The snoRNAs MBII-52 and MBII-85 are processed into smaller RNAs and regulate alternative splicing

Der Naturwissenschaftlichen Fakultät

der Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades

vorgelegt von

Amit Khanna

aus Chandigarh, Indien

2010

Als Dissertation genehmigt von der Naturwissenschaftlichen Fakultät der Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 17.06.2010

Vorsitzender der Promotionskommission: Prof. Dr. Eberhard Bänsch

Erstberichterstatter: Prof. Dr. Andreas Burkovski

Zweitberichterstatter: Prof. Dr. Stefan Stamm

A bud that bulges and blossoms to be a life,

In minutes he has the tenderness and then opts to ripe,

With sway he floats yet holds its roots,

Holding as if saying he will never be too bold,

With days of winds and dust he saps,

Still holding to belief, he will never zap,

Who had the eternity that he shall too?

With every breath he dies but he won't just let go….

He knew he desired more than he deserved,

But eternity was still never heard,

He sang and itched his desire per say,

In all the divine courts even those just hear say,

The enigma lingered with him for long,

Resting was never marked as he wished a prolong,

The day came when he was to go,

He left waning still said number of times the NO,

It never occurred to him he can stay without being toed,

As he lived on immortal with the papers he toned…

ACKNOWLEDGMENTS

"though I owe my life to the soil,

but my sweat appears pale in toil,

my effort though fully enforced,

I labeled myself a learning fawn"

These lines somewhat sum up the feelings and effort that went behind this work. It has been a big learning phase in my life and has given a kind of preparedness for the days to come.

My sincerest gratitude goes to Prof. Stefan Stamm for his confidence in me all these years and in helping me work with an open mind. I owe him most for his patience and understanding.

I would like to thank present and former members of the lab, Zhaiyi Zhang, Manli Shen, Olga Kelemen, David Oliphant, Domnique, Shivendra Kishore, Chiranthani Sumanasekera, Yesheng Tang, Bettina Heinrich, Jingyi Hui and Helen Hager for their support in the lab.

Special thanks to Natalya Benderska for helping me with the paper work at Friedrich-Alexander-Universität Erlangen-Nürnberg.

I would like to thank Prof. Sydney White heart and his lab at University of Kentucky for sharing with us their animal facility and expertise. I would also like to thank Prof. Louis Hersh chairman department of Molecular and Cellular biochemistry for his kind support. I have special thanks for Ms Diana Griffith who helped us with the ordering all through my stay at University of Kentucky.

I appreciate my friends Zubair Karim, Vikas Dukhande, Rohan Kaushal, Nishant Visavadiya, Paras Anand and Aashish Joshi who gave me best of the social support and made my time memorable all these years.

I can never cease to express my gratitude for my family, my parents for their continuous unconditional love and lots of blessings and my sister, brother in law and sweet niece for their love and support. I would also like to thank my parents' in-laws and sisters in-law for their support, concern and care they showed in course of my work.

I reserve a special thanks to my loving wife Aruna for her unconditional love and support all through.

Publications from this work:

Kishore S*, **Khanna A*,** Zhang Z, Hui J, Balwierz PJ, Stefan M, Beach C, Nicholls RD, Zavolan M, Stamm S.The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet.* 2010.

   * Co-first authors

Khanna A and Stamm S Regulation of alternative splicing by short non-coding nuclear
   RNAs. *RNA Biology* 2010.

Kishore S, **Khanna A**, Stamm S. Rapid generation of splicing reporters with pSpliceExpress. *Gene.* 2008.

Hui J, Kishore S, **Khanna A** and Stamm S. Analysis of Alternative splicing with Microarray. *Bioinformatics for Systems Biology* 2008.

2009: Thomas C. Vanaman Graduate student best talk award

2009: As an instructor in EURASNET's workshop on alternative splicing and disease

2008: Poster presentation at IDEA meeting.

Table of Contents

Abbreviations

5-HT$_{2C}$R	serotonin receptor 2C
AMP	adenosine mono phosphate
ASF	alternative splicing factor
ATP	adenosine 5'-triphosphate
bp	base pairs
BP	break point
CBs	Cajal bodies
cDNA	complementary DNA
CMV	cytomegalovirus
dH$_2$O	distilled water
DMEM	dulbeco's modified eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotidtriphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
ESE	exonic splicing enhancer
EST	expressed sequence tag
FCS	fetal calf serum
G-Proteins	guanosine triphosphate binding protein
GPCR	G-Protein coupled receptor
HBII-52	human brain specific snoRNA II-52
HEK	human embryonic kidney
HIV	Human immunodeficiency virus
hnRNP	heterogenous nuclear ribonucleoprotein
kDa	kilodalton
MBII-52	mouse brain specific snoRNA II-52
mRNA	messenger RNA
miRNA	microRNA
N2a cells	Neuro 2a cells
ncRNA	non-coding RNA
Oligos	Oligonucleotides
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFC	prefrontal cortex
PMSF	phenylmethanesulfonyl fluoride

psnoRNA	Processed snoRNA
pSE	pSplice Express
PTC	premature termination codon
PWS	Prader-Willi Syndrome
RBII-52	rat brain specific snoRNA II-52
RBM	RNA binding motif
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RRM	ribonucleic acid recognition motif
rRNA	ribosomal RNA
RT-PCR	reverse transcription followed by polymerase chain reaction
RUST	regulated unproductive splicing and translation
SDS	sodium dodecyl sulfate
SF	splicing factor (1 or 2)
siRNA	small interfering RNA
SmB	small nuclear ribonucleoprotein associated protein B
SMN	survival motor neuron gene (1 or 2)
SmN	small nuclear ribonucleoprotein polypeptide N
snmRNA	small non-messenger RNA
snoRNA	small nucleolar ribonucleic acid
snoRNP	small nucleolar ribonucleoprotein
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein particle
SNRPN	small nuclear ribonucleoprotein polypeptide N
SNURF	SNRPN upstream reading frame
SR-protein	serine-arginine- rich protein
SSRI	selective serotonin reuptake inhibitor
stRNA	small temporal RNA
TBE	tris-borate-EDTA buffer
TE	tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
tRNA	transfer RNA
UPD	uniparental disomy
UTR	untranslated region

Zusammenfassung

Die Entwicklung neuer Techniken, wie deep-sequencing und tiling DNA arrays führte zur Entdeckung einer Vielzahl kurzer, metabolisch stabiler RNAs, die keinerlei Leseraster für Proteine enthalten. Einige dieser RNAs entstehen aus bekannten RNA Klassen, wie snoRNAs, (small nucleolar RNAs) und tRNAs (transfer RNAs). Diese neuen RNAs weisen ähnliche Eigenschaften wie Oligonukleotide auf, die schon seit langer Zeit zur Veränderung der prä-mRNA Prozessierung verwendet werden.

Der Verlust der Expression der snoRNAs HBII-52 und HBII-85 ist die wahrscheinliche Ursache für das Prader-Willi Syndrom. In früheren Arbeiten konnte die Gruppe zeigen, dass die snoRNA HBII-52 die alternative prä-mRNA Prozessierung des Serotonin Rezeptors 5 HT2C reguliert. Dies ist eine neue Funktion von snoRNAs, von denen bisher nur bekannt war, das sie die Methylierung und Pseudouridinylierung von nicht-mRNAs regulieren.

In dieser Arbeit benutzten wir bioinformatische Vorhersagen und experimentelle Untersuchungen um neue Zielgene für HBII-52 und HBII-85 zu identifizieren. Wir fanden fünf mRNA, deren alternative Exons von HBII-52 reguliert wurden (DPM2, TAF1, RALGPS1, PBRM1 und CRHR1). Für HBII-85 identifizierten wir zwei Zielgene: PTBP1 und HAGHL.

Um den Wirkungsmechanismus der snoRNAs aufzuklären, analysierten wir die Expression von MBII-85 und MBII-52 mit RNAse protection. Hierbei stellte sich heraus, dass aus den snoRNAs kürzere RNAs durch einen zusätzlichen Prozessierungsschritt entstehen. Die neuen RNAs sind nicht mit snoRNA-

typischen Proteinen verbunden, sondern binden an hnRNPs, ähnlich wie RNA Oligonucleotide.

Um diese RNAs zu klonieren, entwickelten wir ein neues Verfahren, um geringe Mengen doppelsträngiger RNA (dsRNA) aus RNAse protection Experimenten zu klonieren. Hiermit konnten wir zeigen, dass die am stärksten exprimierten Formen von HBII-52 und HBII-85 keine für snoRNA typischen komplementären Enden aufweisen.

Unsere Daten zeigen, dass nicht der Verlust typischer snoRNAs die Ursache für das Prader-Willi Syndrom ist, sondern der Verlust kürzerer, prozessierter snoRNAs, die wir psnoRNAs nennen.

Die Substitution von psnoRNAs könnte ein therapeutischer Ansatz für PWS sein.

Abstract

Recent results from deep-sequencing and tiling array studies indicated the existence of a large number of short, metabolically stable, non-coding RNAs. Some of these short RNAs are derived from known RNA classes like snoRNA or tRNAs. There are intriguing similarities between short non-coding nuclear RNAs and oligonucleotides used to change alternative splicing events, usually targeting a disease-relevant RNA.

The loss of HBII-52 and HBII-85 related C/D box small nucleolar RNA (snoRNA) expression units have been implicated as a cause for the Prader-Willi syndrome (PWS). We recently found that the C/D box snoRNA HBII-52 changes the alternative splicing of the serotonin receptor 2C pre-mRNA, which is different from the traditional C/D box snoRNA function in non-mRNA methylation. Using bioinformatic predictions and experimental verification, we identified five pre-mRNAs (DPM2, TAF1, RALGPS1, PBRM1 and CRHR1) containing alternative exons that are regulated by MBII-52, and two pre-mRNA (PTBP1 and HAGHL) containing alternative exons that are regulated by MBII-85, the mouse homologs of HBII-52 and HBII-85, respectively.

Analysis of a single copy of MBII-52 and MBII-85 from their respective clusters, by RNase protection and northern blot analysis shows that these expression units generates shorter RNAs that originate from the full-length snoRNA through additional processing steps. These novel RNAs associate with hnRNPs and not with proteins associated with canonical C/D box snoRNAs. Our data indicate that not a traditional C/D box snoRNAs MBII-52 and MBII-85, but a processed version lacking the snoRNA stem are the predominant MBII-52 and MBII-85 RNAs missing in PWS. These processed snoRNAs function in alternative splice-site selection. Their substitution could be a therapeutic principle for PWS.

We developed a new and reliable method to clone small dsRNAs from the RNase protection assay. The advantage of this method over other published methods that cloning can be done even with small amount of starting material. We also developed a faster method to clone minigenes for cell based minigene assays.

1   Introduction

Molecular biology can best be defined by the concept of "The Central Dogma" that describes the flow of genetic information from DNA that gets transformed into individual transportable cassettes composed of messenger RNA or mRNA (transcription) and finally each mRNA decoded as protein(s) (translation). It is now understood and accepted that not all transcribed RNA code for proteins. Despite the fact that a large fraction of human genome is transcribed, only 1.2% of it codes for protein. More than 98% of the transcriptional message in humans and other mammals consist of non-protein-coding RNAs (ncRNAs) mostly from the introns of protein coding genes and the exons and introns of non-protein-coding genes (Mattick, J.S. and Makunin, I.V., 2005).

Ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA), short interfering RNA (siRNA), small nuclear RNA (snRNA), piwiRNA and small nucleolar RNA (snoRNA) constitute the non-coding RNAs (ncRNA) populations. The biological significance of most of the ncRNAs is still not completely understood. The finding that small nucleolar RNAs (snoRNAs) not only guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs, but also target mRNA and regulate splicing has given a new insight to understand how these small ncRNA can have multiple roles (Kishore, S. and Stamm, S., 2006).

This chapter introduces the concepts and the current understanding about pre-mRNA splicing, snoRNAs, types of snoRNAs and their associated proteins. A review about abundant expression of small non-coding RNAs, which was revealed by deep sequencing and array analysis, is being discussed. The role of small oligonucleotides in alternative splicing and pre-mRNA splicing is also discussed in brief. Functional relevance and role of snoRNA in development of Prader-Willi syndrome and as a regulator of alternative splicing of target mRNAs is also discussed that makes the basis of work presented later.

1.1 Pre-mRNA Splicing

Protein coding genes in higher eukaryotes have coding regions known as exons and are interrupted by non-coding sequences known as introns. These introns are spliced out to yield mature mRNA.

Introns are marked by conserved sequence elements which allow intron recognition that is sliced and removed. The conserved sequences are: 5' splice site, branch point and 3' splice site that is preceded by a polypyrimidine stretch (Table 1).

Table1. Sequence features marking major class (GT-AG) intron

Feature	Consensus sequence
5' donor splice site	YRG/GURAGU
3' acceptor splice site preceded by a polypyrimidine stretch	Y_{12}NYAG /G
Branch point located 18-200nt upstream of the 3' splice site	YNYURAY

*Symbols used: Y-Pyrimidine, R-Purine, N-any nucleotide, slash (/) denotes the exon-intron border, Invariant nucleotides are underlined.
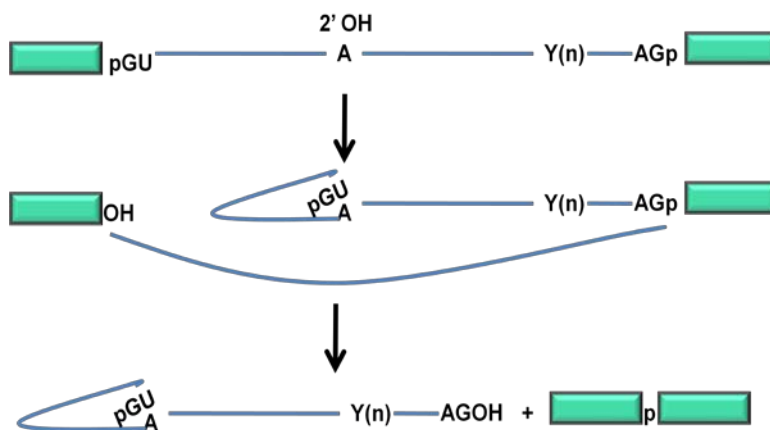
Figure 1: Pre-mRNA splicing reaction. Lariat formation and splicing out of exons

Dinucleotides at the 5' and 3' termini of the introns (GT and AG respectively) are highly conserved for the major class intron i.e. the U2 type. Metazoans can also use another distinct minor intron class (U12 type) that has AT and AC termini. The annotated mammalian genes (Genbank), a bioinformatics analysis of splice junction pairs showed

that 98.71% of splice site conformed to canonical GT-AG, 0.56% to non-canonical GC-AG and 0.73% to other non-canonical splice termini (Burset, M. et al., 2001).

Splicing occurs through two sequential trans-esterification reactions involving a dynamic multi component complex called the spliceosome, assembled from five subcomplexes called small nuclear ribonucleoprotein particles (snRNPs U1, U2, U4, U5 and U6 in major class U2 type spliceosome; snRNPs U11, U12, U4, U5 and U6 in minor class U12 type spliceosome). Junctions between an intron and its flanking exons, namely the 5' and the 3' splice sites are precisely identified by the spliceosome on the pre-mRNA. The splicing reaction proceeds by an orchestrated formation and disruption of RNA-RNA, RNA-protein and protein-protein interactions, which leads to exon ligation and release of intron lariat. In the first step, a nucleophilic attack by the 2' hydroxyl group of a conserved adenosine residue at the branch point, cleaves the 5' exon-intron junction generating a free 3' hydroxyl group on the upstream exon as well as a branched lariat intermediate. In the second step, the 3' intron-exon junction is attacked by the 3' hydroxyl of the 5' exon, displacing a lariat intron and ligating the exons (Figure 1).

In major class introns, commitment of pre-mRNA to splicing pathway occurs upon formation of the E complex (see Figure 2). Assembly of the E complex involves the recognition of 5' splice site, the polypyrimidine tract and 3' splice site by U1 snRNA, U2 auxiliary factor 65 (U2AF65) and U2 auxiliary factor 35 (U2AF35) respectively. The branch point is recognized by the splicing factor1 (SF1). Several non-snRNP splicing factors such as serine/arginine rich (SR) proteins also associate to the pre-mRNA at this step. In addition, U4/U6*U5 tri-snRNP can associate with the first exon near the 5' splice site in the E-complex. This association is ATP dependent. Next, ATP dependent base pairing of U2 snRNP with the branch point leads to the formation of A complex. Next is the formation of complex B, involves major rearrangements of the snRNP components associated with the pre-mRNA. The U6/U4 duplex is disrupted and a new duplex between U6 and the 5' splice site is formed, displacing the U1 snRNP. The 5' splice site is brought close to the branch point and the 3' splice site through U6/U2 snRNA base pairing and interaction of U5 snRNP with both exons near splice sites. At this point, U4 snRNP leaves the complex and the first catalytic step of the splicing occurs, creating the intron lariat. Finally, U5 snRNP base pairs with both 5' and 3' exons, thus positioning the ends of the

3

two exons for the second step of splicing. After the second step has been completed, the ligated exons and the lariat intron are released and the spliceosomal components dissociate and are recycled for further rounds of splicing. Figure 2 schematically depicts the spliceosomal assembly, the formation of catalytic spliceosome and the excision of the intron from the pre-mRNA.
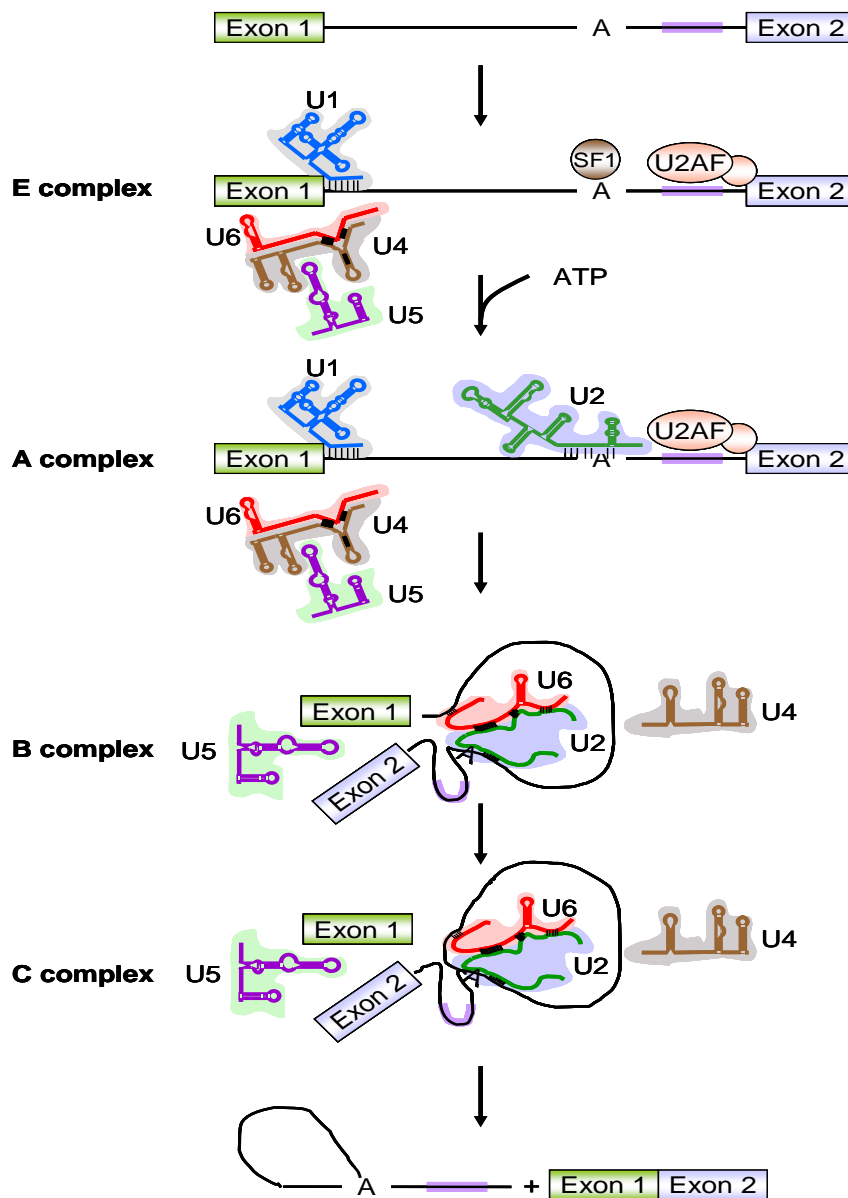
Figure 2: Spliceosome formation and rearrangement during the splicing reactions. (Figure adapted from Patel,A.A. et al., 2003).

1.2   SnoRNAs

Small   nucleolar   RNAs (snoRNAs)   are   defined   as   a   class   of small RNA molecules that guide chemical modifications of other RNAs which include ribosomal RNAs, transfer RNAs and small nuclear RNAs. There are two main classes of snoRNA, the C/D box snoRNAs and the H/ACA box snoRNAs. The C/D box snoRNAs target   methylation   of   the   target   RNA   and   H/ACA   snoRNAs   are   involved   in pseudouridylation of target RNAs. SnoRNAs act as guide RNAs.

In vertebrates, sequences encoding H/ACA and C/D box snoRNAs are present in introns of their host gene and are positioned in the same orientation. An intron can only host one snoRNA gene, leading to a host gene carrying several snoRNA genes in different introns.   It   has   been   reported   that   the   human   C/D   box   snoRNAs U3, U8, U13, mgU2-25/61 and mgU12-22/U4-8 are transcribed by RNA pol II as independent units. Intronic snoRNAs like MBII-52 (SNORD115) are produced by exonucleolytic degradation of the debranched lariat after splicing, the stable part being protected by the binding of snoRNPs. The genes hosting snoRNAs are either protein coding or non-coding, and often belong to the family of 5' TOP (5'-Terminal Oligonucleotide Pyrimidine tract) genes. Coding host genes include those of several ribosomal proteins or proteins associated with ribosome biosynthesis or translation. There are many characterized snoRNAs that reside in genes with unknown function. A growing number of snoRNAs are shown to have a tissue-specific expression, reflecting expression of the host gene.

1.2.1   Structure and function of methylation guide C/D box snoRNA

A C/D box snoRNAs features two short sequence motifs, box C (RUGAUGA) and box D (CUGA) located in the vicinity of their 5' and 3' termini respectively. The snoRNA is generally 60-100 nt long and contains two antisense boxes (Figure 3). Other features include a 5' and 3' stem loop structure which are upstream and downstream of box C and box D motifs respectively, comprising of 4-5 nucleotides that brings two (C/D) motifs together, which are important for snoRNA processing, stability, methylation activity, nucleolar localization and 5' cap hypermethylation (Filipowicz, W. and Pogacic, V., 2002). Apart from C/D motifs, these snoRNAs also contain a less conserved matching set of internal sequence element identical to box C (termed as box C') and box D (box D') motifs. An internal stem-loop is considered to be responsible for bringing C' and D' box

together which are found within 3-9 nucleotides of one another (Kiss-Laszlo, Z. et al., 1998). Upstream from D or D' boxes are one or two antisense elements, that target RNA with varying length of 10-21 nt, which bear complementarity to the rRNA modification site. These antisense elements make an extended duplex with the substrate and the box D or D' is located precisely five nucleotides from the residue which base-pairs to the rRNA nucleotide that is going to be modified (Kiss-Laszlo, Z. et al., 1996). It was experimentally demonstrated that the antisense element associated with the box D or D' is the sole determinant of the site of methylation as an artifical C/D box snoRNA carrying an appropriate antisense element that is sufficient to target a novel ribose methylation on the predicted pre-rRNA nucleotide and also, to a lesser extent, to RNA-polymerase II transcripts (Cavaille, J. et al., 1996).
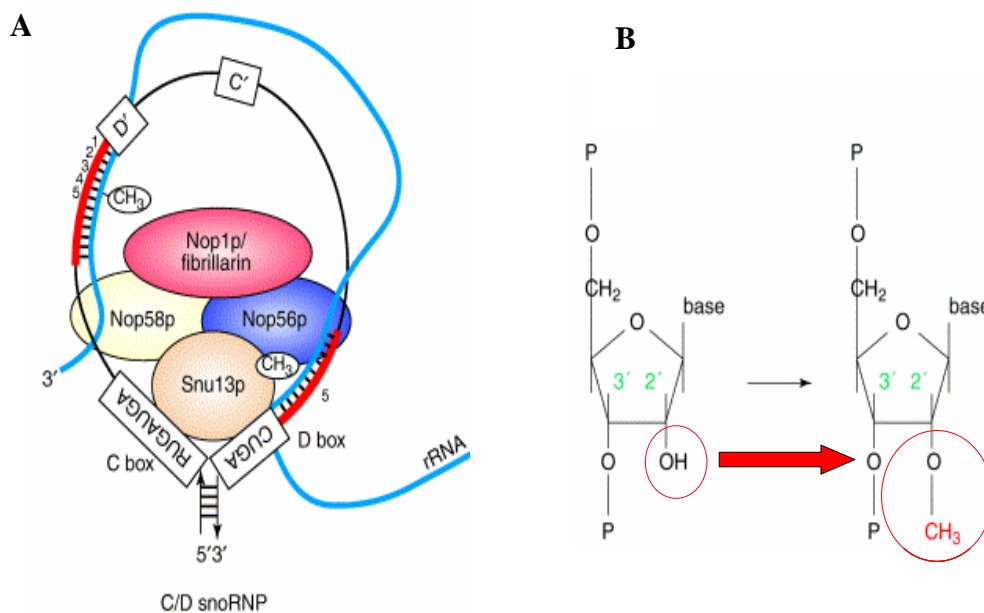
Figure 3: C/D box small nucleolar ribonucleoproteins (snoRNPs). (A) C/D box guide snoRNA and the core associated proteins. **(B)** C/D box snoRNPs guide 2'O-methyl modifications of their target nucleotide. (Figure modified from Brown, J.W.S. et al., 2003)

1.2.2 Structure and function of pseudouridylation guide H/ACA box snoRNA

The H/ACA snoRNAs are 120-140nt long, characterized by two conserved sequence elements H box (ANANNA) and the 3'-terminal ACA sequence (a trinucleotide always found three nucleotides away from the 3' end (Figure 4). They have a common

secondary structure comprising of hairpin-hinge-hairpin-tail structure (Ganot, P. et al., 1997), with the conserved sequences essential for snoRNA processing, stability, pseudouridylation activity and nucleolar localization (Terns, M.P. and Terns, R.M., 2002). The conserved motifs, box H and ACA are located in the hinge and tail region, respectively. Each H/ACA snoRNA contains an appropriate bipartite guide sequence in the internal loop of one or both of the two large hairpin domains. The two stems forming the 9-13 bp bipartite guide duplex precisely flank the substrate uridine, which remains accessible for isomerization. The target uridine and the downstream H or ACA box of the snoRNA are 14-16nt apart. In contrast to C/D box snoRNPs, electron micrographic studies of H/ACA snoRNPs reveal a symmetric bipartite particle with an estimated mass consistent with two copies of each protein (Watkins, N.J. et al., 1998)
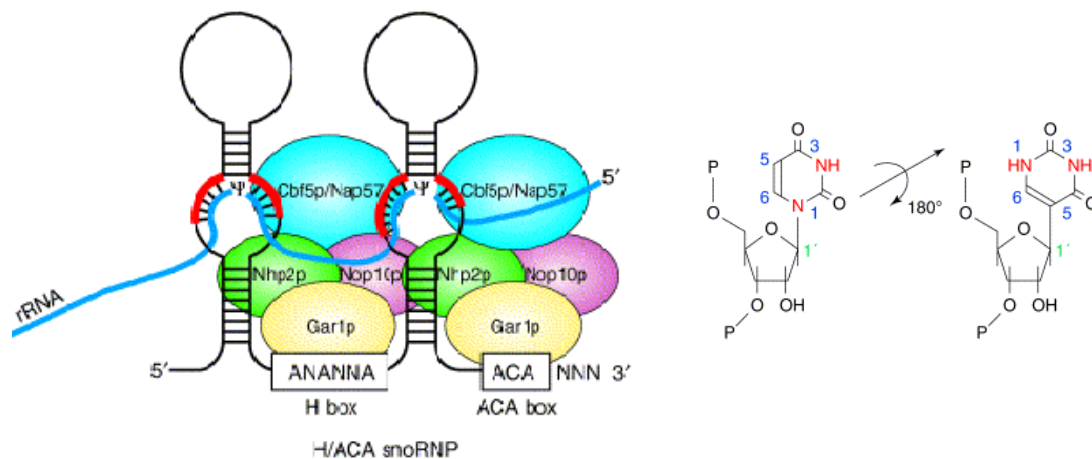
Figure 4: H/ACA box small nucleolar ribonucleoproteins (snoRNPs). (A) H/ACA box guide snoRNA and the core associated proteins. **(B)** H/ACA box snoRNPs guide pseudoridylation of their target Uridine (Figure adapted from Brown, J.W.S. et al., 2003).

1.2.3 Associated proteins

Both types of guide snoRNAs C/D box and H/ACA box, function as small ribonucleoprotein particles (snoRNPs), both consisting of a site-specific snoRNA associated with a small set of proteins common to each guide family.

Four evolutionary conserved core proteins make up the C/D box snoRNPs; fibrillarin (Yeast Nop1p) Nop56p, Nop58p and 15.5kDa protein (Yeast Snu13p). Fibrillarin, which exhibits amino acid sequence motifs characteristic of *S*-adenosyl-Met (SAM)-dependent methyltransferases, is the likely snoRNA-guided modifying enzyme, as point mutations in the methylase-like domain disrupt all rRNA methylations. The 15.5 kDa protein binds specifically to the C/D structural motif, termed as a K-turn. This protein which binds the same structural motif in U4 within the U4/U6.U5 tri-snRNP complex is also an integral component of this spliceosomal complex (Nottrott, S. et al., 1999). This raises the possibility that mRNA splicing and snoRNA synthesis may be linked (Watkins, N.J. et al., 2000).

The core proteins that make up H/ACA snoRNPs, include the proteins dyskerin (Yeast Cbf5p), Gar1p, Nhp2p and Nop10p which are all essential for the pseudouridylation reaction. Cbf5p is likely to correspond to the catalytic component of the H/ACA-snoRNA-guided modification, based on the presence of signature motifs for pseudouridine synthase and deleterious effects of point mutations in these motifs (Lafontaine, D.L. et al., 1998; Zebarjadian, Y. et al., 1999). Based on amino acid sequence homologies, Cbf5p appears closely related to TruB, which catalyzes the pseudouridine formation in the T loops of virtually all tRNAs.

Recently, a pair of highly conserved and related nucleoplasmic proteins was described. They have both DNA helicase activity and are linked with chromatin remodeling and transcription, p50 (Rvb2) and p55, were observed to bind to a model box C/D snoRNA *in vitro*, suggesting that they function at an early stage of snoRNP biogenesis in the nucleoplasm (Newman, D.R. et al., 2000).

1.3 Deep sequencing and array analysis revealed abundant expression of small non-coding RNAs

Deep sequencing of RNAs from human, mouse, chicken, *Drosophila*, *Arabidopsis* and *Schizosaccharomyces pombe* revealed the existence of shorter RNAs that were derived from known H/ACA snoRNAs (Taft, R.J. et al., 2009; Cole, C. et al., 2009). The shorter RNAs are predominantly 20-24 nts in length. These studies were supported by bioinformatic analyses which showed that numerous H/ACA snoRNAs act as precursors

for experimentally confirmed miRNAs (Scott, M.S. et al., 2009). The abundant expression of shorter snoRNAs raised the question of whether they are just degradation products or have a function. A functional role was suggested by the association of H/ACA snoRNA fragments with argonaute proteins (Ender, C. et al., 2008). Again nuclear argonaute proteins have been described as a part of the cytosolic RISC complex which processes miRNAs. The presence of snoRNA fragments in this complex suggest that the fragments assume miRNA-like functions.

The C/D box snoRNAs also give rise to shorter RNA forms, including RNAs that are longer than 27nt. Analysis of deep sequencing data revealed that C/D box derived smaller RNAs fall into two length classes: those of a length of 17 -19 nt and those larger than 27 nt (Taft, R.J. et al., 2009). Two further reports supported the existence of C/D box derived shorter RNAs. The sequencing of small RNAs libraries made from the ancient eukaryote *Giradia lamblia* revealed four RNA fragments that were derived from C/D box snoRNAs (Saraiya, A.A. et al., 2008).

Together these data indicate that snoRNA expressing units give rise to smaller RNAs. It is possible that size differences between shorter RNAs derived from H/ACA or C/D snoRNAs might indicate different processing enzymes.

