universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Experimental verification and functional exploration of intrasplicing"

verfasst von / submitted by

Ismet Srndic, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Renée Schroeder

"Der Ernst, mein Junge, ist eine Angelegenheit der Zeit;
er entsteht, soviel will ich dir verraten, aus einer
Überschätzung der Zeit... In der Ewigkeit aber, siehst du,
gibt es keine Zeit; die Ewigkeit ist bloß ein Augenblick,
gerade lange genug für einen Spaß."[1]

Acknowledgements

Deceived by the illusion of the constant we, for the most part, stay unaware that “whatever has form can disappear in an instant”[2]. This is my attempt in appreciating the form that my life has acquired and the people that contributed to it, including my education that’s being crowned by this piece of work.

My utmost gratitude to Mr. sharp-mind-kind-heart for being the essential link between me and a bunch of life I had since joining the lab. For being the personification of a calming, always flowing river, guiding me and pointing out the red thread in the labyrinth that science occasionally turns out to be. For generously sharing his knowledge and introducing me to spheres of life that I might have stayed unaware of. Captain, it’s been an amazing journey and thanks to you I sailed on the exact ship that I wanted to be on. As life is full of unpredictable opportunities and possibilities I really hope for one to return this tremendous favour.

I am very grateful for the gravity of our group, a woman that has collected a lot of successes in her life of which a particular one I find very special - her enthusiasm and cheerfulness – *heiterkeit*, that smile on her face which I assume was not always easy to keep up. I find it so inspiring that she treats any human she encounters with kindness and respect, no matter if dealing with a servant or a ruler; that she did not compromise her kindness and empathy for power and responsibility. Our achievement based society leaves barely any space for the appreciation of such intrinsic values, but as she probably knows many people, including me are very thankful for hers. It has been an honour and tremendous pleasure being your *labkid*, being exposed to you, your thought-games. Arigatou!

The next round of appreciation goes to my fellow *labkids* and *kids* from the neighbouring labs. I had this revelation on a lonesome day in the lab that I anyway feel very comfortable there, don't even mind the unappealing light, the stench of old bacteria nor the buzzing of centrifuges - it's a huge comfort zone for me. Then again your presence has been adding a thick layer of pleasantness to this whole experience. Thank you for the chats, games, hugs and drinks. For shaping this atmosphere of comfort where my weirdness could flourish and develop into previously unreached dimensions. Even at times when the sweet fruit of science

turned out to taste very bitter, there still was this nice environment to come to, to enjoy and just be. Looking back, it gets obvious that the freedom I was given ended up being so captivating. Not that I mind it the least bit. In fact, I'm utterly thankful for it. I had a blast sharing a time and space with you – thank you.

To get a bit more general I also feel obligated to acknowledge the people who contributed to the accumulation of culture and knowledge that we build our lives upon, who thought the unthought and bothered sharing it. The people developing language and writing. It doesn't cease to amaze me that one can, just by reading ride somebody else's train of thoughts and semantics set aside, demolish the limitations of space and time. The people who created norms, those abolishing them. The weirdos jumping on the impulses, the ascetics learning how to control them. The artists who managed to extract the magnificence of life and make it consumable for us ordinary people. And the people developing humour, they are some of my favourites. As much as I adore science and despise superlatives, I still consider humour to be humans' greatest invention. I am grateful for the people who bothered teaching me something, even though I might not have appreciated it at the moment it was happening. There have been so many people I encountered that left a mark and contributed to the acquisition of the mind-set that I'm forming. I reckon that it might be headed in a different direction without them.

Then again the knowledge would be pretty useless for me without the appropriate cognitive abilities to comprehend it. At this point I acknowledge the universe and life in general for developing these abilities and my ancestors and parents for providing me with them. Speaking of my parents, I owe my deepest gratitude to you and the people closest to me. Your unconditional love and support are the monumental pillars to build an amazing life upon. The empathy that you fostered in me is a great indicator of the excellent care I was given from an infant age. As I'm getting so much from you, you make it so easy for me to invest and waste myself, to absorb life, simply enjoy it. You make my life so easy, so light. You are my wings, my safety net. And since you have to put up with the least pleasant parts of me, to compensate a little bit I dedicate to you (along with this work and anything I do that can be considered good) my smiles, my joy, my cheerfulness. This body just serves as the medium to express, my heart and mind as amplifiers, but the actual source of them are you.

Table of content

Abbreviations

A	adenine
bp	base pair
BP	branch point
BPS	branch point site
C	cytosine
cDNA	complementary DANN
DB	Debranching enzyme
dbr1	debranching enzyme
ddH2O	double-distilled water
del	deletion
DMEM	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
e, ex	exon
EDTA	ethylenediaminetetraacetic acid
Fwd	forward primer
G	guanine
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
HEK	Human Embryonic Kidney cell line
HeLa	cervical cancer cell line
HEP	human hepatocytes
int	intron
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k	kilo
KD	knock down
lar	lariat
lasli	lariat spanning linear split
LB	lysogeny broth
mRNA	messenger RNA
NEB	New England Biolabs
NGS	next generation sequencing
ni	nested inner
NMD	nonsense-mediated decay
no	nested outer
nt	nucleotide
PCI	Phenol stabilized: Chloroform: Isoamyl Alcohol
PCR	polymerase chain reaction
pre	prespliced
pre-mRNA	precursor mRNA
pY	polypyrimidine tract
qPCR	quantitative PCR
Rev	reverse primer
RNA	ribonucleic acid

RNP	ribonucleoprotein
RPMI	Roswell Park Memorial Institute medium
RT	reverse transcription
seq	sequencing
SF	splicing factor
shRNA	short hairpin RNA
sn	small nuclear
SOC	Super Optimal broth with Catabolite repression
SR	serine arginine
ss	splice site
T	thymine
TBE	tris borat EDTA
Tri	TRI Reagent®
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
wt	wild type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
y	pyrimidine

Introduction

The importance of splicing

The discovery that genes are not always continuous, but can be interrupted by intervening sequences, later named introns, was somewhat surprising. In the late seventies they were first found in viral mRNAs [3].

Expectedly the human mind, unable to rationalize occurrences with pure necessity, craved for logic and meaning. So it was not surprising, even before it was obvious that introns are omnipresent in genomes of most eukaryotes that the *why* question arose. It certainly is a legit one: why does a cell bother transcribing the dystrophin precursor mRNA (pre-mRNA) for 15 hours when it could get the exonic sequences in 20 minutes?

The sole existence of introns is not to be expected from the view of the central dogma in molecular biology. Even less to be expected is that a quarter of the human genome codes for introns, while only approximately 1.1% is actually protein coding[4]. While exons are rather short, the average size is 147 nt[5] and even single nucleotide exons have been found[6]; the average intron length in human is 3413 nt[7]. Though prevalently described in eukaryotes, introns have also been found in bacteria, particularly self-splicing group I introns in mRNAs, rRNAs and tRNAs[8]. In humans most genes contain introns, almost all of them multiple.

The next big discovery in the field was alternative splicing – cells' powerful tool to enlarge its proteome without increasing gene quantity. Before sequencing the human genome it was estimated that it contains around 100 000 protein coding genes, afterwards it became clear that many proteins existed due to alternative splicing and the number of genes was reduced to approximately 20 000 genes[9]. It has been estimated that 95% of multiexonic genes in humans undergo alternative splicing[10]. But there is still room for novel surprises because not all introns are involved in alternative splicing so they could have additional functions.

Studies comparing nuclear export of mRNA that has previously been spliced with its intronless analogue showed that introns increase efficiency of the export process, accomplished by binding of the export factors to mRNA during spliceosome assembly[11]. In some cases it seems that this is the exclusive way for

nuclear export proteins to bind mRNA, which may serve as a quality control ensuring that only spliced mRNA are actually exported and translated[12]. It has been shown that components of the splicing machinery interact with the cleavage and polyadenylation specificity factor (CPSF), *in vitro* this contact leads to a more efficient polyadenylation[13]. Mutations in the polyadenylation signal not only abolished the poly-A formation but also lead to a less efficient removal of the terminal intron. The cleavage and polyadenylation machinery can also affect alternative splicing[14]. Its components help defining the last exon and this interaction enhances splicing as well as polyadenylation (reviewed in[15]).

Another attempt to rationalize the omnipresence of introns is that they might function as evolution regulators. Since mutations are evolution's necessity occurring anywhere in the genome at a certain rate, there is a higher probability for them to affect coding parts in compact genomes. Introns might be exposing themselves to the obligatory mutations and protecting the exons from mutations. Of course other influences have to be taken into this equation (e.g. the error rates of polymerases) but a rough hypothesis would be that there is a certain correlation between the percentage of introns in a genome and the evolutionary pace of an organism.

Supporting evidence has been described in two (evolutionary) closely related *Arabidopsis* species: *Arabidopsis lyrata* and *Arabidopsis thaliana*, where the first has 60% more nucleotides in its genome. The consequence of *A. thaliana*'s genome reduction is that 42,3% of the sequences are coding, compared to 28,7% of the *A. lyrata* genome that is coding[16]. Altogether the mutation rate in *A. thaliana* is significantly higher in *A. lyrata*[17], the authors hypothesise that the introns function to minimize the *hazardous* impact of mutations.

Learning that the genome size does not correlate with the complexity of life was a hard pill to swallow for the quantity driven human ego. Some animals - even plants and protist have larger genomes than ours[18]. When it comes to introns the odds seem to be in our favour: human genes contain 9.4 introns per gene[19] which is the highest number in the compared species[20]. It is highly improbable that complex life would have arisen without introns, the cognitive abilities necessary to comprehend this or any other thought are hardly imaginable without complexity provided by introns.

Splicing mechanism

How does a cell differentiate between introns and exons on a pre-mRNA? Another surprise in the history of splicing was that introns have little conservation and very short consensus sequences: most of them have only few nucleotides at the beginning and the end respectively and a polypyrimidine stretch of 10 or more residues, usually 20-50 nucleotides after the branch point site (BPS) – in most cases an adenine (A) residue (**Figure 1**). The distinguishing is completed by ribonucleoproteins (RNPs) – molecules containing both RNA and proteins. The odds were not too bad for RNA and proteins to eventually co-evolve and carry out certain functions together: pre-mRNA is synthesized by a protein, the nascent transcripts are already bound by proteins that will later on determine their future and in general RNA without proteins "naked" RNA, is seldom found in the cell.

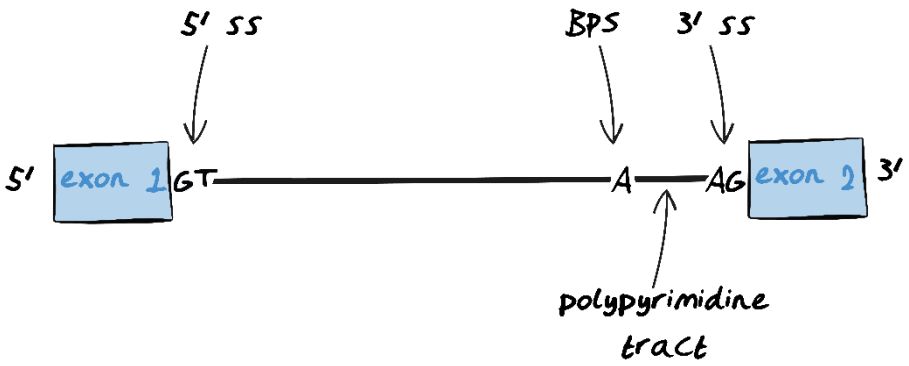


Figure 1 *Schematic overview of consensus sequences in introns, GU (GT on DNA level) at the 5' splice site and AG at the 3'ss. The branch point site (BPS) nucleotide is usually adenine. Between the BPS and the 3'ss is a polypyrimidine tract: 10-12 pyrimidines.*

The ribosome, the textbook RNP, is the executer of mRNA translation into protein, playing a crucial role in the maintenance of life. Its sole nature yearns for precision and efficiency. It contains two subunits, predominantly composed of RNA that perform the two major activities, decoding and peptide bond formation. The role of the ribosomal proteins is to assist the rRNA in folding and assembly and in helping with the whole process. It recognises mRNAs by binding the Shine-Dalgarno sequence in prokaryotes[21] or the elongation initiations factors bound to the 5' cap in eukaryotes[22]. Once bound to these sites it slides to the translation start site and decodes the mRNA, cycling in between three conformational states.

The spliceosome, the other, more sophisticated RNP is one of the biggest machineries in the eukaryotic cell consisting of five different subunits that are

highly dynamic in their assembly and require many additional proteins to execute the splicing reaction. The human genome codes for two distinct spliceosomes: a major and a minor one which are built out of different subunits, recognize different splice sites and are all together assembled in an independent manner. Since the minor spliceosome is responsible for removal of less than 0.5% of human introns[23] and for the sake of simplicity, this introduction will focus only on the major spliceosome.

To build a spliceosome, five different subunits, named small nuclear RNP (snRNPs) are needed and at least 140 proteins[24]. The number seems outrageous, but one should consider that the consensus sequences in introns are pretty short and the intron length can extend million nucleotides (nt). It is of essential importance to recognize the right splice sites. As a general rule they are identified multiple times by different proteins to ensure this exactness and precision. The initial RNA-RNA interactions need additional stabilization that is achieved by proteins. And of course in alternative splicing the appropriate exon in-/exclusion is made possible by the network of splicing associated proteins. In terms of biochemistry splicing is a fairly simple: two consecutive SN-2 transesterification reactions. First the 2' hydroxyl group of the branch-point nucleotide, as a nucleophile, attacks the phosphate backbone between the upstream exon and the 5' splice site, producing a free 5' exon and a lariat intermediate with a 2'-5' phosphodiester bond. In the second step the 3' OH of the freed exon attacks the backbone between the 3' ss of the downstream exon, ligating the exons and releasing a lariat (**Figure 2**). Two magnesium ions are necessary in the catalytic core of the splicing reaction – one activating the attacking nucleophile whilst the other stabilizes the leaving group[24].

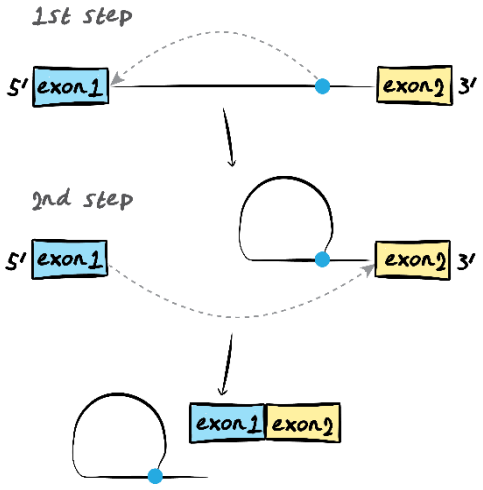
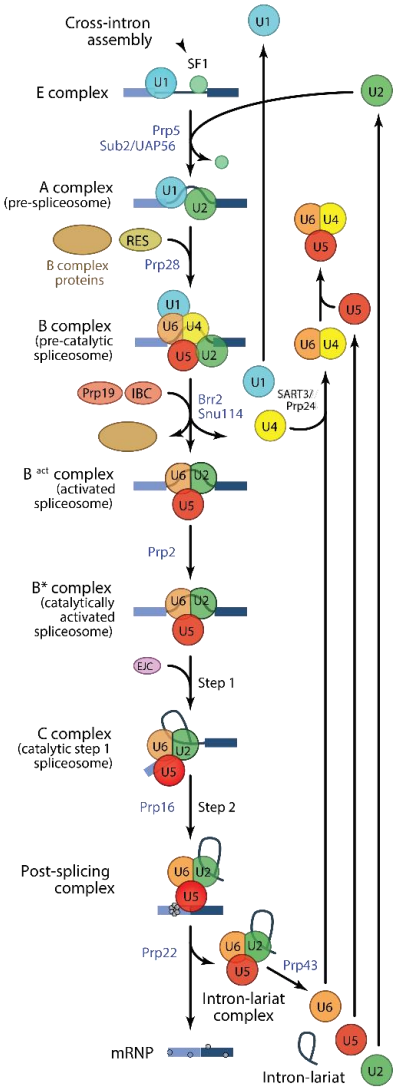


Figure 2 *Biochemistry of splicing: the two splicing steps are depicted, the first transesterification reaction starts with the BPS (depictured as a blue dot) attacking the 5' ss and lariat formation. In the second step exons are ligated and the lariat is released.*

Due to its complexity, dissecting the spliceosome is tedious work (reviewed in[24,25,26,27,28]). The process starts by binding of U1 snRNA to the 5' splice site and splicing factor 1 (SF1) to the polypyrimidine tract. At this state is described as the **E complex**. Next the U2 snRNP substitutes SF1 binding to the branch point site – the **A complex** is formed, also referred to as pre-spliceosome. U4, U5 and U6 join the party as a triple snRNP. By doing so they release U1 and bind to the 5'ss as well as to U2. **Complex B** (pre-catalytic spliceosome) is the description for this spliceosomal state. For the spliceosome to become catalytically active it has to release of U4snRNP (**B^{act} complex** – activated spliceosome) and to undergo rearrangements in which three nucleotides of U6 form a triple helix, partially with U2, partially with distant regions of itself. The special helix is obligatory to coordinate Mg^{++} ions in the catalytic core. **B* complex**, the catalytically activated spliceosome performs the first reaction and is transformed into **C complex** which is



Spliceosomal complex	Minimum number of proteins
E (commitment)	15 +hnRNPs & SR proteins
A (prespliceosome)	48 (+33) +hnRNPs & SR proteins
B (precatalytic spliceosome)	103 (+66,–11) +hnRNPs & SR proteins
B act (activated spliceosome)	56 (+3,–50) +hnRNPs & SR proteins
B* (catalytically activated)	
C (step 1 completed)	83 (+27)

Figure 3 *left: schematic overview of the splicing reaction, the assembly of the respective complexes as described in the text above, also the helicases involved in the restructuring process are indicated, adapted from[25]; above a table of the complexes and the minimal number of proteins required to build each of them, from [24]*

responsible for the second transesterification. The post-splicing complex releases the exons and a lariat whereas the respective snRNPs are being recycled for removal of other introns. transesterification

Besides the five most prominent RNPs for a successful splicing reaction two major groups of proteins are required: heterogeneous nuclear RNPs (hnRNPs) and serine arginine (SR) proteins. As a general rule it is taken that hnRNPs promote exclusion of exons, whereas SR domain containing proteins do the opposite (reviewed in[29]). RNA helicases also play a tremendous role in the spliceosomal assembly.

Alternative splicing

In alternative splicing, identical transcripts are processed in a way that different mRNAs are produced. There are various ways how this process can occur, the most common being ***exon skipping*** where an exon is excluded during splicing. An advanced form of exon skipping are ***mutually exclusive exons*** where two consecutive exons prevent each other in remaining in the matured mRNA. Other types include ***alternative 5' and 3' sites***, depending of the strength of the splicing site and the presence of splicing factors. ***Intron retention*** is also observed when removal of the intron is inhibited (**Figure 4**).

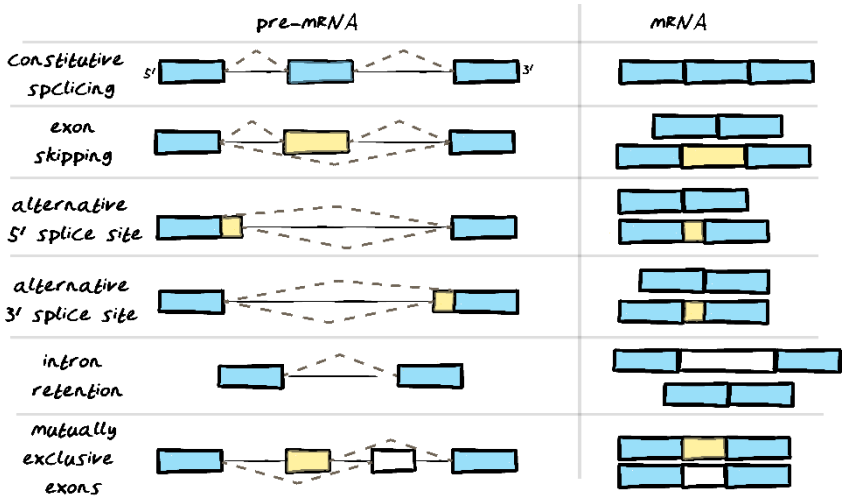


Figure 4 *Graphical overview of alternative splicing mechanisms as described in the text;*

The potential of alternative splicing is nicely demonstrated in the *Drosophila melanogaster* axon guidance receptor gene *Dscam* which codes for 38 016 potential

isoforms[30]. Altogether the gene contains 115 exons of which 95 are alternatively spliced and organized in four clusters. The respective clusters contain 2, 12, 33 and 48 alternative, mutually exclusive exons[31] (2*12*33*48=38016).

Exon definition-intron definition

Considering the previously mentioned low conservation level at splice sites, the average exon length of 147 nt and intron length of 3413 nt, it was not clear what the spliceosome recognizes first: introns or exons. First experiments suggested that introns are recognized and that there is a directionality in this decision, occurring from 5‘ to 3‘ across the intron[32], later termed intron definition model. Contrary to this is the exon definition model, which suggests that the spliceosome recognizes the exon first, starts assembling there and sequentially pairs with its appropriate parts across introns. Supporting this model is the finding that exons that are longer than 300 nt decrease the efficiency of spliceosome assembly[33].

Although both models seem reasonable, it is highly improbable that one model fits all the cases. The current understanding is that the models are applied depending on the length: intron definition for short, exon definition for long introns[34].

