in the PPP6R3 gene.

a) Length and intronic mutants

Following our finding that Extension Sequences exerted context-specific effect on the inclusion of variants, we designed 13 additional sequences of 36 nts. To estimate their enhancing or silencing effect, we used a tool developed by (Ke et al., 2011) that provided scored hexamers based on their predicted impact on splicing in that study. We searched over 10 thousand randomly assembled sequences and selected those with low score for predicted impact on splicing. We then removed stretches of nucleotides longer than 3 nts and AG/GT sequences that could generate cryptic splice sites. We applied the same approach to generate length variants as described in Chapter 3 and generated 3772 sequence insertion and 228 sequence deletion variants. These variants will allow us to further explore the effects of exon length on the sensitivity to SRRM proteins.

It will be interesting to resolve similarities and differences between all 13 sequences (and previous two) and their impact on variant inclusion and sensitivity to SRRM3/4. As discussed in the limitations of high-throughput libraries section, *de novo* sequence motif analysis may potentially lead to the identification of sequences that, in specific sequence contexts, may promote or inhibit splicing.

Within the same library, we also generated mutations in *cis*-acting elements that in the context of microexon regulation may have an important impact on their inclusion and sensitivity to SRRM3/4. We identified and quantified genomic features within 93 nts of upstream introns for the selected 41 exons using *Matt* software (Gohr & Irimia, 2019). The summary of the mutations is provided in the table below.

Genomic feature	Mutations
TGC motif	1) TGC addition and deletion (eg. TAC -> TGC or TGC > TAC) 2) Mutation of extended nTGCn motif
Splice sites	1) Strengthening U1 and U6 sequence consensus of the 5'ss 2) Mutation of the AG-exon boundary at 3'ss
Branch point	1) Inactivation of predicted branch points (in combination) 2) Shuffling of predicted branch points (in combination)
Branch point + PyT	Shifting best predicted branch point and polypyrimidine tract by 10 or 20 nts towards the 3' splice site

Table D- 1. Summary of the intronic variants. Intronic variants were generated through insertion, deletion, sequence mutations or sequence shuffling of specified *cis*-acting elements.

For each type of mutation, we considered several hypotheses. For example, while it has been shown that the TGC motif is required for SRRM3/4 binding to the flanking upstream intronic sequences of a microexon, the relationship between the number of TGCs and their relative distribution remains uncharacterised (Gonatopoulos-Pournatzis et al., 2018; Irimia et al., 2014; Quesnel-Vallières et al., 2015; Torres-Méndez et al., 2019). It has been also suggested (data not shown) that the sequence context of the nucleotides flanking TGC may also contribute to the regulatory potential of this motif.

Microexons are characterised by stronger 5' splice sites and weaker 3' splice sites compared to longer alternative exons and constitutive exons (Irimia et al., 2014). However, considering the insights obtained in this Thesis, it will be interesting to study the interplay between the strength of splice sites and other sequence variants to influence microexon recognition and sensitivity to SRRM3/4.

Previous studies demonstrated that the strength of a branch point correlates with optimal base pairing with U2 snRNA and consequently U2 snRNP binding (Nelson & Green, 1989; Zhuang et al., 1989). However, computational approaches suggested that more than one branch point may be active in the same intron, suggesting more flexible dynamics in branch point selection mechanisms (Corvelo et al., 2010). In line with these studies, we mutated/shifted up to 3 predicted branch points in the 93 nts of upstream intronic sequences corresponding to the microexons tested in our constructs. We

hope that this will allow us to assess: i) positional effects of branch point contribution to SRRM3/4 sensitivity; and ii) branch point usage in the context of microexon and variants inclusion. Moreover, this will provide a first study on the importance of branch point recognition on microexon regulation.

b) Evolutionary conservation

To investigate whether the sensitivity is evolutionarily conserved, we selected 4 primate (gorilla – *Gorilla gorilla,* macaque – *Macaca mulatta,* chimpanzee – *Pan troglodytes,* and marmoset – *Callithrix jacchus*) and 19 non-primate vertebrate species, including platypus (*Ornithorhynchus anatinus*), zebra finch (*Taeniopygia guttata*) and green anole (*Anolis carolinensis*).

We identified the conserved loci for 29 human microexons (15 NeuralMICs, 8 EndoMICs and 6 cryptic microexons), checking for the presence microexon and splice sites. We allowed some nucleotide substitutions in the microexon sequence, provided that the length of the exon matched the length of the human microexon homologue. Additionally, wherever different, we substituted 5' splice sites with the splice site of the homologous human microexon.

Additionally, as our selected targets displayed different sensitivity to SRRM4, we wondered how the substitution of entire intronic regions would affect a given exon. We shuffled entire upstream intronic sequences and 5' splice sites within human targets, hoping to identify combinations of intronic

sequences that alter sensitivity to SRRM3/4 and implement this information in further steps of analysis, discussed below. This library, dedicated to wild-type introns in evolutionary context and intron shuffling, consisted of 5682 constructs.

c) Validation of novel *cis*-regulatory elements underlying microexon regulation

With the combined mutagenesis and evolutionary library, we will assess the effect of *natural* and other mutations and their relative contributions to the sensitivity of microexons to SRRM3/4. To capture the complex relationships that may influence microexon regulation, it will be imperative to use machine learning. Recently, several studies used machine learning and deep learning algorithms to improve splice sites prediction, predict splicing outcome for sequence variants and identify the impact of alternative isoforms on gene expression in clinical data (Albaradei et al., 2020; Gao et al., 2021; Jaganathan et al., 2019). This approach may also prove useful to thoroughly understandi the context-specific effects of Extension Sequences that we used to perturb exon length. Our model will consider the sensitivity to SRRM3/4 of i) wild-type constructs; ii) intronic, length mutants; iii) evolutionary wild-type construct and intron shuffling mutants; iv) sequence motifs and relative distances within all three categories.