SnoRNAs are not the only source of short non-coding RNAs derived from other classes. The analysis of deep-sequencing data also revealed the presence of shorter tRNA-derived RNAs that are on average 19 nt long (Cole, C. et al., 2009). Furthermore, short RNAs can be larger than the 21 nt long fragments generated by dicer, as illustrated by Piwi associated short RNAs, which have lengths between 27 to 31 nt (Malone, C.D. et al, 2009, Brennecke, J. et al., 2007). Finally, analysis of tiling arrays revealed abundant expression of non-coding RNAs, which showed its highest complexity in the nucleus. Most of the non-coding longer RNAs were processed into shorter RNAs of less than 200 nt (Carninci, P. et al., 2005; Willingham, A.T. et al., 2006 and Kapranov, P. et. al., 2007). In summary, these data indicate the existence of numerous short RNAs, most of which are derived from longer precursors and are expressed in the nucleus.

1.4 Natural RNAs that change splice site selection

1.4.1 Regulation of alternative splicing in cis by secondary RNA structures

Single stranded RNAs forms extensive secondary and tertiary structures *in vitro*. A secondary structure is caused by the folding of a short RNA sequence on the same RNA molecule, i.e. an interaction between RNAs in cis. In this sense, secondary structures resemble the interaction between a short RNA and a pre-mRNA molecule in trans.

It is experimentally very difficult to determine the secondary structures formed by pre-mRNAs *in vivo*, as they are transient and are removed when the processing proceeds. Studies in yeast suggested structures between the 5' splice site and branch-points *in vivo* (Charpentier, B. et al., 1996). In addition, studies of *in vitro* systems and of mutations causing human diseases provided evidence for a strong role of RNA secondary structures in alternative splice site selection (Buratti, E. et al., 2004). Secondary structures that include the splice sites or the branchpoint can change exon usage, as found for the tau, HPRT and hnRNPA1 genes (Grover, A. et al., 1999; Tu, M. et al., 2000 and Blanchette, M. et al., 1997).

The recent discovery of alternative splicing regulation by a ribo-switch in *Neurospora crassa* provides further evidence for the regulation of a 5' splice site by a secondary structure *in vivo*. Three *Neurospora crassa* genes contain a thiamine aptamer in introns located at the 5' end of genes involved in thiamine metabolism. Two of these introns have alternative 5' splice sites. The thiamine aptamer sterically blocks one of the alternative 5' splice sites in the absence of thiamine pyrophosphate. Upon binding of thiamine phosphate, the aptamer changes its conformation which makes the alternative 5' splice site accessible. Thus, a metabolite can change alternative splicing by changing the secondary structure of a pre-mRNA (Figure 5 K) (Cheah, M.T. et al., 2007). In contrast to other examples of splicing regulation by secondary structures, this system does not require direct base pairing between the splice site and the regulatory RNA element and is inducible by a metabolite.

The regulation of alternative splicing by secondary structure is not limited to the splice sites. The *Drosophila Dscam* (Down-syndrome cell adhesion molecule) can potentially encode 38,016 proteins due to mutually exclusive exons located at four sites in

the protein. The selection of one of 48 alternative exons at the second site that provides the 6th exon in the mRNA is controlled by the formation of a double stranded RNA structure between a region close to exon 5 (the docking site) and a region close to exon 6 (the selector site). The usage of the 48 exons at the second site is blocked by a protein. Upon interaction between the docking and the selector site, this blockage is interrupted by an unclear mechanism and the selected exon is included in the mRNA (Figure 5 L). Selector and docking sequences exhibit complementarities between 21 and 30 nt. None of the complementarities is perfect and there are mismatches and short bulges on both RNA elements (Graveley, B.R., 2005).

These examples demonstrate that interactions between different parts of a pre-mRNA molecule can change splice site selection. The interacting regions are always small (<50 nt), have numerous mismatches and can be interpreted as a regulation of an RNA in cis.

1.4.2 Regulation in trans

Regulation at the 5' splice site

By far the best understood small RNA in the control of alternative splicing is the interaction between U1 snRNP and the 5' splice site. The 5' end of the U1 snRNA exhibits 9 nt of full complementarity to the 5' splice site (Figure 5 A) (Zhuang, Y. et al., 1986). However, in humans, only 5% of constitutive 5' splice sites show the full 9-nucleotide complementarity. The majority of 5' splice sites, representing about 41% of all constitutive splice sites, have two mismatches towards the U1 snRNA (Carmel, I. et al., 2005). 5% of all constitutive splice sites have 4 mismatches in the 9-nucleotide region (Carmel, I. et al., 2005). The binding of the U1 snRNP is therefore stabilized by contacts to other proteins assembled on the pre-mRNA to stabilize its binding to the 5' splice site. The importance of the correct base complementarity at the 5' splice site is underlined by human mutations that frequently lead to diseases caused by aberrant exon usage. Currently 431 mutations in 5' splice sites that are associated with more than 160 human diseases have been described (Voechovsk, I., 2007). In most cases analyzed, mutating the 5' splice site back into the consensus restored proper exon usage.

The U1 example shows that splice site selection can be influenced by short stretches of complementarity between a nuclear RNA and the pre-mRNA that tolerates an amazing degree of mismatches. The interaction between the short RNA and its target can be stabilized by interaction between associated protein factors.
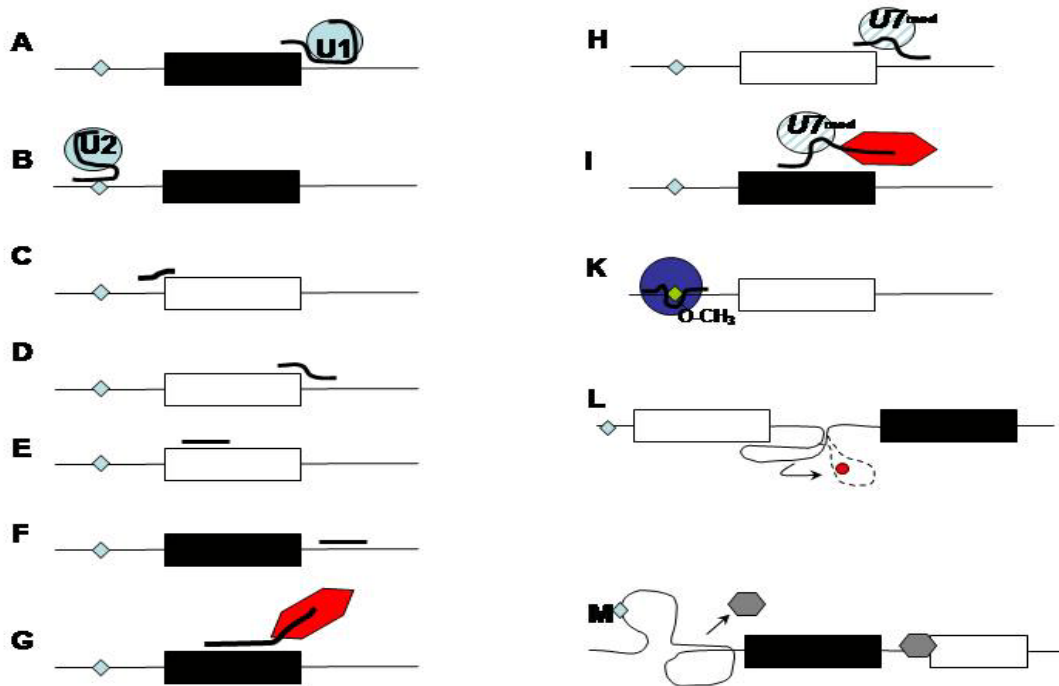
Figure 5: Examples of regulation by short RNAs.

Exons are shown as boxes, introns as lines. The branchpoint is shown as a diamond. Events that generally promote exon inclusion are indicated by black exons, events that lead to exon skipping are indicated by white exons. However, there are numerous exceptions to these rules.

A, B. U1 and U2 snRNP binding promotes exon inclusion

C, D. Oligonucleotides that block 3' or 5' splice sites generally promote exon skipping.

E.   Oligonucleotides that target exonic (shown) or intronic enhancers promote exon skipping.

F.   Oligonucleotides that target exonic or intronic (shown) silencers promote exon skipping.

G.   Bifunctional oligonucleotides can be used to target splicing enhancing proteins to exons, which generally promotes their inclusion. They bind both to RNA and to regulatory proteins (red hexagon).

H.   Modified U7 constructs can be used to target the 5' splice site (shown) or splicing enhancers, which typically blocks exon usage.

I.   Chimeric U7 constructs that bind both the RNA and to splicing regulators can target splicing enhancing proteins to exons, which generally promotes their inclusion.

K.   C/D box snoRNAs and their associated RNPs (blue circle) can be modified to target the branchpoint where they cause 2'-O-methylation, which results in exon skipping.

L.   Riboswitches change conformation upon ligand binding (indicated by the dotted structure and the red dot), which causes release of a previously blocked splice site and subsequent exon inclusion.

M.   RNA can form secondary structures in cis, which can promote exon inclusion by removing repressing proteins (gray hexagon).

1.5   Experimental use of short RNAs

Since the interaction between a short RNA and its target can be easily predicted by its sequence, short DNAs or RNAs have been used for a long time experimentally to influence alternative splice site selection. Oligonucleotides have been used to modify splice site selection by blocking U1 binding to a 5' splice site, by blocking the 3' splice site and masking splicing enhancers necessary for exon inclusion (Figure 5 C-E). These approaches typically lead to exon skipping and are now in clinical trials against Duchenne muscular dystrophy. Other splicing events targeted included APOB, tau, and TNFRSF1B (Aartsma-Rus, A. et al., 2007).

To promote exon inclusion, oligonucleotides can be targeted to silencer regions of the pre-mRNA (Figure 5 F) (Singh, N.N. et al., 2009) or modified to contain binding sites for splicing regulatory proteins that promote exon inclusion (Figure 5 G). In the latter approach, a bifunctional oligonucleotide is used. One part of the bifunctional oligonucleotide targets the pre-mRNA, the other part binds a regulatory protein that promotes exon inclusion (Skordis, L.A. et al., 2003).

The use of bifunctional oligonucleotides shows that a short RNA that is introduced into a cell can assemble RNA binding proteins, which is reminiscent of natural ribonuclear protein complexes. A further development of the oligonucleotide approach is the use of modified U7 snRNPs (Meyer, K. et al., 2009, Asparuhova, M.B. et al., 2007, Liu, S. et al., 2004). The U7 snRNP is necessary for cleaving the 3' end of histone pre-mRNA. The U7 RNP is targeted to histone RNA via an 18 nt long stretch in the U7 RNA that is complementary to the histone pre-mRNA. The binding of U7 marks the histone pre-mRNA 3' end for cleavage. The U7 snRNA contains a Sm protein binding site that is distinct from the Sm site found in the spliceosomal U1, U2, U4-U6 snRNA. If the U7 Sm site is mutated into a spliceosomal Sm site, the resulting snRNP loses its ability to cleave histone RNA, but can still interact with pre-mRNA. However, if the RNA binding site of U7 snRNA is changed to be complementary to splice sites or enhancer sequences, expression of this new U7-RNA causes skipping of targeted exons (Figure 5 H). The U7 snRNA can be further modified to contain a binding site for a splicing activator, like an SR-protein, which resembles the bifunctional oligonucleotide approach. These modified U7 snRNAs promote exon inclusion (Figure 5 I). The modified U7 RNAs have three

functional parts: a targeting site for the pre-mRNA, an Sm binding site that promotes formation of an RNP and a protein-interaction site. These domains allow the RNA to form a scaffold for protein binding and to target these proteins to a pre-mRNA.

Finally, C/D box snoRNAs have been engineered to alter splice site selection by targeting the branchpoint. During the splicing reaction, a RNA-lariat is formed during intron removal. The 'knot' of this lariat is the branchpoint adenosine that forms phosphobonds at its 5', 3' and 2' ribose hydroxyl groups. Blocking the 2'-hydroxyl group by methylation prevents lariat formation, as no phosphodiester bond can be made. C/D box snoRNAs target the methylation of the 2'-OH groups of specific ribose-residues. The ribose to be modified is attached to a base that is complementary to a snoRNA nucleotide that is located 5 bases upstream of its D box. It is therefore possible to engineer snoRNAs that target a particular branch point. For targeting, the antisense box of an existing snoRNA is changed to bind to the RNA of interest. A 2'-O-methylation of the branchpoint adenosine prevents the lariat formation and usage of the downstream exon. Experiments both in yeast (Zhao, X. et al., 2008) and in mammalian systems (Semenov, D.V. et al., 2008) showed that such engineered snoRNAs can change splice site selection.

1.6 Prader-Willi syndrome

Prader-Willi syndrome (PWS) is a genetic syndrome that first manifests as hypotonia and failure to feed in affected newborns giving way to hyperphagia which leads to morbid obesity. Phenotypic features include small stature with small hands and feet, mild to moderate cognitive deficit, and behavioral problems, including obsessive–compulsive disorder (Cassidy, S.B. and Driscoll, D.J., 2009).

PWS results from the absence of expression of the paternally derived alleles of maternally imprinted genes in a critical region of SNRPN locus on chromosome 15q11–q13. A large deletion of paternal chromosome 15q11–q13 contributes to 70% PWS patients; Inheritance of both copies of chromosome 15 from their mother, a condition called maternal uniparental disomy (UPD) affects approximately 25% of PWS patients; and about 5% of PWS are caused by an imprinting defect whereby the individual has inherited one chromosome 15 from each parent, however, the paternally-inherited allele of 15q11–q13 behaves as if it were the maternally-inherited allele. Gene mutation and defective translocation contributes to almost 0.1% of PWS patients.
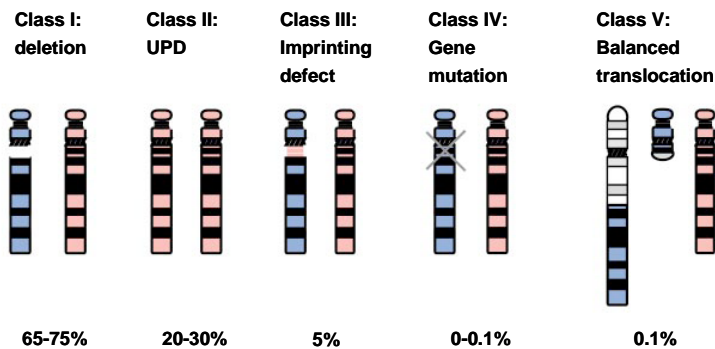
Figure 6: Genetic inheritance in PWS. Shown are the five genetic classes of PWS, with the parental origin (blue, paternal; pink, maternal) and frequencies observed for each. UPD denotes maternal uni-parental disomy (Picture reproduced from Nicholls, R.D. and Knepper, J.L., 2001).

1.7   SnoRNA and PWS

PWS is caused by the loss of gene expression from a maternally imprinted region on chromosome 15q11–q13 (Butler, M.G. et al., 2006). The SNURF–SNRPN locus in the 15q11–q13 region plays a major role in PWS, and its deletion causes PWS-like symptoms in mouse models (Stefan, M. et al., 2005). The SNURF–SNRPN locus spans more than 460 kb and contains at least 148 exons (Runte, M. et al., 2001). Ten exons in the 5' part of the gene are transcribed into a bicistronic mRNA that encodes the SNURF (SmN upstream reading frame) and the SmN (small RNP in neurons) protein. The locus harbors a bipartite imprinting center that silences most maternal genes of the PWS critical region. Owing to this imprinting, the SNURF–SNRPN pre-mRNA is expressed only from the paternal allele. The large 3'-UTR region of the SNURF–SNRPN locus harbors clusters of the C/D box snoRNAs HBII-85 and HBII-52 that are present in 24 and 47 copies, respectively. In addition, the region harbors single copies of other C/D box snoRNAs: HBII-13, HBII-436, HBII-437, HBII-438A and HBII-438B. Recent evidence suggests that the HBII-85 and HBII-52 snoRNA clusters are expressed as two transcriptional units (Vitali, P. et al., 2010). The highly conserved snoRNAs are flanked by poorly conserved non-coding exons, suggesting that the functional relevant products of the locus are snoRNAs, not the flanking exons. The expressions of these snoRNAs are tissue-specific. HBII-52 could be detected only in brain, whereas other snoRNAs from the SNURF–SNRPN locus are also expressed in non-brain tissues (Kishore, S. and Stamm, S., 2006).

Despite the refinement of the PWS critical region to a defined set of transcripts, their function and contribution to the PWS disease mechanism are still not clear.

1.7.1 Deleted SnoRNA clusters lead to PWS

Contrary to conventional understanding that PWS is a contiguous gene syndrome due to multiple genes deletion and is not a single gene mutation, recently there have been reports that PWS may be caused only from the loss of one of the snoRNA clusters that reside in the region (Figure 7).

At least two patients have been reported to lack the HBII-85 (*SNORD116*) snoRNA cluster and have a characteristic phenotype of PWS (de Smith et al., 2009, Sahoo et al., 2008). MBII-52 (SNORD 115) has been shown to regulate exon Vb of the serotonin receptor 2C (HTR2C) (Kishore, S. and Stamm, S., 2006). The HBII-85 (*SNORD116*) snoRNA cluster carries multiple copies of C/D box snoRNAs that have potential complementarity to many different target genes (Bazeley, P.S. et al., 2008).
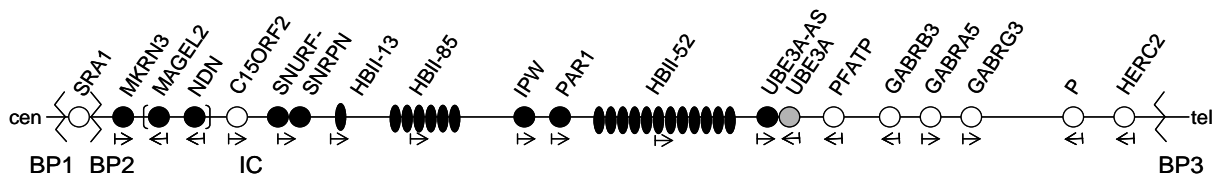
Figure 7: Human Chromosome 15q11-q13 genomic structure. The parental origin (black, paternal; grey, maternal; open circles, biparental) and transcriptional orientation of genes are indicated. The human 4Mb 15q11-q13 region flanked by PWS/AS common deletion breakpoints (BP, zigzag lines). The open bar represents a 1 Mb deletion causing AS only on maternal inheritance. Abbreviations: centromere (cen), telomere (tel) and imprinting center (IC).

Various mouse models have been developed mimicking the human genotype of PWS patients (Table 2). Although the regulation of imprinted gene expression seems similarly regulated between mouse and human, all the murine models of PWS do not show hyperphagia and obesity.

Table 2 (Source Chamberlain S.J. and Lalandea M., 2010)

Mouse Model	Phenotype	References
Prader–Willi syndrome		
Ndn targeted mutation	Three models with conflicting results: 1.) variably penetrant postnatal lethality, skin scraping, increased spatial learning, 2.) postnatal lethality due to respiratory defects, and 3.) no obvious phenotype.	Gerard, L. et al., 1999; Muscatelli, D. et al., 2000 and Tsai, et al., 1999.
Magel2 targeted mutation	Growth deficiency, increased adiposity, hypophagia, altered metabolism, circadian rhythm, and behavior, and reduced fertility	Bischof, J.M. et al., 2007 and Mercer, R.E and Wevrick, R., 2009.
Snrpn targeted mutation	No obvious phenotype	Yang, T. et al., 1998.
Snurf-Snrpn targeted mutation	Two models of different size: 1.) no obvious phenotype 2.) Partial imprinting mutation and variably penetrant postnatal lethality.	Bressler, J. et al., 2001.
Snurf targeted mutation	No obvious phenotype	Tsai, T.F. et al., 1999.
MbII-85/Snord116 targeted	Two models: 1.) strain-dependent postnatal lethality (not fully penetrant),	Ding, F. et al., 2008 and

17

Mouse Model	Phenotype	References
mutation	growth deficiency. 2.) Postnatal growth deficiency, hyperphagia (without obesity), motor learning deficits, increased anxiety	Skryabin, B.V. et al., 2007.
Ipw, *MbII-52/Snord115* radiation induced deletion	Paternal transmission: loss of *MbII-52/Snord115* and *Ube3a* expression. No obvious phenotype.	Ding, F. et al., 2005.
Imprinting center deletion	Completely penetrant postnatal lethality (strain dependent), small stature, and hypogonadism	Chamberlain, S.J. et al., 2004 and Yang, T. et al., 1998.
Snurf-Snrpn to *Ube3a* deletion	Partial postnatal lethality, growth deficiency	Tsai, T.F. et al. 1999.
Snrpn human exon 1 switch	Paternal transmission: postnatal lethality (not fully penetrant), growth deficiency. Loss of *Ndn*, *Magel2* and *Mkrn3* expression.	Johnstone, K.A. et al., 2006.
Transgene insertion/deletion	Paternal transmission: postnatal lethality	Gabriel, J.M. et al., 1999.
Angelman syndrome		

Mouse Model	Phenotype	References
Ube3a targeted mutation	Strain-background dependent seizures, deficits in LTP, poor rotarod performance, loss of plasticity of visual cortex.	Jiang, Y.H. et al., 1998 and Miura, K. et al., 2002.
Snrpn human exon 1 switch	Maternal transmission: loss of *Ube3a* expression, obesity due to increased *MbII-85*/*Snord116* expression.	Johnstone, K.A. et al., 2006.
Transgene insertion/deletion	Maternal transmission: loss of *Ube3a* expression	Gabriel, J.M. et al., 1999.
15q duplication syndrome		
Engineered duplication	Paternal transmission: autism-like behaviors, including a defect in acoustic startle response, obesity	Nakatani, J. et al., 2009.
	Maternal transmission: defect in acoustic startle response	
5-HT2CR mutant		
Transgenic	The relationship between the relative frequency of 5-HT2C receptor RNA editing and steady-state levels of Gαq expression was checked.	Canal, C.E. et al., 2009.

1.8   Role of small oligonucleotides in Alternative splicing

The involvement of short RNAs in splice site selection is evident from the well-established role that U1 and U2 snRNPs play in this process (Luhrmann, R. et al., 1990). The discovery of miRNAs further demonstrated that short non-coding RNAs play a crucial role in RNA metabolism. Common to snRNAs, snoRNAs and miRNAs is that their interaction with their RNA targets is governed by short and often interrupted sequence complementarities. It is therefore difficult to determine their targets bioinformatically. Recent results indicate that snoRNAs and tRNAs are further processed into shorter RNAs. Some of these shorter RNAs could function in pre-mRNA processing, most likely similar to bifunctional oligonucleotides.

1.9   Regulation of alternative splicing by snoRNAs

SnoRNAs are evolutionary highly conserved, whereas the hosting non-coding exons are poorly conserved. HBII-52 is expressed only in brain. A striking feature of HBII-52 is that its antisense box exhibits sequence complementarity to the serotonin receptor 5 HT-2C pre-mRNA. The perfect 18 nt-long sequence complementarity is located in the alternative exon Vb of the receptor. Failing to include this exon into the pre-mRNA causes the generation of a non-functional receptor. Transfection experiments showed that cotransfecting the HBII-52 expression unit with a 5-HT2CR reporter gene promoted exon Vb inclusion. Mutagenesis studies showed that exon Vb contains splicing silencers that normally prevent the inclusion of the exon. Expression of the snoRNA blocks the action of the silencers and promotes exon inclusion. The silencers located on the pre-mRNA can also be modified by RNA editing that changes adenosine to inosine residues (Kishore, S. and Stamm, S., 2006). As a result, there are two ways of generating a full-length serotonin 5 HT2C receptor: blocking the silencers through expression of the snoRNA and weakening the silencers by editing some of its bases. However, the editing events change the amino acid composition of the receptor at three sites. These sites are located in an area critical for protein function, namely in a loop that couples to the G protein. The editing of the receptor pre-mRNA decreases the coupling of the G protein to the receptor and thus reduces its efficacy. The mRNA containing the non-edited version of the 5-HT2C receptor encodes a

receptor that couples optimal to its effecter G protein and shows the highest response to serotonin stimulation. Analysis of limited brain samples from PWS patients showed a reduction of the non-edited isoform (Kishore, S. and Stamm, S., 2006), which has also been observed in mouse models lacking HBII-52 expression (Doe, C.M. et al., 2003). A molecular link between a defect in the 5-HT2CR production and PWS is an attractive hypothesis, as the 5-HT2C receptor plays a crucial role in hunger control and satiety, which is the major problem in PWS. Since HBII-52 promotes the generation of the most active receptor, it acts like a 'genetic agonist' of the serotonin receptor. The administration of selective 5-HT2CR agonists, such as d-fenfluramine, has a strong appetite-suppressing effect (Vickers, S.P. et al., 2001). Underlining the importance of the 5-HT2C receptor for hunger control, the mouse knock-out of 5-HT2CR is hyperphagic and develops obesity. Expression of the 5-HT2CR in the arcuate nucleus, a major hunger control center reverses the hyperphagic phenotype (Xu, Y. et al., 2008). Conversely, when a mutant of the receptor that represents the fully edited 5-HT2CR is expressed in knock-out mice, the resulting mice remain hyperphagic (Kawahara, Y. et al., 2008). Collectively, the data strongly support a model where the loss of HBII-52 causes a loss of the mRNA isoform that encodes the most active form of the receptor, which is necessary for proper hunger control.
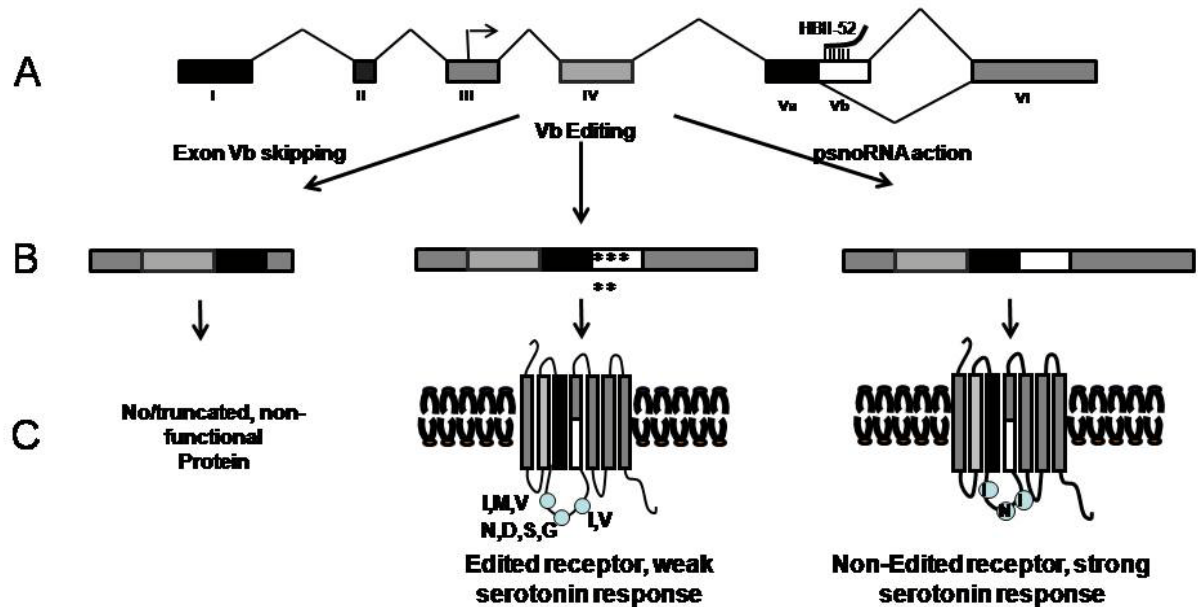
Figure 8: Regulation of serotonin receptor 5-HT2C by HBII-52 derived psnoRNAs.

A. The genomic structure of the 5-HT2C receptor. The arrow in exon III indicated the translational start point. HBII-52 derived psnoRNAs interact with an 18 nucleotide complementarity region in exon Vb.

B. Protein coding parts of the mRNAs derived from different pre-mRNA processing events. Exon Vb skipping results in a shortened mRNA that endodes a truncated protein but is most likely subject to nonsense-mediated mRNA decay. Exon Vb can be edited at five positions (indicated as arrows). The editing event promotes inclusion of the exon, but changes the amino acid sequence at three points. The psnoRNAs cause inclusion of exon Vb without editing, which generates a receptor with the highest agonist efficacy.

C. Structure of the encoded proteins. Editing of exon Vb leads to a change a potentially three amino acids, which are located in the second intracellular loop that couples to the effector G protein. The editing events weaken the receptor-G protein interaction and lead to a weak serotonin response. The non-edited receptor features the amino acids I, N and I at the positions that could be edited and shows the strongest coupling to the G protein and response to serotonin.

The exact molecular mechanism by which a snoRNA-expressing unit changes alternative splicing is difficult to imagine, especially since HBII-52 is mainly localized in the nucleolus.

Therefore we undertook this study to understand the molecular mechanism behind snoRNA mediated regulation of alternative splicing and to find out the processed forms that bring about alternative splicing.

2 Research Overview

SnoRNAs are small nuclear RNAs that can be detected in the nucleolus. They reside in introns from which they are released through nuclease action during the processing of the host pre-mRNA. The C/D box snoRNA HBII-52 (SNORD115) which not expressed in Prader-Willi syndrome (PWS) was shown to be involved in the alternative exon Vb of the serotonin receptor 2C (HTR2C). The overall aim for this work was to identify new target genes for HBII-52 and the related snoRNA HBII-85 to test the hypothesis that the regulation of alternative splicing by snoRNAs is a general mechanism.

Using a bioinformatic approach and experimental verification, the first part of the work identified pre-mRNA targets for SNORD 115 and SNORD 116. The experimental verification was based on splicing of endogenous genes in mammalian cell lines. The findings were confirmed by minigene assays and knock out mouse models. To test numerous splicing events experimentally, we developed a fast and simple recombination-based method to generate splicing reporter genes, using a new vector, pSpliceExpress.

In the second part, we used RNAse protection and showed that the snoRNAs SNORD 115 and SNORD 116 are processed into shorter forms. There is no reliable method available to clone small amounts of dsRNAs, for example from protection analysis. We developed an efficient method to clone dsRNA that can be used to clone smaller dsRNAs. Applying this new technique to SNORD 115 and SNORD 116 we identified the shorter forms of these RNAs and found them to be generated by further nuclease action that removes the terminal RNA stems.

In summary, we found that not canonical snoRNAs appear to be the major missing RNAs in PWS, but their shorter processed forms, that we term psnoRNAs. Bioinformatic analysis indicates that psnoRNAs target double stranded regions in the pre-mRNA. The function of the psnoRNA could therefore be to open up double-stranded RNA structures that mask existing splice sites.

3 Materials and Method

3.1 Materials

3.1.1 Chemicals

Product	Supplier	Product	Supplier
30% Acrylamide/Bis solution	Sigma	Nonidet P-40 / Igepal CA-630	Sigma
Agar	GibcoBRL	dNTPs	Invitrogen
UltraPure agarose	Invitrogen	Paraformaldehyde	Merck
Ampicilin	Sigma	Phenol: Chloroform: Isoamyl Alcohol	Sigma
Aprotinin	Sigma	PMSF	Sigma
$[\gamma^{-32}$ P]-ATP/$[\alpha^{-32}$ P]-CTP	MP Biochem	Potassium chloride	Merck
Boric acid	Roth	Protease Inhibitor Cocktail	Sigma
Brilliant Blue R 250	Sigma	Protein A Sepharose	Amersham
Bromophenol blue	Merck	RNase Inhibitor	Roche
Chloramphenicol	Sigma	SDS	Sigma
Chloroform: Isoamyl alcohol	Sigma	Sepharose CL-4B	Pharmacia
di-Sodiumhydrogen phosphate	Merck	Sodium acetate	Merck
DMSO	Sigma	Sodium chloride	Roth
DTT	Merck	Sodium dihydrogen phosphate	Merck
EDTA	Merck	Sodium fluoride	Sigma
Ethanol	Roth	Sodium hydroxide	Merck
Ethidium bromide	Sigma	Sodium orthovanadate	Sigma
Ficoll 400	Fluka	TEMED	Sigma
Glycerol	Sigma	Tris base	Aldrich
Gel green Dye	Bio-Rad	TBE (10x)	Ambion
Hybridization Buffer	Ambion	TRIzol	Sigma
Magnesium chloride	Merck	Tween 20	Sigma
Magnesium sulfate	Sigma	Xylene cyanole FF	Sigma

3.1.2 **Enzymes**

Product	Supplier	Product	Supplier
T4 PNK	NEB	FastLink T4 DNA Ligase	Biozym
T7 DNA Polymerase	NEB	Platinum Pfx polymerase	Invitrogen
T7/Sp6/T3 RNA Polymerase	NEB	Monster Script RT	Epicentre
AmpliTaq DNA polymerase	Roche	Restriction Enzymes	NEB
DpnI	NEB	RNAse H	USB
DNAseI	Boehringer	SuperScript II	Invitrogen

3.1.3 **Cell lines and media**

Cell Line	Description	ATCC number
HEK293	Human embryonic kidney transformed with adenovirus 5 DNA	CRL-1573
Neuro-2a	Neuroblastoma from mouse brain	CCL-131
NIT1	A pancreatic beta-cell line	CDL- 1789

All cell lines were maintained in DMEM supplemented with 10% fetal calf serum (both from Invitrogen-Gibco).