Intrasplicing

The conventional understanding of splicing claiming that introns are removed in a single piece (one lariat formed from one intron) might not always be applicable, especially for longer introns. A splicing model for long introns has been proposed: recursive splicing. It suggests that long introns are removed in several consecutive steps from one end. After each step the intron is shortened and the splicing sites are being reconstituted. This way a single intron produces more than one lariat (**Figure 5**).

This phenomenon has first been described in the D. melanogaster *Ultrabithorax (Ubx)* gene whose 74 kb intron has been shown to be removed recursively in four fragments[35]. A recent bioinformatics study has identified additional 130 introns in Drosophila which seem to be recursively spliced[36].

A similar approach was performed on vertebrates and 419 putative recursive splicing sites have been found that fit into the motif of recursive splicing, of which 19 located in introns larger than 150 kb, which they define as being recursive splice sites. The same study was able to link recursive splicing to "exon" skipping [36]. An additional paper claims that more than 60% of human transcripts contain partially spliced introns. They also showed that mutating the recursive splice sites in the SMD4a gene in human lead to a decreased splicing efficiency of the whole intron[37].

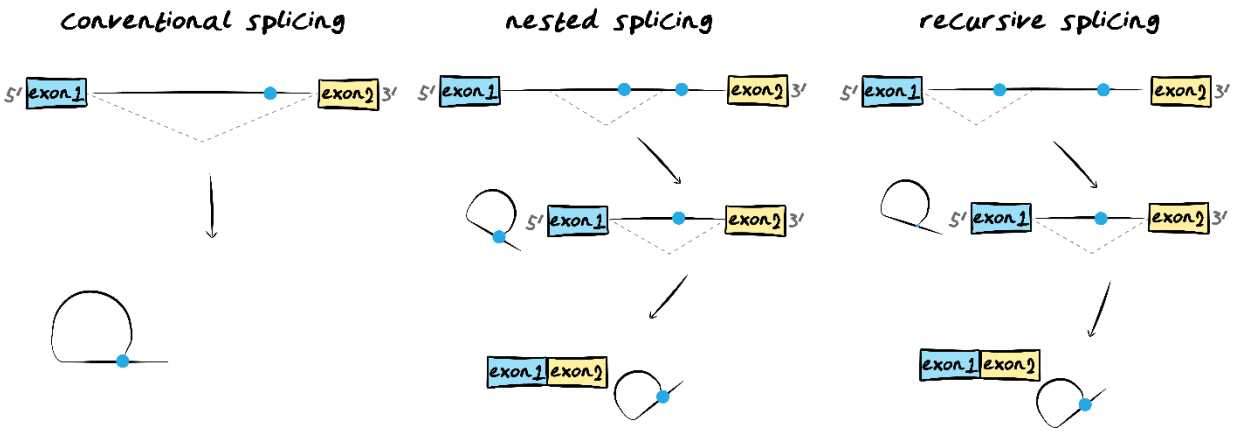


Figure 5 *Schematic overview of splicing, dashed line indicating the intron piece that is about to be removed, blue dot is the branch point; left conventional splicing where the intron is removed in a single piece and one lariat is the by-product of the procedure; middle nested splicing, intron contains multiple branch point sites as well as multiple splice sites; the intron is removed consequently and the final output are the ligated exons as well as two lariats; right recursive splicing: a somewhat ordered nested splicing, instead of a casual intron shortening in the middle, recursive splicing starts either at the 5' or the 3' splice site and recovers the splice site after each step (expect of the last).*

Another splicing model is proposed where the intron is spliced consecutively in a multi-step procedure, not at the canonical splice sites, but at alternative sites from within the intron. Intron shortening occurs somewhere in the middle of the intron. The splice sites are not reconstituted and the removal of the nested intron shortens the conventional intron. This idea was first termed intrasplicing[38] but was later renamed nested splicing. The phenomenon was demonstrated in the human dystrophin gene (DMD). With its 2.5 Mb it is one of the largest annotated genes in the human genome. The spliced transcript is only 14 kb long and contains 79 exons. Intron 7 is 110 199 nt long and contains two nested splicing events identified in a PCR approach and confirmed in an *in-vitro* splicing assay[39].

Project background

The idea of the project was to characterize nested and recursive splicing events in human cells. The fundament of the project was a computational analysis performed in the Schroeder lab, starting with published RNA-Seq datasets from human cells and extracting intronic splicing events using a novel bioinformatic pipeline.

A very convenient method to find introns in these data is to extract split reads, in particular linear and circular split reads. Linear split read can be found when mapping spliced mRNA to its genomic region. Since the intronic part is lacking in the read, once mapped to the genome an obvious interruption in that region is present. Visualizing this gap one can see introns that are found in this approach.

Due to the lariat formation, a huge part of the intron gets a circular (lariat) structure during splicing since the branch point site forms a 5'-2' bond to the 5' ss. When preparing a RNA-Seq library, random primers binding downstream the 5' ss can serve to produce cDNA that covers the region until the upstream ss and continue at the branch point. The enzyme responsible for this job, reverse transcriptase reads through the unconventional BP-5'ss bond although an increased mutation rate around the branch point is observable[40]. Mapping circular split reads is a tricky business: on a genomic DNA it will start after the 5' ss, trace back till the actual splice site, continue at the branch point and end somewhere upstream of the BPS. This feature has successfully been used to map branch points[41,42].

Intronic linear splits are expected to be enriched in poly-A minus RNA-Seq data, circular splits can be found in high quantities in circular RNA-Seq, for which the feature of RNase R to only degrade linear but not circular RNAs is used. The sequences originated from the cell lines that we tested HeLa, HEK293, HepG2 and K562 but also contained samples from different human tissues, at least 12 additional tissues were included in the approach. Single cell sequences were also added in the dataset, particularly important for the circular RNAs are brain cells.

In order to increase the certainty of splicing events we combined the two types of split reads, meaning that if a circular split read fits into a linear split and ends up to 100 nt before the linear split, there is a high probability of a single splicing event happening (**Figure 6**). Those combined linear and circular reads that arose from the same splicing event he termed lariat spanning linear split (***laslis***). In

order to find only nested and recursive splicing we filtered for those combined reads that cover parts of the annotated intron, but not the full length. On a genome wide analysis, we were able to find 128 113 laslis that cover whole annotated introns and 12 010 potential nested splicing events: 3 208 5' recursive, 2915 3' recursive and 5886 nested (see **Figure 7**).

The surprising discovery was that most of them do not span only long introns, ergo there seems to be a lot of nested splicing in short introns as well. Since there is no obvious spatial need for their existence the project also focused on understanding the potential functionality of them.

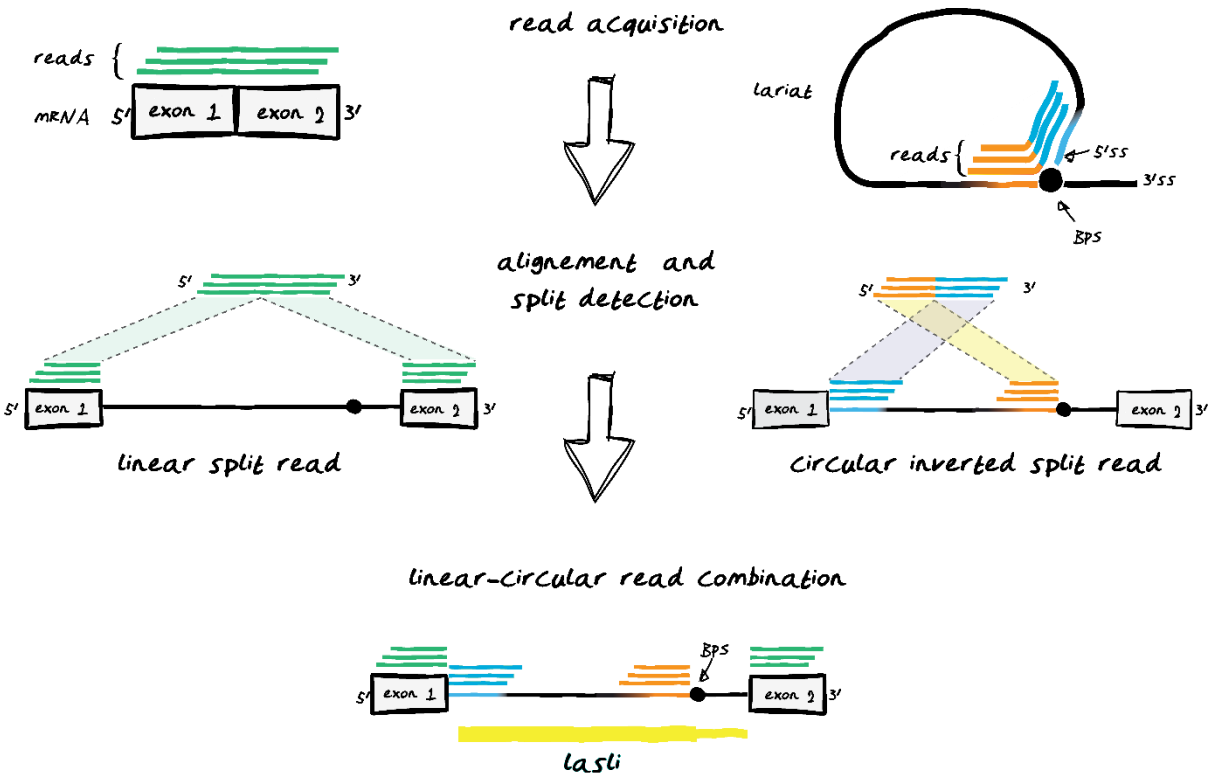


Figure 6 *lasli extraction pipeline: from publicly available RNA-seq data two types of split reads were extracted – linear and circular inverted; when a circular read fits into the linear read and is up to 100 nt shorter than the linear, they are combined and a lasli is annotated; the circular read is interrupted at the branch point site and continues at the 5' splice site;*

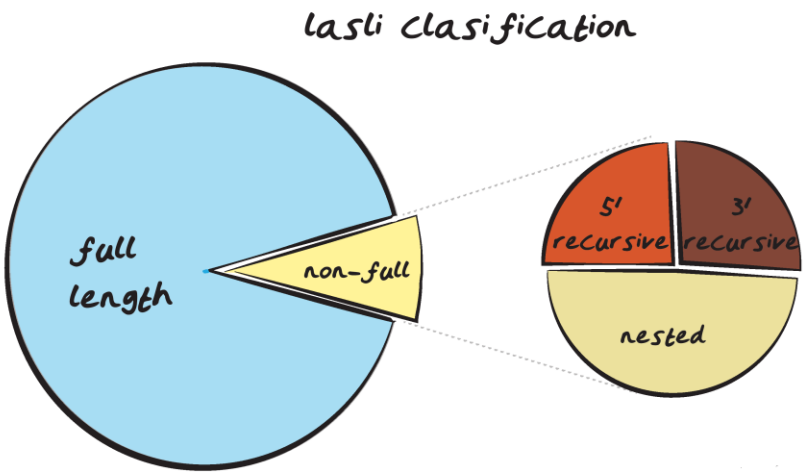


Figure 7 *Classification of laslis: out of 140 123 lalis detected approximately 9% don't cover the full introns, which makes them putative nested lariats; from those 9% circa half are nested lariats, 5' and 3' recursive make up for about a quarter of these splicing events*

Project aims

The first part of the project focused on verification of the *in silico* predicted nested splicing events, which were based on previously published RNA-Seq data and indicating the occurrence of these events with a high frequency, even in short introns. For this validation the corresponding laslis were detected via lariat PCR, a specialized method of amplifying circular RNAs. We also tried todetermine the length of a nested lariat or the splicing intermediates that are produced in a intrasplicing reaction with a virtual northern blot.

As we gathered enough confidence in the predicted nested laslis we wanted to explore the potential functionality of these events, particularly on the splicing efficiency and the influence on exon skipping. To test the effect on the splicing efficiency we developed a reporter system, for which we cloned introns into a luciferase gene, mutated the annotated or the nested splice site and compared the luciferase expression levels.

Another part of the project focused on exon skipping, for which a different, vector with a high expression promotor was used and an intron containing a nested lasli was cloned, but also the surrounding native exons. Mutations in the nested splicing sites were expected to affect the exon skipping patterns.

We explored additional methodologies in order to detect nested splicing events, including northern blotting and high-performance liquid chromatography.

Materials and methods

Standard lab methods

Polymerase chain reaction

PCR reactions were performed with Q5 high fidelity polymerase (NEB) for long amplicons used for cloning or with OneTaq (NEB) for standard PCR applications such as colony pcr, according to manufacturers instructions. Primers designed with *Primerblast* or manually.

Agarose gels

Agarose gels of respective density were prepared in Tris Borat EDTA (TBE) buffer, before casting PeqGreen was added, 2µl/100ml of agarose.

Plasmid preparations

Plasmids were purified with Mini/Midi prep kit *Promega or Qiagen*. For details check manufacturers protocol. The plasmids were eluted with ddH_2O.

Gel extraction

Gel extractions were performed with Qiagen Gel Extraction kit, for details see manufacturers protocol.

Cloning

Cloning was done with the In-Fusion® HD Cloning Kit from *Clonetech*, insert and vector were prepared with respective primers (see Primer list) and Q5, mutations and deletions were inserted with the Q5® Site-Directed Mutagenesis Kit from *NEB*, according to manufacturers instructions. DH5alpha NEB cells transformed with theligation reactions.

Restriction enzyme based cloning was done by adding the restriction sites with respective primers, the insert was amplified by Q5 and purified with PCI. Backbone and insert were digested with restriction enzymes from Thermofisher for 2 h at 37° C, purified with PCI and ligated over night at 16° C with T4 DNA Ligase NEB according to the molar ratios suggested by NEBioCalculator.

Transformation

Competent cells Escherichia coli DH5 (NEB or self-made) were thawed on ice, 50-100 µl cells were added to 5 µl ligation reaction, left on ice for 30', *heat-shocked* at 42° C for 30", put back in ice for 2' and 750 µl SOC was added. A 45' incubation at 37° C followed, plating on LB plates with appropriate antibiotic and overnight incubation at 37° C.

TA cloning

OneTaq PCR products were separated on an agarose gel, bands were cut and extracted with the Qiagen Gel extraction kit, cloned into the pGEM vector (Promega) according to the company's protocol. 5 µl of the ligation product were transformed into DH5l cells as previously described and plated on ampicillin plates with 20 µl IPTG (100 µM), 20 µl X-Gal (20 g/ml) and 40 µl ddH_20, incubated overnight at 37° C. Positive clones were picked for a colony PCR.

Reverse transcription (RT)

Performed on 1 µg RNA with either Protoscript II (NEB) or Superscript III (Thermo Scientific) according to the manufacturers instructions. Random hexamers or nanomers were used.

qPCR

Performed in a 96 well plate, for a single reaction 4 µl Sybergreen master mix was used, 0.5 µl of each primer (10 µM) and 5 µl of 1:5 diluted template in a final volume of 20 µl. The measured signal was normalized with GAPDH.

Cell culture

HeLa, HEK293 and Hep G2 cells were kept in Dulbecco's Modified Eagle's Medium (DMEM, from Sigma Aldrich), K562 cells in Roswell Park Memorial Institute medium (RPMI, Sigma Aldrich) with 10% fetal bovine serum (Sigma Aldrich) at 37° C with 5% CO_2, split at approximate confluency of 100%. Before

splitting they were washed with Dulbecco's Phosphate Buffered Saline (Sigma Aldrich) and trypsinized with Trypsin-EDTA solution (Sigma Aldrich).

Transfection

Transfections were done in 6-well, 24-well or 96-well plates by seeding the cells at the recommended confluency (0.3*10^6; 0.05 …) plasmids were transfected 24 h after with FuGENE HD or the *Lonza* Nucleofection kit for K562. Cells were harvested after 48 h with 500 µl Tri for a well of a 6-well plate.

RNA preparation

RNA was isolated from the Tri harvested cells by adding 100 µl Isopropanol per 500µL of TRI and centrifuging for 15' at full speed in a table top centrifuge at 4° C. The upper aqueous phase was transferred to another 1,5 ml tube, 250 µl isopropanol was added and centrifuged for 30' at 4° C, the pellet was washed with 500 µl of 80% Ethanol, dried and resuspended in 40 µl ddH_2O.

DNase treatment

15 µg RNA was treated with 1 µl DNnaseI (10 Units/µl, Roche) in 50 µl total volume at 37° C for 20', after which another 1 µl DNase was added and incubated for 20'. The reactions were filled up to 100 µl with water, 200 µl PCI was added and transferred to a solid phase gel tube (5prime), centrifuged for 15' at 4° C full speed. Aqueous phase was precipitated in a new Eppi with 3x volume of 96% Ethanol and 1/10th 3 M Sodium phosphate precipitated for one hour at -80 °C and centrifuged for 30' at 4° C, full speed. After a 10' wash with 1 ml 80% Ethanol (full speed, 4° C) and drying, the pellet was resuspended in 40 µl water.

Virtual northern blot

5 or 15 µg RNA from HEK 293 cells were loaded on a 1 % agarose formaldehyde gel, run for 4 hours in MOPS buffer, after which each lane was cut out, split in seven pieces and a single lane was cut in seven pieces: 0,3-0,7 kb (**1**), 0,7-1,3 kb (**2**), 1,3-2 kb (**3**), 2-3 kb (**4**), 3-6 kb (**5**), 6-10 kb (**6**) and larger than 10 kb (**7**). As a control we used RNA from HEK 293 cells, which was not separated on a gel (**total RNA**).

The RNAs were extracted with the Zymoclean™ Gel RNA Recovery Kit (Zymo Research) according to the manufacturers protocol. 9 µl of the recovered RNA were used for a reverse transcription reaction (Protoscript II) and the cDNA for a PCR.

Lariat PCR (larPCR)

Lariat PCR is based on the same principle as the circular split reads: the circular formation of the lariats is used for product acquisition. The primers are designed in a divergent manner, the forward primer binding downstream of the reverse so they cannot form a product from the genomic DNA. Once the lariat is formed the possibility for them to make a product is given, ergo the there is a PCR product only if the lariat is formed. Mapping back the product on the genomic level leads to an interruption of the read, it is again inverted (**Figure 6**).

cDNA obtained from RT reactions on total RNA from HeLa, HEK 293, Hep G2 and K562 was used as a template for lariat PCR. In order to increase the lariat abundance, we knocked down debranchase (dbr1) by transfecting a plasmid coding for a shRNA against it. The knock down was performed with BLOCK-iT™ U6 RNAi Entry Vector Kit (Invitrogen) according to the manufacturers instructions, the primers used in the amplification process are in the primer list at the end of the materials and methods section.

Since lariats are lowly abundant we had to apply nested PCR, with outer and inner primers. The first PCR with the outer primer ran for 25 cycles, the second for 35, reactions in 25 µl, 1 µl of the first was the template for the second. OneTaq (NEB) was the polymerase and the reactions were set up according to manufacturers instructions. The final products were analysed on a 1% agarose gel, the products cloned into a TA vector (pGEM, Promega) and the eventual plasmids sequenced. The sequences were aligned with BLAT tool to hg38.

Luciferase assay

Introns containing laslis were cloned into phRL-TK plasmid at position 414 in the luciferase gene, splice sites were mutated and the luciferase activity was compared between the wild type and mutant plasmids, normalized to a firefly luciferase from the pmirGlo plasmid (whose renilla luciferase was deleted). Performed with the Dual-Luciferase® Reporter (DLR™) Assay System (Promega) in 96-well plates (chimney well, black; Greiner Bio-One) by plating 10k HeLa or 15k HEK 293 cells in 100 µl DMEM, 24 hours before transfection, transfecting 300µg phrlTK (+insert) and 30µg pmirGlo per well. The transfection reagent was FuGENE HD, whereas the DNA-FuGENE ratio was 3:1 for HEK and 3,5:1 for HeLa cells. The transfected cells were incubated for 48 hours at 37° C, 5% $C0_2$, the medium removed and cells lysed with 20 µl KLD buffer for 30', shaking. The substrates were prepared and kept at room temperature shortly before measuring. The luminescence was measured in the plate-luminometer Solaris Robion. First 50 µl of the Dual-Glo reagent were added per well, the plate was shaken for 2" and exposed for 10". Then 50 µl Stop & Glo reagent were added, shaken 2" and exposed 10". All samples in triplicates, the background was subtracted and renilla luciferase signal was normalized to firefly luciferase.

To avoid repetition, all the constructs used in the assay are depicted only in the results section. One of the introns we cloned was PRKAB2 intron7 that has 2069 nt and a 3' lasli spanning from 1362-2069. The whole wild type intron was cloned into the reporter (**p7**), as well as only the lasli (p7 nest, containing intron part from 1362-2069), also the intron where the lasli was prespliced (**p7 pre**, part from 1-1362). The constitutive 5' splice site was mutated in the **p7 1** construct, the GT were substituted with AC; the lasli 5' ss was mutated in the construct **p7 1362**. Our second candidate to test was GNB2L intron 7 which is 566 nt long, contains two laslis, 5' covers the part from 1-243, while the 3' ranges from 384-566. Construct **g7** contains the wild intron, in **g7 5' pre** the 5' lasli has been prespliced (contains nucleotides 245-566 from the original intron), in **g7 3' pre** the downstream nested event already happened (1-384 of the wild type), in **g7 1** the constitutive GT at the 5' ss has been mutated to AC, **g7 242** has the 3' ss of the upstream lasli mutated and **g7 384** is the construct where the downstream laslis 5' ss has been mutated.

