



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Functional characterization and therapy of human pathogenic splicing mutations

Funktionelle Charakterisierung und Therapie von humanen pathogenen Spleißmutationen

Inaugural-Dissertation

zur

**Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf**

vorgelegt von

Linda Hartmann
aus Duisburg

Düsseldorf, April 2012

Aus dem Institut für Virologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. rer. nat. Heiner Schaal,
Institut für Virologie

Ko-Referent: Prof. Dr. rer. nat. Rolf Wagner
Institut für Physikalische Biologie

Tag der mündlichen Prüfung: 30.Mai 2012

Table of contents

LIST OF FIGURES	10
LIST OF ABBREVIATIONS	13
ZUSAMMENFASSUNG	15
SUMMARY	16
1. Introduction	17
1.1. The human genome	17
1.1.1 The split nature of human genes	17
1.1.2 The complex organization of the human genome	18
1.1.3 Genetic factors in human disease	20
1.1.3.1 Cancer susceptibility genes	20
1.1.4 Splicing in human disease	22
1.2. Splicing	22
1.2.1 Sequences of human splice sites	22
1.2.2 The mechanism of splicing	24
1.2.3 The spliceosome	25
1.2.4 Spliceosome assembly and catalysis	28
1.2.5 Splice site and exon recognition	35
1.2.5.1 5'ss recognition	35
1.2.5.2 3'ss recognition	39
1.2.5.3 <i>Cis</i> -active regulatory elements	42
1.2.5.4 Exon recognition	44
1.2.6 Splice site strength and identification of pathogenic splicing mutations	46
1.2.6.1 5'splice strength algorithms	47
1.2.6.2 3'splice strength algorithms	51
2 RESULTS	53
2.1. Requirements for the recognition of human exons with weak splice donor sites within a heterologous splicing minigene	53
2.1.1. Faithful <i>ATM</i> exon 54 recognition and intron removal in a heterologous splicing reporter minigene requires a strong terminal splice acceptor	53
2.1.2. Adjacent genuine downstream intron segment promotes <i>ATM</i> exon 54 definition within the heterologous splicing reporter minigene	56
2.1.3. The adjacent genuine downstream intron sequence contributes to <i>ATM</i> exon 54 definition in a sequence specific manner	58
2.1.4. The sequence of the proximal downstream genuine intron fragment of <i>ATM</i> exon 54 enhances splice donor recognition	59
2.1.5. Identification of proteins bound to the intronic <i>ATM</i> fragments	61

Table of contents

2.1.6. Results from <i>ATM</i> exon 54 are applicable to <i>ATM</i> exon 9	64
2.1.7. An extended genomic context is negligible for <i>ATM</i> exon 54 and <i>ATM</i> exon 9 recognition.....	65
2.2. Functional characterization of putative pathogenic splice donor mutations.....	67
2.2.1. Single point mutations within the splice donor of <i>ATM</i> exon 54 and <i>ATM</i> exon 9 found in ataxia telangiectasia patients cause loss of exon recognition.....	67
2.2.2. The <i>RAD51C</i> c.904+5G>T mutation in familial breast and ovarian cancer pedigree causes loss of <i>RAD51C</i> exon 6 recognition.....	68
2.2.3. The c.145+1G>T mutation within the splice donor site of <i>RAD51C</i> exon 1 resulted in enhanced production of non-functional <i>RAD51C</i> transcripts	69
2.3. The impact of a homozygous micro-deletion in <i>BRCA2</i> exon 6 on splicing	73
2.3.1. The effect of the homozygous micro-deletion on <i>BRCA2</i> exon 6 recognition	74
2.3.2. The biallelic micro-deletion in <i>BRCA2</i> exon 6 causes the generation of an additional in-frame transcript with unique skipping of exon 5 in the male patient	75
2.3.3. HnRNP H1, hnRNP A1 and hnRNP M4 bind to nucleotides deleted in the patient-derived <i>BRCA2</i> exon 6.....	78
2.4. Mechanisms of cryptic splice donor activation upon the <i>FGB</i> IVS7 +1G>T splice donor mutation.....	80
2.4.1 The <i>FGB</i> IVS7 +1G>T splice donor mutation causes activation of a putative splice donor in the downstream intron.....	80
2.4.2. Increasing the complementarity of the cryptic splice site c1 to U1 snRNA exceeding the natural site results in low-level activation of the cryptic site	82
2.4.3. Increased intrinsic strength of the cryptic splice site c3 exclusively activates this cryptic site	84
2.4.4. <i>FGB</i> exon 7 contains multiple splicing enhancer elements.....	86
2.5. Identification and characterization of a non-canonical TT splice donor.....	88
2.5.1. The <i>FANCC</i> c.165 +1G>T splice donor mutation in primary cells of FA-C patients allows correct splicing albeit at a reduced level	88
2.5.2. Increased complementarity to U1 snRNA specifically reconstitutes splicing at the TT dinucleotide in the heterologous splicing reporter minigene.....	91
2.5.3 Artificial TT-adapted U1 snRNAs improve correct mRNA processing at the <i>FANCC</i> TT splice donor within the splicing reporter.....	93
Fig. 26: The U1 snRNA A7 has no effect on usage of the <i>FANCC</i> TT splice donor within the splicing minigene.....	96
2.5.4 Full complementarity of the <i>FGB</i> TT splice donor to U1 snRNA results in activation of close-by GT dinucleotides.....	97