3.1.4 Bacterial strains and media

Strain	Genotype	Reference
XL1-Blue MRF'	Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F′ proAB lacIq ZΔ M15 Tn10 (Tetr)]	(Bullock et al., 1987)
TOP10 cells	F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG	Invitrogen
DH5α-T1R	F⁻ Φ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(r_k^-, m_k^+) phoA supE44 thi-1 gyrA96 relA1 tonA	Invitrogen

LB MEDIUM (1L):

10 g NaCl

10 g tryptone

5 g yeast extract

Agar

LB AGAR (1L):

10 g NaCl

10 g tryptone

5 g yeast extract

20 grams

3.1.5 **Antibiotics**

Antibiotic	Stock solution		Working concentration	
	Concentration	Storage	Liquid culture	Plates
Ampicilin	50 mg/ml in H2O	4°C	100 µg/ml	100 µg/ml
Chloramphenicol	30 mg/ml in ethanol	–20°C	15 µg/ml	30 µg/ml
Kanamycin	20 mg/ml in H2O	4°C	20 µg/ml	20 µg/ml

3.1.6 **Plasmids**

3.1.6.1 **Clones from the Stamm lab collection or outside sources**

Name	Backbone	Description	Reference
5HT$_{2C}$	unknown	T7 promoter, 5-HT$_{2C}$R minigene	A. Hüttenhofer
MBII-52	unknown	CMV promoter, Amp^r	Cavaille, J. et al., 2000.

3.1.6.2 Newly made clones

Name	Backbone	Description
5HT2CEx5LngExonTrap	Exon trap	Sertonin receptor longest minigene w/o flanking constitutive exons.
5HT2CEx5Lng-MG	Exon trap	Sertonin receptor longest minigene with flanking constitutive exons 4 and 6.
5HT2CR-RSSM-MG	Exon trap	Serotonin receptor long minigene with activated distal 5´ splice site.
pSplice Express	Exon Trap	Modified exon trap vector with attp1 and attp2 sites compatible for gateway cloning.
PB1	pSplice express	PB1 minigene with flanking constitutive exon
TAF1	pSplice express	TAF1 minigene with flanking constitutive exons.
CRHR1	pSplice express	CRHR1 minigene with flanking constitutive exons.
RALGPS1	pSplice express	RALGPS1 minigene with flanking constitutive exons.
PTBP1	pSplice express	PTBP1 minigene with flanking constitutive exons.
HAGHL	pSplice express	PTBP1 minigene with flanking constitutive exons.

Name	Backbone	Description
DPM2	pSplice express	DPM2 minigene with flanking constitutive exons.
CLK2	pSplice express	DPM2 minigene with flanking constitutive exons.
PB1 (mut)	pSplice express	PB1 minigene with flanking constitutive exons with mutated target site for MBII-52.
TAF1 (mut)	pSplice express	TAF1 minigene with flanking constitutive exons with mutated target site for MBII-52.
CRHR1 (mut)	pSplice express	CRHR1 minigene with flanking constitutive exons with mutated target site for MBII-52.
RALGPS1 (mut)	pSplice express	RALGPS1 minigene with flanking constitutive exons with mutated target site for MBII-52.
PTBP1 (mut)	pSplice express	PTBP1 minigene with flanking constitutive exons with mutated target site for MBII-52.
HAGHL (mut)	pSplice express	PTBP1 minigene with flanking constitutive exons with mutated target site for MBII-52.
DPM2 (mut)	pSplice express	DPM2 minigene with flanking constitutive exons with mutated target site for MBII-52.
CLK2 (mut)	pSplice express	DPM2 minigene with flanking constitutive exons with mutated target site for MBII-52.

Name	Backbone	Description
HBII-85 Topo	TOPO2.1	HBII-85 w/o constitutive exons as template for RPA probe
MBII-85 Topo +20bp	TOPO2.1	MBII-85 plus 20 bp upstream and downstream intron sequence w/o constitutive exons as template for RPA probe
HBII-85 Topo +20bp	TOPO2.1	HBII-85 plus 20 bp upstream and downstream intron sequence w/o constitutive exons as template for RPA probe
MBII-52 Topo	TOPO2.1	MBII-52 w/o constitutive exons as template for RPA probe
HBII-52 Topo	TOPO2.1	HBII-52 w/o constitutive exons as template for RPA probe
MBII-52 Topo +20bp	TOPO2.1	MBII-52 plus 20 bp upstream and downstream intron sequence w/o constitutive exons as template for RPA probe
HBII-52 Topo +20bp	TOPO2.1	HBII-52 plus 20 bp upstream and downstream intron sequence as template for RPA probe
MBII-85 Topo	TOPO2.1	MBII-85 w/o constitutive exons as template for RPA probe
HBII-52-4 Copies	pDest 12.2	Four tandem copies of HBII-52.
HBII-85-4 Copies	pDest12.2	Four tandem copies of MBII-52.
MBII-52cC-AS	pCDNA3.1+	C/D-Box Mutant MBII-52 for subcloning. expressed directly under CMV promoter.
MBII-52cD-AS	pCDNA3.1+	

3.1.7 **Primers**

3.1.7.1   **Primers used for cloning minigenes**

`Name	Sequence	Target
HAGHLMGF	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGA AGTGCAGACTGGC	HAGHL
HAGHLMGR	GGGGACCACTTTGTACAAGAAAGCTGGGTCAT GGGGAGACTTGTCACTGG	
PTBP1 MG F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCGA GTAGCGAGATGACG	PTBP1
PTBP1 MG R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAT CGATGATCGATGACAGAG	
RALGPSMGF	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAT GACGATGACAGTA	RALGP S1
RALGPSMGR	GGGGACCACTTTGTACAAGAAAGCTGGGTCAT GATTGATCCAGTGACAGT	
TAF1MG F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAT CGATGACGATGGA	TAF1
TAF1MG R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAT CGATGACGTAGACGATGC	
DPM2F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACG ATGACGATGGTAA	DPM2
DPM2R	GGGGACCACTTTGTACAAGAAAGCTGGGTCAT ACGATGACGATGACGATTG	

`Name	Sequence	Target
PB1F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGATGACGATGACGT	PB1
PB1R	GGGGACCACTTTGTACAAGAAAGCTGGGTCATCAGTGACGATGACCAATGCC	
CRHR1F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGTACGTGGACGATG	CRHR1
CRHR1R	GGGGACCACTTTGTACAAGAAAGCTGGGTCATACGATGATGCGATGACCGAT	

3.1.7.2 Oligonucleotides used for pull down

Name	Sequence	Target
MBII-52	[UAAUCCUACUGAGC] BIOTYNYLATED	Antisense box 2
MBII-85	[GAAAAGUUCCUUACAAA] BIOTYNYLATED	Antisense box 2

3.1.7.3 Oligonucleotides used as linkers for cloning dsRNA

Name	Sequence	Target
5' Linker	5AmMC6-GCTCCAGAATTCGGACCCGArGrUrGrCrCrUrArCArrG	5' of dsRNA
3' Linker	5rApp- CTGTAGGCACCATCAATc- 3ddc	3' of dsRNA

3.1.7.4   Primers used for cloning dsRNA

Name	Sequence	Target
Adaptor Linker F	GCTCCAGAATTCGGACCCGAGTG	cDNA from dsRNA with linkers 5' and 3' linkers ligated.
Adaptor Linker R	GATTGATGGTGCCTACAG	

3.1.7.5   Primers used for cloning snoRNA gene for RPA

Name	Sequence	Target
snoRNA_for	TCCCGGGCTGGGTCAATGATGACAACC	MBII-52 snoRNA
snoRNA_rev	TCCCGGGCCTCAGCGTAATCCTAT	
MBII-85F	GCACCTGGGAGACCCAAGAT	MBII-85 snoRNA
MBII-85R	ATGCAATGCACTCCAACGAG	

3.1.7.6   Primers used for Mutagenesis

Name	Sequence	Generated mutants
C-Box mutF	CTGGGTCAAGTTAAGCAACCCAAT	pCMV-MBII52cC
C-Box mutantR	ATTGGGTTGCTTAACTTGACCCAG	
D-box mutantF	TAGGATTACGACGTGGCCCAACCA	pCMV-MBII52cD
D-box mutantR	TGGTTGGGCCACGTCGTAATCCTA	

3.1.7.7 Primers used for RT-PCR

Name	Sequence	Target
MmNEWCRHR1F	CCAGGATCAGCAGTGTGAGA	CRHR1
MmNEWCRHR1R	GATCATCATACGATGACAGT	
MmRALGPS1F	ACGATGACGATGCATGATGC	RALGPS1
MmRALGPS1R	GATGATCTCTGGATGATGCGA	
MmTAF1F	CAGTGATGACGATGAGGCAT	TAF1
MmTAF1R	AGATCGCTAGTAGATGATGA	
MmPB1F	GTAGCTAGGATGAGATGATA	PB1
MmPB1R	CGATGATGAAATAGTACGAT	
MmDPM2F	TCGATAAATACTAGATGATG	DPM2
MmDOM2R	CGTATATTAGGATCGATGG	
5HT2CMG_for	TATGTCTGGCCTTTACCTAGATATTT	Mouse 5-$HT_{2C}R$ minigene
5HT2CMG_rev	TCACCATAATTGTCAACGGGA	
etrap_cDNA_RT	GATCCACGATGCCGCGCT	Human 5-$HT_{2C}R$ minigene
5HT2CMG_for	TATGTCTGGCCTTTACCTAGATATTT	
5HT2Cex6ShtRev	AGTGAGCTCACTTAGTCCAGGCGGTTCCT	
etrap_cDNA_RT	GATCCACGATGCCGCGCT	Exon-trap vector
eTrap-PCR-for	GAGGGATCCGCTTCCTGCCCC	
eTrap-PCR-rev	CTCCCGGGCCACCTCCAGTGCC	

3.1.7.8   Primers used for Real time PCR analysis of pCMV-5HT (5HT₂cR) minigene

Name	Sequence	Target
SeReEx5a-Ex6For	CGCTGGATCGGTGTATCAGT	Exon 5a-6 junction
5HT2Cex5a-5bFor	CGCTGGATCGGTATGTAGC	Exon 5a-5b junction
SeReX6b6ShtRev	GGGATAGGAACTGATACACCTATAGAA	Exon 6

3.1.7.9   Primers used for Sequencing

Name	Sequence	Target
etrap_seq_for	GGATTCTTCTACACACCC	Exon trap vector
etrap_seq_rev	TCCACCCAGCTCCAGTTG	

3.1.8   RNA from animal tissues

RNA from different tissues (Brain, Kidney, Liver, Thymus, Muscle, Adrenal Gland, Testis and Heart) was isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. 15-20 µg of glycogen per 1 ml was added to the reaction prior to ethanol precipitation for better recovery.

3.2   Methods

3.2.1   Plasmid DNA isolation

The high yield of ultra pure plasmid DNA was isolated using QIAGEN Plasmid Maxi kit according to the manufacturer's protocol.

Smaller yield of plasmid DNA was isolated from the alkaline lysis method first described by Birnboim and Doly (Birnboim and Doly, 1979). In brief, bacterial cells carrying the desired plasmid were cultured overnight at 37°C in 5ml LB medium containing the appropriate antibiotics. The cells were harvested by centrifugation for 5 minutes at 12,000 rpm. The pellet was resuspended in 250µl buffer P1. An equal volume of lysis buffer P2 was then added and the solution mixed gently by invertion. The cells were allowed to lyse for 5 minutes, followed by addition of the neutralization buffer P3. The tube was mixed gently by inversion and the solution was maintained on

ice for 15 minutes. After centrifugation for 10 minutes at 12,000 rpm, the resulting supernatant was precipitated by adding 1 volume of isopropanol. Plasmid DNA was pelleted by centrifugation at 12,000 rpm for 10 minutes, washed with 70 % ethanol, air-dried and dissolved in 30µl of TE buffer. All the steps were carried out at room temperature in a conventional tabletop microfuge.

LB MEDIUM:	BUFFER P1:	BUFFER P2:
10g Tryptone	50 mM Tris-HCl, pH 8.0	200 mM NaOH
10g NaCl	10 mM EDTA	1% SDS
5g yeast extract	100 µg/ml RNase A	

Buffer P3:	Buffer TE:
3M Potassium acetate, pH 5.5	10 mM Tris-HCl, pH 8.0
	1 mM EDTA

3.2.2 Electrophoresis of DNA

DNA was resolved on 0.7-2% agarose gels prepared in 1 x TBE buffer. The electrophoresis was run for 80 min at 100 V. The gels were stained for 30 min in 0.5 mg/ml ethidium bromide and visualized under UV light, $\lambda = 260$ nm.

1x TBE:	6x GEL–LOADING BUFFER:
90 mM Tris-borate	0.25% bromophenol blue
20 mM EDTA	0.25% xylene cyanol FF
	15% Ficoll 400 in dH_2O

3.2.3 Elution of DNA from agarose gels

DNA was purified from agarose gels where crystal violet was added to a final concentration of 2 µg per ml to detect DNA under visible light. Individual bands were excised and DNA was extracted using the Qiagen QIAEX II gel extraction kit according to the manufacturer's protocol.

6x CRYSTAL VIOLET GEL–LOADING BUFFER:

0.25% crystal violet

15% Ficoll 400 in dH_2O

3.2.4 PCR amplification of DNA

A standard PCR reaction to amplify DNA from a plasmid template contained 1-10 ng of plasmid DNA, forward and reverse primers (0.5 µM each), dNTPs (200 µM), 1 x Taq polymerase buffer, 1.5 mM MgCl2 and 1 U Taq polymerase in total volume of 25 µl.   When the amplification was made for cloning purposes, a high-fidelity polymerase, i.e. Platinum Pfx polymerase was used instead of Taq polymerase. The amplification was carried out in a Perkin Elmer GeneAmp PCR System 9700 thermocycler under the following conditions: initial denaturation for 2-4 min at 94°C; 25-35 cycles of 15-30 sec at 94°C, annealing at the Tm of the primers pair, extension of 1 min per 1 kb at 72°C (or 68°C for Pfx polymerase). After the last cycle the reaction was held for 5-10 min at the extension temperature to complete the amplification of all products.

3.2.5 DNA Ligation

When the vector ends were blunt or compatible with each other, the vector was dephosphorylated prior to ligation to prevent self-ligation. To remove 5' phosphates from the vector, 2 U of Calf intestinal phosphatase (CIP, Boehringer) was added to 5 µg of linearised vector in 1 x CIP buffer in 20 µl. The reaction was incubated for 1 hour at 37°C. CIP was subsequently inactivated by heating the reaction to 68°C for 20 min. A typical ligation reaction contained vector and insert at a ratio of about 1:3 (500-1000 ng total DNA), 1 x ligase buffer, 1 mM ATP and 200-400 U T4 DNA Ligase (New England Biolabs) in 15 µl. The incubation was carried out at 16°C for 12-16 h. After that, one third to one half of the ligation mixture was then transformed in *E coli* cells.

3.2.6 Preparation of competent *E. coli* cells

5 ml of LB medium were inoculated with a single bacterial colony and grown overnight at 37°C with vigorous shaking. 4 ml of this culture were transferred to 250 ml LB and grown to early logarithmic phase (OD600 = 0.3-0.6). The culture was centrifuged for 10 min at 2500 rpm at 4°C. The bacterial pellet was resuspended in 1/10 volume of cold TSB buffer and incubated on ice for 10 min. Cells were aliquoted into cold Eppendorf tubes and frozen in liquid nitrogen. Competent bacterial cells could then be stored at -80°C for several months.

TSB BUFFER:

10% PEG 3500

5% DMSO

10 mM MgCl$_2$

10 mM MgSO$_4$

in LB medium, pH 6.1

### 3.2.7 Transformation of *E. coli* cells

1-10 ng of plasmid DNA or a ligation reaction were added to 20 µl of 5 x KCM buffer and then the volume was brought with water up to 100 µl. Equal volume of competent cells was added. The reaction mixture was incubated on ice for 20 min followed by incubation at RT for 10 min. Then 1 ml of LB medium was added and the bacteria were incubated for 1 h at 37°C with vigorous shaking. Finally cells were plated on LB Agar plates containing appropriate antibiotic. Plates were incubated at 37°C until colonies were visible.

5x KCM BUFFER:

500 mM KCl

150 mM CaCl$_2$

250 mM MgCl$_2$

3.2.8 Site directed mutagenesis by overlap extension

Four primers were designed to introduce mutations by this method which uses overlapping primers. One set of forward (F) and reverse (R) primer was complementary to the extreme ends of the DNA template (see Figure 9). The other set of forward MF and reverse MR primer carrying the desired mutation, were complementary to each other and target the site where the mutation was desired. Mutant primers had 8-10 bases on either side of the mutation cassette to allow precise annealing. The first PCR was carried out with proofreading polymerase to avoid any A-overhang. Individual PCRs were carried out to amplify fragments with primer sets F/MR and with MF/R respectively. The amplified fragments were gel eluted to free them from any contaminating DNA template. 200 ng of the individual purified fragments were pooled together and allowed to anneal and extend without any addition of primer with dNTPs (200 µM), 1 x Taq polymerase buffer, 1.5 mM MgCl2 and 1 U Taq polymerase in total volume of 25 µl. The amplification was carried out in a Perkin Elmer GeneAmp PCR System 9700 thermocycler under the following conditions: initial denaturation for 5 min at 94°C; 10 cycles of 30 sec at 94°C, annealing at 50°C,

extension of 1 min per 1 kb at 72°C. After the last cycle the reaction was held for 5 min at the extension temperature to complete the amplification of all products. External primers (F and R) were then added and the reaction was again supplemented with 1 U of Taq polymerase. The final PCR was performed with the following conditions: initial denaturation for 5 min at 94°C; 30 cycles of 30 sec at 94°C, annealing at 60°C and extension of 1 min per 1 kb at 72°C. The last cycle was followed by another 5 min of extension at 72°C. A part of the amplified fragment was run on the Agarose gel and the other subcloned into pCR4 TOPO for sequencing.
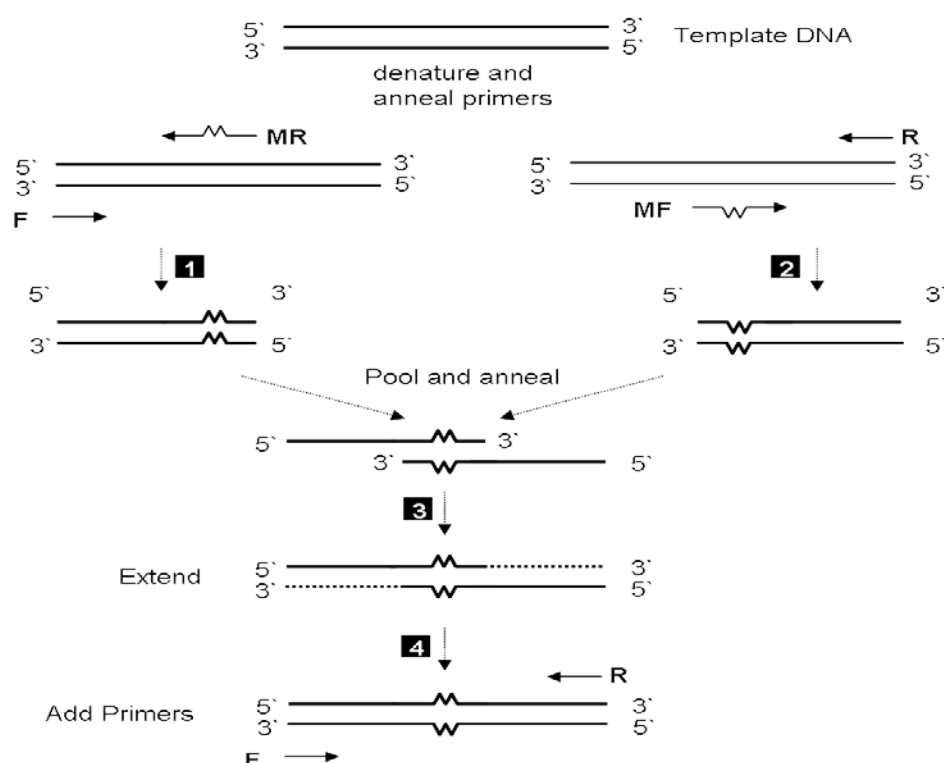
Figure 9: Site-directed Mutagenesis by Overlap Extension. In separate PCR amplification reactions 1 (Primers F and MR) and 2 (Primers MF and R), two partially overlapping fragments of the target gene containing the mutation are amplified. In PCR 3, the denatured products from PCR 1 and PCR 2 anneal at the region of overlap and extend to form full length double –stranded mutant DNA. In PCR 4, the full length mutant DNA is amplified using primers F and R. (adapted from *Molecular Cloning: A Laboratory Manual*, Sambrook and Russel, third edition, 2001).

3.2.9 Radioactive labeling of 5' DNA ends

About 500 ng of purified DNA fragment (or 5-10 pmol oligonucleotide) were mixed with 1 x T4 polynucleotide kinase buffer, 50 µCi of ^{32}P γ ATP (3000 Ci/mmol) and 10 U T4 polynucleotide kinase (New England Biolabs) in 10 µl. The reaction was incubated for 1 hour at 37°C after which the kinase was inactivated at 68°C for 20 min.

3.2.10 Southern Blotting and hybridisation of DNA

DNA was separated on an agarose gel. After electophoresis the gel was placed in denaturation buffer (1.5 M NaCl/0.5 N NaOH) for 30 min. The DNA was transferred to a nylon membrane in a denaturation buffer for 16 h by capillary transfer. The DNA was crosslinked to the membrane on a UV transilluminator. The prehybridisation was carried out for 30 minutes at 68°C in a hybridisation buffer. The hybridisation was done in 10-20 ml of hybridisation buffer for 16 hours at 68°C. The membrane was washed two times in 2 x SSC/0.1% SDS for 10 minutes at 68°C. It was than washed (two times) in 0.5 x SSC/0.1% SDS for 10 minutes at 68°C. Finally, the membrane was washed in 0.1 x SSC/0.1% SDS for 10 minutes at 68°C. The membrane was then exposed to X-ray film or the signal was revealed on a phosphoimager.

Hybridisation buffer:	20x SSC:
0.5M phosphate buffer, pH 7.2	3 M NaCl
7% SDS	0.3 M Na citrate

3.2.11 Freezing, thawing and subculturing of eukaryotic cells

To freeze, cells were grown to mid logarithmic phase (about 75% of confluence) in 10 cm culture dishes. They were collected by trypsinization with 1 x Trypsin/EDTA, resuspended in 1 ml of the freezing medium (90% of the growth medium and 10% of DMSO). Vials were placed in Nalge Nunc Cooler giving a cooling rate of ~1°C/min while at -80°C. Cells were stored later in liquid nitrogen.

To thaw, cells were incubated at 37°C. The entire content of the tube was transferred to a 10 cm culture dish and 10 ml of the growth medium were added. The dish was placed in the incubator at 37°C and 5% CO_2. When cells were attached to the plastic surface, the medium was removed and replaced with fresh one. The cells were maintained in the incubator.

Cells were subcultured after reaching confluence. The monolayer was detached by adding 1x Trypsin /EDTA and incubating at 37°C until a single cell suspension was formed. 1/5 – 1/10 of this suspension was transferred to a new dish and mixed with the growth medium. Cells were maintained in the incubator at 37°C and 5% CO_2.

3.2.12 **Transfection of eukaryotic cells**

The procedure used for HEK293 cells was based on the one published by Chen and Okayama (Chen and Okayama, 1987). Exponentially growing cells were replated at a density of about 3 x 10^5 cells / 8 cm^2. Growth medium was added and the cells were incubated at 37°C, 5% CO_2 for about 24 h, to reach 60-70% of confluence. For most applications cells were grown in 6-well plates, with 2 ml of growth medium per well. The transfection reaction for one well was made the following way: 1 to 5 µg of expression construct were mixed with 25 µl of 1 M $CaCl_2$ in final volume of 100 µl. Equal volume of 2 x HBS buffer was added drop by drop, with constant mixing. In order to form a precipitate, the solution was allowed to stay at RT for 20 min. After that, it was added to the growth medium. To express the transfected plasmid, cells were grown for additional 24 h at 37°C, 3 % CO_2.

2x HBS:

280 mM NaCl

10 mM KCl

1.5 mM $Na_2HPO4.2H_2O$

12 mM Dextrose

50 mM Hepes

pH 6.95

3.2.13 *In vivo* **splicing assay**

To determine the influence of a protein on the splicing of selected minigenes, *in vivo* splicing was performed as described (Stoss et al., 1999; Tang et al., 2005). 1 to 2 µg of the minigene plasmid were transfected in eukaryotic cells together with an expression construct for the protein. Usually a concentration-dependent effect was assessed. The protein was transfected in increasing amounts, in the range of 0 to 3 µg. To avoid 'squelching' effects, the 'empty' parental expression plasmid containing the promoter was added in decreasing amounts, to ensure a constant amount of transfected DNA. Cells were plated in 6-well plates and transfection (section 3.2.11.) was done 24 hours after plating. After incubation for 14-17 hours at 3% CO_2 total RNA was isolated from the cells (section 3.2.14.).

400 ng of RNA were used in a reverse transcription reaction (section 3.2.15.*)*. The reverse primer used for reverse transcription was specific for the vector in which

the minigene was cloned, to suppress reverse transcription of the endogenous RNA. To avoid the problem of the amplification of minigene DNA, *DpnI* restriction enzyme was added into the reverse transcription reaction. *DpnI* cuts GATC sequence in double-stranded DNA when the adenosine is methylated but does not cut non-methylated single-stranded DNA or cDNA. A control reaction with dH_2O instead of RNA was included.

1/8 of the reverse transcription reactions were used for PCR with minigene-specific primers. The primers were selected to amplify alternatively spliced minigene products. A control reaction with no template (RNA instead of cDNA) was included in the PCR. The PCR programs were optimized for each minigene in trial experiments.

PCR reactions were resolved on a 0.3-0.4 cm thick 2 % agarose TBE gel and the image was analyzed using ImageJ analysis software (http://rsb.info.nih.gov/ij/).

3.2.14 Isolation of total RNA

Total RNA was isolated from eukaryotic cells grown in 6-well plates. Cells were washed with 1 x PBS and the RNeasy Mini kit (Qiagen) was used according to the manufacturer's protocol. RNA was eluted from the column in 30 µl of RNase-free dH_2O. However this procedure was applied only when the RNA of interest was larger than 200 bases.

Alternatively, for RNA smaller than 200 bases or when in lower concentrations (as in case of RNA immunoprecipitation), RNA was isolated using TRIzol reagent according to the manufacturer's protocol. 15-20 µg of glycogen blue solution per 1 ml was added to the reaction prior to ethanol precipitation for better recovery. After ethanol precipitation, the RNA pellet was dissolved in 20 µl of RNase-free dH_2O.

3.2.15 RT–PCR

400 ng of total RNA (200 ng/µl), 5 pmol of reverse primer, 40 U of SuperScript II reverse transcriptase, and optionally 4 U of *DpnI* restriction endonuclease were mixed in 5 µl of RT buffer. To reverse transcribe the RNA, the reaction was incubated at 42°C for 45 min.

1/8 of a typical reverse transcription reaction was used to amplify cDNA. The total reaction volume was 25 µl and contained 10 pmol of specific forward and reverse

primers, 200 mM dNTPs, 1x Taq polymerase buffer and 1 U of Taq DNA polymerase. The conditions of the PCR cycles were dependent on the template to be amplified.

RT Buffer:

300 µl 5x First strand synthesis buffer (Invitrogen)

150 µl 0.1 M DTT (Invitrogen)

75 µl 10 mM dNTPs

475 µl dH$_2$O

3.2.16 MBII-52/MBII-85 pull-down assay

3.2.16.1 Transfection

HEK293 cells/N2a cells were transfected with calcium phosphate method. 16 hours after transfection, the cells were washed with serum free DMEM and trypsinised. Trypsinised cells were washed separately with serum free DMEM and isotonic PBS. Finally, the cells were resuspended to a concentration of 8×10^9 cells/ml in appropriate volume of PBS and transferred to a dish such that the depth was 2-3 mm.

3.2.16.2 Isolation of Nuclear extract:

Cells were resuspended in one packed volume of buffer A (10 mM HEPES, pH 8.0, 1.5 mM MgCl2, 10 mM KCL, 1 mM DTT) and to isolate the nuclear extract, cells were trypsinized 24-36 hours after the transfection and washed in 30 volumes of PBS. The pellet was then resuspended in one packed cell volume of buffer A and allowed to swell on ice for 15 minutes. Cells were lysed with a 23G hypodermic needle and nuclei were recovered by centrifugation for 20 sec at 12000g at RT. The crude nuclear pellet was resuspended in two-thirds of one packed cell volume of buffer C and incubated for 30 min at 4°C with stirring. The nuclear debris was pelleted by 5 min centrifugation at 12000g. Collected nuclei were then resuspended in 0.6 ml of NET-Triton, sonicated gently several times on ice and centrifuged. The supernatant contained the nuclear extract.

3.2.16.3 Pull down of RNA using Dynal beads:

M-280 streptavidin coated dynal beads were prepared and the biotinylated 2'-O methyl modified RNA oligonucleotide was immobilized on the beads according to manufacturer's protocol. The precleared nuclear extract was then incubated with the dynal beads immobilized with the RNA oligonucleotide at 4°C for 30 minutes. The

beads were washed with NET-triton buffer twice and captured RNA was isolated using Trizol according to manufacturer's protocol. RT-PCR was done with specific primers to detect the crosslinked snoRNA.

3.2.17 TOPO TA Cloning (Invitrogen)

Fresh PCR product	0.5 to 4 µl
Salt Solution	1 µl
Dilute Salt Solution	1 µl
Water add to a total volume of	5 µl
TOPO® vector 2.1	1 µl
Final Volume	6 µl

1. Add 2 µl of the cloning reaction into a vial of chemically competent *E. coli* and mix gently.

2. Incubate on ice for 5 to 30 minutes.

3. Heat-shock the cells for 30 seconds at 42°C without shaking.

4. Immediately transfer the tubes to ice.

5. Add 250 µl of room temperature S.O.C. medium.

6. Cap the tube tightly and shake the tube horizontally (200 rpm) at 37°C for 1 hour.

7. Spread 10-50 µl from each transformation on a prewarmed selective plate and incubate overnight at 37°C.

8. Pick the clones and isolate plasmid DNA

3.2.18 Gateway cloning (Invitrogen)

The Gateway® Technology is a universal cloning method based on the site-specific recombination properties of bacteriophage lambda. The Gateway® Technology provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression (Hartley et al., 2000)

3.2.18.1   Protocol

Perform a BP recombination reaction between an attB-flanked DNA fragment and an attP-containing donor vector to generate an entry clone.

Figure 10: BP Reaction: Facilitates recombination of an attB substrate (attB-PCR product or a linearized attB expression clone) with an attP substrate (donor vector) to create an attL-containing entry clone (see diagram below). This reaction is catalyzed by BP Clonase™ enzyme mix. (Modified from Invitrogen manual for gateway cloning)

1. Add the following components to a 1.5 ml microcentrifuge tube at room temperature and mix:

attB-PCR product or linearized attB expression clone (40-100 fmol)	1-10 µl
pDONR™ vector (supercoiled, 150 ng/µl)	2 µl
5x BP Clonase™ reaction buffer	4 µl
TE Buffer, pH 8.0 to	16 µl

2. Vortex the BP Clonase™ enzyme mix briefly. Add 4 µl to the components above and mix well by vortexing briefly twice.

3. Incubate reaction at 25°C for 1 hour.

4. Add 2 µl of 2 µg/µl Proteinase K solution and incubate at 37°C for 10 minutes.

5. Transform competent E. coli and select for the appropriate antibiotic-resistant entry clones.

Perform an LR recombination reaction between an attL-containing entry clone and an attR-containing destination vector to generate an expression clone.