RAD54 intron 11 is 1680 nt long, the 5' nested lasli reaches from position 1 to 582. The construct **rad** contains the whole intron, **rad 551** has the pY and the 3' ss deleted;

RMB17 intron is a 1146 nt long intron, three laslis have been found there, lasli1 covers the part 1-233 nt, lasli2 1-539, lasli3 707-1146; the luciferase construct **rbm** contains the whole, unmutated intron, **rbm 213** has the polypyrimidine tract and the 3' ss deleted (from 213-233), in **rbm 514** pY and 3' ss in the part 514-539 were deleted, whereas in **rbm 707** the 5' ss of the 3rd intron was mutated from GT to AA.

Exon skipping analysis

Introns with laslis around exons that are skipped in alternative splicing were manually selected. Three exons with their introns were cloned into pcDNA 3.1 after the in mammalian cells highly active CMV promoter[43,44]. Nested splice sites were mutated, the plasmids transfected into 6-well plates (3.3 µg DNA, 9.9 µl FuGene HD for HEK 293 and 11.55 µl for HeLa) as described above. RNA was treated with DNase I, reverse transcribed (Superscript III) and the cDNA used for a PCR with corresponding primers. The products were loaded on a 1 % agarose gel, the products cut out, purified, cloned into a TA vector and sequenced.

Graphical overview of the respective constructs in **Figure 26**, **27** and **28**. Primers used for to obtain and mutate the constructs are in the primer list. DCNT3 intron 4 contains a 5' lasli after the skipped exon, ranging from nucleotide 1-589, in the polypyrimidine tract deletion nucleotides 550-585 are lacking; NME6 intron 3 contains a 5' lasli before the skipped exon, lasli ranging from nucleotide 1-496, in the deletion mutant nucleotides 467-495 are removed; SACM1L intron 4 has a 3' nested splicing event right before the skipped exon, laslis length is 330 nt, intron 4 length is 1571 nt; laslis 5' ss was mutated;

Primer list

Lariat PCR

L_LASP1_5'lar_no_fwd	GTTTGCCCTGCTTGCCTCTA
L_LASP1_5'lar_no_rev	TTATGCACCCCTGCCCATTC
L_LASP1_3'lar_no_fwd	CCTTGCTCTCCTCTGCATGG
L_LASP1_3'lar_no_rev	TCTGCATGGGCCCCAATTAC
L_LASP1_5'lar_ni_fwd	AGGGTCGTCTCGTCTCTGAA

L_LASP1_5'lar_ni_rev	CGAGGAAAGCCCCCTCTTTT
L_LASP1_3'lar_ni_fwd	GACCTCCATGTGTGGTAGGC
L_LASP1_3'lar_ni_rev	AGATCCCTGACTCACAGGCT
L_PRKAB2_FL_no_fwd	TCTTTTGGAAGTCATCCCTGGT
L_PRKAB2_FL_no_rev	AGAGATGCAAGCACCTGAGT
L_PRKAB2_FL_ni_fwd	AGTAGGGAAGGGAGCAAGGTA
L_PRKAB2_FL_ni_rev	TGCCAGCTTCAGCTTCAATG
L_PRKAB2_3'nest_no_fwd	AAGGGAGCAAGGTAGGAGACT
L_PRKAB2_3'nest_no_rev	TTCCAAGTCCCAAAGCAGGTC
L_PRKAB2_3'nest_ni_fwd	ACCCAAGGCCACAAGTAGAA
L_PRKAB2_3'nest_ni_rev	GGGAACCTAGGAATCCCTCTCT
L_MFN2_ni_fwd	GGTTGGCTCCTTGTCACTGC
L_MFN2_ni_rev	AAGAGACTTGCCACCAGAGGT
L_MFN2_no_fwd	TCATTTGGTCTCTGCTGTGCT
L_MFN2_no_rev	GCAACCAGCGAAATCCTTCAC
L_MFN2_FL_ni_fwd	GCAGTGACCTGAGATCCCCA
L_MFN2_FL_no_fwd	TCTCCCTAGCAGGATGTCCC
L_CTIF_midnest_no_fwd	GTAGGGTTGTGGGGATAGGG
L_CTIF_midnest_no_rev	GGAGCTCTGAGGTCATGAGTG
L_CTIF_midnest_ni_fwd	GATCCACGATGGAGGAGCATT
L_CTIF_midnest_ni_rev	GCGGCTGCAGGTTAACAATc
L_CTIF_FL_no_fwd	CGTTCCTGCAGACTCACCAT
L_CTIF_FL_no_rev	GAGAAACACTCGGCTCAGGC
L_CTIF_FL_ni_fwd	CATCTTCTGCCTTTCCCAGCG
L_CTIF_FL_ni_rev	GCAGGGAATGGGACAAACCCA
L_GNB2L1_FL_int7_ni_fwd	TACAGTCTACCCGGCGTTGT
L_GNB2L1_FL_int7_ni_rev	ATTCACTGGCATGGCAAACG
L_GNB2L1_FL_int7_no_fwd	CAAGCTGGCTGTTGAAATGAGAG
L_GNB2L1_FL_int7_no_rev	CCCTCTGATGCAGAAACACATTC
L_GNB2L1_5nest_int7_no_fwd	TTGATGAGGATGCCCTCGTTT
L_GNB2L1_5nest_int7_no_rev	GGCTTGTCATTTCACTTTATGCC
L_GNB2L1_5nest_int7_ni_fwd	ATGCCAGTGAATGTGTTTCTGC
L_GNB2L1_5nest_int7_ni_rev	TTTATGCCAAGGACACTGCTCT
L_SRRT1_5'nest_fwd	TCCTCCTAAAGCGACGAGTG
L_SRRT1_5'nest_rev	CTGGACAGGGAAAACTGGGA
L_SRRT1_3'nest_fwd	GGTTCCTTTTGGGGTGTTGG
L_SRRT1_3'nest_rev	CAACAGGGAAGAAGGTGGTGT
L_SQ_full_ni_fwd	CACTTAGCTGCTTGTGGGGA
L_SQ_full_no_fwd	TGGGATATGGGGAGAAGCCT
L_SQ_1_ni_fwd	CAGCCTGTCCAAAGGTTGTG
L_SQ_1_ni_rev	CACCAGTGTCAAGGGCATGT
L_SQ_1_no_fwd	GTGATGGCAGATGGATCGCT
L_SQ_1_no_rev	CTGCAACCACAGTCTGGTCA
L_SQ_2_no_fwd	AGGGGAACGAACGAGTCAAC
L_SQ_2_no_rev	TCAGTTCCGAGATGCTGCAC

L_SQ_2_ni_fwd	CTCCTGGTCGTGGGATCTTT
L_SQ_2_ni_rev	TGAGACAGTGTGACCCCAAG
L_SQ_3_no_fwd	GCCACAACTGAGACCCAGG
L_SQ_3_no_rev	CAACCAAAAGGCCTTGGCAC
L_SQ_3_ni_fwd	CCACAACTGAGACCCAGGC
L_SQ_3_ni_rev	AAAGGCCTTGGCACTTCACA
L_SQ_4_no_fwd	TTGGAAGATGGACATGGAGGGA
L_SQ_4_no_rev	TTGTGTAAGCCCTTTCCTGTTAT
L_SQ_4_ni_fwd	GATGGACATGGAGGGAGGGTG
L_SQ_4_ni_rev	GCAGAGGTAAAAGCAGGCTTAACT
L_SQ_5_no_fwd	CTGAGGAAAGCCCCCACACA
L_SQ_5_no_rev	GATCCCACTGAAGCACTAAGGT
L_SQ_5_ni_fwd	CCACACATGGGAGAAAAGGTC
L_SQ_5_ni_rev	ACTAAGGTTGACAGGAAGCCT
L_SQ_6_no_fwd	TGGAACAACTGATGACTCCAGA
L_SQ_6_no_rev	TCCTCAGGGTTCTGACACTT
L_SQ_6_ni_fwd	ACTCCAGAAATCAGTTAAAAAGGC
L_SQ_6_ni_rev	TCAGGGCCGCTTTCAATCTA
L_SQ_7_no_fwd	GGAAACGGGGGACCTAACAG
L_SQ_7_no_rev	TCTGCCCTGATACTTGATCGT
L_SQ_7_ni_fwd	GTAAAGGGGGTCACAGCATCA
L_SQ_7_ni_rev	ACTTGATCGTTTTAGCTGGGT
L_SQ_8_no_fwd	GTCTGCTACTTGACCCAGCA
L_SQ_8_no_rev	TCGCTTTTTGTTTTCCTACCACT
L_SQ_8_ni_fwd	TTAGATAGAGGTCCTAGCTGAAACA
L_SQ_8_ni_rev	CAGTGCCGCAATCTCTCGG
L_SQ_9_no_fwd	GAAAGGGCACAGGGAGAGTT
L_SQ_9_no_rev	CTTGCCTGCTGCTGGAGTTC
L_SQ_9_ni_fwd	GGAGAGTTTGGGGAGTGGGATG
L_SQ_9_ni_rev	AGTTCACATCTGCGCACTTCCT
L_SQ_10_no_fwd	CAGAGCTCCGGCGTTGAG
L_SQ_10_no_rev	GGGGAGAATCACCTGCACTA
L_SQ_10_ni_fwd	TGTCGTCGCACAAAGCCTC
L_SQ_10_ni_rev	CTCCACACCCACAGCTTAGAG
L_SQ_11_no_fwd	TGGAGAAAGAGAAAGGTCCAGT
L_SQ_11_no_rev	ACCTCAGCTAGCATCTAAAGTCA
L_SQ_11_ni_fwd	ATGGGCCCACCAAGAATGTT
L_SQ_11_ni_rev	CTAAAGTCAGCTTCTGAGCCA
L_SQ_12_no_fwd	CTGCATCCACACGCTGAGT
L_SQ_12_no_rev	GTTCTACCCAGCTCGACCC
L_SQ_12_ni_fwd	ACCATCCAGACACTTAGCTGC
L_SQ_12_ni_rev	CAGCTCGACCCCGCAATC

Renilla luciferase assay constructs

Intron in phRL-TK cloning	
HPRT_int7_inFus_414_fwd	CAAGACAAGATCAAG**gtaagtgaattacttttttgtc**
HPRT_int7_inFus_414_rev	AGCATGGACGATGGC**ctaaaaagaatcataattc**
HPRT_int8_inFus_414_fwd	CAAGACAAGATCAAG**GTAAGTAATTGCTTCTTTT**
HPRT_int8_inFus_414_rev	AGCATGGACGATGGC**CTATAAAAAAAAAATCCAGAAATTCAC**
PRKAB2_int7_inFus_414_fwd	CAAGACAAGATCAAG**GTAAGTGGTGGCCCTGGCTC**
PRKAB2_int7_inFus_414_rev	AGCATGGACGATGGC**ctgcAAGGAAAAGAAACATACG**
GNB2L1int7_inFus_414_fwd	CAAGACAAGATCAAGGTAAGTGGGTCTGTCCTCTC
GNB2L1int7_inFus_414_rev	AGCATGGACGATGGCCTGCAGGGAAGAAATGACAG
rbm17_8_inFus_414_fwd	CAAGACAAGATCAAGGTGTGTCCCCAGGGAAGCG
rbm17_8_inFus_414_rev	AGCATGGACGATGGCCTGTATCAAAACATACCATTAGATGTG
rad54l_10_inFus_414_fwd	CAAGACAAGATCAAGGTAATGACCTTAAGCGAAG
rad54l_10_inFus_414_rev	AGCATGGACGATGGCCTAGAGAAAACAGGATAAAAAGGCAAG

Mutation insertion	
h7_1_GT-AC_fwd	CCTACGAGCACCAAGACAAGATCAAGACAAGTGAATTACTTTTTTTGTC
h7_1_GT-AC_B2B_rev	AGTAGTGAAAGGCCAGACAAGCCC
h7_170_AG-GG_fwd	CTAAATGATGAATTATGATTCTTTTTGGGCCATCGTCCATGCTGAGAGTG
h7_170_AG-GG_rev	CACTCTCAGCATGGACGATGGCCCAAAAAGAATCATAATTCATCATTTAG
h8_del1310-1345_Fwd	GCCATCGTCCATGCTGAGAG
h8_del1310-1345_Rev	GTATATTCTGAAATGAAAGGCAAGAATAGC
p7_prespliced_3nest_fwd	GCCATCGTCCATGCTGAG
p7_prespliced_3nest_rev	CTGAGTCAGAACCAGTAAGAAG
p7_3nest_fwd	GTGAAGGTGTTTGATTATTGTGG
p7_3nest_rev	CTTGATCTTGTCTTGGTGC
p7_1_GT-AC_fwd	CCTACGAGCACCAAGACAAGATCAAGACAAGTGGTGGCCCTGGCTCTTC
p7_1_GT-AC_B2B_rev	used h7_1_GT-AC_B2B_rev
p7_1362_2GT-CC_fwd	GTCTTCTTACTGGTTCTGACTCAGCCGAAGCCGTTTGATTATTGTGGC
p7_1362_2GT-CC_B2B_rev	AGCAAGAGGTACCATTTCCCAGGCTGTGTTGTAGACTACAG
g7_3'_pre_fwd	GCCATCGTCCATGCTG
g7_3'_pre_rev	CTCAGCCTCCAGTAGGC
g7_5'_pre_fwd	GTTTTCAAATGTCTGGAGCCTTATC
g7_5'_pre_rev	CTTGATCTTGTCTTGGTGCTCGTAG
g7_242_GT_CC_fwd	CCTTTTCAAATGTCTGGCTCCTTATCACCTTCTGTTTC
g7_242_GT_CC_rev	GGTCAGATTAATTAAGGTATTAATTGAAGGCTGACAGCC
g7_384_GT-CC_fwd	CTGAGCCGGGAGGATCTCTTG
g7_384_GT-CC_rev	CAGTAGGCATGTGCCACA
megf6_del_549-574_fwd	GGCCTAACTGTAGCAGACGA
megf6_del_549-574_rev	GCAATACCCCTCTCATTCAAG
rbm17_i8_del213-233_Fwd	ACGTAGAGAGATGAAGGCTTG
rbm17_i8_del213-233_Rev	TGAGCCTCAGGATTCCTGG
rbm17_i8_del514-539_Fwd	TCTGTTCTGTGTTGAGGGAATTC
rbm17_i8_del514-539_Rev	CACCAACTAGGCATCATTTACC
rbm17_i8_707_GT_AA_Fwd	CATCACGAGCCTGACAGaaGAaaGCCAGAAACCAACAaaTaaC

	CCGACTTG
rbm17_i8_707_GT_AA_Rev	TATCCTAAGCTGTGCATGAACGGGGCTGCCAGAAAAGTTTTCTTC
rad54_i10_del551-582_Fwd	ATGGCAGGACGGGCATGGTG
rad54_i10_del551-582_Rev	CCAGGCAGTTGTTGGGAAGGG

Exon skipping

for acquisition of insert with restriction site	
DCTN3_KpnI_Fwd	TATCggtaccTTGAAGATCTGATCAAGTACCTGG
DCTN3_XhoI_Rev	TATGctcgagCTGCTGCTGAATGTGGATCTG
NME6_KpnI_Fwd	TATCggtaccGAGTGAGATGGCCTCAATCTT
NME6_XhoI_Rev	TATGctcgagCCGAACCATGGGTGGTGTTg
SACML_BamHI_Fwd	TATCggatccTGTTCCTCCTTCAGCTGTCA
SACML_EcoRV_Rev	TATGgatatcACTCATTTCTTGGAATTCAGGACT

Mutagenesis primer	
DCTN_PP_del_fwd	caaggtgctgcagaacccaga
DCTN_PP_del_rev	ggaatccccagagctggggact
NME6_PP_del_fwd	ggtaaggaggtagaagtacagaa
NME6_PP_del_rev	ccttggagtctgacaaacct
SACML_1305_GTAC_Fwd	GCTGGGAAATGATTAGCAGAGacCTacATCCCTTGGGGacGATGAAAGAGA
SACML_1305_GTAC_Rev	AGGCAGCTTAAAGCATCAGGACGTGAGATCTGACACTACAGGGCAATGTC

final PCR	
DCTN_assay_fwd	GGATCCTGAGTACATCGACCG
DCTN_assay_rev	TGGATCTGGGCCAAGCGctgc
NME_assay_fwd	GCTCTCCAGCTCACTCTAGC
NME_assay_rev	CCACGGATAGAATCTGGGGC
SACML_assay_fwd	AGACCAATATTTGGTATACTGGGCA
SACML_assay_rev	GTTGGATAGCCGCTGCAAAG

Reagents and resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
QIAprep Spin Miniprep Kit	Qiagen	Cat# 27106
PureYield™ Plasmid Miniprep System	Promega	Cat# A1222
PureYield™ Plasmid Midiprep System	Promega	Cat# A2495
QIAquick Gel Extraction Kit	Qiagen	Cat# 28706
Dual-Luciferase® Reporter Assay System	Promega	Cat# E1910
Zymoclean™ Gel RNA Recovery Kit	Zymo Research	Cat#R1011
pGEM®-T Vector Systems	Promega	Cat#A3600
BLOCK-iT™ U6 RNAi Entry Vector Kit	Invitrogen	Cat#K4944-00
Phenol stabilized: Chloroform: Isoamyl Alcohol 25: 24: 1	Applichem	Cat#A2279
DNase I recombinant, RNase-free	Roche	Cat#04716728001
RNasin Ribonuclease Inhibitor	Promega	Cat#N2115
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0491
ProtoScript II First Strand cDNA Synthesis Kit	New England Biolabs	Cat#E6560
SuperScript® III Reverse Transcriptase	ThermoFisher Scientific	Cat#18080044
TRI Reagent®	Sigma-Aldrich	Cat#T9424
100 bp DNA Ladder	New England Biolabs	Cat#N3231S
1 kb DNA Ladder	New England Biolabs	Cat#N3232S
0.5-10 Kb RNA Ladder	invitrogen	Cat#15623-200
Dulbecco's Modified Eagle's Medium	Sigma-Aldrich	Cat#D6429
Dulbecco's Phosphate Buffered Saline	Sigma-Aldrich	Cat#D8537
Trypsin-EDTA solution	Sigma-Aldrich	Cat#T3924
RPMI-1640 Medium	Sigma-Aldrich	Cat#R8758
FuGENE® HD Transfection Reagent	Promega	Cat#E2311

Website	Usage
https://www.ncbi.nlm.nih.gov/tools/primer-blast/	Primer design
https://genome.ucsc.edu/cgi-bin/hgBlat	Sequence aligning
http://nebiocalculator.neb.com/#!/	Vector-insert ratio calculation for cloning
http://tmcalculator.neb.com/	Melting temperature for manually designed primers
http://www.promega.com/techserv/tools/FugeneHdTool/default.aspx	Interactive protocol for FuGene HD

Results

Lariat PCR

The verification of the predicted nested lariats was accomplished using a specified PCR approach termed lariat PCR. Contrary to the conventional PCR, the primers are designed in a divergent manner on the genomic DNA but lead to product amplification on lariats due to their circular nature (see **Figure 8**). This way the chances of getting a product from partially undigested DNA are abolished. Still the obtained products were sequenced to exclude the possibility of unspecific amplification. RNA extracted from HeLa, HEK 293, Hep 3b or K562 cells was reverse transcribed and used as a template for the reactions. Since the lariat abundance can be very modest, we also knocked-down the debranching enzyme (debranchase, dbr1) in order to increase the yield of lariats

The following figures contain original gene loci, the nested splicing event (nested lariat, lasli) or in some cases the circular split read, the inner primers used to get the product, the agarose gel of the respective PCR reactions and sequences that were mapped with BLAT.

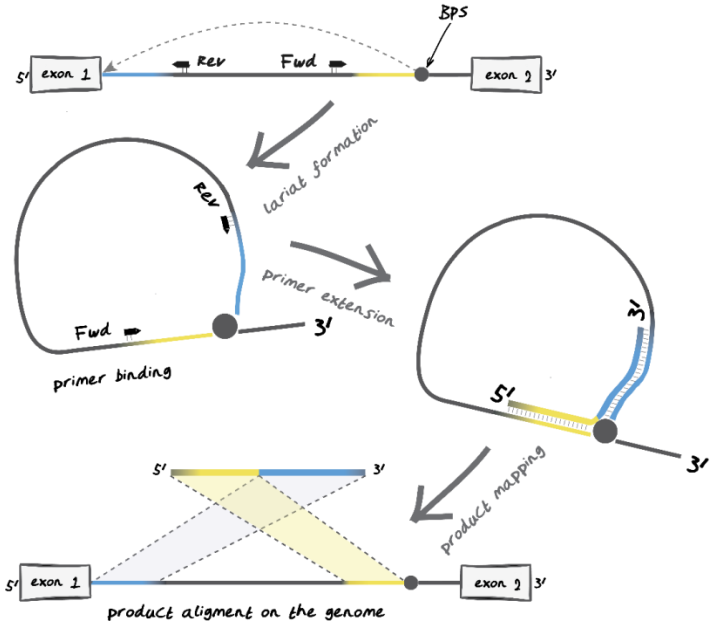


Figure 8 Lariat PCR (larPCR) schematic overview: *the primers are designed to be divergent on genomic DNA, but are able to amplify a product when the lariat is formed; alignment of the product sequence leads to a inverted split;*

MFN2, intron 2 lasli

The MFN2 intron 2 is approximately 8 kb long and has a predicted nested splicing event at its 5' splice site. The lasli length is 455 bp and particularly intriguing is that the upstream exon is skipped (**Figure 9 A**). The nested lariat was detected via PCR, the products loaded on an agarose gel and the indicated bands were sequenced

(**Figure 9B**). The sequences are visualized below (**Figure 9C**), the lasli area is zoomed in. The predicted lasli is confirmed via sequencing, found in cells that were not transfected (in further text termed mock). RNA from cells where the debranching enzyme has been knocked down (in further text dbr1 knock down, DB KD) seems to influence exon skipping since the exon is also excluded from the transcript together with the lasli. This of course raises the question of how the rest of the introns is spliced: it might happen in subsequent splicing step or the mRNA might be degraded in the nonsense mediated decay.