Table of contents

2.5.5. Intrinsic features of the TT splice donor sequence determine the exclusive usage of the TT as splice donor.....	102
2.5.6. The genomic context of FANCC exon 2 enhances splicing at the pathogenic TT splice donor.....	104
2.6. An U1 snRNA based therapy approach for human splice donor mutations.....	106
2.6.1. Ectopic expression of the TT-adapted U1 snRNAs specifically enhances the amount of the endogenous in-frame transcript in FA patient-derived fibroblasts.....	106
2.6.2. Phenotypic correction of FANCC-mutant fibroblasts by integrating lentivirus-mediated expression of TT-adapted U1 snRNAs.....	106
2.6.3. Delivery of extended compensatory U1 snRNA molecules can improve exon recognition within patient-derived FANCC c.456+4A>T fibroblasts	110
2.6.4. Delivery of an extended suppressor U1 snRNA does not produce sufficient amounts of endogenous functional FANCC transcript for phenotype correction of c.456 +4A>T fibroblasts	113
3. Discussion	115
3.1. <i>In vivo</i> analysis of human exon recognition in a heterologous minigene.....	116
3.2. Functional splicing assay contributes to establishment of RAD51C as cancer susceptibility gene	120
3.3. hnRNP H1, A1 and M4 seem to be involved in an exon definition net within the BRCA 2 transcript.....	122
3.4. The local enhancer density and splice donor strength might bring about the decision between exon skipping or cryptic splice site activation.....	124
3.5. Intrinsic features of the 5'ss and the genomic context of FANCC exon 2 allow functional splicing at a mutant +1G>T splice donor.....	127
3.6. A novel U1 snRNA based therapy approach for human splice donor mutations.....	131
4. MATERIALS AND METHODS	134
4.1. Material.....	134
4.1.1. Chemicals and Consumables	134
4.1.2. Enzymes	135
4.1.3. Bacteria	135
4.1.4. Cells	136
4.1.4.1 Human cell lines	136
4.1.4.2 Human primary cells.....	136
4.1.5.1 Oligonucleotides for cloning	137
4.1.6.2 Oligonucleotides for RT-PCR	141
1.6.3 Oligonucleotides for RNA affinity chromatography.....	142

Table of contents

4.1.7. Recombinant plasmids.....	142
4.1.7.1. Three-exon-two-intron splicing reporter minigenes.....	143
4.1.7.2. 1-intron-2-exon splicing reporter minigenes	149
4.1.7.3. SV-env/eGFP reporter plasmids.....	149
4.1.7.4. U1 snRNA expression plasmids	150
4.1.7.5. Control plasmids.....	151
4.1.7.6. Plasmids for protein expression	151
4.1.5.7. Lentiviral vectors.....	151
4.1.8. Antibodies	153
4.1.8.1. Primary Antibodies	153
4.1.8.2. Secondary Antibodies.....	153
4.2. Methods	153
4.2.1. Cloning.....	153
4.2.1.1. Polymerase Chain Reaction (PCR).....	153
4.2.1.2 Restriction and purification of PCR products or plasmid fragments using agarose gel electrophoresis	154
4.2.1.3. Ligation	154
4.2.1.4. Transformation	154
4.2.1.5. Analytical plasmid DNA isolation	154
4.2.1.6. Preparative plasmid DNA isolation	155
4.2.1.7. DNA sequencing.....	155
4.2.2. Eukaryotic cell culture	156
4.2.2.1 Cell Culture and Transfection	156
4.2.4. Flow cytometrical analysis of transiently transfected HeLa cells	156
4.2.3 Lentiviral particle production	157
4.2.5. Reverse transcriptase (RT)-PCR analysis	157
4.2.5.1. Isolation of total RNA using anionic exchange columns.....	157
4.2.5.2. Reverse transcription and PCR analysis.....	157
4.2.5.3. Native gel electrophoresis and EtBr staining to visualize RT-PCR products	158
4.2.5.4. Purification of RT-PCR products from native polyacrylamide gels (PAA)	158
4.2.6. RNA affinity chromatography	158
4.2.6.1. Purification of DNA oligos.....	158
4.2.6.2. Expression and purification of recombinant T7 RNA polymerase	159
4.2.6.3. <i>In vitro</i> transcription	160
4.2.6.4. Protein isolation by RNA affinity chromatography	160
4.2.6.7.1 Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis	161