We hope that our previous study and the future efforts will generate new insights into *cis*-regulatory elements that contribute to the microexon sensitivity to SRRM3/4. Genome

editing techniques may be used to knock in or knock out *cis*-regulatory motifs in neurons and beta cell lines to assess their importance in their proper genomic and physiological context. Thus far we considered only SRRM3/4 in the model, however, as previously described, several other factors have been found to regulate microexon inclusion, such as SRSF11 and RNPS1 (Gonatopoulos-Pournatzis et al., 2018). An evolutionary study on eMIC domain also identified its interactions with early spliceosomal components (U2AF2 and SF1), which adds to the complexity of our model (Torres-Méndez et al., 2019). Transient depletion of these proteins under conditions of SRRM3/4 over-expression will allow to investigate the contribution of these regulators to microexon sensitivity in the context of specific intronic mutations or length perturbations.

4.5. Closing remarks

Here, we have discussed the biological relevance as well as potential mechanisms (and technical caveats) of the results presented in this Thesis. The characterisation of microexons in endocrine pancreas led to an unexpected and possibly general discovery of the “nestedness” of regulatory programs. This discovery generated several hypotheses regarding microexon regulation and their biological functions that we have only started to explore.

The analyses dedicated to EndoMICs in beta cells and pancreatic islets showcased the role of microexons in non-

neural tissues, also providing evidence for their importance in secretory functions and hints about their influence in islet identity and possible roles in beta cell pathologies, including Type 2 Diabetes. The results therefore satisfied Objectives 1 and 2 of this Thesis.

Our high-throughput studies provided first insights into molecular models of microexon sensitivity to SRRM3/4 and led to the design of further, more complex studies. The results satisfied Objective 3 of this Thesis.

It is tempting to imagine the implications and applications of the identified – or at least narrowed down – set of rules that dictate the differential sensitivity of microexons to SRRM3/4. Among these considerations, we could ask: what effects on the transcriptomes and proteomes of beta cells and neurons would microexons have if their sensitivity towards SRRM3/4 levels could be modulated?, or could we modulate specific biological processes by engineering microexon and regulatory sequences conferring predetermined sensitivity to SRRM3/4?

We will continue to ponder these and many more questions with the hope to better understand how the tiniest of exons have a mighty effect in tissues that evolved to use them.

CONCLUSIONS

Conclusion 1: For the first time, microexons were characterised in non-neural tissues, specifically in endocrine pancreas. In case of endocrine pancreatic microexons, that we called EndoMICs, the regulation is achieved by not by SRRM4 but its paralog, SRRM3. EndoMICs form a nested program within the larger neural program of microexons.

Conclusion 2: SRRM3-regulated EndoMICs impact keys aspects required for endocrine biology, such as calcium channel activity, mitochondrial integrity and function, and Golgi-vesicle apparatus. Deregulation of EndoMICs affects insulin secretory capacity, islet development and glucose homeostasis.

Conclusion 3: Regulation of nested programs of microexons is achieved by differential sensitivity to SRRM3/4 encoded in *cis*-acting regulatory elements. Exon length is one of the determinants that influence this sensitivity.

ACKNOWLEDGMENTS

To my Polish and Spanish family, for their support, love, and strength. I cannot thank them enough.

To Guillem. You have been with me for the better and the worse. It has been and will be epic.

To Katerina, her timeless friendship, and good vibes. You are my rock.

To Fabio and Borja, for the good old gym sessions, our “negative motivation”, too much coffee and conversations when we should have been working.

To Hanna, our “Get a life” never gets old. I hope you got one, girl.

To Marcos and Gosia, for our sushi nights, excessive drinking, and banter.

To Jonas and Sophie. Without you my projects would only exist as a Python code. Their ideas and experience shaped this Thesis.

To Imma, for her incredible capacity to listen.

To Manu and Juan, for their trust, almost unlimited patience, guidance, and inspiration that they offered.

BIBLIOGRAPHY

Agrawal, A. A., Salsi, E., Chatrikhi, R., Henderson, S., Jenkins, J. L., Green, M. R., Ermolenko, D. N., & Kielkopf, C. L. (2016). An extended U2AF65–RNA-binding domain recognizes the 3′ splice site signal. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms10950

Albaradei, S., Magana-Mora, A., Thafar, M., Uludag, M., Bajic, V. B., Gojobori, T., Essack, M., & Jankovic, B. R. (2020). Splice2Deep: An ensemble of deep convolutional neural networks for improved splice site prediction in genomic DNA. *Gene*, *763*, 100035. https://doi.org/10.1016/j.gene.2020.100035