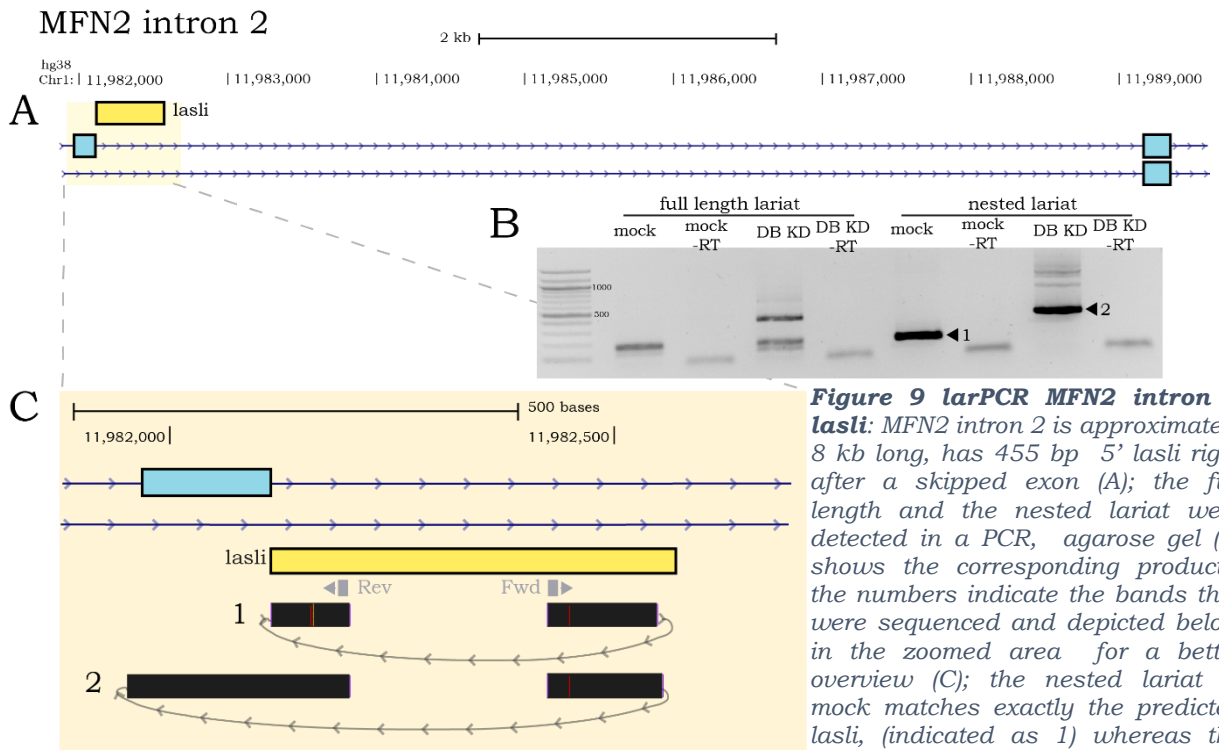


***Figure 9 larPCR MFN2 intron 2 lasli**: MFN2 intron 2 is approximately 8 kb long, has 455 bp 5' lasli right after a skipped exon (A); the full length and the nested lariat were detected in a PCR, agarose gel (B) shows the corresponding products, the numbers indicate the bands that were sequenced and depicted below in the zoomed area for a better overview (C); the nested lariat in mock matches exactly the predicted lasli, (indicated as 1) whereas the nested lariat in the debranchase knockdown (2) includes the alternatively skipped exon;*

PRKAB2, intron 7 lasli

PRKAB2 intron 7 is approximately 2 kb long, contains a 3' nested lariat which stretches 700 bp. The full length intron was found in untransfected cells using the lariat PCR approach. RNA from cells after the dbr1 knockdown shows again an unconventional splicing pattern: the lariat contains, besides the intron, the upstream exon as well as parts of the intron 6 (data not shown).

The nested lariats could be picked up in mock and the knockdown (**Figure 10A**). Surprisingly we did not only find the predicted lasli (band indicated with 2) but also laslis that have a alternative 5' splice site (**Figure 10B**).

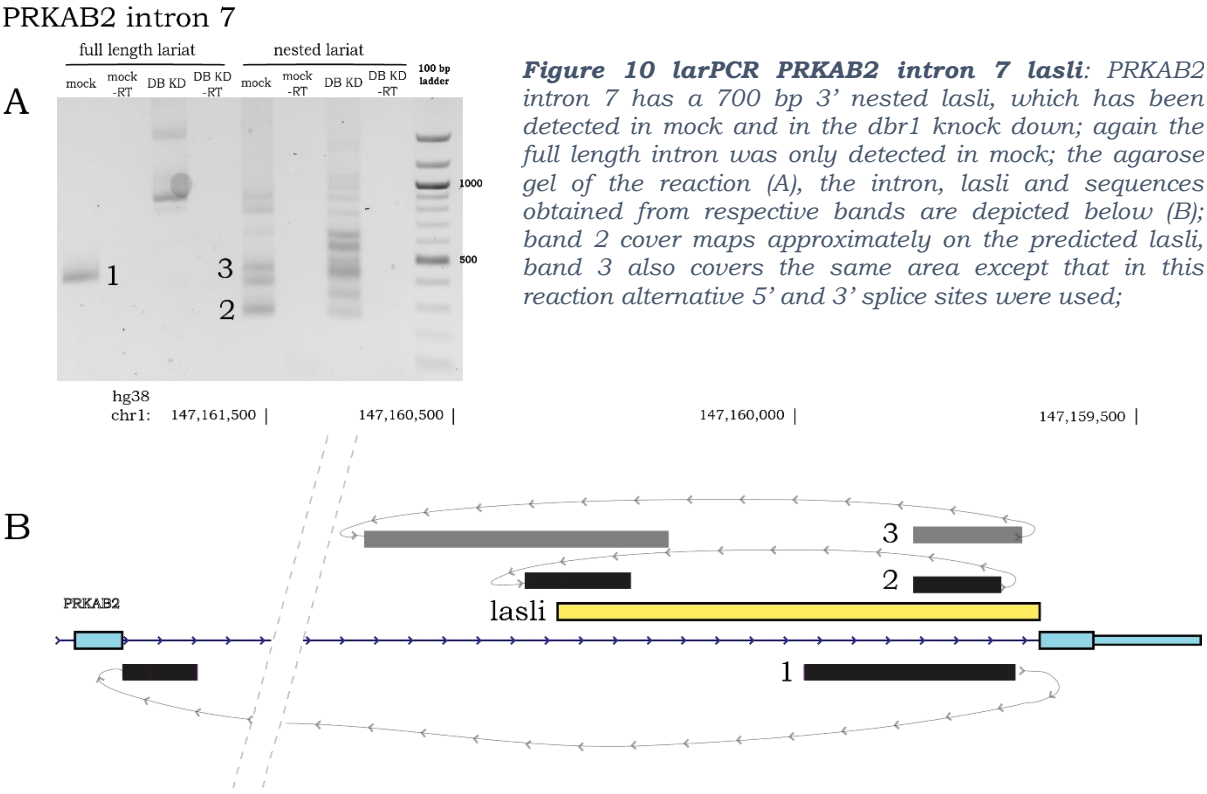


***Figure 10 larPCR PRKAB2 intron 7 lasli**: PRKAB2 intron 7 has a 700 bp 3' nested lasli, which has been detected in mock and in the dbr1 knock down; again the full length intron was only detected in mock; the agarose gel of the reaction (A), the intron, lasli and sequences obtained from respective bands are depicted below (B); band 2 cover maps approximately on the predicted lasli, band 3 also covers the same area except that in this reaction alternative 5' and 3' splice sites were used;*

CTIF, intron 2 lasli

CTIF intron 2 is approximately 80 kb long, has a nested splicing event that spans 1529 bp, no surrounding exons can be found in the conventional annotations as well as in the expressed sequence tags which makes this lasli an

optimal candidate to verify a nested splicing event where the splice sites seem not to be reconstituted.

In the larPCR we were able to detect the nested lariat. The predicted lasli was found in cells with the debranchase knockdown, an alternative variant was also found in untransfected cells.

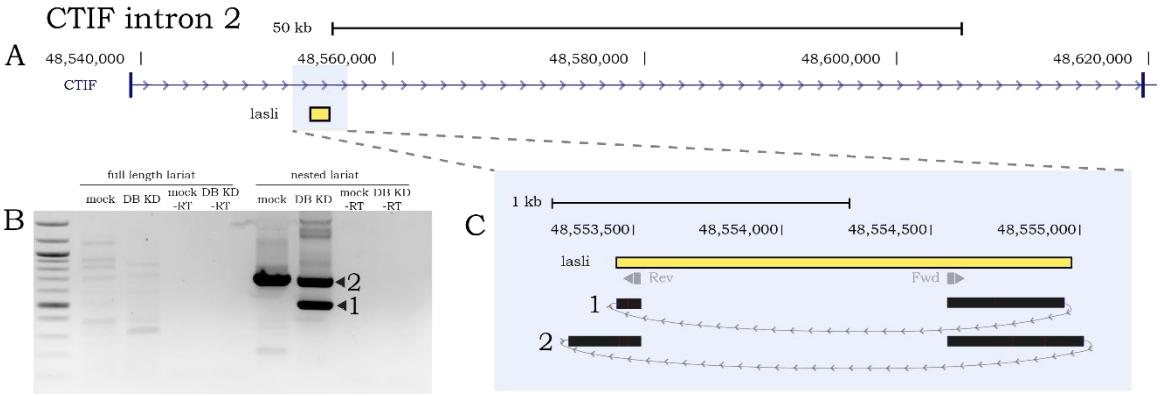


***Figure 11 larPCR CTIF intron7 lasli**; up overview of the native intron and its exons, including the predicted lasli in yellow; below left the agarose gel with the lar PCR products loaded; below right the sequences acquired from the indicated bands; the middle lasli is 1529 bp long, no exons are annotated in its surroundings; the predicted nested lasli was found in HeLa debranchase knockdown, while an alternative lasli with shifted splice sites has been found in untransfected cells;*

LASP1 lasli

LASP1 intron 2 is about 12 kb, has a circular split read at the 5' end that is 9186 bp in length and a 2790 bp lasli at the 3'. Since laslis require a circular and a linear split, the first circular split read is lacking the corresponding linear split read and didn't retain in the newest dataset with additional filtering.

The PCR approach confirmed the circular read as well as the lasli. The PCR product indicated with **5** on the agarose gel in **Figure 12 A** has a GT 5' ss; upstream of it is an adjacent 3' ss which could make these nested splice events potentially recursive. The branch point site of the circular split read indicated with **1** on the gel is approximately 230 nt away from the detected 5' ss of the nested lasli, the other two products (**2,3**) are also not fitting in the recursive model since their 3' ss overlap with the nested lasli. The PCR product **4** could be size wise, part

of a recursive splice pattern, although its 5' splice site is not conserved. There is also the possibility that the exactly matching recursive splice events were not detected in our approach, but there are indications that this intron is spliced in a recursive manner.

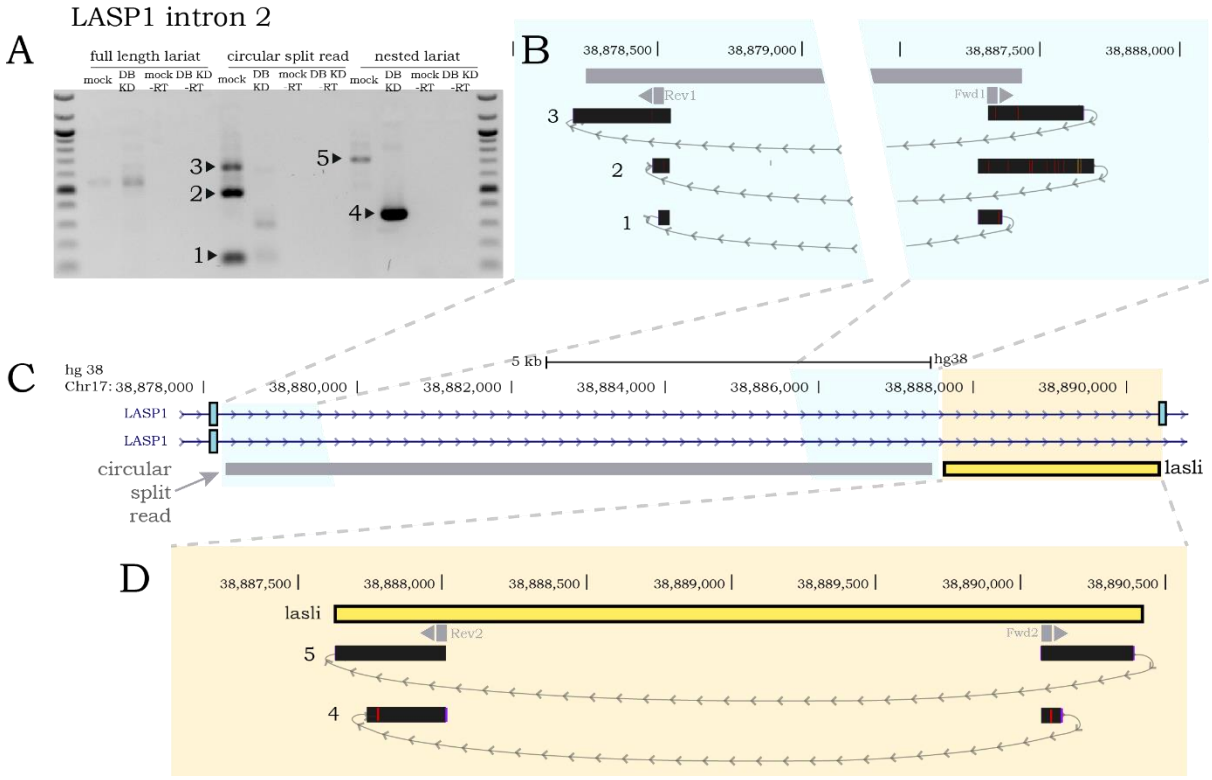


***Figure 12 larPCR LAPSP1 intron 2 lasli and circular split read**; intron and the flanking exons overview in the middle(C), the agarose gel where the full lariat, the circular split and the nested lariat have been found (A); up right are the splice sites of the circular split depicted(B), products 1,2,3 map on the predicted nested lasli with alternative splice sites (D);*

SRRT laslis

SRRT intron 5 spans approximately 1800 bp, contains two laslis, a 5' lasli ranging 526 bp and a 3' lasli that is 981 bp long. An excision of both laslis does not reconstitute the splice sites which might indicate an unannotated exon between them, the data from express sequence tags show no exon in the predicted area. Looking at the conservation level in mammals there is a tendency to keep the sequences between the nested laslis unaltered, compared to the surrounding sequences, even in the evolutionary more distant species as mice. If not removed in

subsequent processing steps, this potential cryptic exon introduces two premature stop codons into the mRNA sequence, which would likely lead to NMD of the RNA.

In the PCR approach we detected the full length lariat, as well as both of the laslis (**Figure 13).**

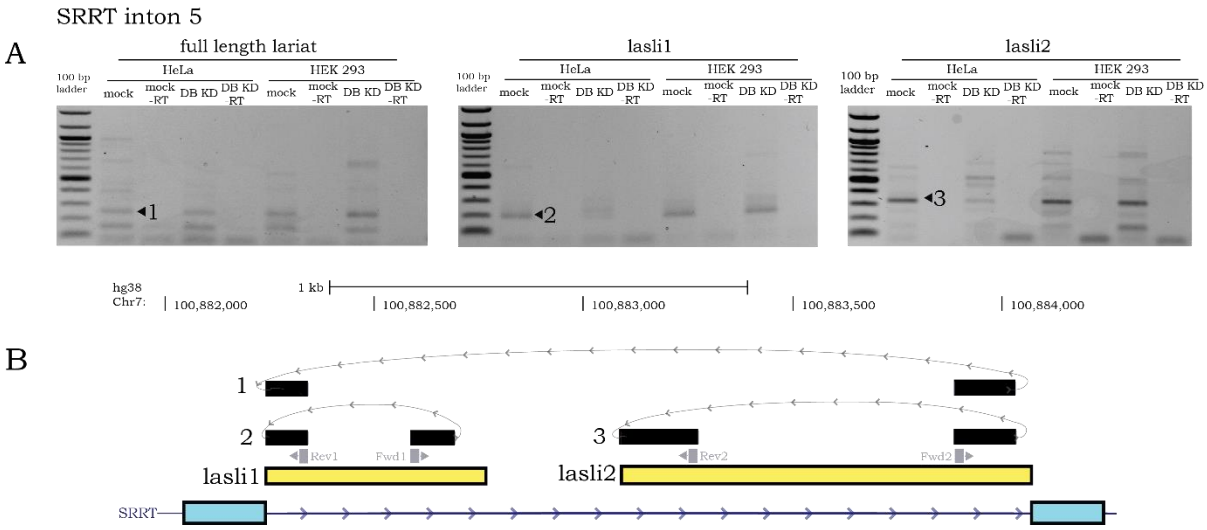


***Figure 13 larPCR SRRT intron 5 and its laslis**;(A) the respective agarose gel of the PCR products; the with numbers indicated bands were sequenced, the sequences obtained in the process were visualized in a genome browser (B); band 1 covers the whole intron, band 2 maps on the predicted 5' lariat, band 3 matches lasli2;*

GNB2L intron 7

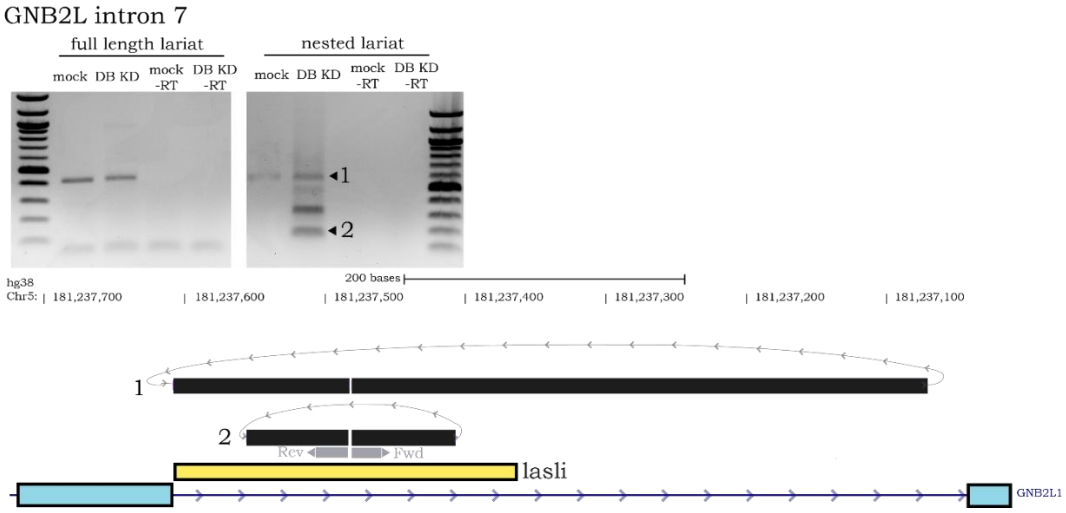


***Figure 14 larPCR GNB2L intron 7 5' lasli:** intron 7 of GNB2L is 566 nt long and has a 5' lasli that spans 245 nt; above is the agarose gel of the PCR reactions where we identified the full length lariat (indicated with 1) as well as the predicted laslis (2) with a alternative 5' splice site; below is the sequence alignment in the UCSC genome browser;*

GNB2L intron 7 is relatively short, spans 566 nt but contains two laslis, a 5' lasli that is 245 nt and a 3' nested splicing event that removes a 182 nt long piece. In the PCR approach we only aimed for the 5' lasli. Due its size and because the primers were not of suitable nucleotide content, the primers suggest by *primerblast* were almost back to back. Nevertheless, we were able to identify the nested lariat in the dbr1 knockdown in HeLa cells, although with an alternative 5' splice site, there is also a predicted branch point site deviation of approximately 30 nt. Interestingly the full length lariat could also be detected with the same primers, also without dbr1 knockdown, where we could not detect the nested splicing event (**Figure 14**).

Comprehensive approach for SQSTM1 intron

In the extended verification of the predicted circular split reads - assumed nested lariats, we picked intron 5 of SQSTM1 that spans approximately 8 kb. It seems to be saturated with nested splicing events. We picked 12 of them, designed outer and inner primer. cDNA acquired from HeLa cells was used as template for the first PCR. Again a knock down of the debranching enzyme (dbr1) was performed on HeLa cells as well (**Figure 15**).