Figure 11: LR Reaction: Facilitates recombination of an attL substrate (entry clone) with an attR substrate (destination vector) to create an attB-containing expression clone (see diagram below). This reaction is catalyzed by LR Clonase™ enzyme mix. (modified from Invitrogen manual for gateway cloning)

1. Add the following components to a 1.5 ml microcentrifuge tube at room temperature and mix:

Entry clone (supercoiled, 100-300 ng)	1-10 μl
Destination vector (supercoiled, 150 ng/μl)	2 μl
5x LR Clonase™ reaction buffer	4 μl
TE Buffer, pH 8.0 to	16 μl

2. Vortex the LR Clonase™ enzyme mix briefly. Add 4 μl to the components above and mix well by vortexing briefly twice.

3. Incubate reaction at 25°C for 1 hour.

4. Add 2 μl of 2 μg/μl Proteinase K solution and incubate at 37°C for 10 minutes.

5. Transform competent E. coli and select for the appropriate antibiotic-resistant expression clones.

3.2.19 Generation of the recombination vectors

pSpliceExpress (Figure 12) was constructed by amplifying the ccdB/CmR element from the pDONR221 vector (Invitrogen) using the primers: ccdB-Cmr-*XhoIF*: aaa*ctcgag*agaaacgcaaaaaggccatc; ccdB-CmR-*XbaIR*: aaa*tctaga*gagctgccaggaaacagcta. The amplicon was cloned into the pCR4 TOPO vector (Invitrogen) and subcloned into the multiple cloning site of Exontrap vector (Mobitec) using *XhoI* and *XbaI* restriction sites. This Exontrap vector system is the best-studied vector for minigene analysis (Tang,Y. et al., 2005). It contains multiple cloning sites in the intron of insulin exons derived from *Rattus norvegicus*.
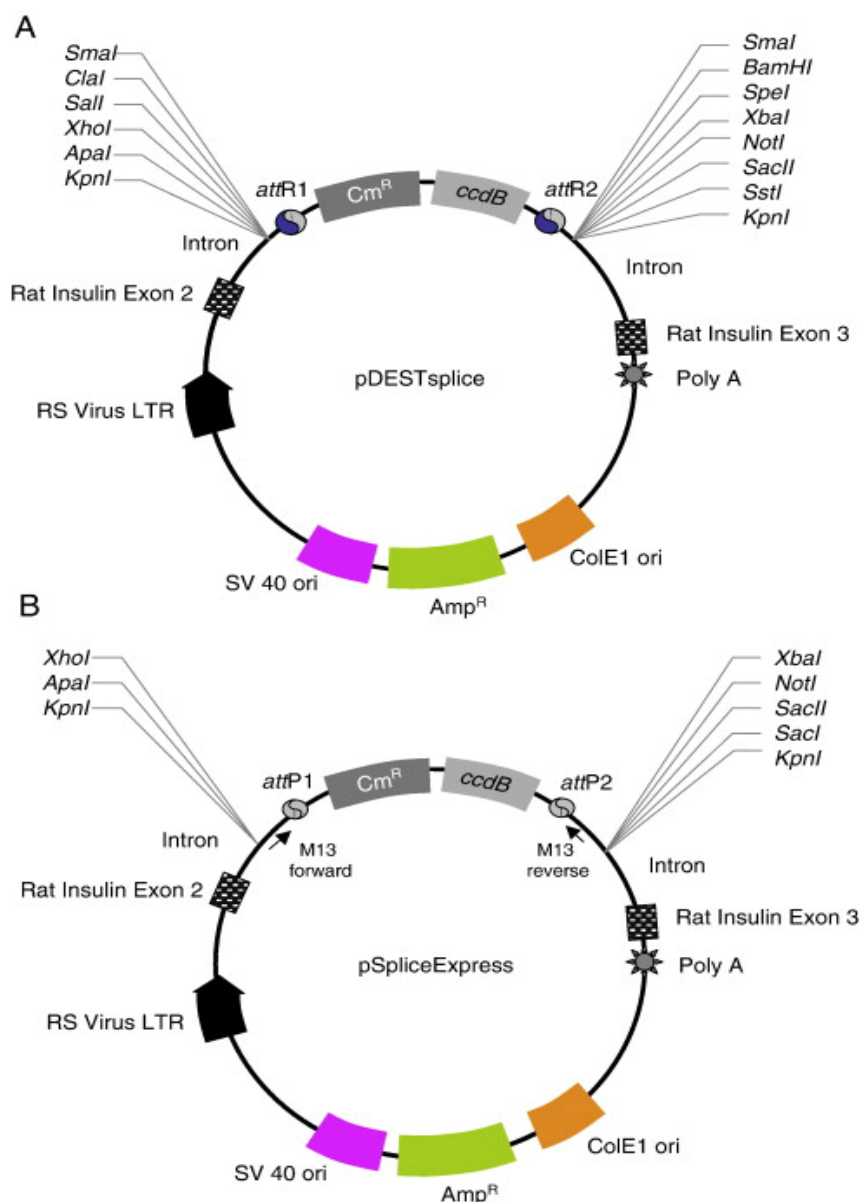
Figure 12: Maps of the cloning vectors pDESTsplice and pSpliceExpress. (A) pSpliceExpress. The vector contains the CmR, ccdB, colE1 ori, AmpR, SV40 ori and RS virus LTR that are indicated. Restriction sites of the multiple cloning sites are indicated. Two m13 sites can be used for sequencing. The selection cassette is flanked by attP sites. (B) pDESTsplice. The vector has similar features as pSpliceExpress. The only difference is the attR attachment sites that are used for recombination.

3.2.20 Method Overview

To analyze an exon using pSpliceExpress, the DNA of interest is first amplified using primers that contain the attB1 and attB2 attachment sites. To determine the splicing regulation of an exon in its genomic context, this exon is usually amplified together with its constitutive flanking exons. The recombination of this PCR product with pSpliceExpress generates a reporter minigene.

To allow usage of cDNA and genomic libraries that contain inserts flanked by attL sites, we constructed pDESTsplice that can recombine with attL sites (Figure 13 D–F). The minigene construct carrying the DNA insert made by either method is then transfected in cell lines of choice. The splicing products are assayed by RT-PCR (Figure 13 G). pSpliceExpress and pDESTsplice contain two constitutively expressed insulin exons, which are spliced together in most cases and serve as a positive control. In addition, the vector generates mRNA species that reflect the splicing pattern of the subcloned genomic fragment (Figure 13 G).
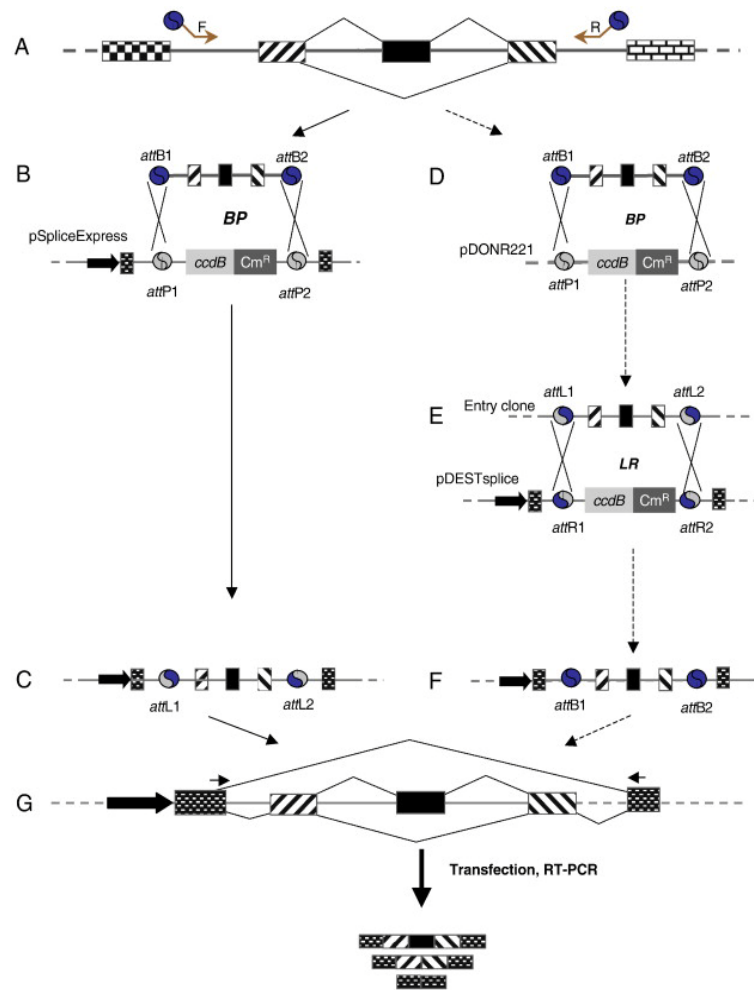
Figure 13: **Overview of the method.** The minigene of interest can be cloned by two methods. Either a PCR product is directly cloned into pSpliceExpress (A, B, C) or it is first cloned into a gateway entry clone (D), which is then recombined with pDESTsplice (steps E–F) to generate the final construct (G). (A) Amplification of the region of interest. Two primers F and R are used to amplify a part of the genomic DNA that harbors the alternative exon (black, splicing patterns are indicated). The primers have recombination sites that are indicated by circles. (B) Construction of the splicing reporter using pSpliceExpress. The PCR fragment is recombined in vitrowith pSpliceExpress vector. The vector contains Cm and ccdB selection markers that are used to isolate recombined clones. (C) Structure of the final construct using pSpliceExpress. The inserted DNA is flanked by two constitutive rat insulin exons, indicated by dotted pattern. The transcript is driven by a RSV LTR promoter (arrow) and the subcloned genomic fragment is flanked by attL sites, generated by the recombination of attB and attP sites. (D) Subcloning of the genomic fragment for use with pDESTsplice. The genomic fragment of interest is generated with attB sites by PCR, which are recombined using any Gateway entry clone that has the ccdB-CmR selection cassette flanked by attL sites. (E) Construction of the reporter gene using pDESTsplice. The attR1 sites of pDESTsplice are recombined in vitro with the attL1 sites of the entry clone. (F) Structure of the final construct using pDESTsplice. The subcloned genomic fragment is flanked by attB1 sites. Except for the recombination sites, the structures of pDESTsplice minigenes are identical to those generated with pSpliceExpress (C). (G) Analysis of the splicing reporter. The splicing reporter construct is transfected into a cell line of choice. The RNA generated is determined by RT-PCR, using the primers indicated (small arrows). The mRNA structures, indicated below the gene structure are expected to be generated by the construct.

3.2.21 **Rapid generation of splicing reporters with pSpliceExpress**

3.2.21.1 **PCR amplification**

The insert can be prepared from genomic DNA by standard PCR using a proofreading Pfx DNA polymerase (Invitrogen). If the PCR template is a plasmid that contains the ampicillin resistance gene, the PCR reaction mixture was treated with Dpn I. Depending on the template DNA used for PCR reaction, 5–10 units of *DpnI* was added to the PCR reaction and incubated at 37 °C for 2 h. This treatment degrades the contaminating plasmid DNA (i.e. *DpnI* recognizes methylated GATC sites) and reduces background in the subsequent BP recombination reaction associated with template contamination. Purification of the PCR-amplified DNA is generally not required. In those cases where there is a high background, PCR purification of the products is performed by agarose gel electrophoresis followed by crystal violet staining and gel purification. This step also removes any contaminating template plasmid (Figure 13).

3.2.21.2 **Performing the BP recombination reaction**

The 5 μl BP recombination reaction consisted of 20–30 fmol of the attB containing PCR product, mixed with 25 fmol of pSpliceExpress vector. 1 μl of 5-fold BP clonase reaction buffer mixture was added. The additional reaction volume was made up with TE buffer, pH 8. The reaction was incubated at 25 °C for 1 h (preferably overnight for fragments larger than 3 kb). To inactivate the enzyme, 0.5 μl of Proteinase K (2 mg/ml) solution was added to the reaction, which was further incubated at 37°C for 10 min.

3.2.21.3 **Transforming the competent cells**

Any recA, endA E. coli strain including OmniMAX™ 2-T1R, TOP10, DH5α™, DH10B™ or equivalent can be used for transformation. Use of E. coli strains that contain the Fepisome (e.g. TOP10F') for transforma tion is not recommended. These strains contain ccdA gene, which prevents negative selection with the ccdB gene. The transformed cells were plated on Ampicillin-supplemented LB plates pre-warmed at 37°C for 30 min.

3.2.21.4   Screening of colonies

Single colonies were inocculated in 5 ml of LB medium overnight and plasmid DNA was isolated. Restriction digestion was performed to confirm the presence of the desired insert. The positive clones were further verified by sequencing.

3.2.21.5   Cloning minigenes into pDESTsplice

The PCR reaction was carried out as described above. The pDESTsplice vector contains attR1–attR2 recombination sites and cloning into pDESTsplice requires the construction of an entry clone containing attL1 and attL2 recombination sites. The entry clone is generated by performing a BP recombination of the PCR product with the pDONR221 vector, which carries the gene for Kanamycin resistance. If the PCR template as well carries a Kanamycin resistance, *DpnI* treatment is performed as described above.

3.2.21.6   LR recombination of the entry clone with the pDESTsplice

5.0μl of a LR recombination reaction was set up similar to the above-described BP recombination. In brief, 25–30 fmol of entry clone were mixed with 25 fmol of pDESTsplice. 1 μl of LR clonase reaction buffer mixture was diluted five-fold from its commercially supplied concentration. The additional volume was made up with TE buffer, pH 8.0. The reaction was incubated at 25 °C for 1 h. For most applications, 1 h will yield a sufficient number of colonies for analysis. The length of recombination reaction can be extended up to 16 h. For large plasmids (> 5 kb), a longer incubation time (i.e. 16 h incubation) yields additional colonies and is recommended.

The reaction was treated with 0.5 μl of Proteinase K (2 mg/ml) for 20 min, transformed and plated on Ampicillin supplemented LB-Agar plates and screened for positive clones. The positive clones were confirmed by sequencing to rule out any undesired mutations.

3.2.22 RNAse Protection Assay and cloning of dsRNA

3.2.22.1   Plasmid Template

1. Plasmid DNA is linearized using appropriate restriction enzyme downstream of the insert to be transcribed. Avoid using restriction enzymes that leave 3' overhangs.

2. Terminate the linearization by adding the following:

   > 1/20th volume of 0.5 mEDTA
   >
   > 1/10th volume of 3M Na acetate
   >
   > 2 volumes of Ethanol

3. Mix well and let it chill at -20°C for at least 30 mins.

4. Pellet the DNA for 15 minutes in a micro-centrifuge at top speed. Dissolve the pellet to 0.5 to 1µg/µl concentration. Run the linearized plasmid DNA along with uncut plasmid DNA to determine the integrity of the restriction digested plasmid.

5. The digested plasmid DNA is proteinase K treated to remove RNAse A by treating with proteinase K (100-200ug/ml) and 0.5% SDS for 30 minutes at 50°C.

6. The DNA was phenol chloroform extracted and was concentrated by ethanol precipitation.

3.2.22.2 Probe Preparation

1. We used Maxi script kit to make labeled probes (Ambion, Catalog no. Am1308).

2. Thaw 10x reaction buffer and ribonucleotide solutions, place 10x reaction buffer at room temperature and ribionucleotide solutions on ice.

3. Add reagents in the following order

4. Nuclease free water to make the final volume 20µl

DNA Template	1 µg
10mM rNTPs(rATP, rCTP, rGTP)	3 µl
^{32}P UTP (3.125 µM)	5 µl
10x Transcription Buffer	2 µl
Enzyme Mix (T7, T3 or SP6)	2 µl

5. Similarly prepare cold probe with cold rUTP was prepared.

6. Nuclease free water to make the final volume 20µl

DNA Template	1 µg
10mM rNTPs (rATP, rCTP, rGTP,rUTP)	4 µl
10x Transcription Buffer	2 µl
Enzyme Mix (T7, T3 or SP6)	2 µl

7. Mix thoroughly and spin the tube to collect the reaction at bottom of the tube and incubate at 37°C for 10 minutes.

8. Add TURBO DNase provided with the kit, mix the reaction well and incubate at 37°C for 15 minutes.

9. Separate the full length probe from shorter fragments by running on a 15% Acrylamide/8 M urea denaturing gel. Gel should be run at 20V/cm of the gel or gel can be run at ~ 20 mA with constant current.

10. Expose the gel for 30 seconds to 1 minute on the X-ray film and cut out bands from the gel corresponding to labeled probe.

11. Elute the probe from the gel by crushing it in the 1.5 ml tube with the tube pestle and resuspending it in 250 µl of elution buffer (Ambion, RPA kit Catalog# AM1414).

12. Elute the RNA by shaking at 37°C overnight.

13. Centrifuge the tubes with probes at 5000rpm for 2 minutes and collect the supernatant. The gel fragments will form a pellet. Take precaution while taking out the supernatant.

14. Precipitate the probe by adding 3 volumes of ethanol. Aliquot and store the probe at -20°C.

3.2.22.3 RNase Protection assay

1. Mix 10µg of sample RNA with labeled probe about 150-600pg or 2×10^4 cpm. And mix 100µg of sample RNA with unlabeled probe. Use yeast RNA provided in the kit to set up two controls for each probe.

2. Co-precipitate the probes with sample RNA by adding

1/10th Volume of 5M NH₄OAc

3. Add equal volumes of ethanol and mix thoroughly and allow RNA to precipitate at -20°C for 30 minutes

4. Pellet the RNA by centrifuging at ≥ 10,000 x g for 15 minutes.

5. Discard the supernatant and air dry the pellet for 5 minutes.

6. Resuspend the pellet in 10 µl of Hybridization buffer and denature at 95°C for 4 minutes.

7. Centrifuge to collect the reaction at bottom of the tube and hybridize overnight at 42°C.

8. Do the RNase digestion using 1:100 dilution of RNaseA/RNaseT1 in 150 µl of RNA digestion buffer (10mM Tris-HCl, pH 7.5, 5mM EDTA, 200mM NaAc) supplied with the kit, for each sample. Leave aside one control tube as RNase digestion control by adding just RNase digestion buffer with the RNases.

9. Incubate the reaction for 30 minutes at 37°C.

10. Perform extraction with 40µl of Phenol/chloroform and concentrate by ethanol precipitation with 1/10th volume of NH₄Ac.

11. Chill the tubes at -20°C for 30 minutes.

12. Centrifuge at ≥ 10,000 x g for 15 minutes, remove the supernatant and air dry the pellets for 10 minutes.

3.2.22.4 T4 Kinase and T4 DNA polymerase treatment

13. Re-suspend the pellet in 35 µl of DEPC treated water and do the T4 Polynucleotide Kinase treatment by adding the following reagents to the re-suspended pellet and T4 DNA polymerase to make it blunt end.

10x T4 PNK Buffer	5 µl
rNTPS	5 µl
T4PNK	5 µl
T4 DNA Polymerase	1 µl

14. Incubate the reaction at 37°C for 10 minutes and stop the reaction by adding 2ul of 0.5M EDTA.

15. Perform extraction with 40µl of Phenol/chloroform and concentrate by ethanol precipitation with $1/10^{th}$ volume of NH_4Ac.

16. Chill the tubes at -20°C for 30 minutes.

17. Centrifuge at ≥ 10,000 x g for 15 minutes, remove the supernatant and air dry the pellets for 10 minutes.

18. Re-suspend the pellet in Gel loading buffer II (provided in the kit).

19. Denture the reactions at 95°C for 3 minutes and run on 15% acrylamide/8M Urea denaturing gel. Let the lower dye to get out of the gel. Run the RNA marker (RNA decade marker, Ambion Catalog# AM7778) for the size control.

20. Expose it over night on X-ray film along with fluorescent ruler.

3.2.23 **Colony Hybridization**

Adapted from Sambrook, Fritsch, and Maniatis Molecular Cloning (1989) by Vinh Pham.

3.2.23.1 **Materials:**

1. Solution A (10% SDS)

2. Solution B (0.5N NaOH, 1.5M NaCl)

3. Solution C (0.5M Tris-Cl, 1.5M NaCl pH 7.4)

4. Solution D (2X SSC)

5. Nylon-backed Membrane, cut to size of 10mm petri dish

6. Whatman 3mm or other Blotting Paper

7. 15 x 150mm petri dish

3.2.23.2 PROCEDURE:

1. Grow cells to approximate concentration of 50-100 colonies per plate. If desired, replica plate before proceeding (highly recommended).

2. Cut a piece of nylon membrane to the size of the petri dish (plate) to be probed.

3. Soak 4 of the pieces of blotting paper into one of the 4 solutions (A, B, C, or D). Avoid overwetting the paper. It is useful to keep each soaked paper in a 15 x 150mm petri dish.

4. Press the piece of nylon membrane onto the plate to be probed. Mark 3 spots on both the membrane and plate for orientation.

5. Put the piece of nylon membrane, cells side up, over the 4 blotting papers soaked in their respective solutions in the order, and for the incubation times, indicated below:

First Solution A	3 minutes
Second Solution B	5 minutes
Third Solution C	5 minutes
Fourth Solution D	5 minutes

6. Place the membrane, DNA side up, over a dry piece of blotting paper and air dry for 30 minutes.

7. Wrap membrane with saran wrap and UV-crosslink (for 3') DNA onto membrane.

8. Probe and wash blot under same conditions as with standard Southern protocol.

9. Expose to phosphorimager plate, scan on Storm machine, and identify candidate colonies.

10. Pick-and-patch candidate colonies.

11. Repeat steps 1-6 for the pick-and-patch plate.

3.3. Databases and computational tools

Database/ software	URL	Description	Reference
ASD	http://www.ebi.ac.uk/asd	The alternative splicing database	Thanaraj, T.A. et al., 2004.
Alternative splicing workbench	http://www.ebi.ac.uk/asd-srv/wb.cgi?	Combines Intron Analysis, Regulatory sequences, Exon Finder, Splicing Rainbow, and Blast analysis	Thanaraj, T.A. et al., 2004.
ClustalW	http://www.ebi.ac.uk/clustalw/index.html	Multiple Sequence alignment program for DNA or proteins	Larkin, M.A. et al., 2007.
Human BLAT search	http://www.genome.ucsc.edu/cgi-bin/hgBlat	Sequence alignment tool similar to BLAST	Kent, W.J., 2002.

4 Results

The loss of SNORD115/116 and related C/D box small nucleolar RNA (snoRNA) expression units have been implicated as a cause for the Prader-Willi Syndrome (PWS). We recently found that the C/D box snoRNA HBII-52 (SNORD115) changes the alternative splicing of the serotonin receptor 2C pre-mRNA, which is different from the traditional C/D box snoRNA function in non-mRNA methylation. Using bioinformatic predictions and experimental verification, we identified five pre-mRNAs (DPM2, TAF1, RALGPS1, PBRM1 and CRHR1) containing alternative exons that are regulated by MBII-52, the mouse homolog of HBII-52. Recently a report has shownthat the loss of HBII-85 (SNORD-116) cluster can lead to PWS. Using the bioinformatic approach we identified two pre-mRNAs (PTBP1 and HAGHL) as targets for MBII-85, the mouse homolog of HBII-85. We could verify the pre-mRNA targets experimentally. Analysis of a single copy from clusters of MBII-52 and MBII-85 snoRNAs by RNase protection and northern blot analysis showed that these snoRNAs generate shorter RNAs that originate from the full-length snoRNA through additional processing steps. Although the smaller RNA form lacking the hair pin structure is more prevalent, MBII-52 gets processed into shorter fragments ranging from 37 nt to 60 nt. These novel RNAs associate with hnRNPs and not with proteins associated with canonical C/D box snoRNAs. Our data indicate that not the traditional C/D box snoRNAs, but a processed version lacking the snoRNA stems are the predominant MBII-52 and MBII-85 RNAs missing in PWS. This processed snoRNAs function in alternative splice-site selection. Its substitution could be a therapeutic principle for PWS.

The most common technique used to analyze the regulation of an alternative exon is through reporter minigene constructs. However, their construction is time-consuming and is often complicated by the limited availability of appropriate restriction sites. We have developed a fast and simple recombination-based method to generate splicing reporter genes, using a new vector, pSpliceExpress. The system allows generation of minigenes within one week. Minigenes generated with pSpliceExpress show the same regulation as displayed by conventionally cloned reporter constructs and provide an alternate avenue to study splice site selection *in vivo.*

There is no reliable method known to clone small dsRNAs. We have developed a fast and efficient method to clone dsRNA that can be used to clone smaller

dsRNAs. In this procedure two modified oligonucleotides were used as linkers. The applications of this technique involve cloning of dsRNA from RNase Protection Assay and can be used to clone any type of small dsRNAs. The efficiency of this technique is quite high.

4.1 New targets for MBII-52 and MBII-85

The recent finding that HBII-52 regulates alternative splicing of the 5-HT2C receptor (Kishore, S. et al., 2006) raised the question whether there are other targets for this snoRNA. The antisense boxes of the 47 human copies of HBII-52 show up to three sequence variations from their 18 nt consensus sequences (Kishore, S. et al., 2006). We tested HBII-52 variants with one, two, three and five mutations in their antisense box for their ability to change alternative splicing of exon Vb of the serotonin receptor. We found that a snoRNA with three mismatches can still promote exon Vb inclusion (Figure 14). There is no statistically significant change when five mismatches are present in the antisense box. This argues that naturally occurring HBII-52 variants with up to three mismatches between antisense box and target region can influence pre-mRNA processing of the serotonin receptor.
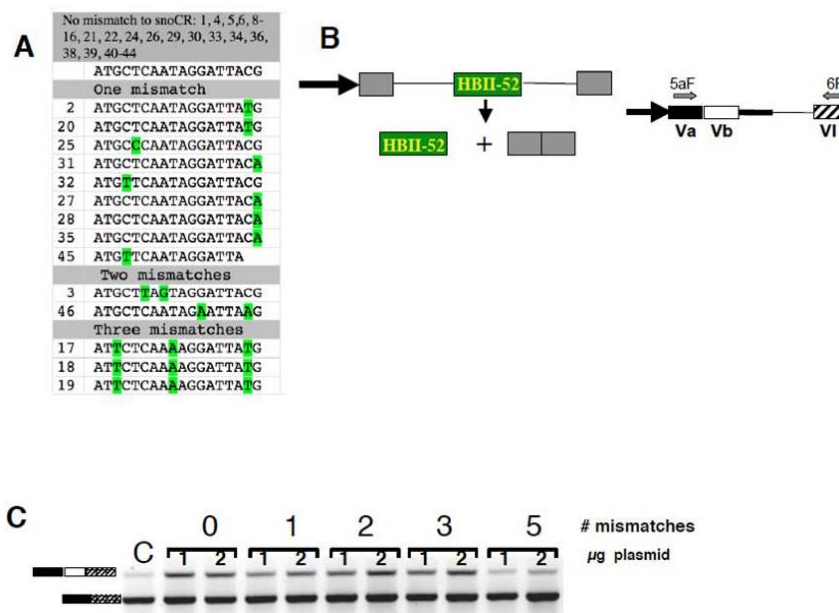
Figure 14: (A) All sequences of the snoRNA antisense boxes from the HBII-52 cluster are shown. Mismatches to the serotonin receptor mRNA are indicated by shading. The first line shows the antisense box with no mismatches to the serotonin receptor 2C pre-mRNA. (B) Test system: (top, left) A CMV promoter-driven construct expresses snoRNAs (box with HBII-52) located in the intron between constitutive, non-coding exons. The effect of the snoRNAs on splice site selection is tested by a splicing reporter, in this case 5-HT2C in cotransfection (right). The amount of snoRNA expression construct is varied in the cotransfections. (C) Influence of snoRNAs with mutations in their antisense box on exon Vb usage. C: control, empty vector.

In order to uncover additional targets of HBII-52 and HBII85, we performed a computational screen in collaboration with Prof. Mihaela Zavolan from Biozentrum, University of Basel. Because the mode of interaction between HBII-52 and its targets is not yet known, we based our analysis on the constraints on snoRNA:rRNA interactions leading to ribose methylation in ribosomal targets (Cavaille, J. et al., 1998). Concretely, we started by extracting an 18-nt-long antisense element upstream of the D box of MBII-52 and 18-nt-long antisense element upstream of the D box and downstream of D' of MBII-85. We defined a putative target site of these snoRNAs a genomic region that can either form a perfect stem of length at least 10 bp or form a duplex of low free energy (below –15 kcal/mol) with the antisense elements, with the duplex satisfying additional constraints. Minimum free energy duplexes were predicted with RNA hybrid (Rehmsmeier, M. et al., 2004) allowing G:U wobble in addition to canonical base pairing.

The constraints on the duplexes were that (i) loops in the duplex were limited to maximum two nucleotides in either the target sequence or in antisense element and (ii) only up to three unpaired nucleotides in any of the sequences was allowed.

Finally, similar to approaches previously employed to predict miRNA targets, we required that the predicted target site be conserved across mammalian species. More specifically, we extracted the regions in the human, rhesus macaque, cow and dog that are orthologous to the predicted HBII-52 and HBII-85 target sites in human and we determined whether they would also be predicted as target sites. As the antisense box of HBII-52 and HBII-85 is highly conserved in mammals, we compared all orthologous genes to the human antisense box sequence. Our final set of predictions included only putative target sites that were conserved in all of these other species. We obtained 457 such sites, 222 of which are in close proximity (200nt) or within known exons. The target mRNAs experimentally verified are listed in Table 3 along with their details. The predictions are available at http://www.mirz.unibas.ch/restricted/piotr/.