Alvelos, M. I., Brüggemann, M., Sutandy, F. R., Juan-Mateu, J., Colli, M. L., Busch, A., Lopes, M., Castela, Â., Aartsma-Rus, A., König, J., Zarnack, K., & Eizirik, D. L. (2021). The RNA-binding profile of the splicing factor SRSF6 in immortalized human pancreatic β-cells. *Life Science Alliance*, *4*(3). https://doi.org/10.26508/lsa.202000825

Andersson, L. E., Valtat, B., Bagge, A., Sharoyko, V. v., Nicholls, D. G., Ravassard, P., Scharfmann, R., Spégel, P., & Mulder, H. (2015). Characterization of Stimulus-Secretion Coupling in the Human Pancreatic EndoC-βH1 Beta Cell Line. *PLOS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0120879

Archibald, J. M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology*, *25*(19). https://doi.org/10.1016/j.cub.2015.07.055

Arntfield, M. E., & van der Kooy, D. (2011). β-Cell evolution: How the pancreas borrowed from the brain. *BioEssays*, *33*(8). https://doi.org/10.1002/bies.201100015

Arous, C., & Halban, P. A. (2015). The skeleton in the closet: actin cytoskeletal remodeling in β-cell function. *American Journal of Physiology-Endocrinology and Metabolism*, *309*(7). https://doi.org/10.1152/ajpendo.00268.2015

Atouf, F., Czernichow, P., & Scharfmann, R. (1997). Expression of Neuronal Traits in Pancreatic Beta Cells. *Journal of Biological Chemistry*, *272*(3). https://doi.org/10.1074/jbc.272.3.1929

Baeza-Centurion, P., Miñana, B., Schmiedel, J. M., Valcárcel, J., & Lehner, B. (2019). Combinatorial Genetics Reveals a Scaling Law for the Effects of Mutations on Splicing. *Cell*, *176*(3), 549-563.e23. https://doi.org/10.1016/j.cell.2018.12.010

Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, *18*(7). https://doi.org/10.1038/nrm.2017.27

Bedi, K., Magnuson, B., Narayanan, I. V., Paulsen, M. T., Wilson, T. E., & Ljungman, M. (2021). Cotranscriptional splicing efficiencies differ within genes and between cell

types. *RNA*, *27*(7), 829–840. https://doi.org/10.1261/rna.078662.120

Berget, S. M. (1995). Exon Recognition in Vertebrate Splicing. *Journal of Biological Chemistry*, *270*(6). https://doi.org/10.1074/jbc.270.6.2411

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences*, *74*(8). https://doi.org/10.1073/pnas.74.8.3171

Black, D. L. (1995). Findig splice sites within a wilderness of RNA. *RNA*, *1*, 763–771.

Bonner, J. T. (1998). The origins of multicellularity. *Integrative Biology: Issues, News, and Reviews*, *1*(1). https://doi.org/10.1002/(SICI)1520-6602(1998)1:1<27::AID-INBI4>3.0.CO;2-6

Calarco, J. A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinas, D., van der Kooy, D., Zhen, M., Ciruna, B., & Blencowe, B. J. (2009). Regulation of Vertebrate Nervous System Alternative Splicing and Development by an SR-Related Protein. *Cell*, *138*(5). https://doi.org/10.1016/j.cell.2009.06.012

Cavalier-Smith, T. (2010). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biology Direct*, *5*(1). https://doi.org/10.1186/1745-6150-5-7

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5′

ends of adenovirus 2 messenger RNA. *Cell*, *12*(1). https://doi.org/10.1016/0092-8674(77)90180-5

Christofi, T., & Zaravinos, A. (2019). RNA editing in the forefront of epitranscriptomics and human health. *Journal of Translational Medicine*, *17*(1). https://doi.org/10.1186/s12967-019-2071-4

Clevers, H. (2006). Wnt/β-Catenin Signaling in Development and Disease. *Cell*, *127*(3). https://doi.org/10.1016/j.cell.2006.10.018

Climente-González, H., Porta-Pardo, E., Godzik, A., & Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Reports*, *20*(9). https://doi.org/10.1016/j.celrep.2017.08.012

Colli, M. L., Ramos-Rodríguez, M., Nakayasu, E. S., Alvelos, M. I., Lopes, M., Hill, J. L. E., Turatsinze, J.-V., Coomans de Brachène, A., Russell, M. A., Raurell-Vila, H., Castela, A., Juan-Mateu, J., Webb-Robertson, B.-J. M., Krogvold, L., Dahl-Jorgensen, K., Marselli, L., Marchetti, P., Richardson, S. J., Morgan, N. G., … Eizirik, D. L. (2020). An integrated multi-omics approach identifies the landscape of interferon-α-mediated responses of human pancreatic beta cells. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16327-0

Corvelo, A., Hallegger, M., Smith, C. W. J., & Eyras, E. (2010). Genome-Wide Association between Branch Point Properties and Alternative Splicing. *PLoS Computational*

Biology, *6*(11), e1001016. https://doi.org/10.1371/journal.pcbi.1001016

Darras, B. T., Chiriboga, C. A., Iannaccone, S. T., Swoboda, K. J., Montes, J., Mignon, L., Xia, S., Bennett, C. F., Bishop, K. M., Shefner, J. M., Green, A. M., Sun, P., Bhan, I., Gheuens, S., Schneider, E., Farwell, W., & de Vivo, D. C. (2019). Nusinersen in later-onset spinal muscular atrophy. *Neurology*, *92*(21). https://doi.org/10.1212/WNL.0000000000007527

Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P., & Reed, R. (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Molecular Cell*, *26*(6), 867–881. https://doi.org/10.1016/j.molcel.2007.05.036

Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, *51*(6). https://doi.org/10.1016/0092-8674(87)90585-X

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1). https://doi.org/10.1093/bioinformatics/bts635

Du, H., & Rosbash, M. (2001). Yeast U1 snRNP–pre-mRNA complex formation without U1snRNA–pre-mRNA base pairing. *RNA*, *7*(1). https://doi.org/10.1017/S1355838201001844

Du, H., & Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. *Nature*, *419*(6902). https://doi.org/10.1038/nature00947

Elkon, R., Ugalde, A. P., & Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics*, *14*(7). https://doi.org/10.1038/nrg3482

Ensembl. (2021, March). *Human assembly and gene annotation*.

Fadista, J., Vikman, P., Laakso, E. O., Mollet, I. G., Esguerra, J. L., Taneera, J., Storm, P., Osmark, P., Ladenvall, C., Prasad, R. B., Hansson, K. B., Finotello, F., Uvebrant, K., Ofori, J. K., di Camillo, B., Krus, U., Cilio, C. M., Hansson, O., Eliasson, L., … Groop, L. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proceedings of the National Academy of Sciences*, *111*(38), 13924–13929. https://doi.org/10.1073/pnas.1402665111

Finkel, R. S., Mercuri, E., Darras, B. T., Connolly, A. M., Kuntz, N. L., Kirschner, J., Chiriboga, C. A., Saito, K., Servais, L., Tizzano, E., Topaloglu, H., Tulinius, M., Montes, J., Glanzman, A. M., Bishop, K., Zhong, Z. J., Gheuens, S., Bennett, C. F., Schneider, E., … de Vivo, D. C. (2017). Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *New England*

Journal of Medicine, *377*(18). https://doi.org/10.1056/NEJMoa1702752

Förch, P., & Valcárcel, J. (2003). *Splicing Regulation in Drosophila Sex Determination*. https://doi.org/10.1007/978-3-662-09728-1_5

Gaisano, H. Y. (2017). Recent new insights into the role of SNARE and associated proteins in insulin granule exocytosis. *Diabetes, Obesity and Metabolism*, *19*. https://doi.org/10.1111/dom.13001

Galloway, A., & Cowling, V. H. (2019). mRNA cap regulation in mammalian cell function and fate. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1862*(3). https://doi.org/10.1016/j.bbagrm.2018.09.011

Gandasi, N. R., & Barg, S. (2014). Contact-induced clustering of syntaxin and munc18 docks secretory granules at the exocytosis site. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms4914

Gandasi, N. R., Yin, P., Riz, M., Chibalina, M. v., Cortese, G., Lund, P.-E., Matveev, V., Rorsman, P., Sherman, A., Pedersen, M. G., & Barg, S. (2017). Ca2+ channel clustering with insulin-containing granules is disturbed in type 2 diabetes. *Journal of Clinical Investigation*, *127*(6). https://doi.org/10.1172/JCI88491

Gao, D., Morini, E., Salani, M., Krauson, A. J., Chekuri, A., Sharma, N., Ragavendran, A., Erdin, S., Logan, E. M., Li, W., Dakka, A., Narasimhan, J., Zhao, X., Naryshkin, N., Trotta, C. R., Effenberger, K. A., Woll, M. G., Gabbeta,

V., Karp, G., … Slaugenhaupt, S. A. (2021). A deep learning approach to identify gene targets of a therapeutic for human splicing disorders. *Nature Communications*, *12*(1), 3332. https://doi.org/10.1038/s41467-021-23663-2

Gohr, A., & Irimia, M. (2019). *Matt* : Unix tools for alternative splicing analysis. *Bioinformatics*, *35*(1), 130–132. https://doi.org/10.1093/bioinformatics/bty606

Gonatopoulos-Pournatzis, T., & Blencowe, B. J. (2020). Microexons: at the nexus of nervous system development, behaviour and autism spectrum disorder. *Current Opinion in Genetics & Development*, *65*. https://doi.org/10.1016/j.gde.2020.03.007

Gonatopoulos-Pournatzis, T., Niibori, R., Salter, E. W., Weatheritt, R. J., Tsang, B., Farhangmehr, S., Liang, X., Braunschweig, U., Roth, J., Zhang, S., Henderson, T., Sharma, E., Quesnel-Vallières, M., Permanyer, J., Maier, S., Georgiou, J., Irimia, M., Sonenberg, N., Forman-Kay, J. D., … Blencowe, B. J. (2020). Autism-Misregulated eIF4G Microexons Control Synaptic Translation and Higher Order Cognitive Functions. *Molecular Cell*, *77*(6). https://doi.org/10.1016/j.molcel.2020.01.006

Gonatopoulos-Pournatzis, T., Wu, M., Braunschweig, U., Roth, J., Han, H., Best, A. J., Raj, B., Aregger, M., O'Hanlon, D., Ellis, J. D., Calarco, J. A., Moffat, J., Gingras, A.-C., & Blencowe, B. J. (2018). Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks

Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Molecular Cell*, *72*(3). https://doi.org/10.1016/j.molcel.2018.10.008

Grabowski, P. J., & Sharp, P. A. (1986). Affinity Chromatography of Splicing Complexes: U2, U5, and U4 + U6 Small Nuclear Ribonucleoprotein Particles in the Spliceosome. *Science*, *233*(4770). https://doi.org/10.1126/science.3638792

Grau-Bové, X., Ruiz-Trillo, I., & Irimia, M. (2018). Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1499-9

Hendrickson, H. L., & Poole, A. M. (2018). Manifold Routes to a Nucleus. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02604

Hollins, C. (2005). U2AF binding selects for the high conservation of the C. elegans 3' splice site. *RNA*, *11*(3). https://doi.org/10.1261/rna.7221605

Horiike, T., Hamada, K., Kanaya, S., & Shinozawa, T. (2001). Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nature Cell Biology*, *3*(2). https://doi.org/10.1038/35055129

Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F., & Krainer, A. R. (2008). Antisense Masking of an hnRNP A1/A2 Intronic Splicing Silencer Corrects SMN2 Splicing in Transgenic Mice. *The American Journal of Human*

Genetics, *82*(4). https://doi.org/10.1016/j.ajhg.2008.01.014

Irimia, M., & Blencowe, B. J. (2012). Alternative splicing: decoding an expansive regulatory layer. *Current Opinion in Cell Biology*, *24*(3). https://doi.org/10.1016/j.ceb.2012.03.005

Irimia, M., Penny, D., & Roy, S. W. (2007). Coevolution of genomic intron number and splice sites. *Trends in Genetics*, *23*(7). https://doi.org/10.1016/j.tig.2007.04.001

Irimia, M., & Roy, S. W. (2008). Evolutionary Convergence on Highly-Conserved 3′ Intron Structures in Intron-Poor Eukaryotes and Insights into the Ancestral Eukaryotic Genome. *PLoS Genetics*, *4*(8). https://doi.org/10.1371/journal.pgen.1000148

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., Barrios-Rodiles, M., Sternberg, M. J. E., Cordes, S. P., Roth, F. P., Wrana, J. L., Geschwind, D. H., & Blencowe, B. J. (2014). A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains. *Cell*, *159*(7). https://doi.org/10.1016/j.cell.2014.11.035

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting Splicing from

Primary Sequence with Deep Learning. *Cell*, *176*(3), 535-548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Juan-Mateu, J., Alvelos, M. I., Turatsinze, J.-V., Villate, O., Lizarraga-Mollinedo, E., Grieco, F. A., Marroquí, L., Bugliani, M., Marchetti, P., & Eizirik, D. L. (2018). SRp55 Regulates a Splicing Network That Controls Human Pancreatic β-Cell Function and Survival. *Diabetes*, *67*(3). https://doi.org/10.2337/db17-0736

Juan-Mateu, J., Rech, T. H., Villate, O., Lizarraga-Mollinedo, E., Wendt, A., Turatsinze, J.-V., Brondani, L. A., Nardelli, T. R., Nogueira, T. C., Esguerra, J. L. S., Alvelos, M. I., Marchetti, P., Eliasson, L., & Eizirik, D. L. (2017). Neuron-enriched RNA-binding Proteins Regulate Pancreatic Beta Cell Function and Survival. *Journal of Biological Chemistry*, *292*(8). https://doi.org/10.1074/jbc.M116.748335

Juan-Mateu, J., Villate, O., & Eizirik, D. L. (2016). MECHANISMS IN ENDOCRINOLOGY: Alternative splicing: the new frontier in diabetes research. *European Journal of Endocrinology*, *174*(5). https://doi.org/10.1530/EJE-15-0916

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, *468*(7324). https://doi.org/10.1038/nature09479

Ke, S., Anquetil, V., Zamalloa, J. R., Maity, A., Yang, A., Arias, M. A., Kalachikov, S., Russo, J. J., Ju, J., &

Chasin, L. A. (2018). Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Research*, *28*(1), 11–24. https://doi.org/10.1101/gr.219683.116

Ke, S., Shang, S., Kalachikov, S. M., Morozova, I., Yu, L., Russo, J. J., Ju, J., & Chasin, L. A. (2011). Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Research*, *21*(8), 1360–1374. https://doi.org/10.1101/gr.119628.110

Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, gkv717. https://doi.org/10.1093/nar/gkv717

Kennedy, C. F., & Berget, S. M. (1997). Pyrimidine tracts between the 5’ splice site and branch point facilitate splicing and recognition of a small Drosophila intron. *Molecular and Cellular Biology*, *17*(5). https://doi.org/10.1128/MCB.17.5.2774

King, N. (2004). The Unicellular Ancestry of Animal Development. *Developmental Cell*, *7*(3). https://doi.org/10.1016/j.devcel.2004.08.010

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., & Peterson, H. (2020). gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, *9*. https://doi.org/10.12688/f1000research.24956.2

Konarska, M. M. (1998). Recognition of the 5’ splice site by the spliceosome. *Acta Biochimica Polonica*, *45*(4).