In this approach we were able to get products for each of the primer pairs, which is not surprising considering that altogether the PCR ran for 55 cycles, allowing for the amplification of lowly abundant RNA species, but the negative –RT is a proper indicator that are products are not coming from genomic DNA. Many of target products are found in this approach since the PCR run for so many cycles and four different primers have been used in the process. Introns tend to have repetitive sequences which make them less suitable for primer design.

Sequencing the respective PCR products, we were able to confirm six of the predicted twelve nested lariats. The sequence alignments are split, as expected, split, confirming the circularity of the RNA it was derived from. There is some discrepancy in the predicted splice sites that can be explained with alternative splice sites and might also indicate that the nested splice sites are *weaker* than constitutive, having a reduced affection towards spliceosomal components thus still attracting it but not with the usual precision that would be required to define exon-intron boundaries.

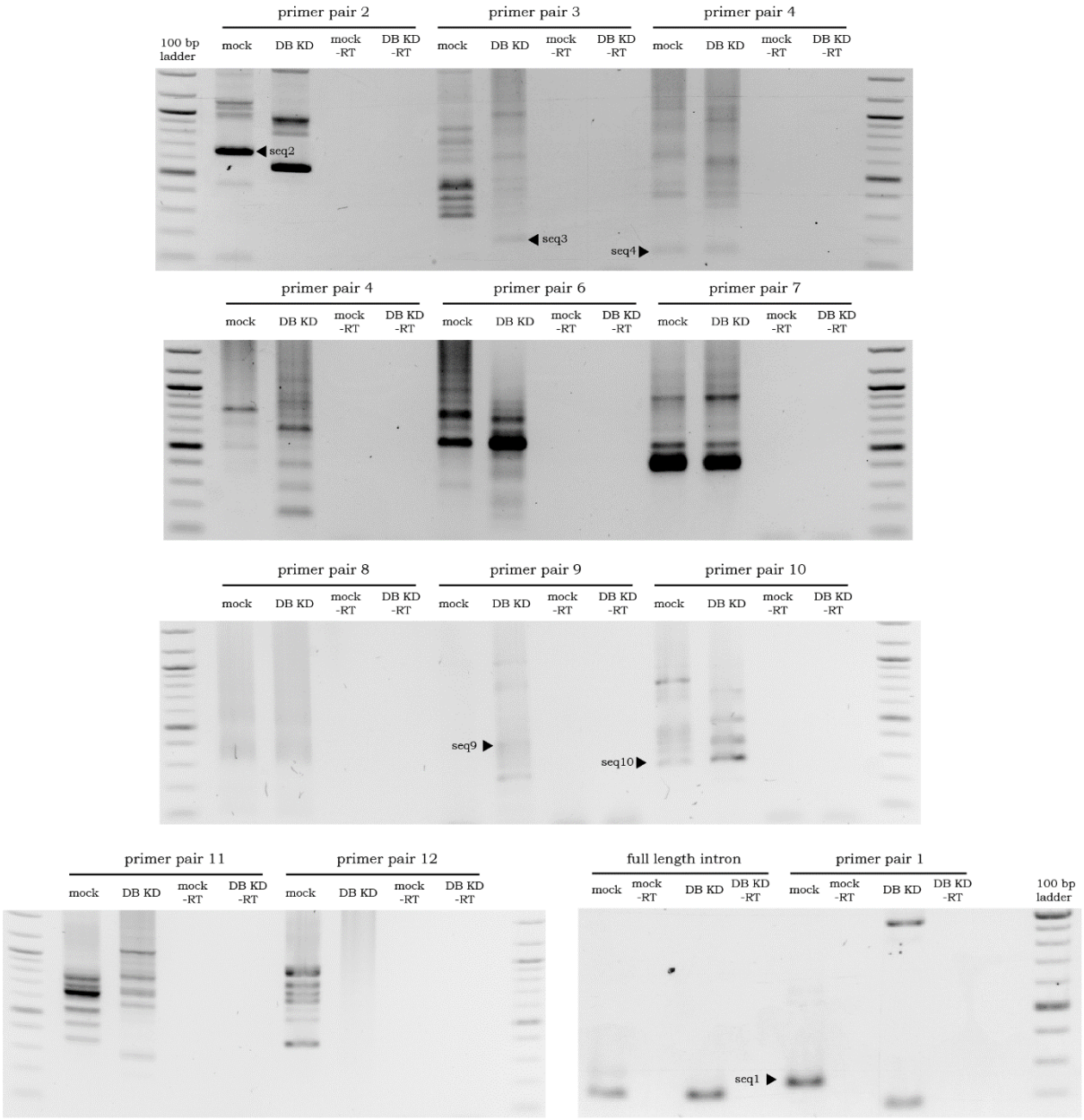


Figure 15 Agarose gel picture of nested lariats in SQSTM1 intron 5*; 12 nested splicing events were predicted as circular split reads (overview in* ***Figure 16****) and detected in a lariat PCR approach, cDNA acquired from HeLa cells (mock) and HeLa cDNA where the debranching enzyme has been knocked down (DB KD) was the template; indicated products were sequenced, sequences obtained in the approach are visualized in the* ***Figure 16****;*

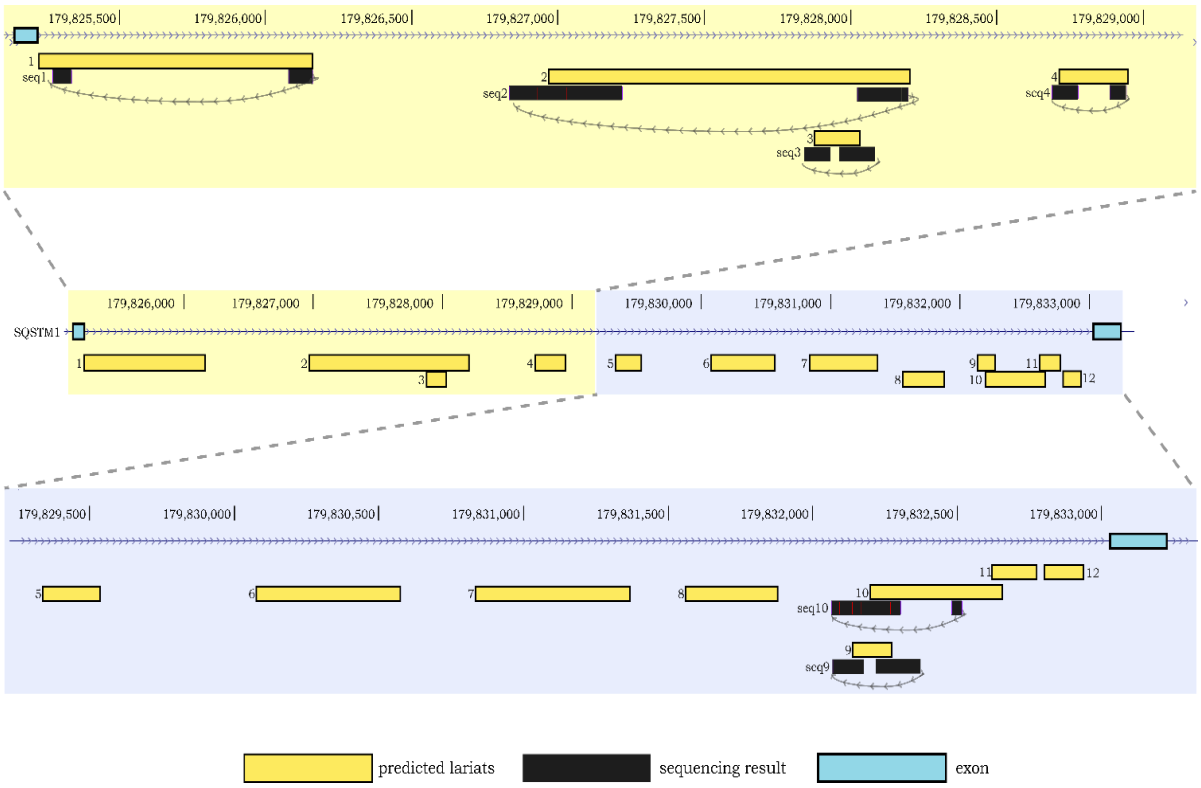


Figure 16 Nested lariats overview of SQSTM1 intron 5 and the sequences of the PCR products from Figure 15: *in yellow the predicted nested lariats, in black the sequencing results, the arrow connecting the black bars indicates the split; six out of twelve predicted circulars split reads were confirmed in HeLa in a single approach;*

To summarize using in the lariat PCR we were able to verify several of nested lariats in different cell lines. Considering the lariat abundance and that for the lasli acquisition datasets of at least 12 different tissues and multiple immortalized cell lines were used, this confirmation rate is remarkably high, especially as we were focusing on unusual and non-constitutive splicing events. The double confirmation of a lasli with a linear and a circular split read can be considered successful in lariat prediction.

Out of eleven tested introns with predicted nested splicing events we verified seven of them, three with multiple nested laslis or circular split reads. For some of the unverified introns we had troubles detecting the full-length lariat as well indicating the low expression in the tested cell lines. Even if a full length intron is detected, the inability to get the nested lariat could be explained by various reasons, for example the absence of splicing factors necessary for the nested splice sites selection or by a mutation of the nested splice site in the particular cell line.

Virtual northern blot

The presence of intrasplicing events within a constitutively spliced intron would result in a shorter lariat from the final splicing reaction. As larPCR is unsuitable to measure lariat length, virtual northern blotting was used to obtain this information by size-fractionation of RNA and subsequent PCR amplification. The advantage of this method compared to a conventional northern blot is the amplification step in the PCR, since the instable, lowly abundant lariats tend to be difficult to detect as we learned from the lariat PCRs. By Lariat PCR we for the most part could identify the 5' ss and the branch point, but could not make any conclusions about the rest of the lariats, if there are parts that have been pre-spliced. The idea behind the virtual northern blot was to identify the splicing intermediates in intrasplicing.

RNA from HEK cells was separated on formaldehyde agarose gels and a single lane was fractionated in seven different pieces: fraction 1 contains the shortest, fraction 7 the longest RNAs (sizes in the materials and methods section). RNA from the respective gel pieces was extracted and after a reverse transcription used as a template for a PCR detection of particular RNAs in respective fractions. RNA that has not been gel separated was used as a positive control (total RNA).

To verify the functionality of the whole approach we tested human GAPDH which has 11 transcript variants ranging 390-1875 bp, according to the Ensembl Genomes database. We were able to pick them up in 3 fractions between 0.7 and 3 kb (**Figure 17**) and of course in the total RNA, but not in the other fractions.

Detecting a highly abundant mRNA seems to work properly, but looking for lowly abundant lariats and even less abundant nested lariats turned out to be tricky. Since our results from these approaches were very inconclusive we tried to find highly abundant circular RNAs, although we could detect them in the total RNA (**Figure B**), they did not show up in any of the fractions. Reasons for this include diverging electrophoretic properties of circular RNA molecules as compared to linear ones and inefficient RNA extraction form agarose slices. These difficulties also apply for RNA lariats, which, compared with circular RNAs, are expected to be present in extremely low quantities, even further minimizing their detection by virtual northern blotting. Thus, for determination of lariat length, we had to rely on conventional larPCR and subsequent product sequencing.

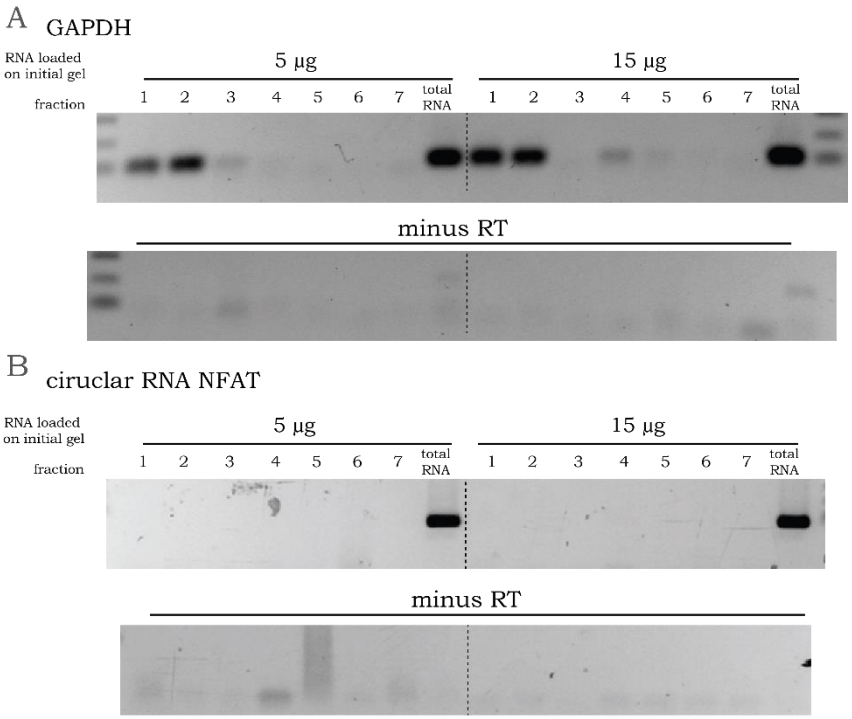


Figure 17 Virtual northern blot fractions analysed on a PCR agarose gel *(A) GAPDH agarose gel loaded with the respective fractions from the virtual northern blot; either 5 or 15 µg RNA were loaded per slot; GAPDH is detected in the fractions 1,2 and 3, naturally the signal tends to be more saturated the more RNA is loaded; minus RT is the lower gel; (B) circular RNA NFATC2 agarose gel of the respective fraction acquired in the virtual norther blot; the circular RNA is only detected in the total RNA but not in the respective fractions;*

Luciferase splicing assay

Luciferases are widely accepted reported genes, due to the luminescence only emitted in presence of a very specific substrate and the absence of luciferases in many organisms, like mammals. As this reporter system provides easily quantifiable readout, we cloned our introns of interest into the renilla luciferase minigene and checked if any expression differentiation could be detected, when intronic splice sites are modified. The firefly luciferase was used to normalize the signal since the transfection efficiency can vary from sample to sample.

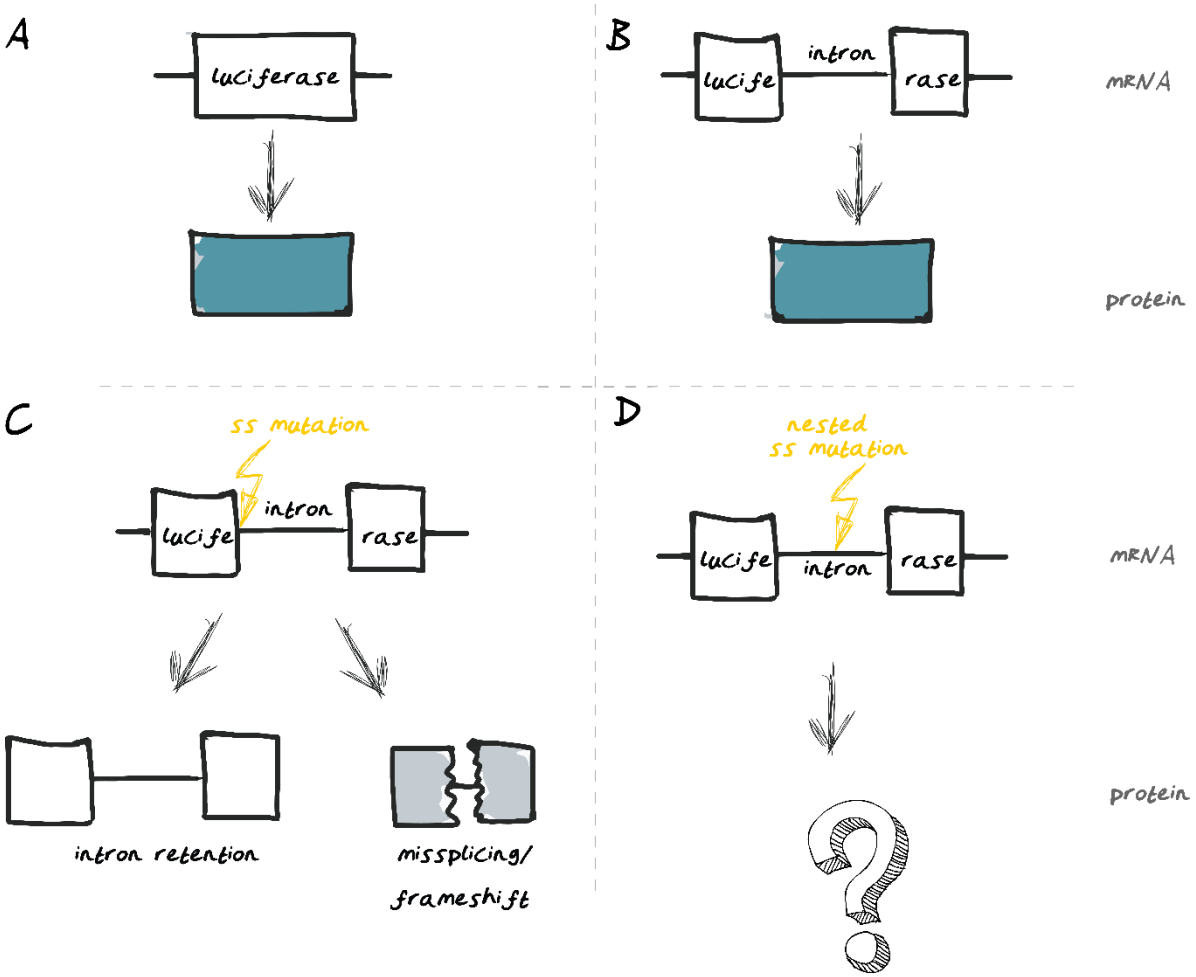


***Figure 12 Overview of the luciferase splicing assay**: (A) luciferase from the phRL-TK plasmid, after transfection, translation and substrate addition emits detectable light; (B) cloning an intron into the luciferase on the plasmid leads to splicing and a functional luciferase; (C) mutating the conventional splicing site is expected to produce intron retention or selection of a proximate splice site, which will most probably lead to exon shifting or non-sense mediated decay and the lack of luminescence; (D) we wanted to test a potential function of nested splicing events by mutating their splice sites;*

In order to test the functionality of the renilla luciferase splicing assay we cloned HPRT intron 7 and 8 and PRKAB2 intron 7 respectively into the phRL-TK plasmid (constructs **h7, h8, p7**, depicted in **Figure 19A**). To test which mutation

abolishes splicing most efficiently we mutated the 5'ss in p7, the GT was exchanged for AC and the construct is labelled **p7 1**. We altered the terminal nucleotides in the intron h7 from AG to GG and termed the construct **h7 170**. It has been shown that the removal of the pY tract of an intron abolishes the splicing more efficiently than the 3' ss mutation[45,46], so we made another mutant **h8 del**, where the polypyrimidine tract and the according 3' ss were removed. The intronless phRL-TK plasmid (**no insert**) and the phRL-TK plasmids containing the constructs described above were transfected into HEK 293 cells. To each transfection reaction pmirGlo-RL plasmid with a firefly luciferase was added, the pmirGlo-RL to phRL-TK ratio was 1:10. The transfected cells were incubated and the luciferase activities were measured as described in the methods section. The measured renilla luciferase activity was normalized to firefly luciferase signal. The respective constructs and the corresponding luciferase luminescence are visualized in **Figure 19**.

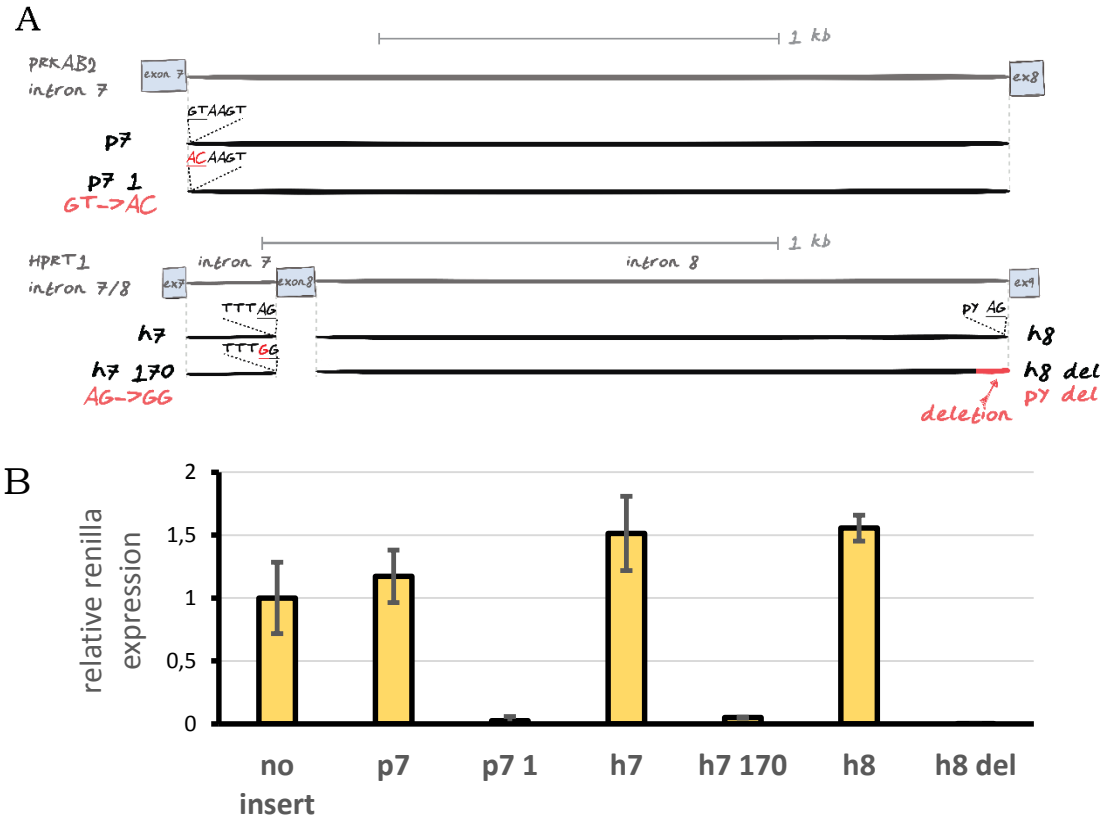


Figure 19 Renilla luciferase assay *(A) overview of the introns tested in the luciferase splicing assay and the responding mutants; (B) luciferase activity of the intronless renilla luciferase (**no insert**) compared intron containing luciferase (**p7**, **h7**, **h8**) as well as the effects of mutations in conventional splice sites; 5' ss was mutated in prkab2 intron 7 (**p7**), GT was exchanged with AC, 3' ss was mutated in HPRT1 intron 7 (AG substituted with GG, construct **h7 170**); in HPRT1 intron 8 the pY tract was removed along with the 3' ss (**h8 del**); the ss mutation and pY tract deletion abolishes proper translation of the luciferase and therefore its activity;*

Unexpectedly the intron containing luciferases had a higher emission intensity than the intronless version. It has already been shown that spliced mRNA interacts with the cleavage and polyadenylation machinery and that they are exported from the nucleus more efficiently, therefore leading to a higher transcript abundance in the cytoplasm, ergo increasing the chances for its translation. Our constructs with the mutated splice sites or lacking the pY tract have no renilla luciferase activity. A PCR approach showed that the pY deprived introns still get spliced, but an alternative 3'ss was selected, which introduced a frameshift in the protein and presumably lead to degradation of the mRNA via nonsense mediated decay.