Table of contents

4.2.4.7.2 Immunoblotting	161
4.2.8 Protein sequencing by mass spectrometry	162
4.2.8.1. In gel digestion and sample preparation.....	162
4.2.8.2. Mass spectrometry	162
4.2.9. FANCD2 immunoblotting	163
4.2.10 Foci assay	164
4.2.11. Cell cycle analysis.....	164
REFERENCES.....	166
ANHANG	196
PUBLIKATIONEN (MIT EIGENER BETEILIGUNG).....	196
DANKSAGUNG.....	197
ERKLÄRUNG	198

LIST OF FIGURES

Introduction:

- Fig. I1: Discovery of splicing in adenovirus messenger RNA
(taken from Sharp, 1994)
- Fig. I2: Sequences of human splice sites
- Fig. I3: The splicing reaction itself ensues by two consecutive trans-esterification reactions
- Fig. I4: Assembly cycle of the human major spliceosome
- Fig. I5: The network of RNA interactions in the pre-catalytic and catalytically activated spliceosome
- Fig. I6: Recognition of the 5' splice site by RNA duplex formation between U1 snRNA
- Fig. I7: 3' splice site recognition during early spliceosomal complex formation
- Fig. I8: Exon recognition in the human genome
- Fig. I9: H-Bond Score distribution of 7,849 real 5' splice sites
- Fig. I10: Sequence motifs for 3' splice site cluster
(taken from Yeo & Burge, 2004)

Results:

- Fig. 1: Faithful *ATM* exon 54 exon recognition and intron removal in a heterologous splicing reporter minigene requires an optimized terminal splice acceptor
- Fig. 2: Presence of the proximal downstream genuine intron fragment promotes *ATM* exon 54 definition in the heterologous splicing reporter minigene
- Fig. 3: The proximal downstream genuine intron fragment contributes to *ATM* exon 54 definition
- Fig. 4: Sequence of the proximal downstream genuine intron fragment of *ATM* exon 54 enhances splice donor recognition
- Fig. 5: HnRNP A2/B1 and hnRNP A1 bind *ATM* intron fragment II
- Fig. 6: Analysis of *ATM* exon 9 recognition in the heterologous splicing reporter minigene
- Fig. 7: Analysis of *ATM* exon 54 and *ATM* exon 9 recognition in subgenomic minigenes
- Fig. 8: Single point mutations within the splice donor of *ATM* exon 54 and *ATM* exon 9 found in telangiectasia patients cause loss of exon recognition

List of Figures

- Fig.9: The *RAD51C* c.904+5G>T mutation in a familial breast and ovarian cancer pedigree causes loss of *RAD51C* exon 6 recognition
- Fig.10: The c.145+1G>T mutation within the splice donor site of *RAD51C* exon 1 resulted in enhanced production of non-functional *RAD51C* transcripts
- Fig.11: Non-functional *RAD51C* transcripts 008 and 009 in peripheral blood leukocytes are produced by usage of alternative splice donor sites in *RAD51C* exon 1
- Fig.12: Splicing minigene demonstrates that mutant allele failed to produce a functional transcript in case of the *RAD51C* c.145+1G>T mutation
- Fig.13: The effect of the homozygous micro-deletion on *BRCA2* exon 6 recognition
- Fig.14: Splicing pattern of the *BRCA2* mRNA in normal and patient-derived (del 707-716) fibroblasts and lymphoblastoid B-cell lines
- Fig.15: Splicing patterns of the *BRCA2* pre-mRNA in EBV immortalized lymphoblastoid B-cell lines of healthy male and female controls
- Fig.16: HnRNP H1, hnRNP A1 and hnRNP M4 bind to the *BRCA2* exon 6 sequence affected by the 10bp micro-deletion
- Fig.17: The *FGB* IVS7+1G>T mutation causes activation of the putative splice donor site p1 in the downstream intron in addition to activation of cryptic splice sites in *FGB* exon 7
- Fig. 18: Increasing the complementarity of the cryptic splice site c1 to U1 snRNA exceeding the complementarity of the natural site results in low-level activation of the cryptic site
- Fig.19: An increased intrinsic strength of the cryptic splice site c3 exclusively activates this cryptic splice site
- Fig. 20: *FGB* exon 7 contains multiple splicing enhancer elements
- Fig.21: Homozygous c.165 +1G>T splice donor mutation in *FANCC* allows correct splicing at low level.
- Fig.22: Additional maternally inherited genomic deletion of *FANCC* exon 2 and 3 in family 640 on the second allele
- Fig.23: Improved complementarity to U1 snRNA reconstitutes splicing at the TT dinucleotide in the heterologous splicing reporter construct.
- Fig. 24: TT-adapted U1 snRNAs restored usage of the *FANCC* TT 5'ss within the minigene splicing reporter
- Fig.25: U1 snRNA α TT did not influence the ratio of splicing at the TT dinucleotide and splicing at the GT at position -1 within the improved *FANCC* TT 5'ss but increased the overall level of *FANCC* exon 2 inclusion