Name (exon size)	Description	Exon sequence	Function	Complementarity

DPM2 (91 nt)	Dolichol phosphate-mannose biosynthesis regulatory protein	gagacctccctttctccagGGCCACGGGGACAGACCAGGTGGTGGGACTCGGCCTCGTCGCCGTTAGCCTGATCATCTTCACCTACTACACCGCCTGG<u>GTGATTCTCTTG</u>gtatgtcattctccccgtccgctgctcaccttccccgagccctggcaccgccagagcaactactatata	Frame shift	**target 5'** G C G **3'**
				    GUGAUUCU UUG GUAU
				    CAUUAGGA AAC CGUA
				snoRNA 3' G U U **5'**

TAF1 (157 nt)	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa	tgacccccaactggtctcattcagGAAGGTGAAGATGGAGATGGTGATCTTGCAGATGAAGAGGAAGGAACTGTACAACAGCCTCAAGCCA<u>GTGTCCTGTATGAGGAT</u>TTGCTTATGTCTGAAGGAGAAGATGATGAGGAAGATGCTGGGAGTGATGAAGAAGGAGACAATCCTTTCTCTGgtaggcc	Unknown domain,	**target 5'** A A G **3'**
				    GU GUCCUGU UGAG AU
				    CA UAGGAUA ACUC UA
				snoRNA 3' G U G **5'**

RALGPS1 (158 nt)	Ral guanine nucleotide exchange factor RalGPS1A	ctttccattttccagTATGA<u>TGTGTCAGTTGAGTGT</u>AGTTGAGAGTAAAAGTGCGACATTCCCATCGGAGAAAGCAAGGCACCTACTGGACGACAGTGTCCTAGAGTCCCGCAGCCCCCGAAGGGGCCTGGCTCTGACCTCCTCCTCTGCTGTCACCAATGGACTCTCCCTAGgtaagcg	Frame shift	**target 5'** A A **3'**
				    UGU GUC GUUGAGUGU
				    GCA UAG UAACUCGUA
				snoRNA 3' U GA **5'**

PBRM1 (165 nt)	Polybromo 1	tgtttttact<u>agtcctgttgaatgca</u>atggatggtattttgaattcctgggttaaaaacagaattgaaaatctgaaatgcctttacagAGCGGGCAGCTAAAGTTGCTGAGCAGCAGGAGAGAGAGCGAGCAGCACAGCAACAGCAGCCGAGTGCTTCTCCCCGAGCAGGCACCCCTGTGGGGGCTCTCATGGGGGTGGTGCCACCACCAACACCAATGGGGATGCTCAATCAGCAGTTGACACCTGTTGCAGgtaaaaacaggagctaag	Unknown domain	**target 5'** C AU A **3'**
				    UAGUCCUGUUGA GCA
				    AUUAGGAUAACU CGU
				snoRNA 3' GC A **5'**

CRHR1 (120nt)	Corticotrophin hormone releasing receptor	ctgtgccttaccagccgtctctgcccccagGACTGCAGTGCAACGCAT<u>CCGTGGACCTCATTGGCA</u>CCTGCTGGCCCCGCAGCCCTGCGGGGCAGCTAGTGGTTCGGCCCTGCCCTGCCTTTTTCTATGGTGTCCGCTACAATACCACAAgtaagga	Deletion of hormone binding domain	**target 5'** C A C **3'**
				    CGUGG CCU AUUG GCA
				    GCAUU GGA UAAC CGU
				snoRNA 3' A U A **5'**

Table 3: Genes that showed a dependency on MBII-52 expression both on endogenous and reporter gene level are listed using their HUGO nomenclature (columns 1 and 2). Numbers in parentheses indicate the exon length. The sequence of the regulated exon and its surrounding sequence is shown in column 3. Introns are in small letters, exons in capital letters. The snoRNA complementary region is highlighted in grey and underlined. Column 5 shows the alignment between the MBII-52 antisense box (snoRNA) and its target region

We next tested more than 100 computational predictions experimentally. Neuro2A cells were transfected with either MBII-52 or MBII-52mut, an MBII-52 variant with a scrambled antisense box and the isolated RNA was analyzed by RT–PCR, using primers in the flanking constitutive exons. As shown in Figure 15, we observed a change in alternative splicing patterns in the DPM2, TAF1, RALGPS1, PBRM1 and CRHR1 pre-mRNAs.
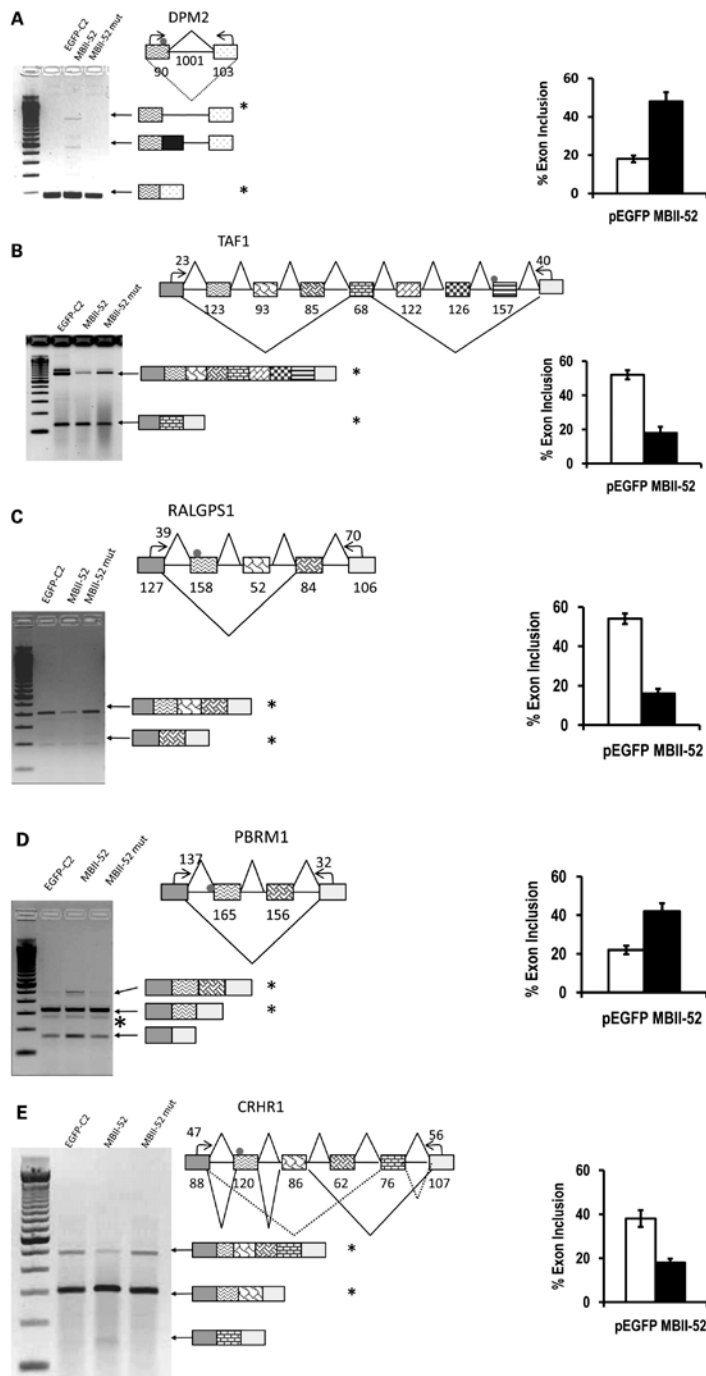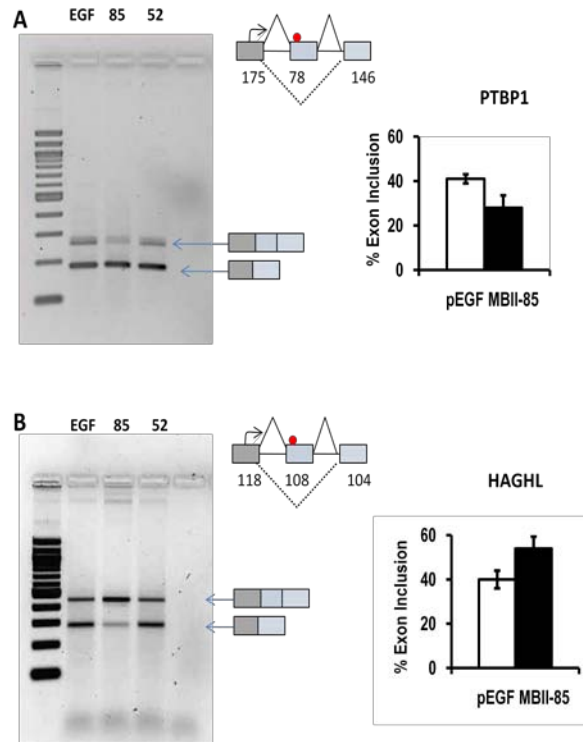

Figure 15: MBII-52 changes the alternative splicing pattern of predicted targets. Computationally predicted MBII-52 target genes expressed in Neuro2A cells were analyzed by RT–PCR. Cells were transfected with 1 µg pEGFP-C2, 1 µg of the MBII-52 expressing construct pCMV/MBII-52 (MBII-52) and 1 µg MBII-52 consensus box mutation, MBII-52 cm, (MBII-52 mut) expressing an antisense box mutation of MBII-52. A representative ethidium bromide-stained agarose gel is shown. The adjacent diagram shows the part of the genes that was analyzed. Small arrows indicate the location of the primers used. The MBII-52 complementarity region is indicated by a dot. Numbers in boxes indicate the length of the exons and numbers next to PCR primers indicate the length of the amplified exon fragment. The structure of the PCR products is indicated by similar shading of exons in cDNA and genomic DNA. The statistical analysis of at least four independent experiments is shown on the right. Stars indicate the bands that were used for quantification.

We identified two target genes PTBP1 and HAGHL as targets for MBII-85. Over expression of the MBII-85 expression construct promoted either inclusion or skipping of the different exons (Figure 16). The validated list of MBII-85 mRNA targets is listed in table 4.

Name (exon size)	Description	Exon sequence	Function	Complementarity

PTBP1 (78 nt)	Polypyrimidine tract binding protein 1	gtctcatttatttctagGTGCACCTGGTATAATCTCAGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCCCACCTTTGCCATTCCTCAAGCTGCAGgtattcaaacgcttggt	Unknown domain	<pre>target 5' C U CA 3'
				           UCCG AUG GGAGC UGG
				           AGGC UAC UCUUG GCC
				snoRNA 3' A CU 5'</pre>

HAGHL (108 nt)	Hydroxyacylglutathione hydrolase-like	caccatcacagTTTGGGGCCATCCACGTGCGATGCCTCTTGACGCCAGGCCACACCTCTGGCCACATGAGCTACTTCCTGTGGGAGGACGACTGCCCGGACTCACCAGCCCTCTTCTCTGgtaccgtg	Unknown domain,	<pre>target 5' C U 3'
				            UUCC GUGGGAGGACG
				            AAGG UACUCUU CUGC
				snoRNA 3' C G C 5'</pre>

Table 4: Genes that showed a dependency on MBII-85 expression both on endogenous and reporter gene level are listed using their HUGO nomenclature (columns 1 and 2). Numbers in parentheses indicate the exon length. The sequence of the regulated exon and its surrounding sequence is shown in column 3. Introns are in small letters, exons in capital letters. The snoRNA complementary region is highlighted in grey and underlined. Column 5 shows the alignment between the MBII-85 antisense box (snoRNA) and its target region.

The cell based over expression assay with MBII-85 expression construct was done. The pEGFP-c2 (EGF) was used as transfection control and MBII-52 was used as negative control (Figure 16). We found that MBII-52 and MBII-85 over expression changes alternative splicing of several endogenous pre-mRNAs.

Figure 16: MBII-85 changes the alternative splicing pattern of predicted targets. Computationally predicted MBII-85 target genes expressed in Neuro2A cells were analyzed by RT–PCR. Cells were transfected with 1 µg pEGFP-C2, 1 µg of the MBII-85 expressing construct pCMV/MBII-85 (MBII-85) and 1 µg MBII-52 expression construct as a control. A representative ethidium bromide-stained agarose gel is shown. The adjacent diagram shows the part of the genes that was analyzed. Small arrows indicate the location of the primers used. The MBII-85 complementarity region is indicated by a line. Numbers below the boxes indicate the length of the exons and numbers next to PCR primers indicate the length of the amplified exon fragment. The structure of the PCR products is indicated by similar shading of exons in cDNA and genomic DNA. The statistical analysis of at least four independent experiments is shown on the right.

4.2 Minigenes construction using pSpliceExpress

pSpliceExpress was constructed by amplifying the ccdB/CmR element from the pDONR221 vector (Invitrogen) and subcloned into the multiple cloning site of Exontrap vector (Mobitec) using *XhoI* and *XbaI* restriction sites as described in section 3.2.19.

The BAC clones (Bacterial artificial chromosome), commercially available through Invitrogen were used to amplify the alternatively spliced exons and their flainking exons. They were cloned into pSpliceExpress as described in section 3.2.20.

The minigenes were used in the minigene assays for validation of target mRNA for snoRNAs MBII-52 and MBII-85. A combination of gene specific and pSpliceExpress construct specific primers were used for PCR amplification.

4.3 MBII-52 and MBII-85 change alternative splicing of targeted pre-mRNAs in reporter gene assays

In the next step of the analysis, we determined whether alternative exons that are influenced by snoRNA expression show this dependency also in a heterologous system, where they are surrounded by a different RNA context. We cloned the regulated exons into an exon-trap vector, where they were flanked by constitutively spliced insulin exons. All constructs were cloned into pSpliceExpress, a system that we developed as described in the section 3.2.20 (Kishore, S., Khanna, A. et al., 2008).

The reporter genes were co-transfected with snoRNA expression constructs into Neuro2A cells and the splicing patterns were analyzed by RT–PCR. As shown in Figure 17, we observed for the splicing events identified in endogenous genes a similar dependency on MBII-52 expression. In Figure 17 the expression of MBII-85 snoRNA and C and D box mutants of MBII-52 (MBII-52cC, cD) did not show a significant effect on the alternative exons, suggesting that the effect is specific for MBII-52. With the exception of PBRM1, the reporter minigenes followed the splicing pattern of the endogenous genes. In the endogenous PBRM1 gene, MBII-52 promoted both inclusion and skipping of two exons located in a cluster of alternatively spliced cassette exons. In the heterologous system, we observe only the skipping event for PBRM1. This difference is most likely due to the presence of strong insulin exons in pSpliceExpress that interfere with the arrangement of regulatory sequences in this cluster of multiple alternative cassette exons. Finally, we created a series of compensatory mutations in the antisense box of MBII-52 and the snoRNA complementarity regions (snoCR) of its targets. These experiments proved inconclusive, as in most cases mutating the snoCR resulted in strong exon activation that was no longer susceptible to regulation.
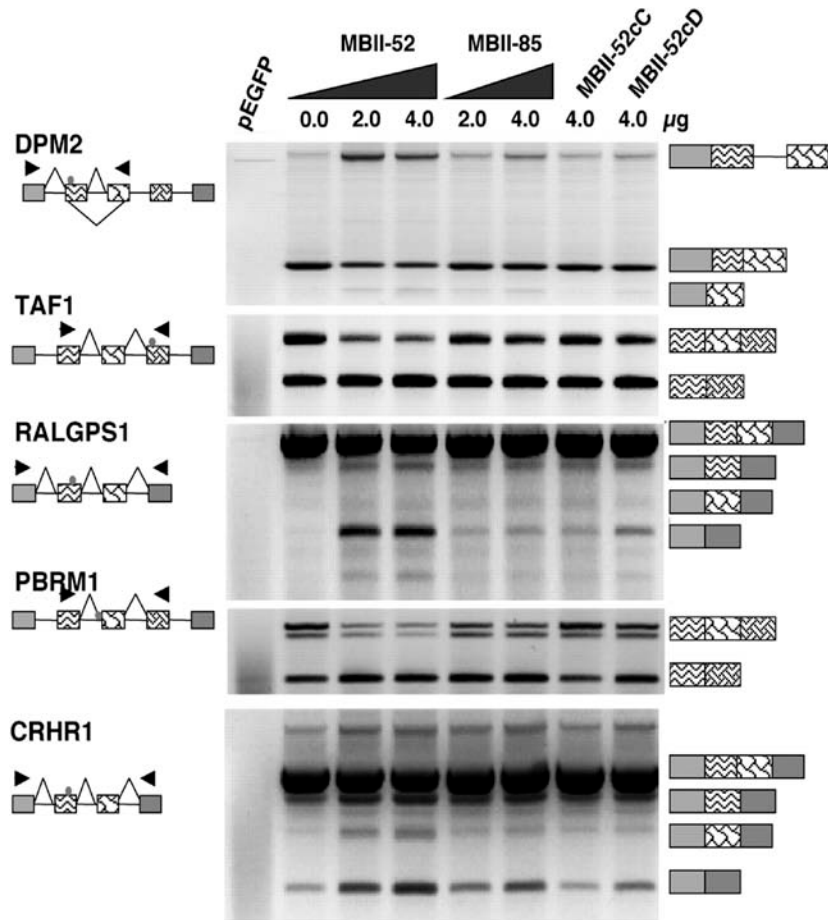
Figure 17: Minigene analysis of MBII-52 target genes. The exons harboring the MBII-52 complementary region were subcloned into the exon trap vector pSpliceExpress. The structure of the resulting constructs pSE-RALGPS1, pSE-CRHR1, pSE-DPM2, pSE-PBRM1 and pSE-TAF1 as well as the location of the primers used for RT–PCR analysis is indicated on the left. pEGFP: only an expression construct for GFP is transfected. All other lanes contain 1 µg of pSE-reporter. MBII-52: cotransfection with 2 and 4 µg of MBII-52 expression construct, MBII-85: cotransfection with 2 and 4 µg of an MBII-85 expression construct, MBII52cC: cotransfection with 4 µg of a C-box mutant of MBII-52: MBII52cD: cotransfection with 4 µg of a D-box mutant of MBII-52. The structure of the products is shown schematically on the right, using the same shading scheme as in Figure . The usage of alternative exons indicated with a triangle was statistically evaluated. The comparison between MBII-52 and MBII-85 transfected cells showed statistically significant differences, the P-values of the Student's t-test were: DPM2: 0.001, TAF1: 0.023; RALGPS: 0.021; PBRM1: 0.076 and CRHR1: 0.002; (n = 4).

To test MBII-85 target genes, minigene constructs of PTBP1 and HAGHL were made using pSplice express. These mingene constructs were contransfected with MBII-85 expression construct and MBII-52 was used as negative control. As shown in Figure 18 we observed the splicing events following the pattern observed in over expression studies.
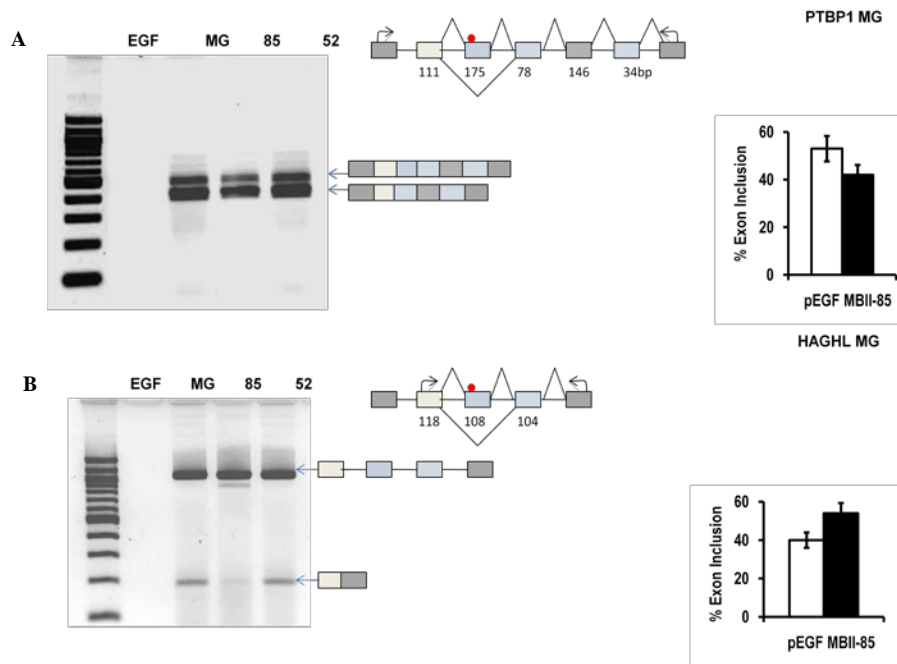
Figure 18: Minigene analysis of MBII-52 target genes. The exons harboring the MBII-85 complementary region were subcloned into the exon trap vector pSpliceExpress. The final clones being pSE PTBP1 and pSE HAGHL. Exons structures are shown as colored blocks and red line shows the target comeplemtarity with MBII-85. Arrows siginifies the position of the primers for RT –PCR analysis.

Together, these data suggest that after being transferred into a heterologous gene context at least five alternative exons are influenced by MBII-52 expression and two alternative exons are influenced by MBII-85 expression.

4.4 A mouse model of PWS shows changes in the predicted exons

To address the physiological significance of our data, we asked whether MBII-52 influences alternative splicing of the identified target genes *in vivo* and analyzed RNA samples from the TgPWS mouse model (Stefan, M. et al., 2005). TgPWS mice have a paternally derived deletion of the PWS critical region that contains the SNURF–SNRPN locus. They show hormonal and metabolic defects resembling those of human newborns with PWS (Stefan, M. et al., 2005). As a larger locus is deleted, in addition to MBII-52, the mice do not express MBII-85 and other snoRNAs from the Prader–Willi critical region.

We compared RNA from newborn TgPWS mice with RNA from littermates expressing the region. As shown in Figure 19, we found that the mouse knockout system recapitulates a dependency of alternative splicing on the presence of MBII-52. However, the overall splicing patterns of the endogenous genes are different in mouse brain and

Neuro2A cells. This most likely reflects the presence of numerous cell types in brain that show different splicing patterns. Despite this limitation, the presence of MBII-52 promotes exon inclusion in the alternative exons with a complementarity to MBII-52 of the DPM2 and PBRM1 pre-mRNAs and promotes skipping of the RALGPS1 and TAF1 exons, similar to the effect seen in Neuro2A cells. The only discrepancy between the MBII-52 effects in brain and Neuro2A cells was an alternative exon of CRHR1 that showed an increase in exon usage in brain tissue, whereas it showed a decrease in response to MBII-52 in Neuro2A cells. The regulated alternative CRHR1 exon is in a cluster of alternative exons and the discrepancy could be due to differences in splicing regulators between brain and Neuro2A cells. Collectively, the data suggest that the loss of MBII-52 expression influences alternative splicing of target genes in a physiological context.
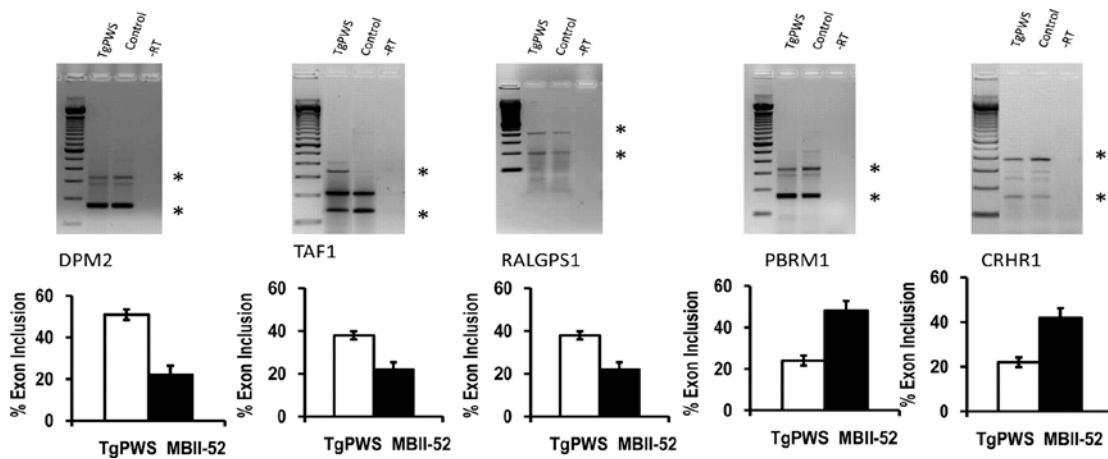
Figure 19: Comparison of RNA from TgPWS mice lacking MBII-52 expression and control littermates. Total brain samples from TgPWS mice lacking expression of the Prader–Willi critical region were compared with normal littermates expressing all the snoRNAs from the PWS critical region (control).

We compared RNA from 21 days old MBII-85 Knockout mice RNA from littermates expressing the MBII-85 snoRNAs. As shown in Figure 20, we found that the mouse knockout system shows a dependency of alternative splicing of MBII-85 target mRNA just like over expression in endogenous system. However, the overall splicing patterns of the endogenous genes are different in mouse brain and Neuro2A cells. This most likely reflects the presence of numerous cell types in brain that show different splicing patterns. Despite this limitation, the presence of MBII-85 promotes exon inclusion in the alternative exons with a complementarity to MBII-85 of the PTBP1 and promotes skipping of the HAGHL exon, similar to the effect seen in Neuro2A cells.
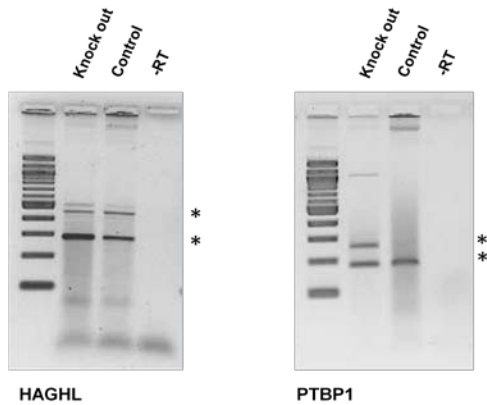
Figure 20: Comparison of RNA from knockout mice lacking MBII-85 expression and control littermates. Total brain samples from MBII-85 Knockout mice were compared with normal littermates expressing MBII-85 snoRNAs from the PWS critical region (control). RNA sample was used as RT control to rule out DNA contamination.

4.5 MBII-52 is processed into smaller RNAs

The data indicate that MBII-52 expression influences usage of multiple exons that contain regions with sequence complementarity to the antisense-box of MBII-52. Four recent studies reported that H/ACA snoRNAs give rise to smaller RNAs (Ender, C. et al., 2008; Saraiya, A.A. et al., 2008; Scott, M.S. et al., 2009; Hutzinger, R. et al., 2009). We therefore tested whether the C/D box snoRNA MBII-52 also gives rise to other RNAs by RNase protection assay.

Whereas humans have 47 HBII-52 copies, there are at least 130 copies of MBII-52 snoRNAs in mouse. We used an antisense probe against the MBII-52 copy employed in transfection experiments described above. This isoform is 87 nt in length and its sequence is shown in Figure 23 as form A. In silico analysis shows that this copy shares only an uninterrupted stretch of 20 nt in the antisense box region with other snoRNA isoforms of the MBII-52 cluster. All other regions show single nucleotide differences that prevent longer protected fragments. For the analysis, we used an *in vitro* transcribed, ^{32}P labeled RNA-antisense probe that detects the 87 nt encompassing the full-length snoRNA. Together with linker and vector sequences, the probe is 175 nt in length. After hybridization, RNase A and T1 digestion, the fragments were separated on 15% acrylamide/TBE/8M urea gels. As shown in Figure 21, lane 1, we observed additional fragments when total mouse brain RNA was analyzed with this probe. In agreement with earlier studies, we do not detect expression in liver (Cavaille, J.et al., 2000) (Figure 21, lane 9). We then asked whether the snoRNA expression construct used in Figure 15 is processed in a similar way. We analyzed total RNA from Neuro2A cells transfected with the pCMV/MBII-52 expression construct (Figure 21, lane 2) and found a similar RNA pattern. Importantly, the most abundant RNA species from both the expression construct

and brain is shorter than 80 nt (form B, Figure 21). SnoRNAs contain C and D boxes that stabilize the snoRNP. Mutation of these RNA elements abolished the effect on splicing (Figure 15). We therefore tested expression from constructs expressing this mutants and could not detect any RNA expression (Figure. 21, lanes 6 and 7), suggesting that the smaller RNAs (form B, C, D) derive from a precursor with intact C and D boxes.
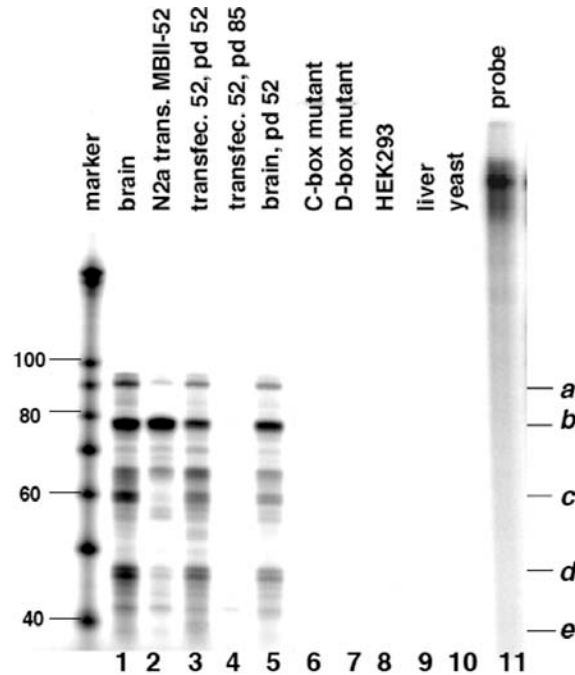
Figure 21: RNase protection analysis using a probe detecting the MBII-52 copy used in transfection studies in Figures 12 and 13. Five microgram of the following total RNAs was hybridized to an MBII-52 antisense probe: (1): total mouse brain, (2): Neuro2A cells transfected with pCMV/MBII-52. Lanes (3–5) are protections from RNPs captured with oligonucleotides against the antisense box (Figure. 5). (3): Affinity captured RNA from Neuro2A cells expressing MBII-52 using a MBII-52 probe for pull down (pd), (4): affinity captured RNA from Neuro2A expressing MBII-52 using an MBII-85 probe (negative control) and (5): affinity captured RNA from brain using an MBII-52 probe for pull down. Lane 6: RNA from Neuro2A cells transfected with an expression construct containing a C-box mutant, (7): RNA from Neuro2A cells transfected with an expression construct containing a D-box mutant, (8): HEK293 cells non-transfected, (9, 10): RNA from liver and yeast. (11): Undigested probe. The marker is a 100 nt RNA base ladder.

It is possible that MBII-52 undergoes nucleotide modifications that would result in mismatching of an RNase protection probe and subsequent generation of smaller fragments. To rule out this possibility, we performed northern blot analysis, using denaturing 15% PAGE gels. Total RNA from brain, liver and spleen was probed with MBII-52 antisense RNA corresponding to the sequence in Figure 23, form A. Even after stringent washing, we see cross-hybridization of MBII-52 with RNAs from liver, spleen and HEK293 cells (Figure 22 (A)). This is to be expected, as there are numerous copies of

sequence-related snoRNAs expressed in these tissues (Cavaille, J. et al., 2000). To detect the specific hybridization between MBII-52 form A and brain RNA, we treated the membrane with RNase A and RNase T1. The RNase treatment reduced the overall signal strength, as we had to use a 3-fold longer exposure time. As shown in Figure 22 (B), this treatment abolishes the cross-hybridization with non-brain RNAs. However, this treatment does not abolish the signal from brain RNA tissue that corresponds in length to RNA forms B–D. Similar to the RNA protection experiment; the major RNA species is shorter than 80 nt. This indicates that the protection pattern observed in the protection assay is due to shorter RNAs and not the result of nucleotide editing. Unexpectedly, we observe a distinct pattern of shorter RNAs and not a continuous smear of bands. This finding implies that all of the estimated MBII-52 copies are processed in a similar way, giving rise to specific metabolically stable short RNAs.
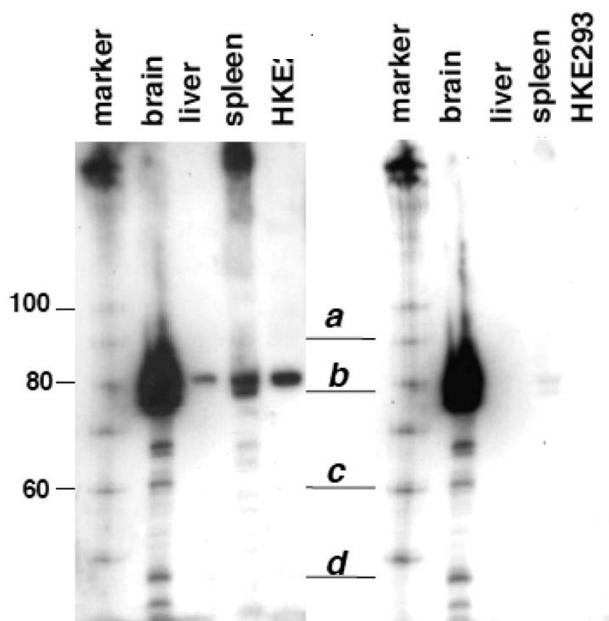
Figure 22: (A) Northern blot analysis of MBII-52. Fifteen microgram total RNA from brain, liver, spleen and HEK293 cells was separated on 15% polyacrylamide gels and probed with a 32P labeled probe for MBII-52. After stringent washing, cross-hybridizing bands in liver, spleen and HEK293 cells still remain. Exposure was overnight. (B) The filter from (B) was treated with RNase A/T1 and again washed. The cross-hybridizing bands disappear, but the signals corresponding to smaller bands remain. Exposure was for 3 days.

To determine the identity of the novel short RNAs, we cloned the protected fragments using the method we developed, described in section 4.8). The positive clones are shown in Figure 23. All shorter RNAs lack the sequences forming the stem of the snoRNA, but contain the C and C' box. The stem conveys complementarity between the snoRNA ends and stabilizes the snoRNP. The remaining cloned RNAs are shortened from the 5' and 3' ends, indicating that they are generated by 3'-5' and 5'-3' exonuclease activity that stops at the C and C' boxes.

|	Stem	C box	D'box	C' box	Antisense box	D-box Stem
form A (87 nt)	ACTGGGTCAAUGAUGACAACCCAAUGUCAUGAAGAAAGGUGAUGACAUAAAAUUCAUGCUCAAUAGGAUUACGCUGAGGCCCAACCA
form B (73 nt)	AAUGAUGACAACCCAAUGUCAUGAAGAAAGGUGAUGACAUAAAAUUCAUGCUCAAUAGGACUACGCUGAGGCC
form C (60 nt)	GGUCAAUGAUGACAACCCAAUGUCAUGAAGAAAGGUGAUGACAUAAAAUUCAUGCUCAAU
form D (44 nt)	UGGGUCAAUGAUGACAACCCAAUGUCAUGAAGAAAGGUGAUGAC
form E (38 nt)	GGUCAAUGAUGACAACCCAAUGUCAUGAAGAAAGGUGA
form E (37 nt)	AUGAUGACAACCCAAUGUCAUGAAGAAAGGUGAUGAC

Figure 23: Sequences of the shorter RNAs. The stems and functional boxes are indicated. The clones are ordered according to their length. Form A corresponds to the published snoRNA MBII-52. Underlined nucleotides in forms C and D indicated predicted stems.

Together, these data suggest that the expression unit consisting of MBII-52 and its flanking intron and exons gives rise to several RNAs. These RNAs include the previously described MBII-52 snoRNA (form A), as well as shorter RNA species. The major RNA species (form B) expressed from the MBII-52 cluster lacks the stem box, but still contain C and D boxes.