Prespliced nested introns

As previously mentioned PRKAB2 intron 7 and GNB2L intron 7 contain laslis. To check if the lasli itself gets spliced and if after this splicing step a novel splice site is reconstituted, as required for recursive splicing, deletion mutants were designed.

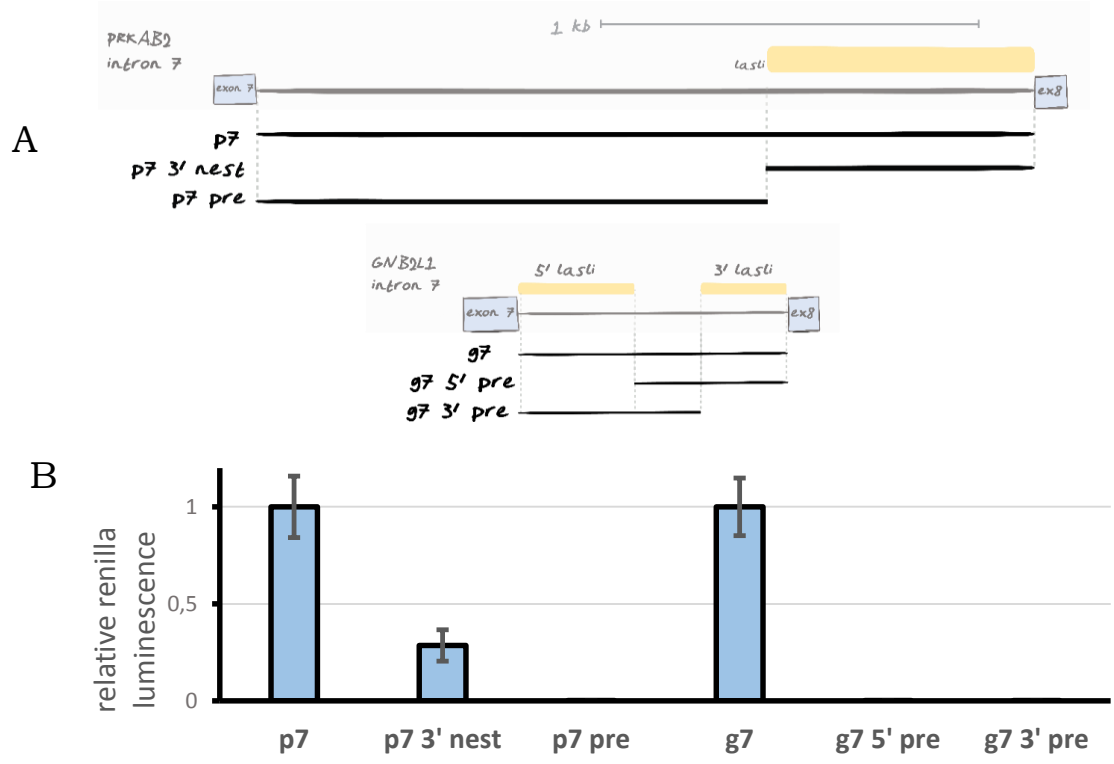


Figure 20 luciferase splicing assay for nested lasli of p7 and prespliced introns p7 and g7*; A) overview of the constructs: PRKAB2b intron 7 - p7, its nested lasli p7 3' nest and the intron part that is retained after the lasli has been prespliced, p7 pre; GNB2L1 intron 7 and parts of it where the nested laslis have been prespliced; (B) the renilla luciferase luminescence; the nested lasli is spliced on its own, the renilla stays functional, the prespliced introns cannot be spliced properly and have no luminescence;*

The nested lasli of p7 was cloned into the luciferase (**p7 nest**) and a construct where the lasli has been prespliced was created (**p7 pre**). Luciferase containing GNB2L intron 7 was also used (**g7**) and two constructs – one where the 5' lasli was prespliced (**g7 5' pre**) and the other had the 3' lasli removed (**g7 3' pre**). The overview of the constructs and the assay results are depicted in **Figure 20A**. The lasli alone gets spliced without disrupting the luciferase activity, although not as efficiently as the whole p7 intron. P7 pre, g7 5' pre and g7 3' pre fail to produce a functional luciferase, indicating that the p7 and g7 are not recursively spliced (**Figure 20B**).

Examples of nested introns that have no influence on the luciferase activity

RAD54 intron 11 and MEG6 intron 11 both have a 5' nested laslis. In order to check their functionality, we cloned them into our reporter (constructs **rad** and **meg**) and removed their polypyrimidine tracts (**rad del** and **meg del**). The luciferase splicing assay was performed and the results are presented in **Figure 21** along with the respective construct.

The laslis in both RAD54 intron 11 and MEG6 intron 11 seem not to affect the luciferase acivity, since the removal of the pY tract shows no significant difference in the luminescence compared to the unalterted intron. From this we can assume that the nested splicing events do not happen in these introns or might serve as a part of a potentional recurisve splicing pattern which serves as an optional pathway in the intron removal.

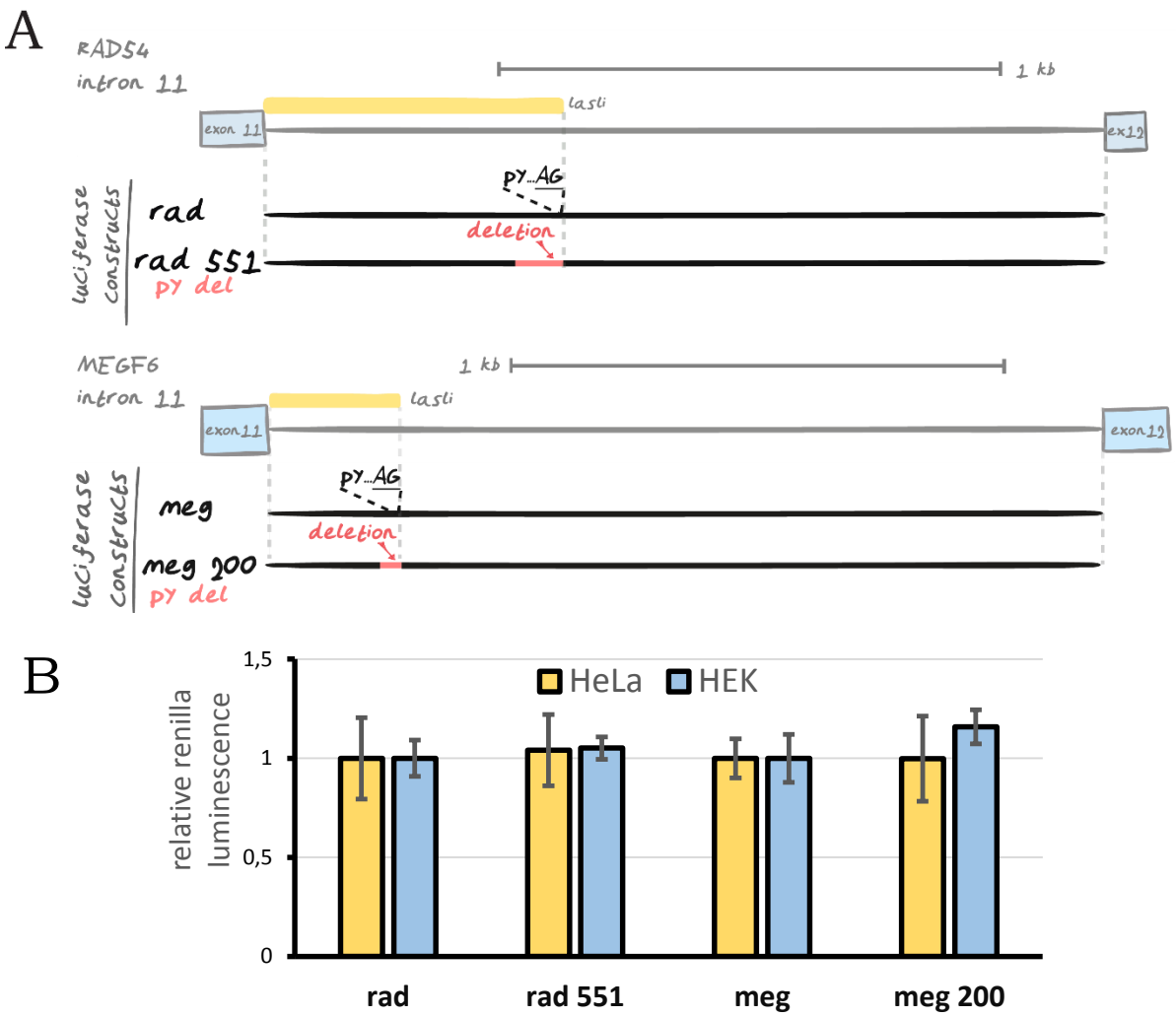


Figure 21 renilla splicing assay for RAD54 intron 11 and MEG6 intron 11*; (A) overview of the constructs that were cloned into the renilla luciferase (**rad** and **meg**) and the mutants where the pY tract was deleted (**rad del** and **meg del**); there is no significant difference between wild type and the altered introns suggesting that the nested laslis have no effect on the full intron splicing;*

Examples of nested introns that have an inhibiting effect on the full intron splicing efficiency

GNB2L intron 7 has two nested laslis and PRKAB2 intron 7 has one as previously mentioned. To investigate their effect on full intron splicing, we modified the constructs **g7** and **p7** and mutated the intra-intronic splice sites. The 5' nested lasli of g7 spans 243 nt and we exchanged the AG at the 3' ss with a CC (**g7 242**) and to avoid regular splicing of the 3' nested lasli we mutated its 5' ss by mutating the GT at position 384 to a CC (**g7 384**). The 3' nested lasli of p7 starts at position 1362 of the intron, we mutated its 5' ss by exchanging the GT with a AC (**p7 1362**).

The splicing assay was performed and the results of it along with the constructs cloned into our reporter are depicted in **Figure 22**. By inhibiting the splicing of GNB2L intron 7 5' nested lasli **g7 242**, the luciferase activity was increased 1.5-fold, inhibiting splicing of the 3' nested lasli **g7 384** also increased the luciferase abundance, this time 2-fold. The same pattern can be observed once the splicing of the PRKAB2 intron 7 nested lasli was inhibited in **p7 1362**, the luciferase activity is increased 2-fold. From this evidence we conclude that the nested splicing events in these particular introns have an inhibitory effect on the full intron splicing.

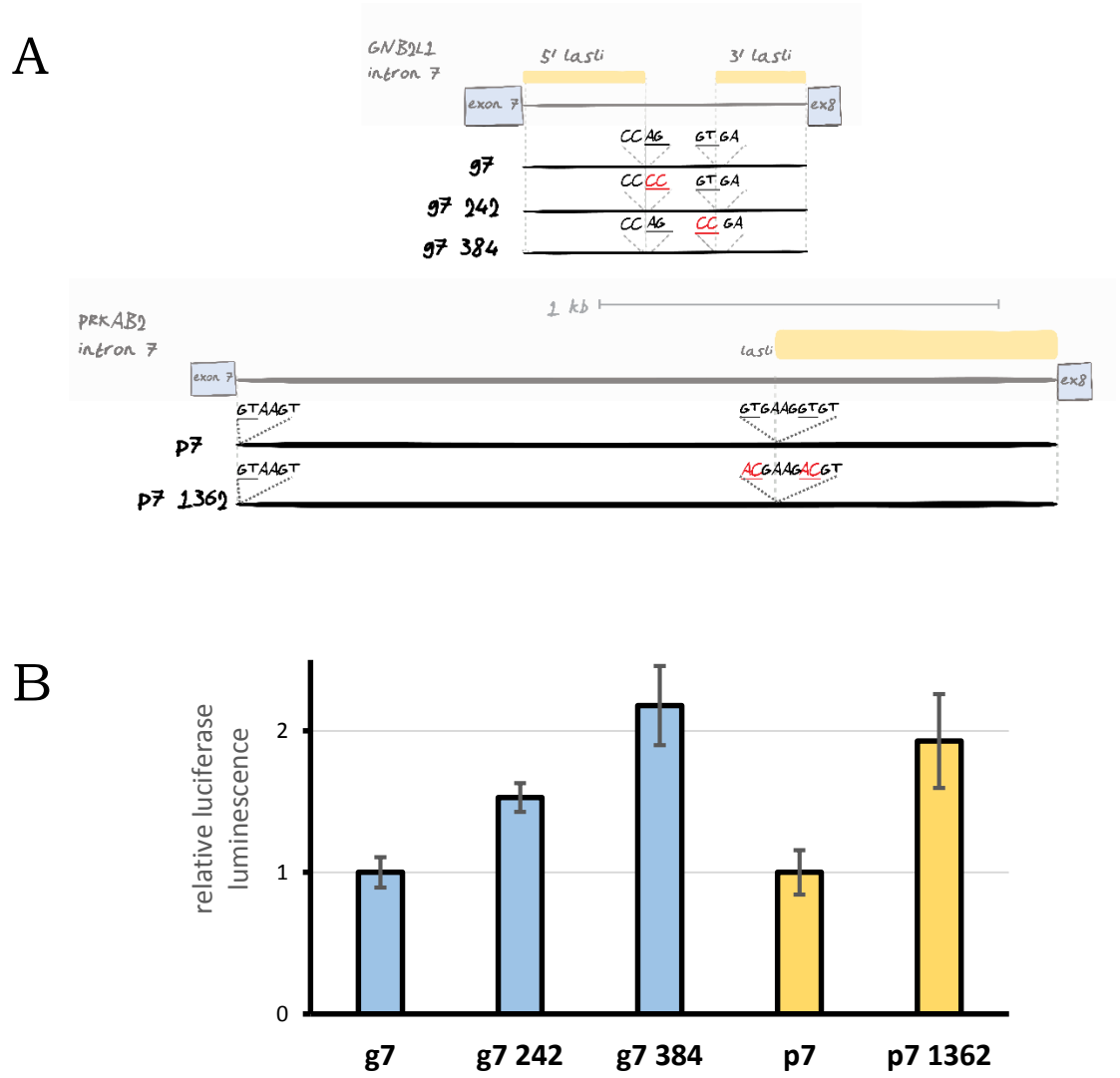


Figure 22 renilla splicing assay of GNB2L intron 7 and PRKAB intron7*; (A) overview of the constructs that were cloned into the renilla luciferase and the respective mutations; (B) renilla luciferase emission normalized to firefly luciferase emission; mutating the splice sites of the nested introns in* ***g7*** *(****g7 242*** *and* ***g7 384)*** *increases the luciferase activity, suggesting a functionality of these events; same can be observed for* ***p7****, mutating the 5' ss of the nested intron leads to an increased luciferase activity;*

Examples of nested introns that increase the full length intron splicing efficiency

Rbm17 intron 8 has three potential nested laslis. In order to examine the function of these splicing events we cloned the wild type intron into the renilla luciferase reporter (construct termed **rbm**) and mutated the respective nested splice sites. For lasli 1 and 2 we removed the polypyrimidine tract (**rbm 213** and **rbm 514, respectively**) and since lasli3 pY tract is necessary for regular full intron splicing, we mutated the 5‘ ss of the nested lasli, exchanging GTs with AAs (**rbm 707**).

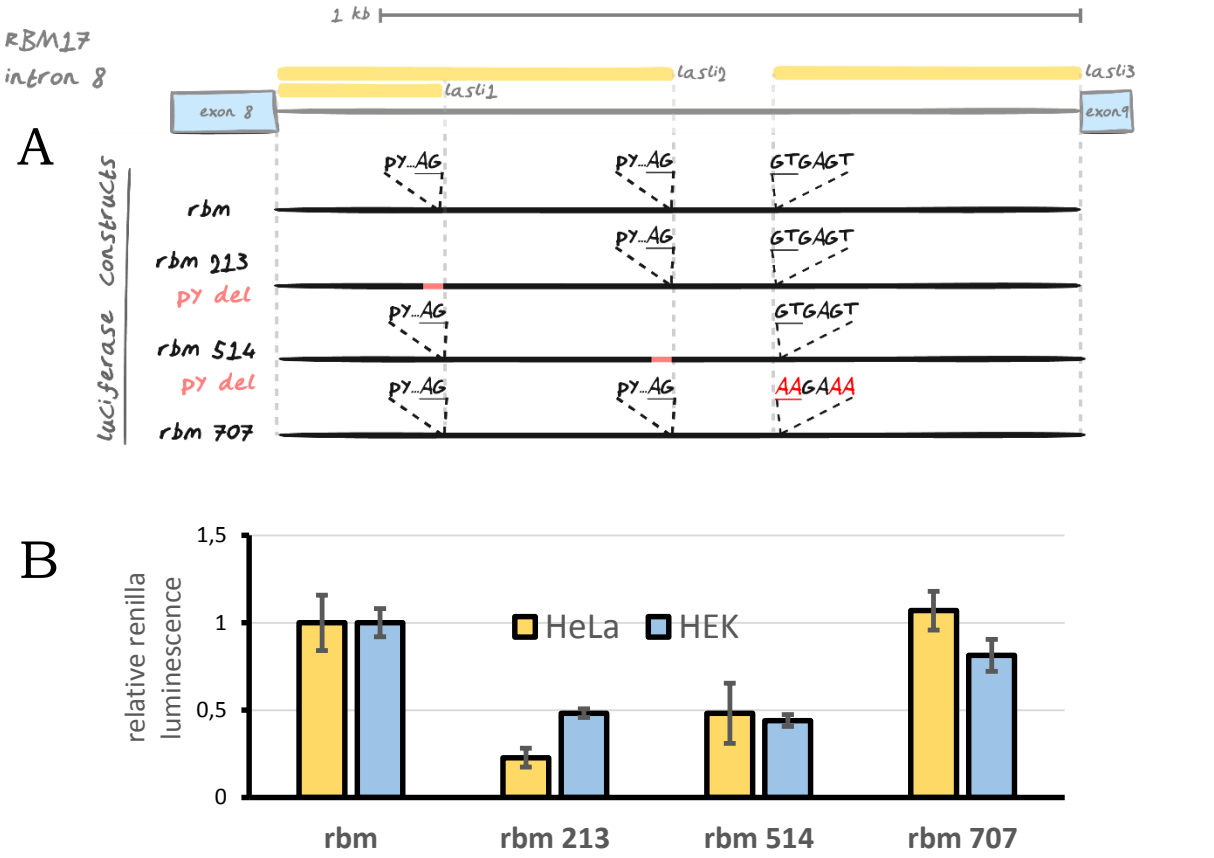


Figure 23 renilla splicing assay for RBM17 intron 7*; (A) overview of the constructs that were cloned in the renilla luciferase; the wild type intron (**rbm**), lasli1 pY deletion is labelled **rbm 213**, lasli2 pY deletion **rbm 514** and 5’ ss deletion of lasli3 **rbm 707**; (B) relative luciferase luminescence of the respective constructs, **rbm 213** and **rbm 514** have a significant depletion of the luciferase emission, whereas the effect is not as strong in **rbm 707**; expression of HeLa cells visualized in yellow, HEK in blue;*

We transfected HeLa and HEK 293 cells with the constructs and with a plasmid containing firefly luciferase whose signals was used for normalization of the transfection efficiency. The calculated luminescence is depicted in **Figure 23**. In HEK the mutants **rbm 213** and **rbm 514** have a significantly lower luciferase luminescence compared to the wild type. From this data we conclude that the nested laslis 1 and 3 in RBM17 intron 8 benefit the splicing of the full length intron. HeLa also shows a similar pattern, although **rbm 707** has a similar luciferase luminescence as the native intron.

qPCR data corresponds to these findings: the mutants have a lower renilla luciferase mRNA abundance than the cells that have been transfected with a construct where the nested splice sites have not been changed.