Lake, J. A. (1989). Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. *Canadian Journal of Microbiology*, *35*(1). https://doi.org/10.1139/m89-017

Lamond, A. I., Konarska, M. M., Grabowski, P. J., & Sharp, P. A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proceedings of the National Academy of Sciences*, *85*(2). https://doi.org/10.1073/pnas.85.2.411

Lassar, A. B., Paterson, B. M., & Weintraub, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T12 fibroblasts to myoblasts. *Cell*, *47*(5). https://doi.org/10.1016/0092-8674(86)90507-6

Lees, J. A., Messa, M., Sun, E. W., Wheeler, H., Torta, F., Wenk, M. R., de Camilli, P., & Reinisch, K. M. (2017). Lipid transport by TMEM24 at ER–plasma membrane contacts regulates pulsatile insulin secretion. *Science*, *355*(6326). https://doi.org/10.1126/science.aah6171

Lenoir, O., Flosseau, K., Ma, F. X., Blondeau, B., Mai, A., Bassel-Duby, R., Ravassard, P., Olson, E. N., Haumaitre, C., & Scharfmann, R. (2011). Specific Control of Pancreatic Endocrine β- and δ-Cell Mass by Class IIa Histone Deacetylases HDAC4, HDAC5, and HDAC9. *Diabetes*, *60*(11), 2861–2871. https://doi.org/10.2337/db11-0440

Leung, Y. M., Kwan, E. P., Ng, B., Kang, Y., & Gaisano, H. Y. (2007). SNAREing Voltage-Gated K+ and ATP-Sensitive K+ Channels: Tuning β-Cell Excitability with Syntaxin-1A and Other Exocytotic Proteins. *Endocrine Reviews*, *28*(6). https://doi.org/10.1210/er.2007-0010

Li, Y., & Blencowe, B. J. (1999). Distinct Factor Requirements for Exonic Splicing Enhancer Function and Binding of U2AF to the Polypyrimidine Tract. *Journal of Biological Chemistry*, *274*(49). https://doi.org/10.1074/jbc.274.49.35074

Li, Y. I., Sanchez-Pulido, L., Haerty, W., & Ponting, C. P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Research*, *25*(1). https://doi.org/10.1101/gr.181990.114

Lin, L., Zhang, M., Stoilov, P., Chen, L., & Zheng, S. (2020). Developmental Attenuation of Neuronal Apoptosis by Neural-Specific Splicing of Bak1 Microexon. *Neuron*, *107*(6). https://doi.org/10.1016/j.neuron.2020.06.036

Lorson, C. L., Hahnen, E., Androphy, E. J., & Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences*, *96*(11). https://doi.org/10.1073/pnas.96.11.6307

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq

data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Malakar, P., Chartarifsky, L., Hija, A., Leibowitz, G., Glaser, B., Dor, Y., & Karni, R. (2016). Insulin receptor alternative splicing is regulated by insulin signaling and modulates beta cell survival. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep31222

Marselli, L., Piron, A., Suleiman, M., Colli, M. L., Yi, X., Khamis, A., Carrat, G. R., Rutter, G. A., Bugliani, M., Giusti, L., Ronci, M., Ibberson, M., Turatsinze, J.-V., Boggi, U., de Simone, P., de Tata, V., Lopes, M., Nasteska, D., de Luca, C., … Marchetti, P. (2020). Persistent or Transient Human β Cell Dysfunction Induced by Metabolic Stress: Specific Signatures and Shared Gene Expression with Type 2 Diabetes. *Cell Reports*, *33*(9), 108466. https://doi.org/10.1016/j.celrep.2020.108466

Martin, D., & Grapin-Botton, A. (2017). The Importance of REST for Development and Function of Beta Cells. *Frontiers in Cell and Developmental Biology*, *5*. https://doi.org/10.3389/fcell.2017.00012

Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1678). https://doi.org/10.1098/rstb.2014.0330

Maslon, M. M., Braunschweig, U., Aitken, S., Mann, A. R., Kilanowski, F., Hunter, C. J., Blencowe, B. J., Kornblihtt,

A. R., Adams, I. R., & Cáceres, J. F. (2019). A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *The EMBO Journal*, *38*(9). https://doi.org/10.15252/embj.2018101244

Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, *15*(2). https://doi.org/10.1038/nrm3742

Miguel-Escalada, I., Bonàs-Guarch, S., Cebola, I., Ponsa-Cobas, J., Mendieta-Esteban, J., Atla, G., Javierre, B. M., Rolando, D. M. Y., Farabella, I., Morgan, C. C., García-Hurtado, J., Beucher, A., Morán, I., Pasquali, L., Ramos-Rodríguez, M., Appel, E. V. R., Linneberg, A., Gjesing, A. P., Witte, D. R., … Ferrer, J. (2019). Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nature Genetics*, *51*(7), 1137–1148. https://doi.org/10.1038/s41588-019-0457-0

Mikl, M., Hamburg, A., Pilpel, Y., & Segal, E. (2019). Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries. *Nature Communications*, *10*(1), 4572. https://doi.org/10.1038/s41467-019-12642-3

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., …

Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3). https://doi.org/10.1038/ng1180

Muñoz, M. J., Santangelo, M. S. P., Paronetto, M. P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J. J., Bird, G., Bentley, D., Bertrand, E., & Kornblihtt, A. R. (2009). DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. *Cell*, *137*(4), 708–720. https://doi.org/10.1016/j.cell.2009.03.010