4.6 Processed MBII-85 RNA patterns different from MBII-52

RNA protection assay was performed using MBII-85 specific probe to check the processing pattern of MBII-85. We used 100 µgs of RNAs of different mouse tissues (Adrenal gland, Thymus, Heart, Muscle, Spleen, Testis and Kidney). Whereas humans have 29 HBII-85 copies, there are at least 20 copies of MBII-85 snoRNAs in mouse. We used an antisense probe against the MBII-85 copy employed in transfection experiments described above. This isoform is 97 nt in length and its sequence is shown in Figure 26 as form A. In silico analysis shows full sequence similarity with all the snoRNAs MBII-85 clusters with the exception of two copies. These two snoRNA copies show single nucleotide differences that prevent longer protected fragments. For the analysis, we used an *in vitro* transcribed, ^{32}P labeled RNA-antisense probe that detects the 97 nt encompassing the full-length snoRNA. Together with linker and vector sequences, the probe is 175 nt in length. After hybridization, RNase A and T1 digestion, the fragments were separated on 15% acrylamide/TBE/8 M urea gels. As shown in Figure 24, (brain), we observed additional fragments when total mouse brain RNA was analyzed with this probe. In agreement with earlier studies when we analyzed MBII-52 but the intensity of the smaller band is different, we do not detect expression in thymus and kidney (Cavaille, J. et al., 2000) (Figure 24). Then we used the snoRNA expression construct used in Figure 16 is to see if it does get processed in a similar way. We analyzed total RNA from Neuro2A cells transfected with the pCMV/MBII-52 expression construct (Figure 24) and found a similar RNA pattern. Importantly, the most abundant RNA species from both the

expression constructs and brain is shorter than 97 nt (form B, Figure 26). SnoRNAs contain C and D boxes that stabilize the snoRNP.
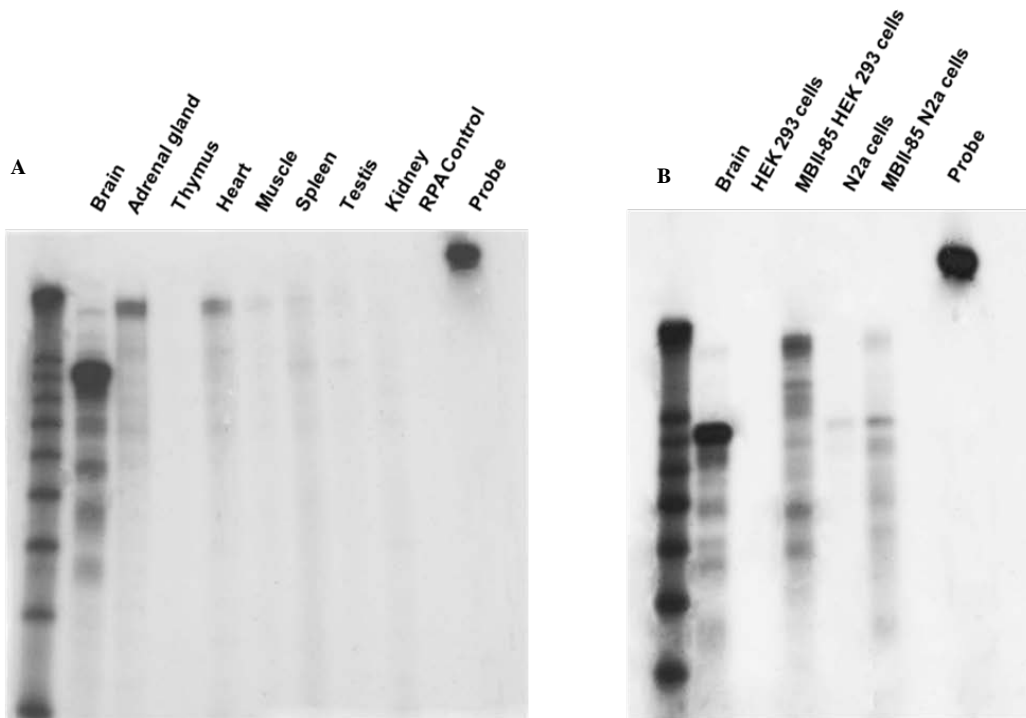
Figure 24: Auto radiographs showing RNase protection assays using MBII-85 antisense probe. (A) 10 micrograms of the following total RNAs from brain, adrenal gland thymus heart muscle spleen testis kidney were hybridized to an MBII-85 antisense probe. (B) 10 micrograms total RNAs from brain, HEK293 cells, MBII-85 over expressed in HEK 293 cells, N2a cells and MBII 85 over expressed in N2a cells were hybridized to an MBII-85 antisense probe.

As in case of MBII-52, to rule out the possibility that MBII-85 undergoes nucleotide modifications that would result in mismatching of an RNase protection probe and subsequent generation of smaller fragments. We performed northern blot analysis, using denaturing 15% PAGE gels. Total RNA from brain, liver and spleen was probed with MBII-85 antisense RNA corresponding to the sequence in Figure 26, form A. Even after stringent washing, we see cross-hybridization of MBII-85 with RNAs from spleen testis and heart (Figure 25 A). This is to be expected, as there are numerous copies of sequence-related snoRNAs expressed in these tissues (Cavaille, J. et al., 2000). To detect the specific hybridization between MBII-85 form A and brain RNA, we treated the membrane with RNase A and RNase T1. The RNase treatment reduced the overall signal strength, as we had to use a 3-fold longer exposure time. As shown in Figure 25 B, this treatment abolishes the cross-hybridization with non-brain RNAs. However, this treatment does not abolish the signal from brain RNA tissue that corresponds in length to RNA forms

B. Similar to the RNA protection experiment; the major RNA species is shorter than 97 nt. This indicates that the protection pattern observed in the protection assay is due to shorter RNAs and not the result of nucleotide editing. On the same lines like MBII-52, we observe a distinct pattern of shorter RNAs and not a continuous smear of bands. The smaller protected RNAs were cloned using the technique described in section 4.8. This finding implies that all of the estimated MBII-85 copies are processed in a similar way, giving rise to specific metabolically stable short RNAs.
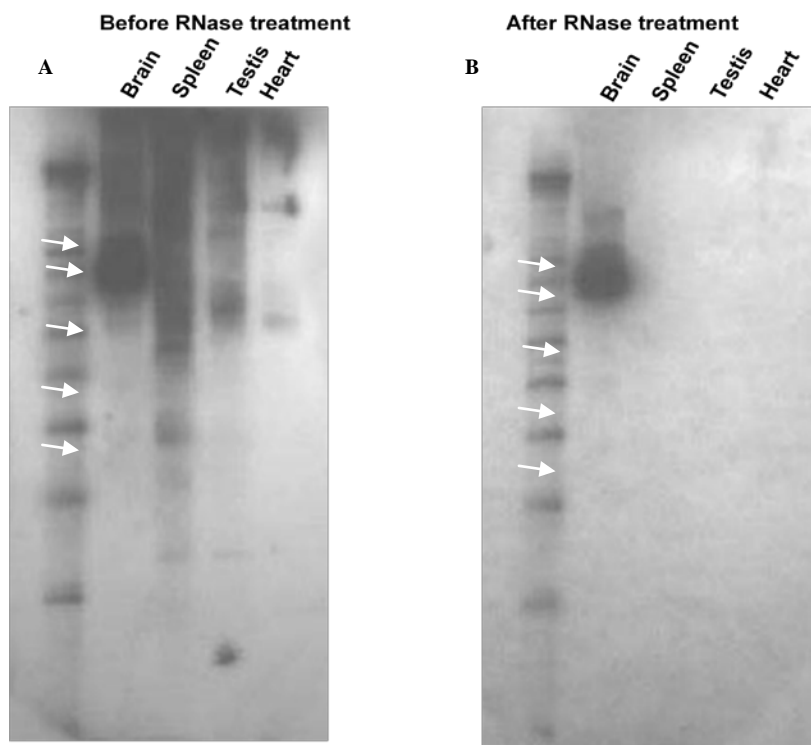
Figure 25: **Northern blot analysis of MBII-85.** Fifteen microgram total RNA from brain, spleen, testis and heart was separated on 15% polyacrylamide gels and probed with a ^{32}P labeled probe for MBII-85. (A) Shows an auto- radiograph of the northern blot before RNase A/T1 treatment. (B) showing the autoradiograph showing northern blot after RNase A/T1 treatment.

The positive clones are shown in Figure 26. The shorter RNA lack the sequences forming the stem of the snoRNA, but contain the C and C' box. The stem conveys complementarity between the snoRNA ends and stabilizes the snoRNP. The cloned RNAs are shortened from the 5' and 3' ends, indicating that they are generated by 3'->5' and 5'->3' exonuclease activity that stops at the C and D boxes.
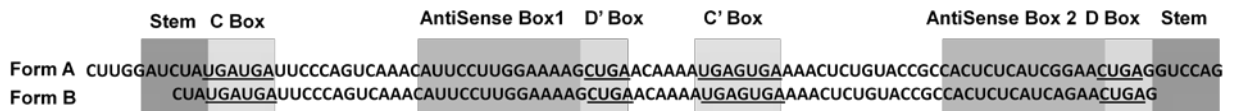
Figure 26: Sequences of the shorter RNAs. The stems and functional boxes are indicated. The clones are ordered according to their length. Form A corresponds to the published snoRNA MBII-85. The dark shaded area at the extreme ends shows the predicted stem.

Together, these data suggest that the expression unit consisting of MBII-85 and its flanking intron and exons gives rise to two smaller RNA forms. These RNAs include the previously described MBII-85 snoRNA (form A), as well as shorter RNA species. The major RNA species (form B) expressed from the MBII-85 cluster lacks the stem box, but still contains C and D boxes.

4.7 MBII-52 and MBII-85 derived RNAs do not bind to classical snoRNA-associated proteins

As we found that the MBII-52 locus gives rise to previously not described products, we identified the proteins that associate with these RNAs. We used the affinity between a biotinylated 2'-O-methylated oligonucleotide and the antisense box of MBII-52 to isolate RNAs derived from the MBII-52 locus (Figure 27). Using streptavidin beads, we isolated the MBII-52 snoRNA particle (snoRNP) from nuclear extracts generated from cells transfected with the MBII-52 expression construct. Nuclear extract was generated by a scaled-down Dignam procedure (Lee, K.A. et al., 1988). After washing with 100 and 200 mM NaCl, the captured material was separated by SDS–PAGE and proteins were identified by mass spectrometry and database matching. An oligonucleotide against the snoCR of MBII-85 was used as the control. As shown in Figure 28, we found that hnRNPs were associated with the expressed snoRNA. Similar results were seen with samples obtained from mouse brain nuclear extracts. Repeated experiments using different washing and isolation methods to find canonical snoRNP proteins, such as fibrillarin or NOP56, in pulled-down material from MBII-52 affinity material failed to identify known snoRNP-associated proteins.

Figure 27: Experimental strategy: a biotinylated oligonucleotide is used to capture the snoRNP complex. The oligonucleotide is shown as a line, the complementarity is dashed, biotin is shown as a circle. The RNP complex (boxes) is isolated by streptavidin (half-circle)-capture from extracts expressing MBII-52 and washed in non-denaturing buffer. The extracts were prepared by transfecting MBII-52 expression constructs and performing Dignam mini-extracts.

We determined which RNAs are present in the pulled-down material and performed RNase protection. As shown in Figure 21, lane 4 and 5, we found that the isolates contained the smaller MBII-52-related RNAs, as well as the full-length MBII-52 snoRNA. No MBII-52 RNA was pulled-down with the probe against MBII-85, suggesting the selectivity of the pull-down.
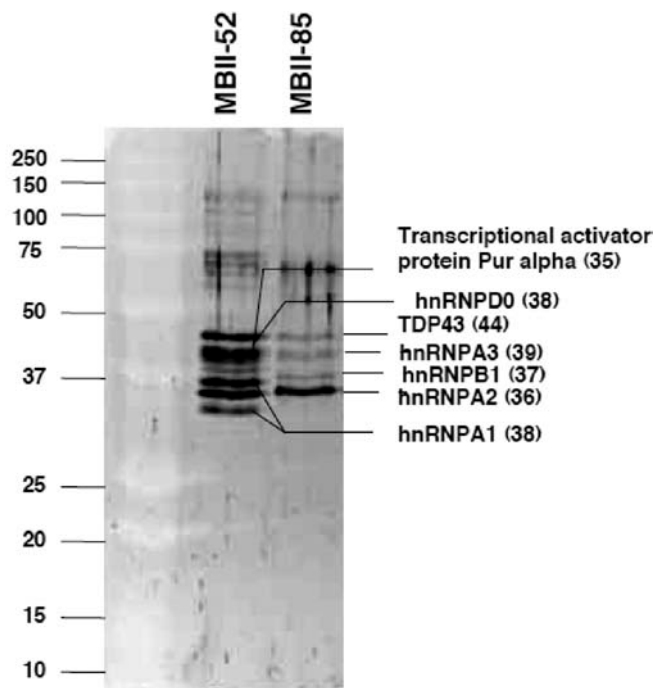
Figure 28: Silver stain of a gel with affinity purified RNPs. MBII-52: the RNPs were isolated with a capture-oligonucleotide against MBII-52. MBII-85: the RNP was isolated with a capture oligonucleotide against MBII-85. Proteins were determined by mass spectrometry and are indicated. Sizes in kDa are shown in parentheses.

In summary, the findings indicate that the shorter RNAs assemble with hnRNPs, but not with proteins that have previously been described to associate with C/D box snoRNAs. Although the major RNA isoform B contains C and D boxes, structural

hallmarks of C/D box snoRNAs, the composition of the RNP formed is different from a C/D box snoRNP.

## 4.8 Cloning of Double stranded RNAs using Modified oligonucleotides

4.8.1 Objective

Double stranded RNA (dsRNAs) have been cloned only from the the viral systems, where the dsRNA represents the largest amount of nucleic acids. Currently there is no system available that allows cloning of small amounts of dsRNAs. We tested one published method and found it could not detect small amounts of dsRNA present in a RNase protection assay. Due to unavailability of an efficient method to clone dsRNA a fast and efficient cloning procedure for double stranded RNAs (dsRNA) was developed. The known methods to clone single stranded RNA involve a prerequisite that the information of the RNA sequence should be known or at least it should be a processed mRNA that has a polyA tail. A polyT primer can be used to reverse transcribe such RNAs and subsequently amplified using PCR. The PCR product can be cloned and sequenced. However this method cannot be used to clone dsRNA. The methods used to clone dsRNA viral genome have a limitation that it cannot clone short dsRNA and is ineffective if the dsRNA amount is small (Potgieter, A.C. et al., 2002). So the objective was to develop a technique which can be used to clone dsRNA of any length. The method should be sensitive to allow cloning of small amount of dsRNA.

4.8.2 Method overview

Total mouse brain RNA was subjected to RNase protection to isolate dsRNA. Subsequently, the RNases were removed by Proteinase K treatment and phenol extraction. The double-stranded RNA was phosphorylated using T4 kinase and polished using T4 DNA polymerase and an adenylated-linker was ligated in the absence of ATP (Lau, N.C. et al., 2001). After gel purification and isolation, an adapter linker was ligated using T4 DNA ligase. The reaction was subsequently reverse transcribed, amplified and cloned. In this procedure two modified oligonucleotides were used as linkers. Linker 1 (3'linker) was an activated 5' adenylated-oligonucleotide. It had 3' end blocked with Dideoxycytidine (ddC) to prevents self ligation. The linker 2 (5' linker) had a 5' modification with a primary amino group with a standard (C6) spacer arm attached to prevent ligation to its 5'end.

Linker 2 is a chimeric oligonucleotide with 21 RNA bases at the 3'end to allow efficient ligation to RNA (Figure 29).
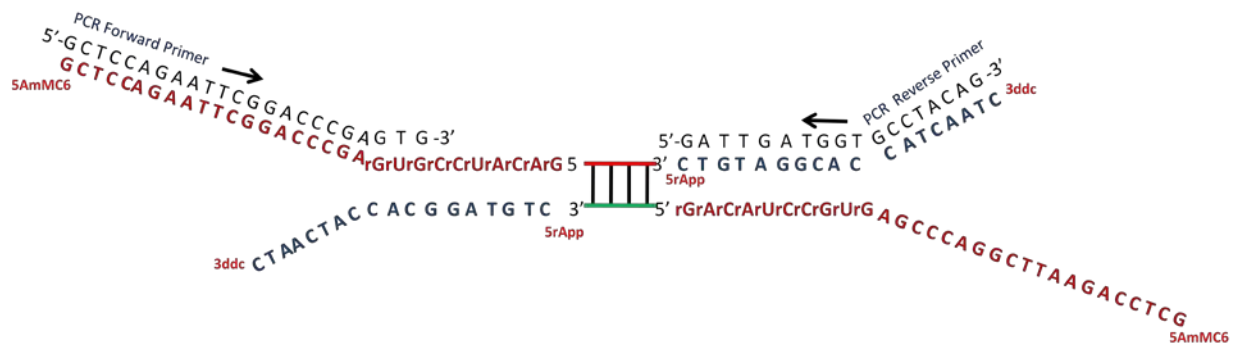
Figure 29: Modified oligonucleotides: red and green lines separated by black lines depict the dsRNA and the oligonucleotides shown are the modified oligonucleotides used to ligate on to the ds RNA

4.8.3 Procedure

4.8.3.1 Modified RNase Protection assay and dsRNA elution

Total mouse RNA was hybridized overnight with the ^{32}P αUTP labeled MBII-85 antisense probe. After RNase A/T1 digestion the reaction was phenol/ chloroform extracted and concentrated by addition of 1/5 volume of ammonium acetate with 2.5 volume of ethanol. The precipitation was done at -20°C for 1 hr. The hybridized RNAs were pelleted and air dried. The pellet was dissolved in DEPC treated water and was T4 PNK treated. The objective is to remove RNases. The PNK treated hybridized RNAs were run on the 15% denaturing Acrylamide gel to remove single nucleotides. The gel was exposed overnight on X-ray film and a flurosent marker was aligned to obtain the gel orientation.

The X-ray film was developed and the protected fragments were cut out of the gel. The gel fragments were crushed using a RNase free pestle. The elution buffer was added and the reaction was incubated at 37°C with shaking (Figure 24). Tubes were spun to pellet the gel fragments. The dsRNA was in the supernatant. The dsRNAs from the supernatant were concentrated by adding an equal volume of ethanol. The pellet was air dried and re-suspended in DEPC treated water.

The stepwise procedure is as follows:

1. Mix 10µg of sample RNA with labeled probe about 150-600pg or $2x10^4$ cpm. And mix 100µg of sample RNA with unlabeled probe. Use yeast RNA provided in the kit to set up two controls for each probe (RPAIII Kit, Ambion Catalog. No. AM 1414).

2. Co-precipitate the probes with sample RNA by adding

   $1/5^{th}$ Volume of 5M NH_4OAc

3. Add equal volumes of ethanol and mix thoroughly and allow RNA to precipitate at -20°C for 30 minutes

4. Pellet the RNA by centrifuging at $\geq$10,000 x g for 15 minutes.

5. Discard the supernatant and air dry the pellet for 5 minutes.

6. Resuspend the pellet in 10µl of Hybridization buffer and denature at 95°C for 4 minutes.

7. Centrifuge to collect the reaction at bottom of the tube and hybridize overnight at 42°C.

8. Perform the RNase digestion using 1:100 dilution of RNaseA/RNaseT1 in 150 µl of RNA digestion buffer (10mM Tris-HCl, pH 7.5, 5mM EDTA, 200mM NaAc) supplied with the kit, for each sample. Leave aside one control tube as RNase digestion control by adding just RNase digestion buffer with the RNases.

9. Incubate the reaction for 30 minutes at 37°C.

10. Perform extraction with 40µl of Phenol/chloroform and concentrate by ethanol precipitation with $1/10^{th}$ volume of NH_4Ac and add 2-3 µg of glycogen solution 1ug/ul (glycol blue Ambion Catalog. No. AM9515).

11. Chill the tubes at -20°C for 30 minutes.

12. Centrifuge at$\geq$10,000 x g for 15 minutes, remove the supernatant and air dry the pellets for 10 minutes.

13. Re-suspend the pellet in 35 µl of DEPC treated water and do the T4 Polynucleotide Kinase treatment by adding the following reagents to the re-suspended pellet. T4 DNA polymerase is used for blunt end polishing.

10x T4 PNK Buffer	5 µl
rNTPS (10mM)	5 µl
T4PNK 1U	0.5 µl
T4 DNA polymerase	1µl

14. Incubate the reaction at 37°C for 10 minutes and stop the reaction by adding 2 µl of 0.5M EDTA.

15. Perform extraction with 40µl of phenol/chloroform and concentrate by ethanol precipitation with $1/10^{th}$ volume of NH_4Ac.

16. Chill the tubes at -20°C for 30 minutes.

17. Centrifuge at≥10,000 x g for 15 minutes, remove the supernatant and air dry the pellets for 10 minutes.

18. Re-suspend the pellet in Gel loading buffer II (95% formamide. 0.025% xylene cyanol, 0/025% bromophenol blue, 18 mM EDTA and 0.025% SDS).

19. Denture the reactions at 95°C for 3 minutes and run on 15% Acrylamide/8M Urea denaturing gel. Let the lower dye to get out of the gel. Run the RNA marker (RNA decade marker, Ambion Catalog# AM7778) for the size control.

20. Expose it over night on X-ray film along with a fluorescent ruler.

21. Develop the overnight exposed film. The lanes in which hot probe was used will be visible. Cut out the gel pieces of the protected bands and also cut out the corresponding length gel pieces from the lanes where cold probe was used.

22. Crush the gel pieces in the tube with tube pestle and suspend it in 250 µl of elution buffer.

23. Incubate it with shaking overnight at 37°C.

24. Precipitate by adding 3 volumes of ethanol and glycogen blue solution (10µg/µl, Invitrogen) and chilling at -20°C for 30 minutes.

25. Centrifuge the tubes at≥ 10,000 x g for 15 minutes. Remove the supernatant and air dry the pellet for 10 minutes.

26. Re-suspend the pellet in 10 µl of DEPC treated water.

4.8.3.2   Ligation of the 3' Linker and purification

The re-suspended and precipitated dsRNA was ligated with linker 1 (3'linker). In this ligation reaction T4 RNA ligase was used to ligate the 3'linker. The final concentration of the linker was 4µM and 1U of RNA ligase was used. The ligation buffer was without ATP as the linker itself is adenylated and activated. The ligation reaction was supplemented with 10% PEG3500 which improves the ligation efficiency. The ligation reaction was incubated at room temperature for 2 hours and was run on the 15% denaturing acrylamide gel (Figure 30 A). The ligation reaction was run in parallel with a control RPA reaction of mouse brain RNA using MBII-85 probe to observe the shift in the band length corresponding to linker length. The gel was exposed to X-ray film overnight and a florescent marker was aligned with the gel for orientation. The X-ray film was developed and the shift in the band in the lane (Figure 30 A, RPA+ linker) showed the 3' linker addition reaction was successful. The purpose of running this gek is two fold (i) A mobility shift indicates a successful ligation and (ii) This removes the surplus linker. The shifted bands from the gel were cut out and crushed using a tube pestle. The elution buffer was added and incubated overnight at 37°C with shaking.

Stepwise procedure is as follows:

1. Ligate the Linker1 by using RNA ligase. Set up the 15 µl reaction as follows:

RNA Ligation buffer (without ATP)	2 µl
Adenylated 3' Linker (Linker 1)	2 µl
Re-suspended pellet	10 µl
T4 RNA ligase 1U	1 µl

2. Incubate at 37°C for 2 hrs.

3. Add RNA gel loading buffer II and denature at 95°C for 4 minutes

4. Run the reactions on 15% Acrylamide/8M urea TBE gel.

5. Expose the gel on X-ray film overnight.

6. Develop the film; look for the shift in the mobility when the linker is present. If the shift is visible cut out the bands.

7. Crush the cut out gel pieces in a 1.5 ml tube with tube pestle and suspend in the gel elution buffer.

8. Centrifuge at 5000 rpm for 2 minutes and collect the supernatant.

9. Let the tubes shake at 37°C overnight.

10. Briefly spin the tubes and precipitate the supernatant with 3 volumes of ethanol and add 1μl of glycogen solution.

11. Chill the tubes at -20°C for 30 minutes.

12. Centrifuge the tubes at≥ 10,000 x g for 15 minutes. Remove the supernatant and air dry the pellet.

13. Re-suspend the pellet in DEPC treated water

4.8.3.3   Ligation of the 5' Linker

The eluted dsRNA with the 3'linker ligated dissolved in DEPC treated water was used for the 5'Linker ligation. In the 5' ligation reaction T4 DNA ligase and T4 RNA ligase were both used and supplemented with PEG3500. The ligation buffer used was the standard ligation buffer with ATP.

The stepwise procedure is as follows:

1. Ligate the second linker (Linker 2). Set up a 20 μl reaction and incubate overnight at 18°C.

Re-suspend pellet from (section 4.7.3.2)	16 µl
10X DNA ligation buffer	2 µl
Linker 2	1 µl
T4 RNA Ligase	0.5 µl
T4 DNA Ligase	0.5 µl

2. The whole reaction is used for cDNA synthesis.

4.8.3.4   Reverse transcription of the linkered dsRNA

The dsRNA with both the linkers attached was reverse transcribed using 3' linker specific primer. A thermostable reverse transcriptase was used to make cDNA as the substraste was dsRNA. The reverse transcription reaction was spiked with ^{32}P αCTP and was run on a 15% denaturing acrylamide gel. The gel was exposed overnight on an X-ray film. The auto-radiograph showed that different sized ligated dsRNA fractions gave different corresponding cDNA bands. (Figure 31C).

The stepwise procedure is as follows:

1. A thermostable MonsterScript™ Reverse Transcriptase (Epicentre Catalog# MS040910) was used.

2. Anneal the reverse primer to the RNA. Set up the following reaction

DEPC treated water	1.0 µl
RNA	8.0 µl
RT primer	1.0 µl
Buffer	4.0 µl

3. Heat at 65°C for 1 min and immediately chill on ice for 2 mins

4. Add following reagents on ice. It is recommended to set up a cold chase reaction.

5. For cold chase add:

DEPC treated water	1.1 µl
5mM dATPs	0.8 µl
5mM dGTPs	0.8 µl
5mM dTTPs	0.8 µl
^{32}P αdCTPs	1.5 µl
RT MonsterScript 1U	1.0 µl

6. Incubate at 42°C for 5 minutes and then 60°C for 20 minutes.

7. Then add

5 mM dCTPs	0.8 µl

8. Incubate at 60°C for 20 minutes.

9. Add 1 µl RNase H (1 Unit) and incubate for 20 minutes.

10. Use 2 µl of the cDNA for each PCR

11. Run 5 µl of reaction on 15% acrylamide/8 M urea denaturing gel. Expose it on the X-ray film overnight.

12. Develop the film and see whether the different sized fragments indicate that you have cDNA of different lengths corresponding to the initial protected fragments from RNase Protection assay.
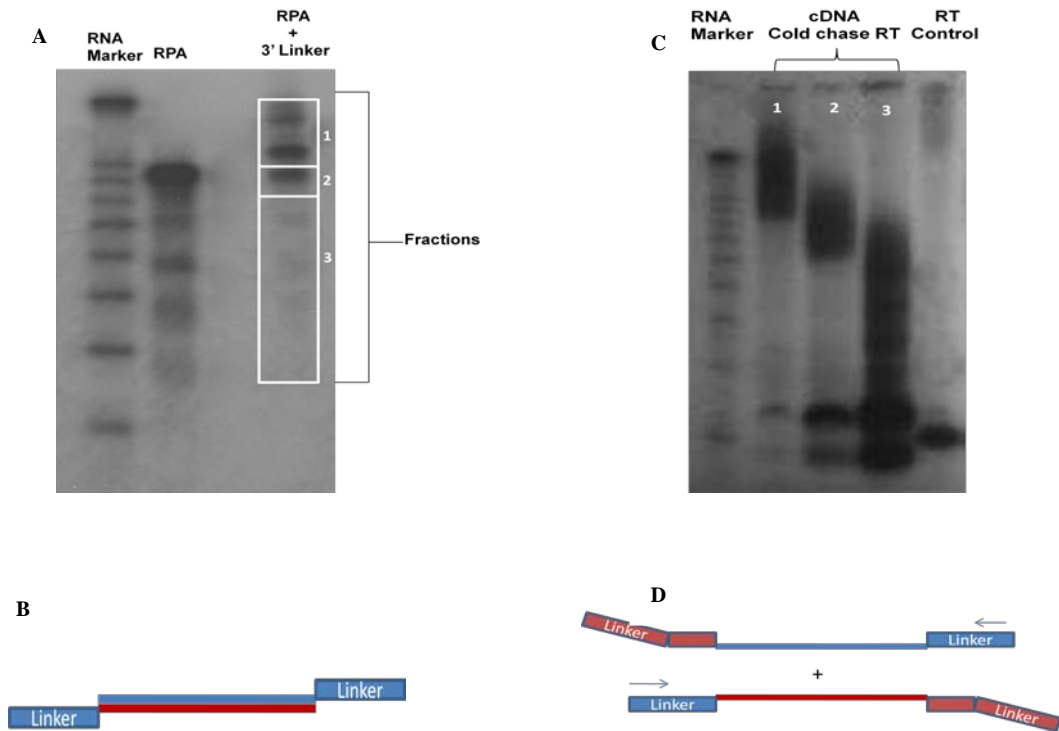
Figure 30: (A) Auto-radiograph of RPA lane 1 is the RNA decade marker, lane 2 has the RPA from total mouse brain RNA hybridized with MBII- 85 antisense probe, lane 3 is empty and lane 4 has the 3' linker addition reaction. A shift in bands from the RPA in lane 4 as compared to lane2 shows the success of the linker reaction. The gel pieces containing three fractions were cut out and eluted overnight and 5' linker was added and the whole reaction was used for reverse transcription. (B) Picture showing the structure of dsRNA with 3' linker attached (C) The auto radio graph of the spiked reverse transcription reaction was reaction run on 15% denaturing acrylamide gel. Lane 1 is fraction 1 and Lane 2 is fraction 2 and lane 3 shows reverse transcription reaction from fraction 3. (D) Picture showing the structure of linkered RNA as blue and red lines depicts dsRNA separated during reverse transcription reaction. The boxes depicting the linker attached. Blue arrow indicates the reverse transcription primer.

4.8.3.5   PCR and Southern Blot

A simultaneous reverse transcription reaction without radiation was made and the reaction was used to perform PCR amplification using linker specific Primers (Figure 31). The touch-down PCR conditions were followed and the PCR products were ran on two different 2% agrose gels. The gels were stained with a gel green dye and viewed under blue light. The PCR amplicons were of the expected size with reference to the cDNA fragments (Figure 31). The first gel was used for southern blot; probes used spanned the whole MBII-85 snoRNA and were labeled with ^{32}P γATP. The first two lanes derived from fraction A and B (Figure 31 A) showed positive signal for MBII-85 specific probes and the

third lane with short fragments did not come out positive for MBII-85 probes. MBII-85 (SNORD116) expression construct was used as a positive clone.
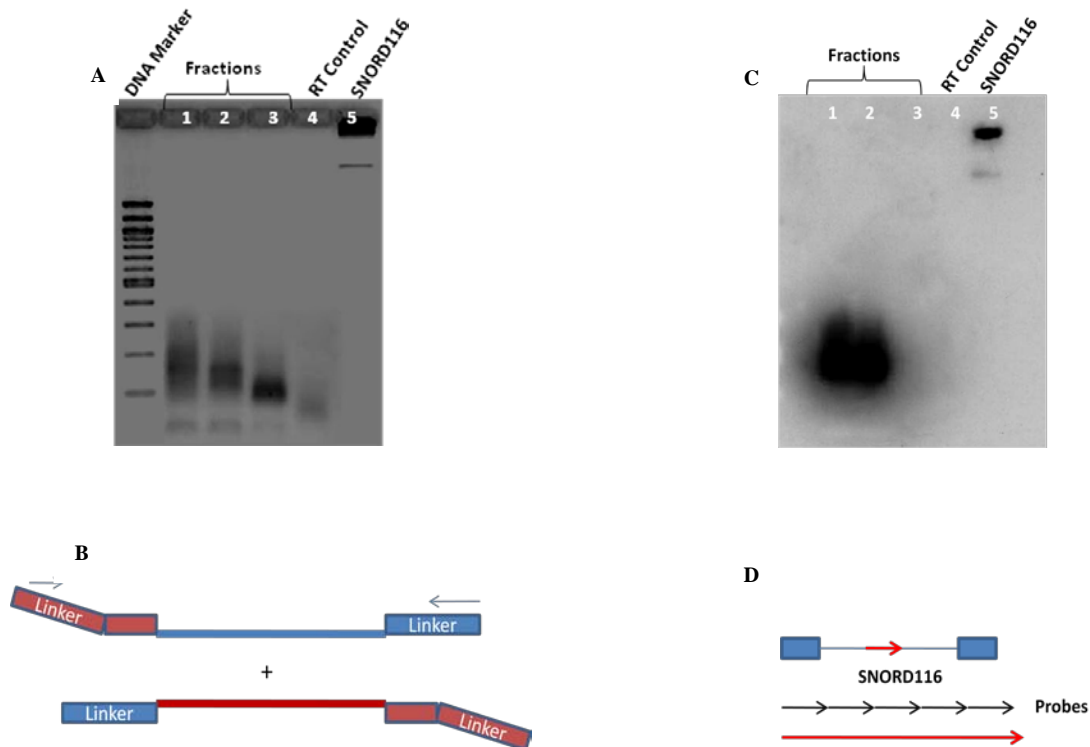
Figure 31: (A) The cDNA from the linkered RNA was amplified using linker specific primers and was run on a 2% agrose gel and stained with Gel green. The lanes 1-3 have the PCR amplicons from three fractions described above, lane 4 is the reverse transcription control and lane 5 has MBII-85 expression construct as positive clone. (B) Picture depicting RNA as a blue line and boxes depicting linkers and arrows show the position of the primers. (C) An auto radiograph of the southern blot from the 2% agrose gel shown in Figure 26 (A) lane one and two corresponding to fraction 1 and 2 that gave positive signal with MBII-85 probes. Lane 6 a positive control. (D) Picture describing ^{32}P labeled probes used for southern blot and colony hybridization; five probes were used each was 20nts. These five oligonucleotides spanned the whole MBII-85 (SNORD116) snoRNA.