The removal of the pY tract of lasli1 (**rbm 213**) leads to a less effective splicing of the whole intron **Figure 24**. The partial intron retention stressed the importance of the nested laslis splicing, without it occurring functional transcript abundance is lowered. The retained intron has an alternative 3' ss once the pY tract of the nested lasli is removed, which probably means that the 5' ss of the adjacent intron is not as strong the 5' ss approximately 50 nt downstream.

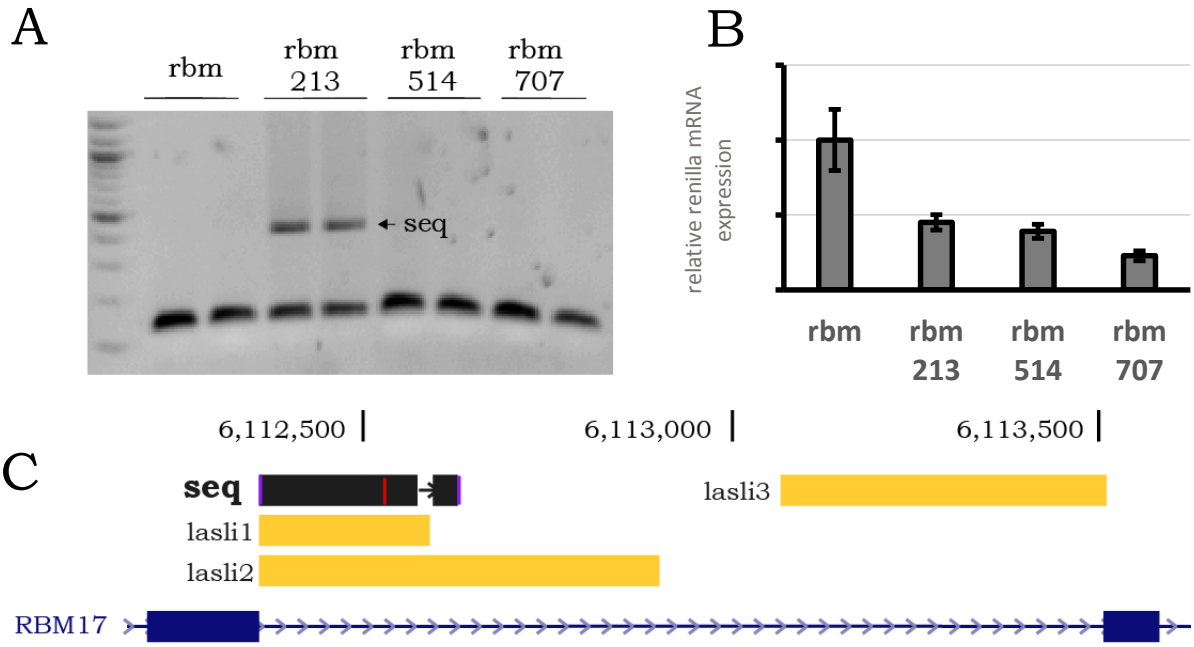


Figure 24 (B) qPCR data for rbm17 intron 11 and its mutants; *(A) agarose gel of the products; mutations in the lasli1(**rbm 213**) has a retained part of the intron 8, indicated with an arrow; sequencing results of the indicated band visualized in BLAT (C); the retained part of the intron does not exactly match the predicted lasli1 3' ss, indicating that the removal of the pY tract in lasli selects an alternative 5' ss in splicing of the adjacent downstream intron;*

Nested splicing events and Exon skipping

We wanted to test if nested splicing events have any influence on exon skipping in an explorative approach. The idea was genuinely simple: cloning of exons and the adjacent introns into a plasmid, mutating the nested splice site in such a way that they do not disrupt the conventional splice sites, transfect cells with wild type and mutated plasmid and look for the splicing variants in a PCR approach.

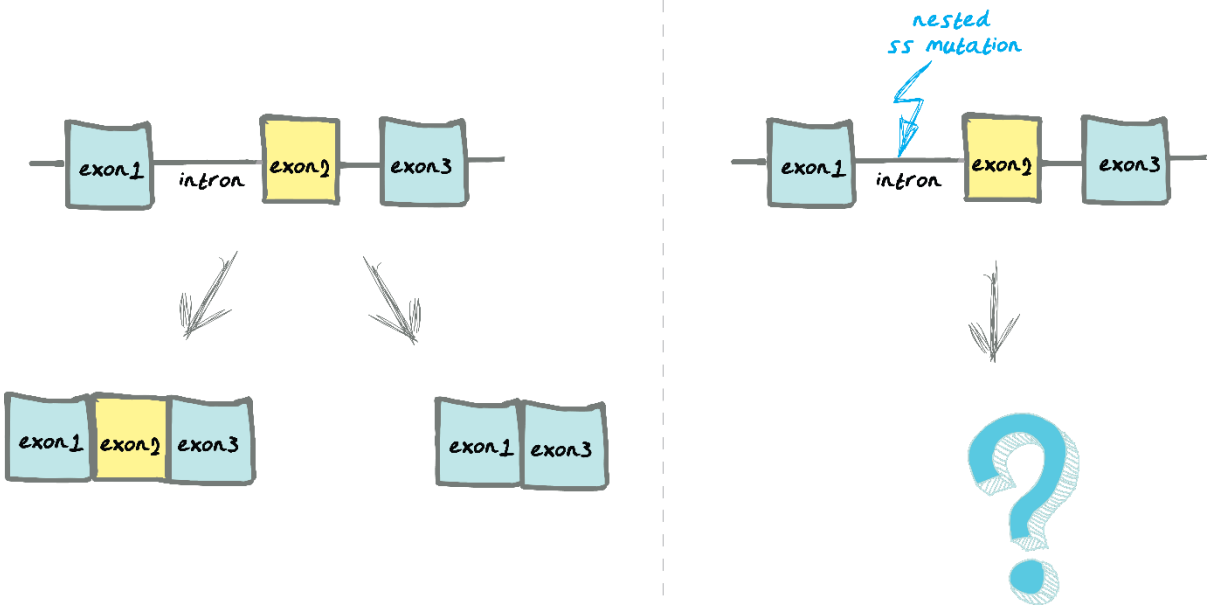


Figure 25 Scheme of exon skipping, *constitutive exons in blue, skipped exon in yellow; in a wild type scenario the exon is skipped or include; we were particularly interested in skipped exons that have nested splicing events happening in their proximity; in order to understand the function of these events we mutated the nested splice sites, transfected the wild type and mutant construct and compared the results*

Exon skipping of NME6 exon 5

NME6 intron 3 has a nested lasli in the intron right before an exon that is annotated as alternative/cassette exon which makes it a good candidate to understand role of nested laslis in alternative splicing. Exons 3,4,5 and 6 including the introns between them were cloned into pcDNA 3.1 vector after the highly

expressed CMV promotor[44]. The nested laslis pY tract and the 3‘ ss AG were deleted and the analysis executed as described in the methods section.

The splicing pattern of the mutant in HEK 293 as well as in HeLa are very similar, both being altered by the nested laslis mutation. Splicing variant containing exons 3 and 6, indicated with **1** seems to be highly abundant in the mutant. Variant indicated with **3** containing exons 3,4,5, and 6 only appears in wild type (**Figure 26**). The exon downstream of the nested lasli has an alternative, unannotated 5’ ss giving rise to the apparent non-full length lasli. Besides detecting this novel isoform, the mutational analysis elucidates the existence of an exon exclusion event in this gene: the strong increase of isoform **1**, instead of isoform **2** indicates that splicing of intron 4 is a pre-requirement of exon 4 and 5 skipping.

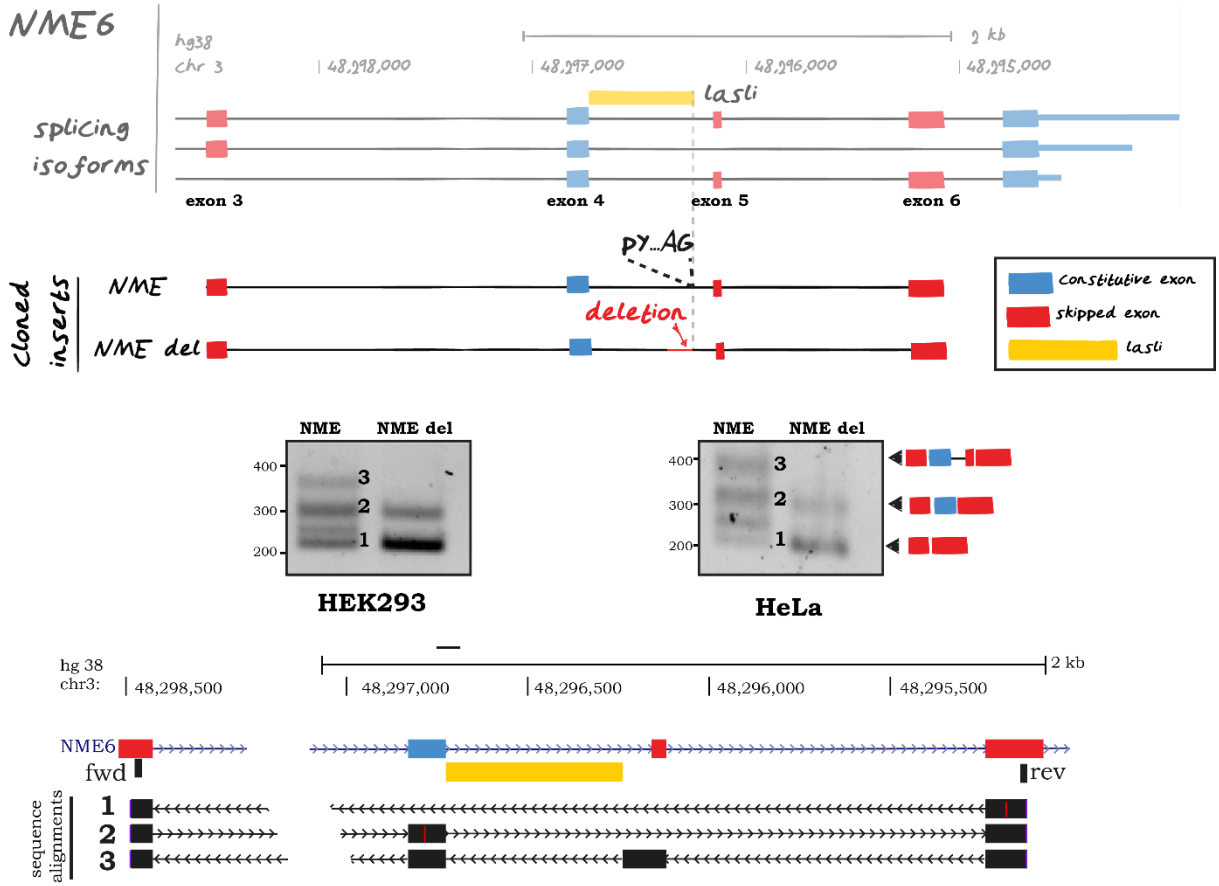


Figure 26 exon skipping analysis of NME6*; (up) the genetic locus of the gen with its constitutive exons in blue and skipped exons in red, nested lasli in yellow, below are the constructs that were cloned into our vector; (middle) agarose gels of the respective PCR products; (bottom) the sequence alignments of the band indicated with numbers;*

Exon skipping of DCTN3 exon 4

DCTN3 intron 4 contains a nested lasli adjacent to the skipped exon (**Figure 27**). To examine the ability of the nested lasli to infulence splicing variants we removed its pY tract and the 3' ss. One of our constructs contained exons 3, 4 and 5 with the introns between them (**DCTN)**, the other was the pY deletion (**DCTN del**).

The splicing patterns show no obvious difference in HEK 293, the unaltered construct and the pY deletion include all of the 3 exons (indicited with **1**) . A minor change is observable in HeLa, the wild type has splicing variants that disappear once the nested pY tract is removed (**Figure 27**, product indicated with **2**). The lowly abundant product indictated with **2** includes a cryptic exon, that uses the lasli defined intronic 3'splice site. Wether this exon is part of a functional mRNA, if it is a recursive exon or triggers NMD is yet to be determined.

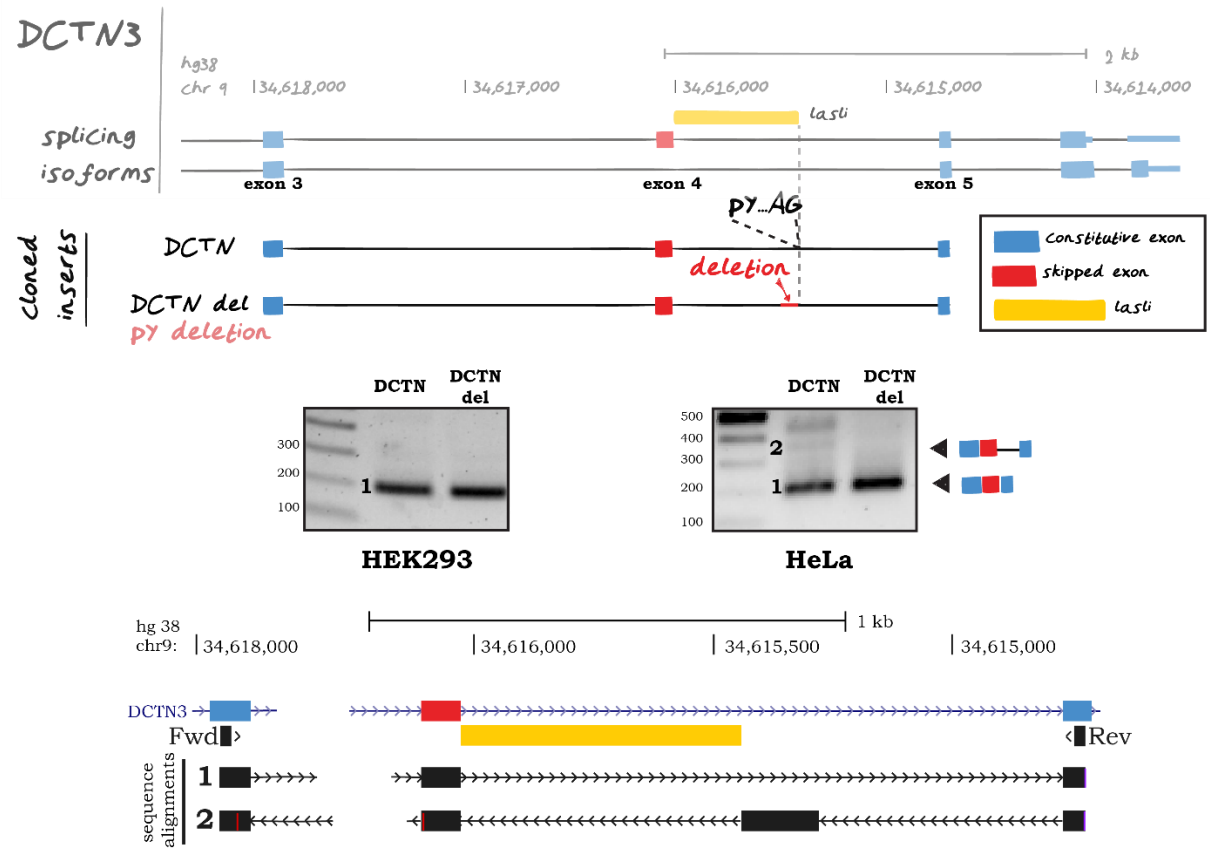


Figure 27 exon skipping analysis of DCTN3*: (up) the genetic locus of the gen with its constitutive exons in blue and skipped exon in red, nested lasli in yellow, below are the constructs that were cloned into our vector; (middle) agarose gels of the respective PCR products; (bottom) the sequence alignments of the band indicated with numbers*

Exon skipping of SACML1 exon 5

SACML1 has a lasli at the 3' in intron 4. Exon 5 is alternatively skipped, which makes it an appropriate candidate for the exon skipping analysis. We cloned exons 4, 5 and 6 and the corresponding introns, construct named **SACML** and we mutated the nested laslis 5' ss by exchanging the GT with a AC - **SACML mut**, **Figure 28**.

We were able to detect only one slicing variants in both cell lines that we tested, suggesting that the splicing variant with the exon 5 skipped cannot be detected in these cell lines, even after overexpression under a highly expressed promotor.

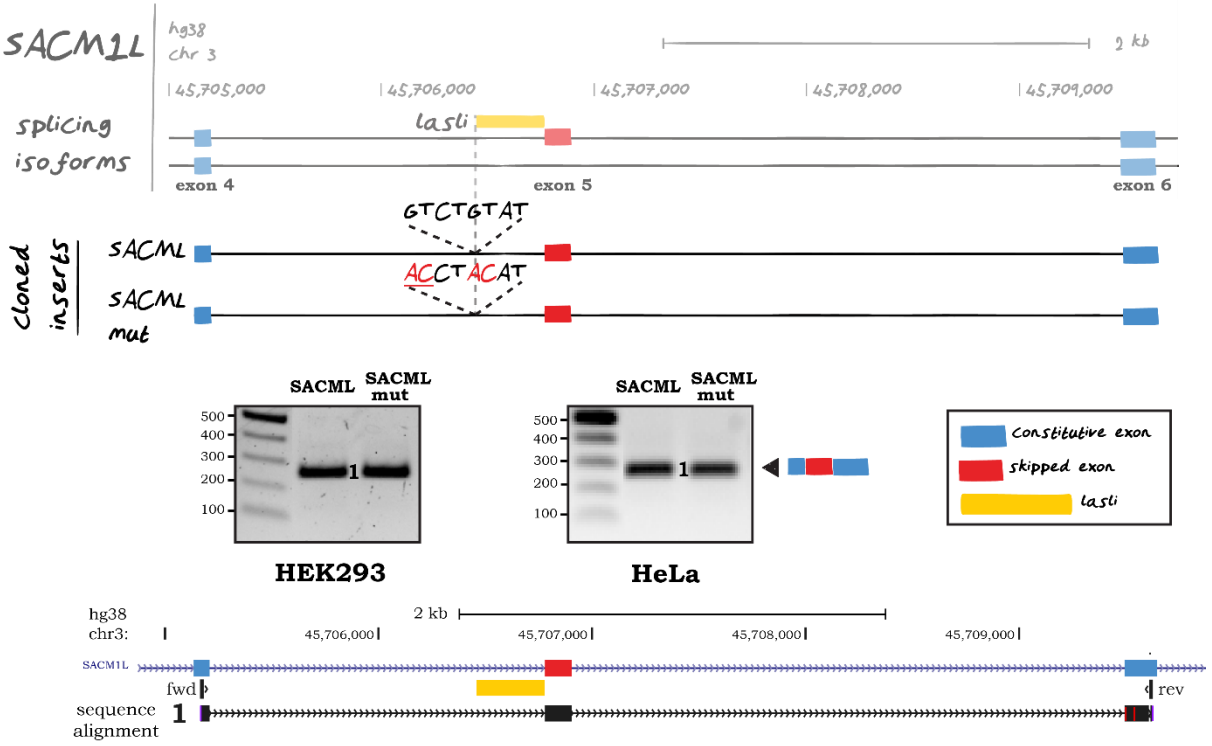


Figure 28 exon skipping analysis of SACML1*: (up) the genetic locus of the gene with its constitutive exons in blue and skipped exon in red, nested lasli in yellow, below are the constructs that were cloned into the vector; (middle) agarose gels of the respective PCR products; (bottom) the sequence alignment of the band indicated with* ***1****;*

Discussion

Lariat PCR confirmed the nested lariats predicted by bioinformatics

The aim of my master thesis was to perform an in depth verification of non-canonical splicing events predicted in the human transcriptome. A precise analysis in search for rare and unusual splicing events, like nested and recursive splicing, predicted these events to be quite frequent. Splicing of long introns was thought to occur in pieces occasionally but this was not known for short introns. In contrast the bioinformatic analysis of existing NGS data from human transcriptomes suggested these events to occur quite frequently even in short introns. My task was to verify these splicing events experimentally.

With our work we were able to show that the splicing procedure in human cell line is not always as simple as previously thought, especially that not every intron gets excised in a single piece. By using lariat PCR, we could confirm these splicing-in-pieces events and we therefore suggest that splicing in human cells is much more dynamic than previously assumed.

Given the short lariat life-time, defined by the usually swift degradation, the amount of excited introns in the cell is of very low abundance and it is therefore genuinely tricky to detect rare splicing events. In order to increase the stability of excised intron lariat we knocked down the debranching enzyme dbr1. This did increase the amount of excised introns in the cell but unfortunately also altered the splicing pattern in many of the studied cases. The lariats in dbr1 knockdown cells are expected to be more stable, therefore accumulating to a detectable quantity. Even though it did not always enrich the lariats above detection limits as expected. One also has to take into the account that the lariats were detected in cell lines, which have a very specific transcriptome and the laslis were acquired from various datasets, which included cells and tissues we were not investigating *in vivo*. The altered splicing pattern in the dbr1 knockdown can be explained by stabilized lariats serving as sponges for splicing factors, thus depriving their availability in the cell and making the recruitment onto other introns less efficient.

The repetitive sequences and the overall nucleotide composition in introns often make them almost unsuitable for primer design, especially for a nested PCR

where two forward and two reverse primer have to be designed for a single lariat. Still we managed to confirm plenty of intrasplicing events, although not always in the exactly predicted positions. Nested splicing events seem to have less strictly defined splice sites, which is to some extent logical for non-recursive intrasplicing since the constitutive splice sites are supposed to be *stronger* than the nested ones in order to give rise to a properly spliced transcript. Splice site deviation in intrasplicing events does also not impact on the reading frame of the fully spliced transcript.