Naftelberg, S., Schor, I. E., Ast, G., & Kornblihtt, A. R. (2015). Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Annual Review of Biochemistry*, *84*(1), 165–198. https://doi.org/10.1146/annurev-biochem-060614-034242

Nakano, Y., Jahan, I., Bonde, G., Sun, X., Hildebrand, M. S., Engelhardt, J. F., Smith, R. J. H., Cornell, R. A., Fritzsch, B., & Bánfi, B. (2012). A Mutation in the Srrm4 Gene Causes Alternative Splicing Defects and Deafness in the Bronx Waltzer Mouse. *PLoS Genetics*, *8*(10). https://doi.org/10.1371/journal.pgen.1002966

Nakano, Y., Wiechert, S., & Bánfi, B. (2019). Overlapping Activities of Two Neuronal Splicing Factors Switch the GABA Effect from Excitatory to Inhibitory by Regulating REST. *Cell Reports*, *27*(3). https://doi.org/10.1016/j.celrep.2019.03.072

Nelson, K. K., & Green, M. R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes & Development*, *3*(10), 1562–1571. https://doi.org/10.1101/gad.3.10.1562

Nica, A. C., Ongen, H., Irminger, J.-C., Bosco, D., Berney, T., Antonarakis, S. E., Halban, P. A., & Dermitzakis, E. T. (2013). Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Research*, *23*(9). https://doi.org/10.1101/gr.150706.112

Papasaikas, P., & Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences*, *41*(1). https://doi.org/10.1016/j.tibs.2015.11.003

Quesnel-Vallières, M., Dargaei, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J. Y., Wu, M., Sterne-Weiler, T., Nakagawa, S., Woodin, M. A., Blencowe, B. J., & Cordes, S. P. (2016). Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Molecular Cell*, *64*(6). https://doi.org/10.1016/j.molcel.2016.11.033

Quesnel-Vallières, M., Irimia, M., Cordes, S. P., & Blencowe, B. J. (2015). Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes & Development*, *29*(7). https://doi.org/10.1101/gad.256115.114

Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z.-Y., Chen, G. I., Easton, L. E., Ule,

J., Gingras, A.-C., Eyras, E., & Blencowe, B. J. (2014). A Global Regulatory Mechanism for Activating an Exon Network Required for Neurogenesis. *Molecular Cell*, *56*(1). https://doi.org/10.1016/j.molcel.2014.08.011

Ravassard, P., Hazhouz, Y., Pechberty, S., Bricout-Neveu, E., Armanet, M., Czernichow, P., & Scharfmann, R. (2011). A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *Journal of Clinical Investigation*, *121*(9). https://doi.org/10.1172/JCI58447

Regazzi, R., Wollheim, C. B., Lang, J., Theler, J. M., Rossetto, O., Montecucco, C., Sadoul, K., Weller, U., Palmer, M., & Thorens, B. (1995). VAMP-2 and cellubrevin are expressed in pancreatic beta-cells and are essential for Ca(2+)-but not for GTP gamma S-induced insulin secretion. *The EMBO Journal*, *14*(12).

Robberson, B. L., Cote, G. J., & Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology*, *10*(1). https://doi.org/10.1128/mcb.10.1.84-94.1990

Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Current Biology*, *27*(21). https://doi.org/10.1016/j.cub.2017.09.015

Rorsman, P., Braun, M., & Zhang, Q. (2012). Regulation of calcium in pancreatic α- and β-cells in health and

disease. *Cell Calcium*, *51*(3–4). https://doi.org/10.1016/j.ceca.2011.11.006

Rosenberg, A. B., Patwardhan, R. P., Shendure, J., & Seelig, G. (2015). Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. *Cell*, *163*(3), 698–711. https://doi.org/10.1016/j.cell.2015.09.054

Ruskin, B., Zamore, P. D., & Green, M. R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*, *52*(2). https://doi.org/10.1016/0092-8674(88)90509-0

Sadoul, K., Lang, J., Montecucco, C., Weller, U., Regazzi, R., Catsicas, S., Wollheim, C. B., & Halban, P. A. (1995). SNAP-25 is expressed in islets of Langerhans and is involved in insulin release. *Journal of Cell Biology*, *128*(6). https://doi.org/10.1083/jcb.128.6.1019

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*(3). https://doi.org/10.1016/0022-5193(67)90079-3

Scharfmann, R., Staels, W., & Albagli, O. (2019). The supply chain of human pancreatic β cell lines. *Journal of Clinical Investigation*, *129*(9). https://doi.org/10.1172/JCI129484

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, *17*(1). https://doi.org/10.1038/nrg.2015.3

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A. S., Nguyen, V. T., Hébert, C., Schmidt, U., Bertrand, E., Darzacq, X., &

Bensaude, O. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *Journal of Cell Science*, *123*(12), 2085–2093. https://doi.org/10.1242/jcs.061358

Steiner, D. J., Kim, A., Miller, K., & Hara, M. (2010). Pancreatic islet plasticity: Interspecies comparison of islet architecture and composition. *Islets*, *2*(3). https://doi.org/10.4161/isl.2.3.11815

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43). https://doi.org/10.1073/pnas.0506580102

Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., Cozzuto, L., Wang, X., Gómez-Velázquez, M., Rayon, T., Manzanares, M., Ponomarenko, J., Blencowe, B. J., & Irimia, M. (2017a). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Research*, *27*(10). https://doi.org/10.1101/gr.220962.117

Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., Cozzuto, L., Wang, X., Gómez-Velázquez, M., Rayon, T., Manzanares, M., Ponomarenko, J., Blencowe, B. J., & Irimia, M. (2017b). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Research*, *27*(10). https://doi.org/10.1101/gr.220962.117

Tarasov, A., Dusonchet, J., & Ashcroft, F. (2004). Metabolic Regulation of the Pancreatic Beta-Cell ATP-Sensitive K+ Channel: A Pas de Deux. *Diabetes*, *53*(Supplement 3). https://doi.org/10.2337/diabetes.53.suppl_3.S113

Tarasov, A. I., Griffiths, E. J., & Rutter, G. A. (2012). Regulation of ATP production by mitochondrial Ca2+. *Cell Calcium*, *52*(1). https://doi.org/10.1016/j.ceca.2012.03.003

Tetenbaum-Novatt, J., & Rout, M. P. (2010). The Mechanism of Nucleocytoplasmic Transport through the Nuclear Pore Complex. *Cold Spring Harbor Symposia on Quantitative Biology*, *75*(0). https://doi.org/10.1101/sqb.2010.75.033

Tokarz, V. L., MacDonald, P. E., & Klip, A. (2018). The cell biology of systemic insulin function. *Journal of Cell Biology*, *217*(7). https://doi.org/10.1083/jcb.201802095

Torres-Méndez, A., Bonnal, S., Marquez, Y., Roth, J., Iglesias, M., Permanyer, J., Almudí, I., O'Hanlon, D., Guitart, T., Soller, M., Gingras, A.-C., Gebauer, F., Rentzsch, F., Blencowe, B. J., Valcárcel, J., & Irimia, M. (2019). A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nature Ecology & Evolution*, *3*(4). https://doi.org/10.1038/s41559-019-0813-6

Ule, J., & Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, *76*(2). https://doi.org/10.1016/j.molcel.2019.09.017

van Arensbergen, J., Garcia-Hurtado, J., Moran, I., Maestro, M. A., Xu, X., van de Casteele, M., Skoudy, A. L., Palassini, M., Heimberg, H., & Ferrer, J. (2010). Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program. *Genome Research*, *20*(6). https://doi.org/10.1101/gr.101709.109

van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., & Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, *44*(2). https://doi.org/10.1016/0092-8674(86)90756-7

Villarreal, D., Pradhan, G., Wu, C.-S., Allred, C. D., Guo, S., & Sun, Y. (2019). A Simple High Efficiency Protocol for

Pancreatic Islet Isolation from Mice. *Journal of Visualized Experiments*, *150*. https://doi.org/10.3791/57048

Villate, O., Turatsinze, J.-V., Mascali, L. G., Grieco, F. A., Nogueira, T. C., Cunha, D. A., Nardelli, T. R., Sammeth, M., Salunkhe, V. A., Esguerra, J. L. S., Eliasson, L., Marselli, L., Marchetti, P., & Eizirik, D. L. (2014). Nova1 is a master regulator of alternative splicing in pancreatic beta cells. *Nucleic Acids Research*, *42*(18). https://doi.org/10.1093/nar/gku861

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, *136*(4). https://doi.org/10.1016/j.cell.2009.02.009

Wang, Z., & Burge, C. B. (2008). Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*, *14*(5), 802–813. https://doi.org/10.1261/rna.876308

Wang, Z., Hoffmann, H. M., & Grabowski, P. J. (1995). Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA (New York, N.Y.)*, *1*(1).

Watson, J. D., & Crick, F. H. C. (1953). A Structure for Deoxyribose Nucleic Acid. *Nature*, *4355*. https://doi.org/10.1038/248765a0

Wilhelmi, I., Neumann, A., Jähnert, M., Ouni, M., & Schürmann, A. (2021). Enriched Alternative Splicing in Islets of Diabetes-Susceptible Mice. *International Journal*

of Molecular Sciences, *22*(16). https://doi.org/10.3390/ijms22168597

Witten, J. T., & Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. *Trends in Genetics*, *27*(3), 89–97. https://doi.org/10.1016/j.tig.2010.12.001

Wong, M. S., Kinney, J. B., & Krainer, A. R. (2018a). Quantitative Activity Profile and Context Dependence of All Human 5′ Splice Sites. *Molecular Cell*, *71*(6), 1012-1026.e3. https://doi.org/10.1016/j.molcel.2018.07.033

Wong, M. S., Kinney, J. B., & Krainer, A. R. (2018b). Quantitative Activity Profile and Context Dependence of All Human 5′ Splice Sites. *Molecular Cell*, *71*(6), 1012-1026.e3. https://doi.org/10.1016/j.molcel.2018.07.033

Zhang, I. X., Raghavan, M., & Satin, L. S. (2020). The Endoplasmic Reticulum and Calcium Homeostasis in Pancreatic Beta Cells. *Endocrinology*, *161*(2). https://doi.org/10.1210/endocr/bqz028

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, *35*(12). https://doi.org/10.1093/bioinformatics/bty895

Zhuang, Y. A., Goldstein, A. M., & Weiner, A. M. (1989). UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proceedings of the National Academy of Sciences*, *86*(8), 2752–2756. https://doi.org/10.1073/pnas.86.8.2752