List of Figures

- Fig. 26: The U1 snRNA A7 has no effect on usage of the *FANCC* TT splice donor within the splicing minigene.
- Fig.27: An increased complementarity to U1 snRNA of the *FGB* IVS7 +1G>T splice donor reconstitutes splicing at the TT dinucleotides but also activates close-by GT dinucleotides
- Fig. 28: Decreased intrinsic strength of the GT dinucleotide at position +5/+6 within the optimized *FGB* TT 5'ss results in splicing at the GT dinucleotide one position upstream (position -1) of the TT dinucleotide in addition to splicing at the TT dinucleotide
- Fig.29: Assessment of the intrinsic strength of close-by GT dinucleotides within the *FGB* and *FANCC* TT splice donor
- Fig.30: Exclusive cleavage at the TT splice donor depends on intrinsic sequence features of the TT splice donor
- Fig. 31: Dependency of splicing at the TT dinucleotide on the genomic *FANCC* context
- Fig.32: Ectopic expression of the TT-adapted U1 snRNAs specifically enhanced the amount of the endogenous in-frame transcript in fibroblasts from the index patient in pedigree 526
- Fig.33: Lentivirus-mediated expression of TT-adapted U1 snRNAs is capable to improve *FANCC* exon 2 inclusion and *FANCD2*-monoubiquitination in patient-derived fibroblasts
- Fig. 34: Phenotypic correction of primary biallelic c.165 +1G>T *FANCC* fibroblasts of the index patient from pedigree 526 by lentivirus-mediated expression of TT-adapted U1 snRNAs
- Fig. 35: Function of *FANCD2* in biallelic *FANCC* c.165 +1G>T fibroblasts
- Fig. 36: The +4A>T splice donor mutation in *FANCC* exon 5 completely abolishes exon inclusion
- Fig. 37: Enhanced complementary by extension of U1 snRNA increases the amount of the transcript with *FANCC* exon 5 inclusion
- Fig. 38: Lentivirus-mediated expression of gene specific extended full complementary U1 snRNAs does not generate sufficient amounts of the functional transcript for phenotypic correction of the +4A>T splice donor mutation in *FANCC* exon 5.

Discussion

- Fig. D1: Model for functional exon recognition.
- Fig. D2: Model for functional splicing at a non-canonical TT splice donor
- Fig. D3: Transduction of the patients' cells with the TT adapted U1 snRNA molecules did not cause off-target effects like cryptic splice site activation

List of Abbreviations

LIST OF ABBREVIATIONS

3' ss	3' splice site
5' ss	5' splice site
Ac	Acetate
Amp	Ampicillin
AP	Alkaline Phosphatase
ATP	adenosine-5'-triphosphat
BSA	bovine serum albumin
CTP	cytidine-5'-triphosphate
ddH ₂ O	deionised and distilled water
DMDC	dimethyldicarbonate
DMEM	Dulbecco's modified Eagle's medium
DNA	desoxyribonucleidacid
DNase	desoxyribonuclease
DTT	dithiotreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
<i>env</i>	gene for the viral membrane protein (envelope)
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EtBr	ethidium bromide (3,8-Diamino-6-ethyl-5-phenylphenatridiumbromid)
FCS	fetal calf serum
<i>Gag</i>	gene for the viral structural proteins (group specific antigen)
gp	glycoprotein
GTP	guanosine-5'-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGH	human growth hormone
HIV-1	Human Immunodeficiency Virus Type 1
hnRNP	heterogeneous nuclear ribonucleoprotein particle
LB	Luria Broth base
LTR	long terminal repeat
mRNA	messenger ribobnucleidacid
ORF	open reading frame
ori	origin of replication
pA	polyadenylation signal
PBS	phosphate buffered saline
PBSdef	Dulbecco's phosphate buffered saline deficient in Ca ²⁺ and Mg ²⁺
PCR	polymerase chain reaction
PMSF	phenylmethane-sulfonyl-fluoride
<i>pol</i>	gene for the viral enzymes (polymerase)
poly	(A) ⁺ polyadenylated
<i>rev</i>	gene for the viral protein Rev (regulator of viral protein expression)
RNA	ribonucleidacid
RNase	ribonuclease
RRE	Rev-responsive element
RS	arginine/serine-rich
SA	splice acceptor
SD	splice donor
SELEX	Systematic Evolution of Ligands by Exponential enrichment
SDS	sodiumdodecylsulfat
SR	serine-arginine-rich