The stepwise procedure is as follows:

1. For PCR set up the following reaction

dH20	17.8 µl
10x PCR buffer	2.5 µl
Primer 1	1.0 µl
Primer 2	1.0 µl
dNTPs	0.5 µl

Taq Enzyme 0.2 µl

cDNA 2.0 µl

PCR Conditions: Alternatively touchdown PCR can also be done (Tm ranging 65°C to 58°C).

Cycle 1 (1x)	95°C	2 Minutes
Cycle 2 (30x)	95°C	30 Seconds
	58°C	15 Seconds
	72°C	30 Seconds
Cycle 3 (1X)	72°C	5 Minutes
Hold	4°C	

2. Run the PCR reactions on a 2.5 % agarose gel.

3. Stain with Gel green and visualize the gel on the Blue light box.

The procedure for the Southern Blotting and hybridisation of DNA is as follows:

After electophoresis the gel was placed in denaturation buffer (1.5 M NaCl/0.5 N NaOH) for 30 min. The DNA was transferred to a nylon membrane in a denaturation buffer for 16 h by using capillary transfer. The DNA was crosslinked to the membrane on a UV transilluminator. The prehybridisation was carried out for 30 minutes at 42°C in hybridisation buffer. The hybridisation was done in ~5 ml of hybridisation buffer for 16 hours at 42°C. The membrane was washed two times in 2 x SSC/0.1% SDS for 10 minutes at 45°C. It was than washed (two times) in 0.5 x SSC/0.1% SDS for 10 minutes at 45°C. Finally, the membrane was washed in 0.1 x SSC/0.1% SDS for 10 minutes at 45°C. The membrane was then exposed to X-ray film or the signal was revealed on a phosphoimager.

Hybridisation buffer:	20x SSC:
0.5M phosphate buffer, pH 7.2	3 M NaCl
7% SDS	0.3 M Na citrate

4.8.3.6   Cloning and sequencing

The PCR amplicons from second agrose gel were eluted and 2 µl of the purified eluted DNA was cloned into a cloning vector and rest of the DNA was sent for deep sequencing (Co-factor genomics). The agarose gel was stained with gel green instead of ethidium bromide as ethidium bromide causes dimerization of thymines and renders the TA cloning inefficient.

The stepwise procedure is as follows:

1. The eluted DNA should be run on the gel to see the integrity and should have a 260/280 wavelength ratio reading of 1.8.

2. The pure and high quality DNA should be sent for Deep Sequencing (Co-Factor genomics).

3. The eluted DNA can also be cloned using TOPO TA cloning kit.

4. Set up the cloning reaction

PCR Product	4 µl
Salt solution ($MgCl_2$)	1 µl
Vector	1 µl

5. Incubate at room temperature for 10 minutes.

6. Heat shock the reaction by putting the tubes in water bath set at 42°C for 30 seconds and immediately putting it on ice for 2 minutes.

7. Add

    SOS medium 250 µl

8. Put the tube in 37°C on revolving wheel for 1hr.

9. Plate the cells on LB plates with appropriate antibiotic and incubate over night.

10. Pick the colonies and prepare Plasmid DNA and send it for sequencing.

4.8.3.7 Colony hybridization

Colonies of the clones from the eluted DNA that was cloned into the cloning vector were picked individually. These colonies were re-streaked onto fresh plates in a grid fashion. The plates were incubated overnight at 37°C and colony hybridization was performed using same probes as used for southern blotting and described in Figure 31 D. Positive clones were picked and plasmid DNA was isolated and subsequently sequenced. The sequencing results showed the two forms of MBII-85 form A and form B (Figure 26).
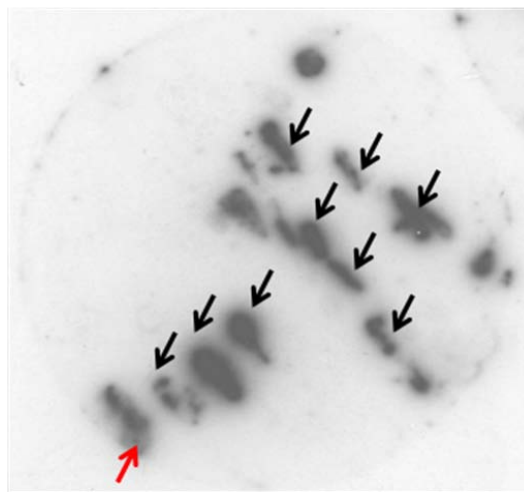

Figure 32: An auto radiograph of Colony hybridization. The elongated shaped signals indicated by black arrows point to the MBII-85 positive clones and rounded signal indicated by red arrow indicate the positive control on the blot. The MBII-85 expression construct was used as positive control.

4.9 Detection of abundance of Double stranded RNAs

As can be seen in Figure 32, most of the PCR products did not contain RNA protected by the hybridization probe, but represented endogenous RNAs that could hybridize. We therefore analysed all the sequences by deep sequencing. A bioinformatic analysis of the deep sequencing data was performed and the sequences were mapped on the whole mouse genome. The data set of sequences that overlapped with other sequences having hits on the opposite strand of chromosome was visualized on the UCSC browser (Figure 33). This indicated the presence of naturally occurring double stranded RNA sequences. Most of the chromosomes have the abundant hits indicating the dsRNA being

distributed all over the mouse genome. The results of mapping are available on the link given below.

http://regulatorygenomics.upf.edu/Projects/StefanStamm/RNA_Reads_overlapping/track_index.html
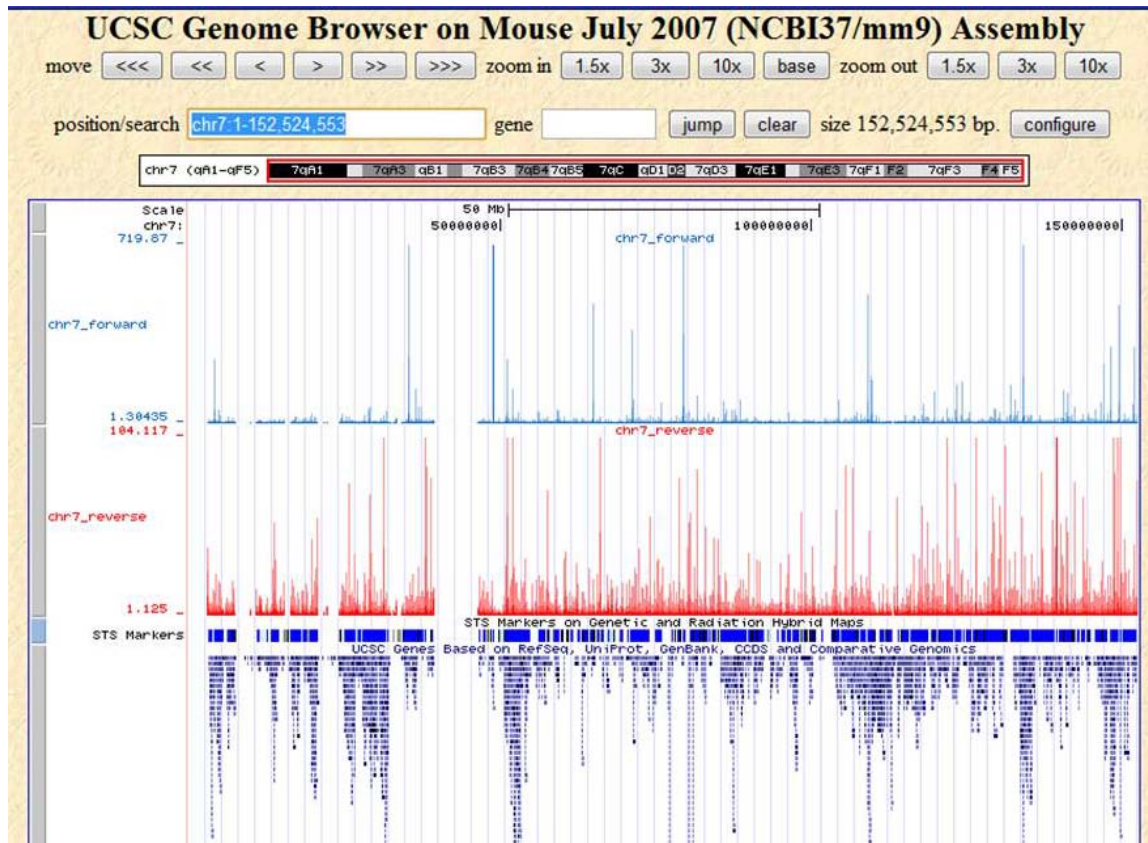
Figure 33: Deep sequence data analysis: A screen shot of the UCSC browser shows the chromosome 7 positive and negative strands. Peaks indicate the hits for the amplicons sequenced using deep sequencing Roche SoLid platform. Blue peaks are for hits on the positive strand and red peaks for hits on the negative strand. Overlapping red and blue peaks indicate the presence of dsRNA.

5 Discussion

In the first part of this work, new mRNA targets for MBII-52 and MBII-85 snoRNAs were identified using a bioinformatic approach. The targets were experimentally verified using snoRNAs over expression assays in mammalian cell lines. These targets were further verified using minigene assays. Different mutants of snoRNAs were used to validate the specific effect of snoRNAs. The regulation of these target exons was cross checked using two types of knock out mouse brain samples. For MBII-52, the transgenic mouse samples were used that had whole Prader willi critical region deleted. For MBII-85, the knock out mouse samples were used that had only the MBII-85 cluster missing. The regulation of all these target exons was consistent in all the studies. This verified the role of these snoRNAs as regulators of alternative splicing.

The second part of the work dealt with the identification of the functional units of these snoRNAs. To begin with we identified that these snoRNAs do not associate with conventional snoRNPs but they recruit hnRNPs. The association with hnRNPs cements role of snoRNAs in regulation of alternative splicing. The RNase protection assay was done to identify the functional snoRNA and was confirmed using northern blot. Both these assays identified the processed form of these snoRNAs. We analysed the function of these smaller RNAs and their role in snoRNA function.

We also developed a new method to clone these small dsRNAs as the known published methods have their limitation either with the size of the dsRNA or the amount available as starting material. We also developed a faster and effiecient method to clone minigene for cell based assays.

5.1 The MBII-52 expression unit generates processed snoRNAs (psnoRNAs)

MBII-52 snoRNAs are expressed from a cluster containing multiple copies of tandemly arranged snoRNA expression units. Each unit contains phylogenetically poorly conserved exons that flank an intron which hosts the snoRNA (Runte, M. et al., 2001). Humans contain 47 HBII-52 copies and mice at least 130 copies. Using RNase protection assays, we analyzed the mouse MBII-52 copy that is most closely related to the copy 27 of human HBII-52 snoRNA cluster. There is enough sequence heterogeneity between the different MBII-52 snoRNA copies that allows their discrimination in protection assays. Unexpectedly, the RNase protection assay indicates that the snoRNA gives rise to other

smaller RNAs and that the full-length C/D box snoRNA is a minor form. The presence of the smaller RNAs could be verified by northern blot analysis, which further rules out that signals corresponding to shorter RNAs are caused by the protection of unrelated RNAs or are caused by RNA editing events that introduce mismatches to the probe. Finally, we tested ectopical expression of MBII-52 in HEK293 cells that do not express this snoRNA. The expression construct gives a similar pattern of shorter RNAs, indicating that they are derived from the transfected single MBII-52 expressing unit. The cloning of the shorter RNAs indicates that the major RNA form expressed from the MBII-52 expression unit is a 73 nt long RNA (form B) that lacks the sequences that form the snoRNA stem. However, this RNA contains other C/D box snoRNA elements, such as the C box, D box and antisense box. This RNA appears to be further shortened by exonuclease trimming, giving rise to smaller RNAs. The shorter RNAs can be detected both by northern blot and RNase protection analyses, indicating that they are metabolically stable. It is possible that these RNAs are protected from further endonuclease action by predicted secondary structures. The RNA form D forms a 5 bp stem on its 5' and 3' ends and RNA form C contains a short stem at its 5' end (Figure 23, underlined region). In addition, the formation of protein complexes is likely to stabilize the RNAs.

Ectopic expression of snoRNA mutants suggests that the formation of shorter RNAs depends on intact C and D boxes, which suggests that they derive from a C/D box snoRNA or pre-snoRNA structure. A possible scenario is that an unknown RNase initially removes the stem of the C/D box RNA, which gives rise to the predominant form B. This form is stabilized by the presence of the C and D boxes, most likely by binding to other proteins. Form B is shortened by exonucleases; giving rise to forms C, D and E that are most likely stabilized by another stem-loop structure and/or associated proteins.

We named these shorter RNAs psnoRNAs for processed small nucleolar RNAs. PsnoRNAs represent a new class of nuclear small RNAs. The psnoRNAs described here are the first to be derived from C/D box snoRNAs.

5.2   pSnoRNA recruits non-conventional RNPs

To obtain insight into proteins associated with these novel RNAs, we isolated them by affinity purification of RNP complexes, using a probe that is complementary to the antisense box of MBII-52. We used nuclear extract generated by the Dignam procedure as a starting material. In this method, most of the nucleolar material is separated in a high-speed centrifugation step. As the MBII-52-derived snoRNAs are present in this fraction, they are most likely present in the nucleoplasm. The major form RNA form B derived from MBII-52 does not contain the characteristic k-turn, which most likely prevents its association to Snu13p/15.5 kDa (Watkins, N.J., et al., 2000). In agreement with this RNA structure, we could not detect C/D box snoRNA-associated proteins, such as fibrillarin, or NOP56 (Matera, A.G., et al., 2007) in the isolated material. In contrast, we identified hnRNPs, including hnRNP A1, A2, TDP-43 and D0 that have been reported to be involved in splice-site selection. Unexpectedly, in the pull-downed material, we could still detect RNA forms C and D. These RNAs lack a complete snoCR that is complementary to the pull-down probe. Relative to the starting brain material, the RNA forms C and D are reduced in the pulled down material (Figure 21, lanes 1 and 5), but are still detectable. This suggests that the different RNA forms could be present within a complex.

5.3   pSpliceExpress as a tool to rapidly construct splicing reporter minigenes

To analyse the effect of psnoRNAs, we developed a high throughput system to test if these snoRNAs regulate alternative splicing in cell based minigene assays. The development of pSpliceExpress allows the fast generation of splicing reporter constructs using direct recombination between PCR products and a modified Exontrap vector. The Exontrap vector contains two constitutively spliced insulin exons that serve as an internal control. The system allows fast cloning of DNA fragments containing an alternative exon. If the fragments are shorter than 4000 nts, we observe more than a 50% recombination frequency and routinely obtain more than 100 clones per reaction. This frequency drops when the fragments are longer. We also devised a second system that is compatible with DNA fragments that are flanked by attL sites. PCR fragments up to 4000 nts length can be cloned into the commercially available donor vectors with high efficiency (Figure 34).

The constructs made with these vectors show the same splicing regulation as similar constructs made by conventional restriction cloning when used in transfection-based splicing assays.

The major advantages of the system are its speed and the lack of any need for restriction enzymes. Since no restriction sites are used, any fragment can be cloned and analyzed for splicing regulation.

The speed of the system allows its usage in medically relevant application. An increasing number of synonymous or intronic mutations appear to be associated with human diseases. The rapid cloning from PCR products, as described here, will greatly facilitate investigations aimed to determine whether such changes have an effect on pre-mRNA splicing.
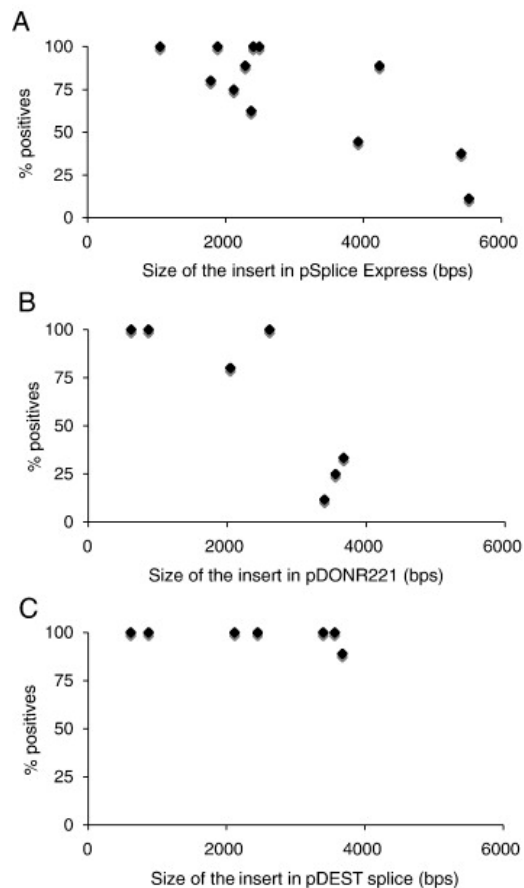
Figure 34: Cloning efficiency of vectors used. The in vitro recombination typically generates more than 50 clones. The graphs show the percent of clones with an insert of the expected size, determined by PCR and restriction digest. Each point represents the percent of successful recombinations as a function of the insert length. (A) Cloning efficiency of PCR products into pSpliceExpress. (B) Cloning efficiency between PCR fragments and pDONR221. (C) Cloning efficiency between pDONR221 inserts and pDESTSplice.

5.4 MBII-52 and MBII-85 derived psnoRNAs regulate splicing of several pre-mRNAs

We previously found that the expression of the snoRNA HBII-52 promotes inclusion of exon Vb of the serotonin receptor 5-HT2C. To investigate whether this

represents a special, unique case or is part of a new regulatory mechanism, we developed a computational screen that predicted more than 400 putative snoRNA targets. We tested some of these predicted targets by RT–PCR in transfection assays and further concentrated on five splicing events that showed consistent dependency on MBII-52 and MBII-85 expression. In contrast to the 5-HT2C receptor pre-mRNA, the pre-mRNAs harboring the MBII-52 and MBII-85 dependent exons are expressed in Neuro2A and HEK293 cells, which allowed us to determine the influence of these snoRNA expressions on the endogenous genes. Also in contrast to the neuron-specific 5-HT2C system where a splice site had to be optimized to detect the dependency on MBII-52 (Kishore, S. et al., 2006), the new alternative exons showed the dependency on these snoRNAs expression when analyzed in their endogenous gene context.

The alternative exons were next tested in a heterologous exon trap system and showed the dependency on these snoRNAs when flanked by insulin exons that are controlled by a CMV promoter. These experiments suggest that MBII-52 and MBII-85 RNAs act on defined parts of the pre-mRNA, in a mechanism that is independent of the promoter usage and genomic context. Together, these data strongly suggest that snoRNA expression influences alternative pre-mRNA splicing events.

Expression of MBII-52 and MBII-85 causes small, but statistically significant changes in multiple targets. This modest influence on numerous targets has been observed for other splicing factors, such as SMN (Zhang, Z. et al., 2008) and NOVA (Licatalosi, D.D. et al., 2008). Detailed work in the NOVA system (Licatalosi, D.D. et al., 2008) suggested that a splicing factor can control biological processes by coordinating numerous small changes and it is possible that MBII-52 fulfills a similar function. An alignment of the antisense box of HBII-52 and its experimentally confirmed targets is shown in Figure 35. The complementarity between the MBII-52 antisense box and its targets can be interrupted in multiple positions. With the exception of the serotonin receptor 5HT2C, there are always three mismatches in the alignment of the snoCR and the MBII-52 antisense box. It is interesting that the serotonin pre-mRNA can be edited at three positions within the snoCR. Taking these editing events into account, the data suggest

that preferably 15 of the 18 nucleotides of the antisense box show complementarity towards its target. It is noteworthy that we initially concentrated on targets with only one or two mismatches, but did not find a dependency of these exons on MBII-52 expression. The data indicate that MBII-52 and MBII-85 derived RNAs need a defined degree of sequence complementarity towards their targets. This scenario is reminiscent of the action of U1 snRNP on the 5' splice site, where natural occurring exons rarely show 100% complementarity towards the U1 snRNA, but usually have several mismatches, which cluster in certain position of the 5' splice site (Stamm, S. et al., 2000).

HBII-52	A	U	G	C	U	C	A	A	U	A	G	G	A	U	U	A	C	G
5HTC	+	+	+	+	+	+	+	+	E	+	+	+	+	E	E	+	+	+
PB1	L	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	M	M
DPM2	+	+	+	+	L	+	+	+	M	+	+	+	+	+	+	+	M	M
RALGPS	+	+	+	+	+	+	+	+	+	L	M	+	+	+	L	+	+	+
TAF1	+	+	M	+	+	+	+	+	+	+	+	+	+	+	L	+	+	M
CRHR1		+	+	+	G	+	+	+	+	+	+	+	M	+	+	+	+	+
CONSENSUS	4	6	5	6	4	6	6	6	4	5	5	6	5	6/5	3	6	4	3

Figure 35: Sequence complementarity between the antisense box of MBII-52 and its experimentally confirmed targets. The top line shows MBII-52 antisense box in 5' to 3' orientation. Lower lines show the base-pairing to the complementary sequence in target exons. +: base-pairing, including G-U base pairing, E: nucleotide can be changed by editing, M: mismatch, L: gap ("loop") in alignment. Positions in the antisense box of HBII-52 that are invariant in all human copies are highlighted in brown. The positions that display complementarity in all target genes are indicated in yellow in the consensus column

The existence of psnoRNAs could explain the influence of MBII-52 and MBII-85 expression on splice-site selection in a model illustrated in Figure 36. We postulate that these snoRNAs expressing unit consisting of two non-coding exons flanking an intron that hosts a snoRNA gives rise to several RNAs. The major form derived from the expression unit is form B that lacks the snoRNA stem-structure and is associated with hnRNPs, but not C/D box snoRNA-typical proteins. Form B contains the antisense box that targets it to other RNAs, including pre-mRNAs identified in this study. Form B RNA can influence splice-site selection by competing with existing splicing regulatory factors on the pre-mRNA or by bringing the associated hnRNPs to the targets, similar to a bifunctional oligonucletide. The longest RNA (form A) has all the hallmarks of a traditional C/D box snoRNA, but is only a minor fraction of the RNA expressed. It is likely that this RNA is transported into the nucleolus, where it can be detected by *in situ* hybridization (Vitali, P. et al., 2005). It is not clear what function this RNA has in the nucleolus, but it could

represent a storage form for the formation of the active RNA form B that is released from the nucleolus according to the physiological needs.
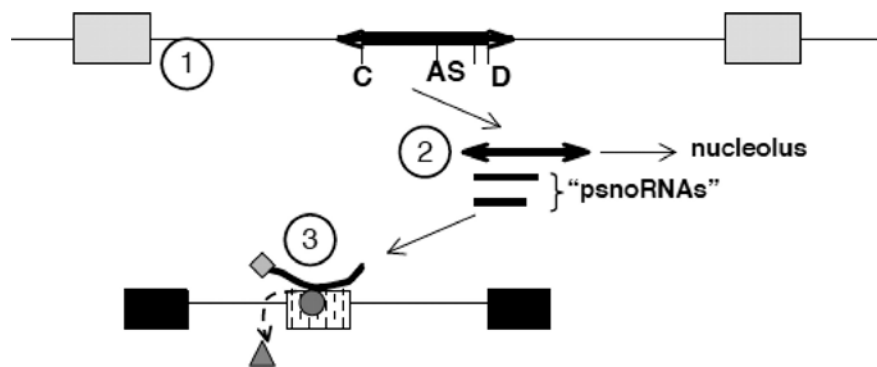

Figure 36: Model for MBII-52 and MBII-85 action on RNA processing. (1) The PWS critical region contains snoRNAs (thick line) located in introns between non-coding exons (grey boxes). The snoRNA is characterized by a C box (C), D box (D) and an antisense box (AS), as well as stem-forming sequences (arrows). (2) This unit generates several RNAs, including the full-length snoRNA that shows its highest concentration in the nucleolus and Cajal bodies (35) as well as several shorter psnoRNAs (for processed snoRNAs). PsnoRNAs are present in the nucleoplasm where they associate with hnRNPs. (3) PsnoRNAs can change splice-site selection, most likely by binding to complementary sequences. We postulate that they either remove regulatory proteins from their targets (triangle) or bring in associated proteins to the exon recognition complex (diamond associated with the small RNA).

5.5   psnoRNA as functional units

The observation that the snoRNAs are processed into smaller RNAs prompted us to look for secondary structures of all the mRNA targets of these snoRNAs. We used the NIPU server to identify motifs and secondary structure (http://www.bioinf.uni-freiburg.de/Software/NIPU/). Probability unpaired (PU) values were calculated for each base in the sequence (Hiller, M. et al., 2007). Higher PU values indicate higher single-strandedness of the motif and a zero PU value indicates complete pairing. The PU value is based on an average of all the possible structures.

As *in vivo* secondary structures of pre-mRNAs are likely to be local rather than global, the length of the flanking sequence has an impact on the PU score. The secondary structures are made cotranscriptionally in which short-range base pairing is favored over long-range base paring (Schroeder, R. et al., 2002). Considering this fact 100 nts upstream and downstream of the mRNA region that had the snoRNA target region was analysed. As shown in Table 5 and Table 6, the snoRNA antisense target regions in all the target genes are fairly paired, indicating they are within a secondary structure. The snoRNA target regions are highlighted with box.
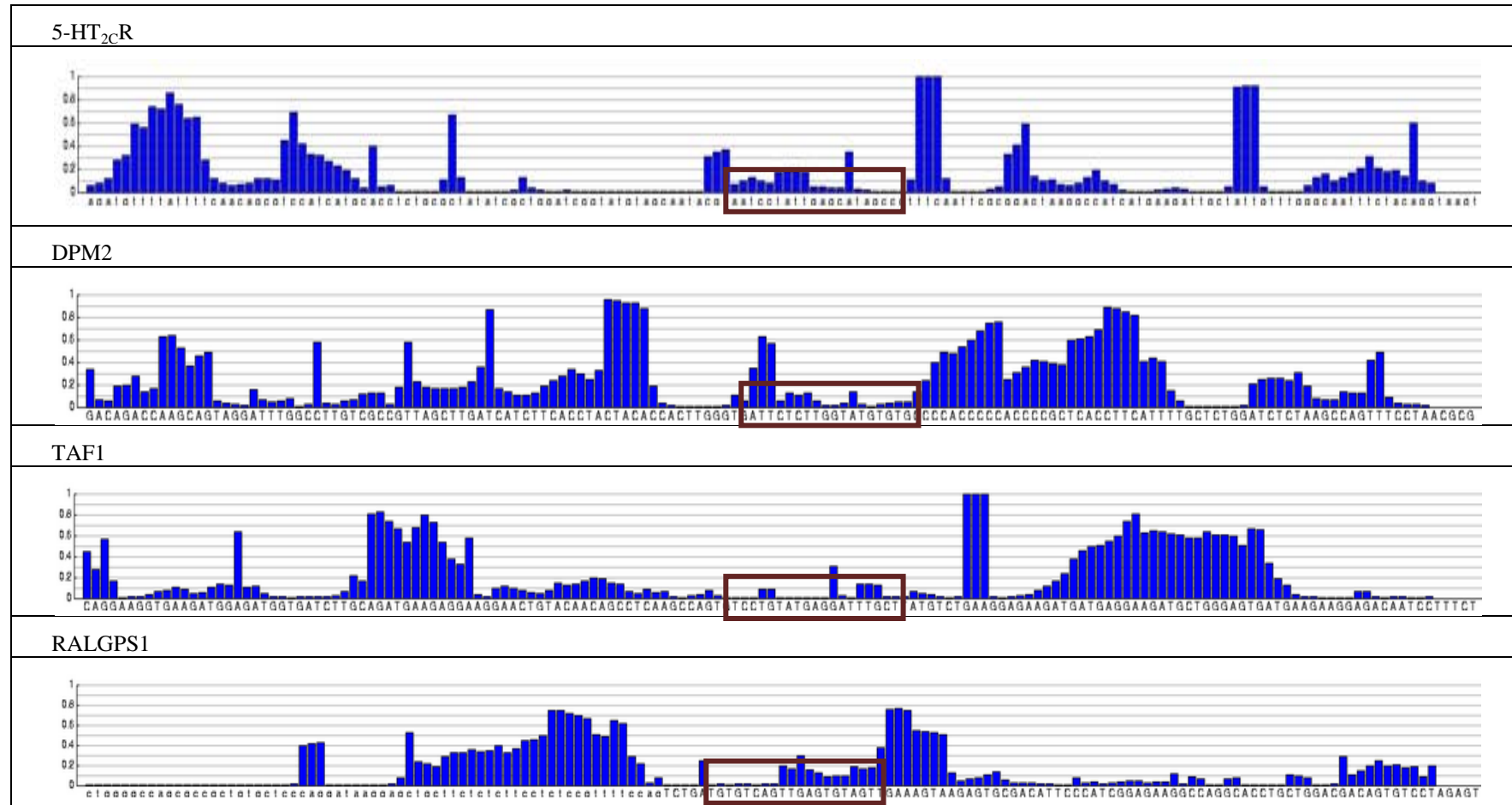
Table 5: We computed the probability that the target mRNA region for snoRNA antisense is paired denoted as the PU value. PU values range between 0 (completely base-paired) and 1 (completely unpaired). The PU value determined as an average of all local folding windows that comprise a context up- and downstream of 11 to 30 nt. Four target genes for MBII-52 are shown here 5-HT$_{2C}$R, DPM2, TAF1 and RALGPS1.
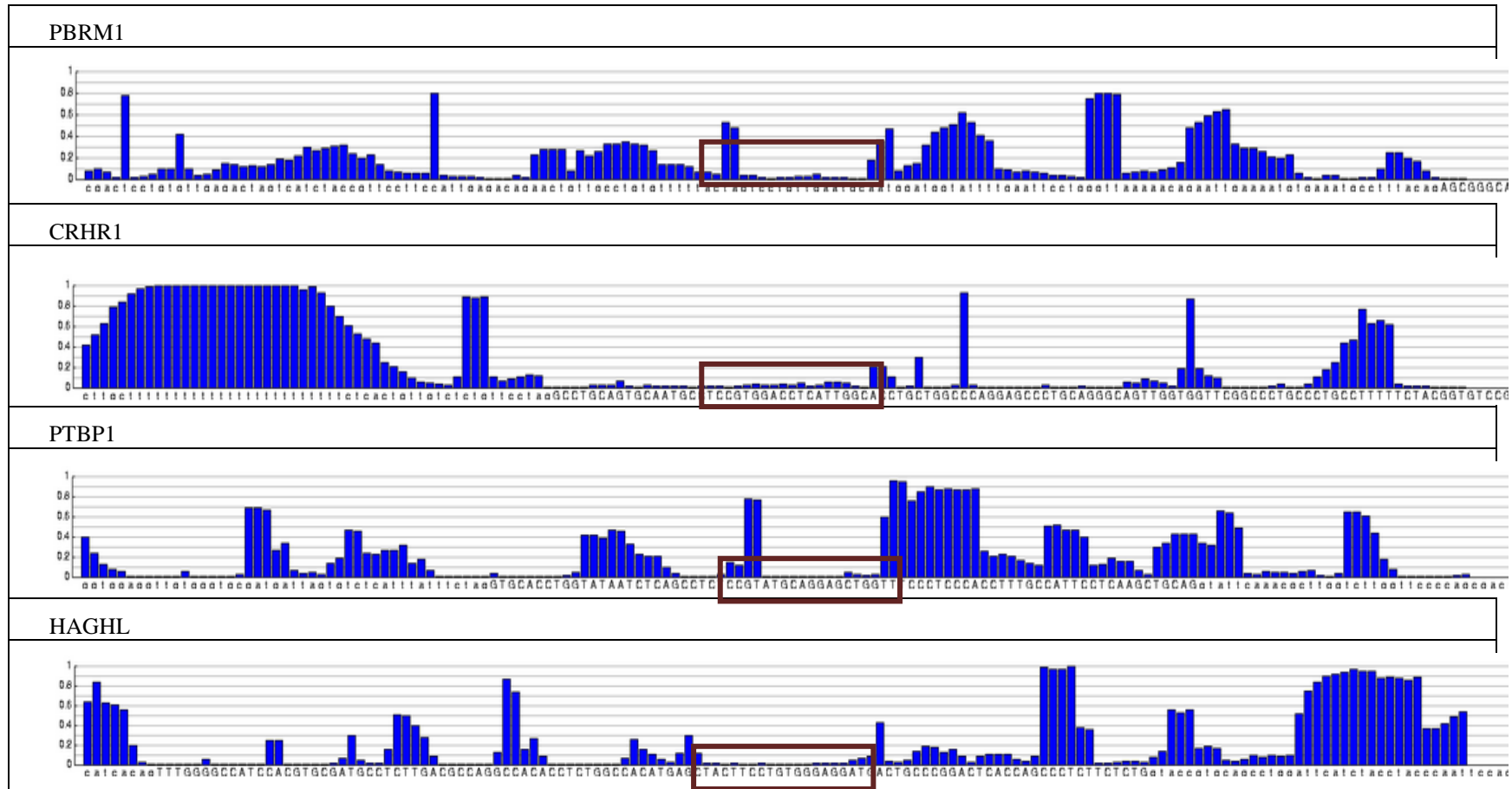
Table 6: We computed the probability that the target mRNA region for snoRNA antisense is paired denoted as the PU value. PU values range between 0 (completely base-paired) and 1 (completely unpaired). The PU value determined as an average of all local folding windows that comprise a context up- and downstream of 11 to 30 nt. Two target genes for MBII-52 are shown here PBRM1and CRHR1. Two target genes for MBII-85 are shown here PTBP1 and HAGHL.