Particularly interesting is the MFN2 intron tested which had the exon, adjacent to the nested intron in one of the detected lariats once the debranching enzyme was knocked down. A logical consequence is to try to find a link between intrasplicing and exon skipping which we addressed in our exon skipping experiments. A simple explanation would be that the lariat, which included the exon is being a missplicing event stabilized or induced by the knock down.

In SQSTM1 we detected 6 out of the predicted 12 nested lariats only in HeLa, our prediction is that in a more extensive approach more of the putative nested lariats would be picked up. As we also identified the full length lariat, an assumption can be made that the nested lariats are not obligatory in the removal of the whole intron, at least not in a recursive manner, which raises the question of their functionality. The fine-tuning of the expression might be one of them, still the possibility of missplicing cannot be excluded.

From the lariat PCR data one cannot make many conclusions regarding the function of intrasplicing, but there are certain indications, such as positioning or absence/presence of recursive splice site motifs, that are indicative in which directions the functions may be. The exon inclusion in some of the nested lariats is pointing towards exons skipping, although this could be considered a dbr1 knockdown artefact. The LASP1 intron 2 circular split and nested lasli could be part of a recursive splicing scheme.

In the lariat PCRs we were identifying the splice sites and the branch points but we also wanted to make statements about the size of the respective lariats and the temporal order of the recursive splicing events. Cases such as SQSTM1 also raised the question of how many of the observed splicing events actually occur. By determining the size (range) of the full length lariat, one could estimate how many preceding intrasplicing events have occurred before full intron removal. The experimental approach to tackle this question was virtual northern blotting. The idea was to get the intermediates in the different fractions, but the tactic turned

out to be unsuccessful due to insufficient RNA yields and presumably bad separation of circular RNA molecules. The challenge could be solved by overexpressing the genes of interest, stabilizing their lariats and separating them in a way that allows for clear measurement of circular RNA sizes.

Nested introns can influence full intron splicing efficiencies

To address a potential function of intra-splicing events we designed and developed a reporter system into which an intron, for which intrasplicing events had been demonstrated, was inserted into the renilla luciferase gene. In addition mutation of intrasplicing sites and pre-splicing constructs allowed the analyses of their impact on full intron spicing.

The renilla luciferase assay showed that the nested laslis can influence gene expression in multiple introns that we have tested. Splicing of the nested lariats of GNB2L 1 intron 7 and PRKAB2 intron 7 decreases the luciferase luminescence. Mutations in the nested splice sites lead to abolishing of these splicing events, resulting in a higher luciferase luminescence. A simple explanation for this occurrence is that the nested splice sites compete with the constitutive splice site selection and, if winning this competition, generate a partially retained intron. Mutants increase the luciferase abundance because this competition does not occur (**Figure 29**). We also could show that the PRKAB2 intron 7 nested lasli (**p7 nest**) can be spliced out and produce a functional luciferase, but the quantity is lower compared to a full length intron cloned in the reporter. Still the fact that the nested lasli gets spliced outside of its native environment is adding weight to our hypothesis. The prespliced versions, lacking the nested lariat could not produce a functional full intron splicing event. The consequence of such a splicing event is a partially retained intron that is incompetent for splicing as necessary sequence elements such as the pY tract, BP and 3' ss are removed in the preceding, nested splicing step. The transcript might have the intron retained or the splice sites might be selected in a way that they introduce an exon shift in the final protein. Of course the transcript might not even get the chance to be translated if it contains premature stop codons and gets turned over by the nonsense-mediated decay.

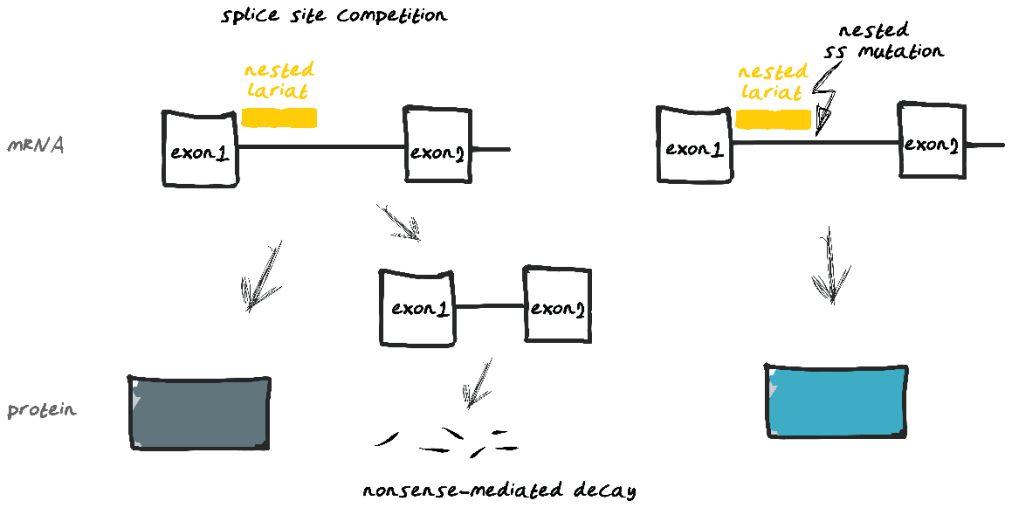


***Figure 29 model of the influence of the nested splicing event on protein expression**; the splicing factors compete for the constitutional and nested splice sites, if the spliceosome gets assembled on the nested splice site and removes it, the rest of the transcript is being detected by the nonsense-mediated decay components and gets degraded; once the nested splice site is blocked the spliceosome assembles properly increasing protein abundance;*

The nested laslis from rbm17 intron 11 have a stimulating effect on the whole intron splicing. The luciferase abundance in the mutants **rbm 213** and **rbm 514** is decreased compared to the wild type intron favours this hypothesis (**Figure 23**). The observed partial intron retention in mutant rbm 213 is also very intriguing. The nested lasli at the 5' of the intron seems to play an important role in the splicing of the whole intron, as the removal of its pY tract partially abolishes splicing of that part and leads to an alternative 5' ss selection at the adjacent downstream nested lasli. The retained part does not completely overlap with the upstream nested lasli because there is an additional splice site, which outcompetes the predicted recursive site between lasli1 and lasli2 (**Figure 24**). This alternative 5' ss of the nested downstream lasli might only be active if the predicted upstream 3' ss nested lasli is not used. The precise mechanisms remain to be elucidated, but so far, there is a good probability for recursive splicing happening in a relatively short intron, but further investigations would be necessary. This case also allows for an estimate of the importance of intrasplicing events for full intron removal: according to the splicing assay almost one third of transcripts retain the 5' part of the intron, leading to the assumption that in one third of cases, the 5' nested splicing event is preceding full intron removal. This is a rate that implies regulatory implications for this splicing step. Yet, it is particularly delicate to get a temporal resolution of the whole event happening due the low abundance of the products of nested splicing events and the intermediate RNA species.

We also found that some of the nested splicing events seem to not influence the transcript abundance, or at least the translated product. To quantify the frequency of a splicing event happening would be very interesting to do. The recently published tool SplicePie, in combination with targeted RNAseq allows for the detection of cryptic and recursive splicing events as well as estimating their sequentiality[47].This approach could be used to deepen insights into the mechanism and impact of the splicing events that we tested here. We are also working on a double and triple mutant to test if the inhibitory effects of the mutations in nested laslis are cumulative.

In summary, we can state that intrasplicing events have diverse effects in full intron splicing: they can be neutral, be detrimental or in contrary promote full intron splicing. Therefore, we can clearly state that splicing efficiency can be regulated by intrasplicing events.

Intrasplicing and exon skipping

The fact that some of the tested splicing events resulted in splicing incompetent, retained introns which could only be rescued by subsequent alternative splicing, it was a logical step to test if there is some correlation between exon skipping and nested lariats. The constructs were cloned after a strong promoter to ensure sufficient expression and detection of lowly expressed transcripts.

The splicing pattern of NME6 skipped exon 5 indeed differs once the nested splice site is mutated. Although a PCR provides only semi quantitative insights, a difference in the abundance of the variants containing exons 3 and 6 is very obvious (**Figure 26).** The product **containing exons 3, 4, 5 and 6** disappears from the mutant, but a closer look reveals that exon 5 is included in both products **2** and **3** (**Figure 26**), just with an alternative 5' splice site in the wild type version which overlaps with our lasli, thus making the lasli not nested. Even though this does not confirm our hypothesis on the influence of nested splicing on exon skipping, we still uncovered a novel case of alternative splice site mediated isoform definition.

In DCTN3 a slight alteration in the splicing patter was also observed. We were only able to detect one splicing variant in both cell lines that we tested. Although for the most part all three exons were included in the spliced transcript

(**Figure 27**), alternative transcripts were detected as well. Unexpectedly, a novel, cryptic exon was detected, that uses the tested lasli to be spliced into a minor isoform. SACML1 exon 5 is a good candidate for the test since it is adjacent to a nested lasli upstream of it. Still in the cell lines we tested exon skipping we were not able to see any alterations in the splicing pattern.

One should keep in mind that the proteomes of cells differ and the amount of proteins that can be involved in the splicing reaction, thus defining alternative splicing patterns. The issue with SACLML1 could be that HeLa and HEK 293 cells are simply not expressing the factors necessary for exon 5 to be skipped, so additional experiments are required to make further conclusions. Another explanation could be that the transcript lacking exon 5 is low abundant and therefor it could not be picked up by the PCR.

In DCTN3 the nested lasli turning out to be an unannotated constitutional lasli shows how difficult it is to make a comprehensive annotation of a genome and that a highly expressed vector and specific conditions of a cell line were necessary to get a transcript with still a relatively low abundance compared to the dominant transcript.

Final conclusions

To sum up this study: we have managed to experimentally verify the laslis suggested from the bioinformatic data. Most of our tested nested lariats were detected in the approach, although a tendency towards alternative splicing and branch point sites is observable, indicating that the strength of these splice site is not as strong as the constitutive splice sites. Testing the effect of nested laslis on the expression levels of a luciferase we found that they indeed can stimulate or inhibit their expression. Although our exon skipping experiments had a good foundation to be involved in the intrasplicing story, the genes tested were found to be more complex than annotation implied. This makes selection of suitable candidates difficult. The now available data on novel splicing events could help to improve available annotations and allow for a more considerate experimental design. Additional experiments where the nested lariats are tested in an endogenous environment would be a good confirmation for our hypothesis.

From our results we can imply functionalities of intrasplicing:

-inhibiting as observed in GNB2L1 intron 7 and PRKAB2 intron 7, based on partial intron retention and potential degradation by NMD

-stimulating as observed in RBM17 intron 11, based on a presumptive recursive splicing pattern

Future outlook

We hypothesised that splicing of some nested lariats produces a non-functional transcript which is recognised by non-sense mediated (NMD) decay proteins and degraded. It has been shown that the NMD can be inhibited by cyclohexamide[48], yet the effect in vivo was inefficient to allow a clear detection of NMD dependent RNA species. A more sophisticated approach to investigate NMD effects on transcripts is to knock down one of its main factors, UPF1 in human cells[49]. The data from previously published genome wide sequencing studies on the UPF1 knock down shows a similar patter we observed in the luciferase assay for GNB2L1 intron 7.

Since our expression experiments are performed with a reporter system, the next step would be investigating the effect of nested splicing in the endogenous context. Previous studies showed that 2'O-methyl oligo RNAs can mask splice sites and abolish splicing *in vivo*[50,51]. In our first experiments we managed to transfect the fluorescently labelled oligo but the results were inconclusive. Additional optimization in antisense oligo design is required to efficiently mask splice sites.

The experiments performed in particular cell lines might not ensure a sufficient expression of the tested genes, so what one could do is test the expression levels of the desired genes and pick a cell line where the genes are highly expressed. Especially for the detection of minor isoforms and rare splicing events, high gene expression is favourable.

Since we are living in a world where a biotechnological revolution is happening, one cannot avoid thinking about the application of the sequence specific genome editing tool CRISPR/Cas9 in order to test the functionality of the nested lariats on gene expression. A simple idea is to mutate the nested splice sites and compare the endogenous expression levels of the mutated versus wild-type transcript.

References

1. Hesse, H. *Der Steppenwolf*. (S. Fischer Verlag, 1927).
2. Murakami, H. *South of the Border, West of the Sun*. (Kodansha, 1992).
3. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5 ' terminus of adenovirus 2 late mRNA *. *PNAS* **74,** 3171–3175 (1977).
4. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–51 (2001).
5. Gelfman, S. *et al.* Changes in exon-intron structure during vertebrate evolution affect the splicing pattern of exons. *Genome Res.* **22,** 35–50 (2012).
6. Guo, L. & Liu, C.-M. A single-nucleotide exon found in Arabidopsis. *Sci. Rep.* **5,** 18087 (2015).
7. Deutsch, M. & Long, M. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res.* **27,** 3219–3228 (1999).
8. Reinhold-Hurek, B. & Shub, D. a. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357,** 173–176 (1992).
9. Pennisi, E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* **337,** 1159, 1161 (2012).
10. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40,** 1413–5 (2008).
11. Luo, M. J. & Reed, R. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14937–14942 (1999).
12. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407,** 401–405 (2000).
13. Lutz, C. S. *et al.* Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev.* **10,** 325–337 (1996).
14. Evsyukova, I., Bradrick, S. S., Gregory, S. G. & Garcia-Blanco, M. A. Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. *RNA* **19,** 103–115 (2013).
15. Martinson, H. G. An active role for splicing in 3′-end formation. *Wiley Interdiscip. Rev. RNA* **2,** 459–470 (2011).
16. Hu, T. T. *et al.* The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43,** 476–81 (2011).
17. Yang, Y. F., Zhu, T. & Niu, D. K. Association of intron loss with high mutation rate in arabidopsis: Implications for genome size evolution. *Genome Biol. Evol.* **5,** 723–733 (2013).

18. Gregory, T. R. *et al.* Eukaryotic genome size databases. *Nucleic Acids Res.* **35,** (2007).

19. Hattori, M. Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso.* **50,** 162–168 (2005).

20. Roy, S. W. & Gilbert, W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* **7,** 211–21 (2006).

21. Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 1342–6 (1974).

22. Kozak, M. The scanning model for translation: An update. *J. Cell Biol.* **108,** 229–241 (1989).

23. Turunen, J. J., Niemelä, E. H., Verma, B. & Frilander, M. J. The significant other: Splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA* **4,** 61–76 (2013).

24. Papasaikas, P. & Valcárcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences* **41,** 33–45 (2016).

25. Wahl, M. C. & Lührmann, R. SnapShot: Spliceosome Dynamics I. *Cell* **161,** 1474–1474.e1 (2015).

26. Wahl, M. C. & Lührmann, R. SnapShot: Spliceosome Dynamics II. *Cell* **161,** 1474–1474.e1 (2015).

27. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136,** 701–718 (2009).

28. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15,** 108–21 (2014).

29. Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514,** 1–30 (2013).

30. Schmucker, D. *et al.* Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101,** 671–684 (2000).

31. Celotto, A. M. & Graveley, B. R. Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated. *Genetics* **159,** 599–608 (2001).

32. Lang, K. M. & Spritz, R. A. RNA splice site selection: Evidence for a 5'->3' scanning model. *Science (80-.).* **220,** 1351–1355 (1983).

33. Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10,** 84–94 (1990).

34. De Conti, L., Baralle, M. & Buratti, E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA* **4,** 49–60 (2013).

35. Hatton, A. R., Subramaniam, V. & Lopez, A. J. Generation of Alternative Ultrabithorax Isoforms and Stepwise Removal of a Large Intron by Resplicing at Exon–Exon Junctions. *Mol. Cell* **2,** 787–796 (1998).

36. Sibley, C. R. *et al.* Recursive splicing in long vertebrate genes. *Nature* **521,** 371–5 (2015).

37. Kelly, S. *et al.* Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res.* **43,** 4721–4732 (2015).

38. Ott, S., Tamada, Y., Bannai, H., Nakai, K. & Miyano, S. Intrasplicing--analysis of long intron sequences. in *Proc. of the Pacific Symposium on Biocomputing 2003 (PSB 2003)* 339–350 (2003).

39. Suzuki, H., Kameyama, T., Ohe, K., Tsukahara, T. & Mayeda, A. Nested introns in an intron: Evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA. *FEBS Lett.* **587,** 555–561 (2013).

40. Vogel, J., Hess, W. R. & Börner, T. Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res.* **25,** 2030–2031 (1997).

41. Taggart, A. J., DeSimone, A. M., Shih, J. S., Filloux, M. E. & Fairbrother, W. G. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat. Struct. Mol. Biol.* **19,** 719–21 (2012).

42. Coombes, C. E. & Boeke, J. D. An evaluation of detection methods for large lariat RNAs. *RNA* **11,** 323–31 (2005).

43. Andersen, C. R., Nielsen, L. S., Baer, A., Tolstrup, A. B. & Weilguny, D. Efficient expression from one CMV enhancer controlling two core promoters. *Mol. Biotechnol.* **48,** 128–137 (2011).

44. Xia, W. *et al.* High levels of protein expression using different mammalian CMV promoters in several cell lines. *Protein Expr. Purif.* **45,** 115–124 (2006).

45. Coolidge, C. J., Seely, R. J. & Patton, J. G. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25,** 888–895 (1997).

46. Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3,** 2113–2123 (1989).

47. Pulyakhina, I. *et al.* SplicePie: A novel analytical approach for the detection of alternative, non-sequential and recursive splicing. *Nucleic Acids Res.* **43,** (2015).

48. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446,** 926–929 (2007).

49. McGlincy, N. J. *et al.* Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics* **11,** 565 (2010).

50. Ullu, E. & Tschudi, C. 2'-O-methyl RNA oligonucleotides identify two functional elements in the trypanosome spliced leader ribonucleoprotein particle. *J. Biol. Chem.* **268,** 13068–13073 (1993).

51. Okabe, K. *et al.* Real time monitoring of endogenous cytoplasmic mRNA using linear antisense 2'-O-methyl RNA probes in living cells. *Nucleic Acids Res.* **39,** (2011).

Abstract

Ever since the unexpected discovery of introns, our understanding of the splicing process has been constantly growing. The current model suggests that the spliceosome recognizes the branch point, 5' and 3' splice sites and removes the intron in one piece. We are trying to challenge this model by saying that introns, at least some of them, show a much more dynamic splice site selection, resulting in a number of transcripts, not all of which are intended for export and translation.

Currently there are two unconventional splicing models, namely nested and recursive splicing, which we summarized as intrasplicing. Recent publications focused on intrasplicing in long introns, but our data suggests that short introns might be spliced in this way as well. Our analysis identified more than 12.000 distinct intrasplicing events happening in human and this thesis aims to verify them experimentally in different cell lines which can be challenging due the low stability and abundance of lariats. We performed a knock down of the lariat debranching enzyme dbr1 to increase the stability and thus detectability of lariats via lariat PCR. We also explored the potential functionality of intrasplicing in a luciferase reporter system. We found that nested splicing can stimulate or inhibit the removal of the whole intron, thus influencing the transcript abundance. Intrasplicing events increase the transcript abundance as a part of a potentially recursive splicing scheme but can have the opposite effect and generate NMD targeted transcripts, if the splice sites are not reconstituted after the removal of the nested intron.

Zusammenfassung

Seitdem die Introns unerwarteter weise entdeckt worden sind, ist unser Verständnis über den Spleißgprozess ständig gewachsen. Das jetzige Model suggeriert dass das Spleißeosom den sogenannten branch point, die 5‘ und 3‘ Spleißstelle erkennt und das Intron in einem Stück rausschneidet. Wir versuchen dieses Model zu hinterfragen und behaupten dass Introns, zumindest einige, eine viel dynamischere Spleißstellen Selektion zeigen, zahllose Transkripte produzieren, wobei nicht alle für den Export und die Translation bestimmt sind.

Zurzeit gibt es zwei unkonventionelle Spleißmodele, nämlich das genestete und rekursive Spleißen, die wir beide als Intrasplicing bezeichnen. Jüngste Publikationen fokussieren sich an das Intrasplicing in langen Introns, doch unsere Daten suggerieren dass kurze Introns auch auf diese Weise gespleißt werden können. Die Analyse identifizierte mehr als 12 000 einzelne Intrasplicing Vorkommnisse im Menschen und diese Arbeit befasst sich mit der experimentellen Verifizierung solcher Vorkommnisse in unterschiedlichen Zelllinien, dass sich wegen der geringen Stabilität und der Abundanz der Lariate als herausfordernd rausstellte. Wir haben das Debranching Enzym dbr1 runterreguliert um die Stabilität und damit die Detektion der Lariate mittels der lariat PCR zu erhöhen. Wir haben auch die potentiale Funktionalität des Intrasplicing in einem Luziferase Reportersystem untersucht. Wir haben herausgefunden das Intrasplicing das rausschneiden des ganzen Introns beeinflussen kann und somit auch einen Einfluss auf die Transkriptmenge hat. Intrasplicing erhöht die Transkriptmenge als Teil einer rekursiver Spleißreaktion, kann aber auch den gegenteiligen Effekt haben und Transkripte produzieren die vom NMD aufgesucht werden wenn die Spleißstellen nach dem Intrasplicing nicht wieder hergestellt werden.