The above mentioned bioinformatics analysis predicted the prevelance of base paired regions in the target regions of the snoRNAs. Next we analysed the secondary structure of 5-HT$_{2C}$R, the first ever reported mRNA target of MBII-52 (Kishore, S. et al. 2006). We compared the reported 5-HT$_{2C}$Rsecondary structure with the structure of psnoRNAs to identify the processed snoRNA (psnoRNA) forms c, d and e target regions (Figure 39). To predict if these shorter snoRNA forms can act as guiding oligonuleotides we aligned the snoRNA sequence and the target exon region of serotonin receptor.

The HBII-52 anti sense box 2 has a complimentarity region on serotonin receptor pre-mRNA. The target mRNA region is well conserved across different species (Figure 37). A multiple alignment was done for MBII-52 target region on serotonin receptor gene across different species (Human, Orangutan, Dog, Horse, Rat and Mouse)

Human	CGTAATCCTATTGAGCATAG	20
Orangutan	CGTAATCCTAT-GAGCATAG	19
Dog	CGTAATCCTATTGAGCATAG	20
Horse	CGTAATCCTATTGAGCATAG	20
Rat	CGTAATCCTATTGAGCATAG	20
Mouse	CGTAATCCTATTGAGCATAG	20
	* * * * * * * * * * * * * * * * *	

Figure 37: Multiple alignment across different species of sequence from serotonin receptor that makes a target for HBII-52 anstisense box 2.

To figure out the function of smaller RNAs that are derived from full length snoRNA, a bioinformatics analysis of a small stretch of serotonin mRNA (ACTTTC) shows 100% complimentarity with stretch of HBII-52 which is present in psnoRNA form c, d and e and this shorter target for HBII-52 is also well conserved across different species (Figure 38). These shorter snoRNAs binds to the target complementary region and disrupts the mRNA secondary structure and makes the splice site available for splicing to take place.

|		
Human	AATTGCAGCGGCTATGCTCAATACTTTCGGATTATGTAC	39
Orangutan	AATTGCAGCGGCTATGCTCAATACTTTCGGATTATGTAC	39
Dog	AATTGCAGCGGCTATGCTCAATACTTTCGGATTATGTAC	39
Horse	AATTGCAGCGGCTATGCTCAATACTTTCGGATTATGTAC	39
Mouse	AATTGCAGCGGCTATGCTCAATACCTTCGGATTATGTAC	39
Rat	AATTGCAGCGGCTATGCTCAATACCTTCGGATTATGTAC	39
	* ** *** * * * * * * * * *	

Figure 38: Multiple alignment across different species of sequence from serotonin receptor that makes a target for psnoRNA HBII-52 form c, d and e indicated by circles in Figure39 A

This bioinformatic analysis suggests that different processed forms of snoRNA can act as independent elements or in conjunction with the antisense box to change alternative splicing of the target exons.
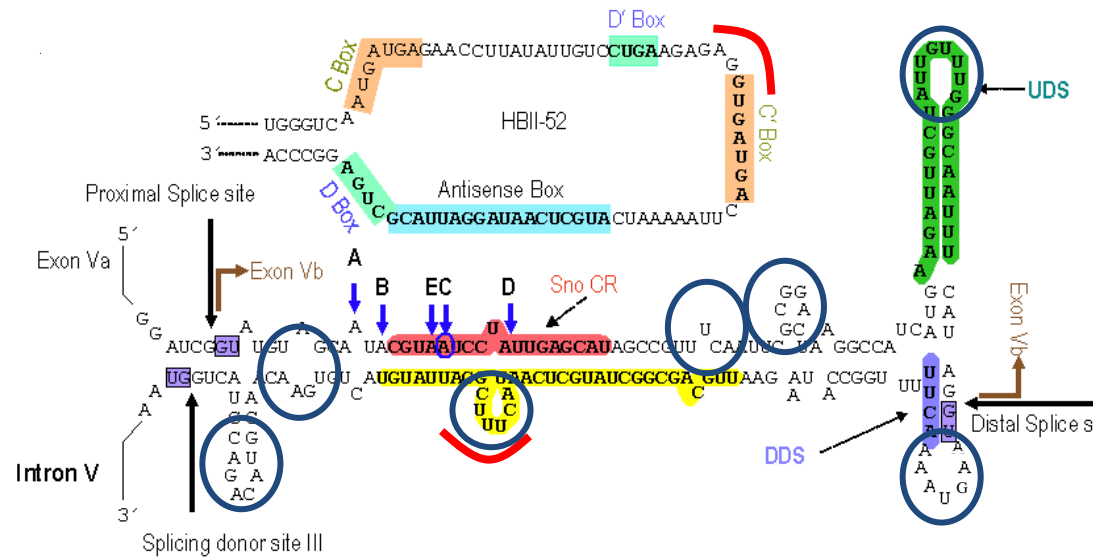
Figure 39: Structure of snoRNA HBII-52 and secondary structure of human serotonin mRNA. The highlighted pink region of the serotonin mRNA shows the target region for HBII-52 antisense box 2 and blue outlined box shows the putative target of psnoRNA form c,d or e (sequence from psnoRNA highlighted with a thick red line). (**B**) We computed the probability that the target mRNA region for snoRNA antisense is paired denoted as the PU value. PU values range between 0 (completely base-paired) and 1 (completely unpaired). The PU value determined as an average of all local folding windows that comprise a context up- and downstream of 11 to 30 nt. The $5\text{-HT}_{2C}R$ region complementary to HBII-52 antisense Box and the region that base pairs with it in secondary structure formation is highlighted with a box. Circles indicate the single stranded stretches in the secondary structure.
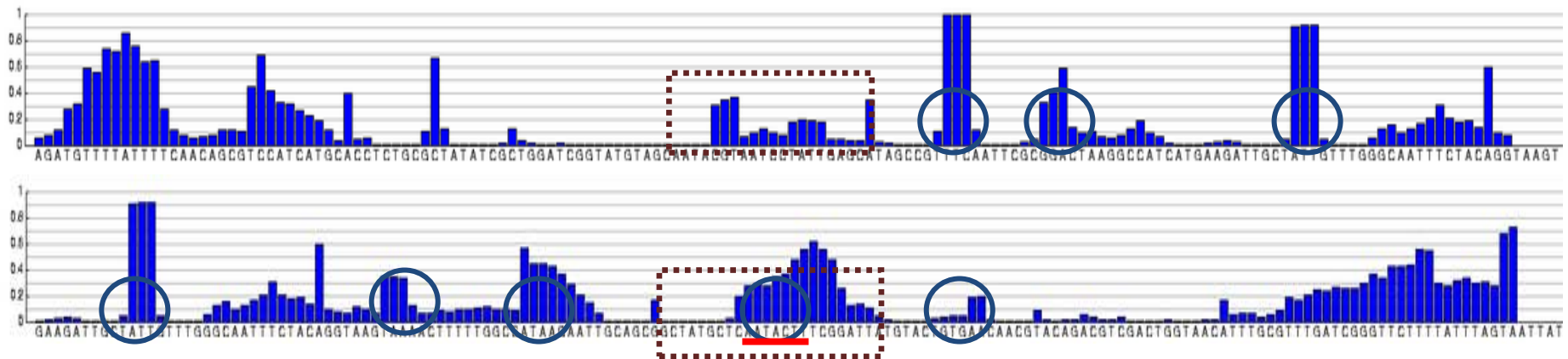
Next we wanted to see if the secondary structure of 5-HT$_{2C}$R corresponds to PU score when we use NIPU server. The information based on PU score did overlap with the most stably predicted secondary structure of 5-HT$_{2C}$R. Since the PU score prediction is based on the average values of shifting frame across the structure, the single stranded stretches cannot be seen as sharp peaks. The single stranded stretches and peaks are highlightes by a circle in Figure 39 A and B.

5.6   MBII-85 and MBII 52 similarities and dissimilarities

Just like MBII-52, MBII-85 snoRNAs are expressed from a cluster containing multiple copies of tandemly arranged snoRNA expression units. Each unit contains phylogenetically poorly conserved exons that flank an intron which hosts the snoRNA (Runte, M. et al., 2001). Unlike MBII-52, MBII-85 snoRNAs are quite conserved. Out of 20 mouse copies 18 copies are identical. Bioinformatics screen yielded more than 300 targets for the two antisense boxes of MBII-85. We could experimentally prove two alternative splicing events of these genes in cell based assays, using over expression of MBII-85 expression constructs. We could confirm these alternative splicing events in the minigene assays. The genes that are spliced are HAGHL and PTBP1. The similarity between these two snoRNAs cluster ends with their role in alternative splicing. The processed snoRNA from MBII-85 cluster yield forms (A and B) and lesser amount of smaller forms, unlike MBII-52 that yields forms (C, D and E) (Figure 23). The proteins associated with MBII-85 that were pulled down using biotinylated MBII-85 specific RNA probe showed a different protein profile. The major difference in the proteins pulled down by MBII-52 specific probe is of hnRNPA1 and Transcriptional activator factor Pur alpha (Figure 28). The processing of both these snoRNAs is similar to the extent that both of them have shorter form (psnoRNA) as they have their stem hair pin structure deleted (Figure 30).

5.7   Cloning of dsRNAs: a novel tool to understand the RNA world

The dsRNAs are difficult to clone as there are no restriction eznzymes that can be used for cloning. DNA can be digested with restriction enzyme and ligated into any desired digested vector. DNA can also be PCR amplified and cloned. Unlike DNA, RNA needs to be reverse transcribed and cloned. The unknown RNAs as in Figure 22 need to be identified to illustrate their function. With the advent of deep sequencing and plethora of information about non-coding RNAs (Taft, R.J. et al., 2009; Cole, C. et al., 2009). It

becomes all the more important to have a reliable technique to sequence the small RNAs. The method that we developed for dsRNA cloning is fast and reliable considering the fact that these small RNAs small are low in abundance. Colony hybridization gives us around 20% positive clones in the final step of this technique.

A detailed bioinformatic analysis of the deep sequencing data of the dsRNA gave a new insight into the presence of naturally occurring dsRNA. Though their function is yet to be understood, their presence itself is intriguing enough to look into their distribution across the genome. We analyzed the data set from the sequences which had hits on both strands of chromosome. The presence of EST (expressed sequence tags) supports the fact that they are naturally occurring and are stable, ruling out this data as not an experimental artifact (Figure 33).

The usefulness of information to understand role and presence of small RNAs makes this new method a very reliable tool to analyse small RNAs by cloning them effieciently.

5.8 Relevance of PWS

The loss of C/D box snoRNA expression has been postulated as the underlying mechanism for the development of PWS (Ding, F. et al., 2005). This hypothesis was recently supported by a patient with a 1,74,584 bp large microdeletion that encompassed only snoRNAs and their flanking hosting introns and exons. The deletion led to a Prader–Willi phenotype (Sahoo, D. et al., 2008). To date, the only published RNAs expressed from the 1,74,584 bp region are snoRNAs and fragments of their surrounding non-coding exons.

The idea that the loss of snoRNA expression is central to PWS is further supported by genetic evidence that ruled out proteins encoded by MKRN3, MAGEL2 and NDN genes expressed in the Prader–Willi critical region (Kanber, D. et al., 2009). The 1,74,584 bp microdeletion removes the snoRNAs HBII-438A, -85 and 23 of the 47 HBII-52 copies from the paternally expressed allele. The only snoRNA that was totally removed by the microdeletion was MBII-85, which led to the suggestion that MBII-85 loss is the major reason for PWS. However, there is evidence that HBII-85 and HBII-52 are expressed by two transcriptional units (Vitali, P. et al., 2010). As the 1,74,584 bp

micordeletion contains the 5' end of the HBII-52 cluster, it could harbor transcriptional elements necessary for proper HBII-52 expression. Furthermore, in the majority of cases, the complete SNURF–SNRPN locus is deleted (Butler, M.G. et al., 2006). The contribution of HBII-85 and HBII-52 loss to PWS is therefore not clear.

Our findings indicate that the SNURF–SNRPN locus not only gives rise to typical C/D box snoRNAs, but generates shorter psnoRNAs. The northern blot analysis indicates that all of the at least 130 MBII-52 copies are processed in a similar manner. The major RNA form from the MBII-52 cluster is not the canonical C/D box snoRNA, but a shorter RNA form, most likely similar to psnoRNA form B. psnoRNAs are associated with hnRNPs and could have multiple functions by targeting these proteins to other RNAs. It is not clear whether several psnoRNAs lacking the antisense box use other RNA parts for targeting or have non-related functions.

The loss of the regulatory psnoRNAs could be a significant contribution to the etiology of PWS and substitution of the short psnoRNAs could be a therapeutic principle for the disease.

6   References

Aartsma-Rus, A. and van Ommen, G.J. Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. *Rna* **13**, 1609-24 (2007).

Asparuhova, M.B., Marti, G., Liu, S., Serhan, F., Trono, D., Schümperli, D. Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. *J Gene Med* **9**, 323-34 (2007).

Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. The expanding snoRNA world. *Biochimie*, **8**, 775-90 (2002).

Bazeley, P.S., Shepelev, V., Talebizadeh, Z., Butler, M.G., Fedorova, L., Filatov, V. and Fedorov, A. snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene* **408,** 172–79 (2008).

Bischof, J.M., Stewart, C.L. and Wevrick, R. Inactivation of the mouse Magel2 gene results in growth abnormalities similar to Prader–Willi syndrome. *Hum. Mol. Genet.* **16**, 2713–2719 (2007)

Blanchette, M. and Chabot, B. A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *Rna* **3**, 405-19 (1997).

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., Hannon, G.J. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-103 (2007).

Bressler, J., Tsai, T.F., Wu, M.Y., Tsai, S.F., Ramirez, M.A., Armstrong, D. and Beaudet, A.L. The SNRPN promoter is not required for genomic imprinting of the Prader-Willi/Angelman domain in mice. *Nat. Genet.* **28,** 189-205 (2001).

Brown, J.W., Echeverria, M. and Qu, L.H. Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.* **8,** 42-9 (2003).

Buratti, E. and Baralle, F.E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* **24**, 10505-14 (2004).

Burset, M., Seledtsov, I.A., Solovyev, V.V. SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.* **29**, 255-9 (2001).

Butler M.G., Hanchett, J.M., Thompson, T.E. Clinical findings and natural history of Prader–Willi syndrome. In: Managment of Prader–Willi Syndrome—Butler M.G., Lee P.D.K., Whitman B.Y., eds. *Springer.* **48**, 3–48 (2006).

Carmel, I., Tal, S., Vig, I. and Ast, G. Comparative analysis detects dependencies among the 5' splice-site positions. *Rna* **10**, 828-40 (2004).

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63 (2005).

Cassidy, S.B and Driscoll, D.J., Prader–Willi syndrome. *Eur. J. Hum. Genet.* **17**, 3–13 (2009).

Cavaille, J. and Bachellerie, J.P. SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.*, **26**, 1576–1587 (1998).

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., Huttenhofer, A. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci.* USA **97**, 14311–14316 (2000).

Cavaille, J., Nicoloso, M. and Bachellerie, J.P. Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature*, **383**, 732-5 (1996).

Chamberlain, S.J. and Lalandea, M. Neurodevelopmental disorders involving genomic imprinting at human chromosome 15q11–q13 *Neurobiol Dis.* **7**, 210-19 (2010).

Chamberlain, S.J., Johnstone, K.A., DuBose, A.J., Simon, T.A., Bartolomei, M.S., Resnick, J.L. and Brannan, C.I. Evidence for genetic modifiers of postnatal lethality in PWS-IC deletion mice. *Hum. Mol. Genet.* **13,** 123-08 (2004).

Canal, C.E., Mahautmr, K.C., Cao,C., Sanders-Bush, E. and Airey, D. C. RNA editing of the serotonin 2C receptor and expression of Gαq protein: genetic mouse models do not support a role for regulation or compensation. *J Neurochem.* **108**(5),1136–1142 (2009).

Charpentier, B. and Rosbash, M. Intramolecular structure in yeast introns aids the early steps of *in vitro* spliceosome assembly. *Rna* **2**, 509-22 (1996).

Cheah, M.T., Wachter, A., Sudarsan, N. and Breaker, R.R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497-500 (2007).

Chen, T., Cote, J., Carvajal, H.V. and Richard, S. Identification of Sam68 Arginine Glycine-rich Sequences Capable of Conferring Nonspecific RNA Binding to the GSG Domain. *J. Biol. Chem.*, **276**, 30803-11 (2001).

Cole, C., Sobala, A., Lu Shawn, C., Thatcher, R., Bowman, A., Brown, J.W.S., Green, P.J., Barton, G. J. and Hutvagner, G. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *Rna* **15**, 2147-60 (2009).

De Smith, A.J., Purmann, C., Walters, R.G., Ellis, R.J., Holder, S.E., Van Haelst, M.M, Brady, A.F., Fairbrother, U.L., Dattani, M. and Keogh, J.M. et al. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum. Mol. Genet.* **18**, 3257–3265 (2009).

Ding, F., Li, H.H., Zhang, S., Solomon, N.M., Camper, S.A., Cohen, P. and Francke, U. SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* **3,** 435-42 (2008).

Ding, F., Prints, Y., Dhar, M.S., Johnson, D.K., Garnacho-Montero, C., Nicholls, R.D., Francke, U. Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader–Willi syndrome mouse models. *Mamm. Genome* **16,** 424–31 (2005).

Doe, C.M. et al. Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. *Hum Mol Genet* **18**, 2140-8 (2009).

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., Meister, G. A human snoRNA with microRNA-like functions. *Mol. Cell* **32**, 519–28 (2008).

Filipowicz, W. and Pogacic, V. Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol.* **3**, 319-27 (2002).

Gabriel, J.M., Merchant, M., Ohta, T., Ji, Y., Caldwell, R.G., Ramsey, M.J., Tucker, J.D., Longnecker, R. and Nicholls, R.D. A transgene insertion creating a heritable chromosome

deletion mouse model of Prader–Willi and Angelman syndromes. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 545-62 (1999).

Ganot, P., Bortolin, M.L. and Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799-809 (1997).

Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X. and Kiss, T. Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol.* **10**, 6906-17 (1999).

Gardner, J.M., Nakatsu, Y., Gondo, Y., Lee, S., Lyon, M.F., King, R.A. and Brilliant, M.H. The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman syndromes. *Science,* **257**, 1121-4 (1992).

Gerard, M., Hernandez, L., Wevrick, R. and Stewart, C.L. Disruption of the mouse necdin gene results in early post-natal lethality, *Nat. Genet.* **23,** 458-54 (1999).

Graveley, B.R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65-73 (2005).

Grover, A. et al. 5' splice site mutation in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **274**, 15134-43 (1999).

Hiller, M., Zhang, Z., Backofen, R., Stamm, S. Pre-mRNA Secondary Structures Influence Exon Recognition. *PLoS Genet.* **3** (11): e204 (2007).

Hutzinger, R., Feederle, R., Mrazek, J., Schiefermeier, N., Balwierz, P.J., Zavolan, M., Polacek, N., Delecluse, H.J., Huttenhofer, A. Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome. *PLoS Pathog.* **5,** 78-83 (2009).

Jiang, Y.H., Armstrong, D., Albrecht, U., Atkins, C.M., Noebels, J.L., Eichele, G., Sweatt, J.D. and Beaudet, A.L. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation, *Neuron* **21**, 123-32 (1998).

Johnstone, K.A., DuBose, A.J., Futtner, C.R., Elmore, M.D., Brannan, C.I. and Resnick, J.L. A human imprinting centre demonstrates conserved acquisition but diverged

maintenance of imprinting in a mouse model for Angelman syndrome imprinting defects. *Hum. Mol. Genet.* **15,** 230-42 (2006).

Kanber, D., Giltay, J., Wieczorek, D., Zogel, C., Hochstenbach, R., Caliebe, A., Kuechler, A., Horsthemke, B., Buiting, K. A paternal deletion of MKRN3, MAGEL2 and NDN does not result in Prader–Willi syndrome. *Eur. J. Hum. Genet.* **17**, 582–90 (2009).

Kapranov, P. et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-8 (2007).

Kawahara, Y. et al. Dysregulated editing of serotonin 2C receptor mRNAs results in energy dissipation and loss of fat mass. *J Neurosci.* **28**, 12834-44 (2008).

Kent, W.J. BLAT - the BLAST-like alignment tool. *Genome Res.* **12**(4),656-64 (2002).

Kishore, S. and Stamm, S. Regulation of alternative splicing by snoRNAs. *Cold Spring Harb. Symp. Quant. Biol.*, **LXXI**, 329–34 (2006).

Kishore, S. and Stamm, S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**, 230–32 (2006).

Kishore, S., Khanna, A., Stamm S. Rapid generation of splicing reporters with pSpliceExpress. Gene **427**, 104–10 (2008).

Kiss-Laszlo, Z., Henry, Y. and Kiss, T. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* **3**, 797-807 (1998).

Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M. and Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **7**, 1077-88 (1996).

Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. and Tollervey, D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **4**, 527-37 (1998).

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins D.G. ClustalW and ClustalX version 2. *Bioinformatics* **23**(21), 2947-2948 (2007).

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294,** 858–62 (2001).

Lee, K.A., Bindereif, A., Green, M.R. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Tech.* **5**, 22–31 (1988).

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer A.C., Blume J.E., Wang X., Darnell J.C., Darnell R.B. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–69 (2008).

Liu, S. et al. Inhibition of HIV-1 multiplication by antisense U7 snRNAs and siRNAs targeting cyclophilin A. *Nucleic Acids Res* **32**, 3752-9 (2004).

Luhrmann, R., Kastner, B. & Bach, M. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim Biophys Acta* **1087**, 265-92 (1990).

Malone, C.D. et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522-35 (2009).

Matera, A.G., Terns, R.M., Terns, M.P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 209–20 (2007).

Mattick, J.S. and Makunin, I.V. Small regulatory RNAs in mammals. *Hum Mol Genet.* **14**, 121-32 (2005).

Mercer, R.E. and Wevrick, R. Loss of magel2, a candidate gene for features of Prader–Willi syndrome, impairs reproductive function in mice, *PLoS One* **4** (2009).

Meyer, K. et al. Rescue of a severe mouse model for spinal muscular atrophy by U7 snRNA-mediated splicing modulation. *Hum Mol Genet* **18**, 546-55 (2009).

Miura, K., Kishino, T., Li, E., Webber, H., Dikkes, P., Holmes, G.L. and Wagstaff, J. Neurobehavioral and electroencephalographic abnormalities in Ube3a maternal-deficient mice, *Neurobiol. Dis.* **9,** 321-29 (2002)

Muscatelli, F., Abrous, D.N., Massacrier, A., Boccaccio, I., Le Moal, M., Cau, P. and Cremer, H. Disruption of the mouse necdin gene results in hypothalamic and behavioral

alterations reminiscent of the human Prader–Willi syndrome. *Hum. Mol. Genet.* **9,** 732-48 (2000).

Nakatani, J., Tamada, K., Hatanaka, F., Ise, S., Ohta, H., Inoue, K., Tomonaga, S., Watanabe, Y., Chung, Y.J. and Banerjee, R. *et al.* Abnormal behavior in a chromosome-engineered mouse model for human 15q11–13 duplication seen in autism. *Cell* **137,** 236-53 (2009).

Newman, D.R., Kuhn, J.F., Shanab, G.M. and Maxwell, E.S. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861-79 (2000).

Nicholls, R.D. and Knepper, J.L. Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet.* **2**, 153-75 (2001).

Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R. and Luhrmann, R. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.* **21**, 6119-33 (1999).

Patel, A.A. and Steitz, J.A. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol.***12**, 960-70 (2003).

Potgieter, A.C., Steele, A. D. and van Dijk1, A. A. Cloning of complete genome sets of six dsRNA viruses using an improved cloning method for large dsRNA genes. *Journal of General Virology* **83**, 2215–23 (2002).

Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–17 (2004).

Runte M., Huttenhofer A., Gross S., Kiefmann M., Horsthemke B., Buiting K. The IC-SNURF–SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* **10**, 2687–2700 (2001).

Sahoo, D. del Gaudio, J.R. German, M. Shinawi, S.U. Peters, R.E. Person, A. Garnica, S.W. Cheung and A.L. Beaudet, Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.* **40,** 167-74 (2008).

Saraiya, A.A. and Wang, C.C. snoRNA, a novel precursor of microRNA in Giardia lamblia. *PLoS Pathog*. **4,** 230-46 (2008).

Schroeder, R., Grossberger, R., Pichler, A., Waldsich, C. RNA folding *in vivo*. *Curr Opin Struct Biol.* **12**, 296–300 (2002).

Schule, B., Albalwi, M., Northrop, E., Francis, D.I., Rowell, M., Slater, H.R., Gardner, R.J. and Francke, U. Molecular breakpoint cloning and gene expression studies of a novel translocation t(4;15)(q27;q11.2) associated with Prader-Willi syndrome. *BMC Med Genet.,* **6**, 18-26 (2005).

Scott, M.S., Avolio, F., Ono, M., Lamond, A.I., Barton, G.J. Human miRNA precursors with box H/ACA snoRNA features. *PLoS Comput. Biol.* **5,** e1000507 (2009).

Semenov, D.V., Vratskih, O.V., Kuligina, E.V. and Richter, V.A. Splicing by exon exclusion impaired by artificial box c/d RNA targeted to branch-point adenosine. *Ann N Y Acad Sci* **1137**, 119-24 (2008).

Singh, N.N., Shishimorova, M., Cao, L.C., Gangwani, L. and Singh, R.N. A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy. *RNA Biol* **6**, 341-50 (2009).

Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C. and Muntoni, F. Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc Natl Acad Sci* USA **100**, 4114-9 (2003).

Skryabin, B.V., Gubar, L.V., Seeger, B., Pfeiffer, J., Handel, S., Robeck, T., Karpova, E., Rozhdestvensky, T.S. and Brosius, J. Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet.* **3,** e1002707 (2007).

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., Zhang, M.Q. An alternative-exon database and its statistical analysis. *DNA Cell Biol.* **19**, 739–756 (2000).

Stefan, M., Ji, H., Simmons, R.A., Cummings, D.E., Ahima, R.S., Friedman, M.I., Nicholls, R.D. Hormonal and metabolic defects in a Prader–Willi syndrome mouse model with neonatal failure to thrive. *Endocrinology* **146**, 4377–85 (2005).

Stoss, O., Novoyatleva, T., Gencheva, M., Olbrich, M., Benderska, N. and Stamm, S. p59(fyn)-mediated phosphorylation regulates the activity of the tissue-specific splicing factor rSLM-1. *Mol Cell Neurosci.* **27**, 8-21 (2004).

Taft, R.J., Glazov, EA, Lassmann, T, Hayashizaki, Y, Carninci, P, Mattick, JS. Small RNAs derived from snoRNAs. *Rna* **15**, 1233-40 (2009).

Tang, Y., Novoyatleva, T., Benderska, N., Kishore, S**.**, Thanaraj, A. and Stamm, S. Analysis of Alternative Splicing *In vivo* using Minigenes. *Handbook of RNA Biochemistry,* WILEY-VCH Verlag GmbH & Co. KHaA, Weinheim (2005).

Terns, M.P. and Terns, R.M. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr.*, **10**, 17-39 (2002).

Thanaraj T.A., Stamm S., Clark F., Riethoven J.J.M, Le Texier V., and Muilu J. ASD: the Alternative Splicing Database. *Nucl. Acids. Res.*, **32**, D64-D69 (2004).

Tsai, T.F., Armstrong, D. and Beaudet, A.L. Necdin-deficient mice do not show lethality or the obesity and infertility of Prader–Willi syndrome. *Nat. Genet.* **22,** 52-58 (1999).

Tsai, T.F., Jiang, Y.H., Bressler, J., Armstrong, D. and Beaudet, A.L. Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader–Willi syndrome, *Hum. Mol. Genet.* **8,** 348-61 (1999).

Tu, M., Tong, W., Perkins, R. & Valentine, C.R. Predicted changes in pre-mRNA secondary structure vary in their association with exon skipping for mutations in exons 2, 4, and 8 of the Hprt gene and exon 51 of the fibrillin gene. *Mutat Res* **432**, 15-32 (2000).

Vickers, S.P., Dourish, C.T. and Kennett, G.A. Evidence that hypophagia induced by d-fenfluramine and d-norfenfluramine in the rat is mediated by 5-HT2C receptors. *Neuropharmacology* **41**, 200-9 (2001).

Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaille, J., Huttenhofer, A. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.* **169**, 745–753 (2005).

Vitali, P., Royo, H., Marty, V., Bortolin-Cavaille, M.L., Cavaille, J. Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J. Cell. Sci.* **123**, 70–83 (2010).

Voechovsk, I. Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res* **35**, 4250-63 (2007).

Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant C., Luhrmann, R. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457–66 (2000).

Willingham, A.T. and Gingeras, T.R. TUF love for "junk" DNA. *Cell* **125**, 1215-20 (2006).

Xu, Y. et al. 5-HT2CRs expressed by pro-opiomelanocortin neurons regulate energy homeostasis. *Neuron* **60**, 582-9 (2008).

Yang, T., Adamson, T.E., Resnick, J.L., Leff, S., Wevrick, R., Francke, U., Jenkins, N.A., Copeland N.G. and Brannan, C.I. A mouse model for Prader–Willi syndrome imprinting-centre mutations. *Nat. Genet.* **19,** 136-41 (1998).

Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L. and Carbon, J. Point mutations in yeast CBF5 can abolish *in vivo* pseudouridylation of rRNA. *Mol Cell Biol.* **11**, 7461-72 (1999).

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., Dreyfuss, G. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell **133**, 585–600 (2008).

Zhao, X. & Yu, Y.T. Targeted pre-mRNA modification for gene silencing and regulation. *Nat Methods* **5**, 95-100 (2008).

Zhuang, Y. & Weiner, A.M. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827-35 (1986).

Curriculum Vitae

Name:	**Amit Khanna**
Date of Birth:	13th February, 1981.
Place of Birth:	Chandigarh, India.

Education:

1996	Primary schooling, GMSS School Chandigarh, India. (75%)
1999	Higher Secondary, SGGS School. (72%)
2002	Bachelor's degree in Biotechnology Punjab Univsersity (66%)
2004	Master's degree in Human Genomics Punjab Univsersity (66%)
Jan' 2007-	Graduate student at Stamm's lab, University of Erlangen-Nürnberg, Germany.

Professional Experience:

August 2004- December 2006	Worked as senior research assistant with Post graduate institute of medical education and research (PGIMER) Chandigarh in a DST funded project named "Transcriptional regulation of PPAR γ".

Honors and awards:

2009	Thomas C. Vanaman Graduate student best talk award
2009	As an instructor in EURASNET workshop on alternative splicing and disease
2002	Secured second rank in all India entrance exam for masters in human genomics

Publications:

Kishore S*, **Khanna A*,** Zhang Z, Hui J, Balwierz PJ, Stefan M, Beach C, Nicholls RD, Zavolan M, Stamm S.The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet.* 2010.

* Co-first authors

Khanna A and Stamm S Regulation of alternative splicing by short non-coding nuclear RNAs. *RNA Biology* (In communication) 2010.

Kishore S, **Khanna A**, Stamm S. Rapid generation of splicing reporters with pSpliceExpress. *Gene.* 2008.

Hui J, Kishore S, **Khanna A** and Stamm S. Analysis of Alternative splicing with Microarrays. *Bioinformatics for Systems Biology* 2008

Kaul D, Anand PK, **Khanna A.** Functional genomics of PPAR-gamma in human immunomodulatory cells. *Mol Cell Biochem.* 2006.

Kaul D, **Khanna A**, Suman Detection of novel miRNA in HIV-1 genome and its role in cd4T-cell deletion. *PINAS India* 2006.