List of Abbreviations

SSC	standard saline citrate
SU	viral surface envelope protein
SV40	Simian Virus 40
TE	Tris-EDTA buffer
TM	viral transmembrane envelope protein
Tris	Tris-(hydroxymethyl)-aminomethan
TTP	thymidine-5'-triphosphate
UV	ultraviolet

v/v	volume per volume
w/v	weight per volume
bp	base pairs m meter
°C	degree Celsius
min	minutes
M	molar
g	gramm
n	nano (10 ⁻⁹)
h	hour nt nucleotide
kb	kilobases
RLU	relative light units
kDa	kilodalton
rpm	rotations per minute
l	liter sec second
μ	micro (10 ⁻⁶)
U	unit
m	milli (10 ⁻³) V volt

ZUSAMMENFASSUNG

Humane Genmutationen, die ein präzises Spleißen der prä-mRNA verhindern, werden zunehmend als ein wichtiger krankheitsauslösender Mechanismus anerkannt. Die akkurate Erkennung von kodierenden Exon-Sequenzen ist eine Grundvoraussetzung für die Generierung von intakten und funktionsfähigen Proteinen. Auch wenn computergestützte Analysen die Wahrscheinlichkeit für ein fehlerhaftes Spleißen mit beachtlichem Erfolg vorhersagen können, erfordert eine vertrauenswürdige Diagnose des Spleiß-Phänotyps immer noch funktionelle Spleiß-Assays.

In der vorliegenden Arbeit wurde ein auf einem Spleiß-Reporter Minigen basierender funktionaler Spleiß-Assay genutzt, um die Voraussetzungen für die Exon-Erkennung in einem heterologen Kontext zu ermitteln. Es zeigte sich, dass die Erkennung der humanen *ATM* Exons 9 und 54 nicht nur von jeweiligen Exon-Sequenz und den flankierenden Spleißstellen sondern auch von der Stärke der Spleißstellen im Spleiß-Reporter Konstrukt abhängig war. Hierbei kann die natürliche Intron-Sequenz unmittelbar stromabwärts von der Spleißdonor-Stelle einen ausschlaggebenden Einfluss auf die Exon-Erkennung im Minigenkonstrukt haben. Es wurde bestätigt, dass zwei putativ pathogene Spleißdonor Mutationen im *RAD51C* Gen den Verlust der Exon-Erkennung oder die Aktivierung von kryptischen Spleißstellen verursachen, wodurch diese eindeutig mit einem erhöhten Risiko für ein Mamma- oder Ovarkarzinom assoziiert werden konnten. Mittels RNA-Affinitätschromatographie und Massenspektrometrie wurde nachgewiesen, dass die spleißregulatorischen Proteine hnRNP H1, A1 und M4 an die wildtypische Sequenz von *BRCA2* Exon 6 binden, jedoch nicht, wenn eine bei an Fanconi Anämie erkrankten Patienten gefundene Mikro-Deletion in diesem Exon vorliegt. Dieser Befund stand im Einklang einer Veränderung des *BRCA2* Spleißmusters durch die Mikro-Deletion. Weiterhin wurde der Mechanismus der Aktivierung von kryptischen Spleißstellen am Beispiel einer in *FGB* Exon 7 beschriebenen +1G>T Spleißdonor-Mutation untersucht. Die Ergebnisse ließen die Schlußfolgerung zu, dass die lokale Dichte an spleißfördernden Enhancer-Elementen und auch die Spleißdonor-Stärke entscheidend dafür sein könnte, ob eine Spleißdonor-Mutation zum Verlust der Exon-Erkennung oder zur Aktivierung von kryptischen Spleißstellen führt.

Da die am häufigsten vorkommende Mutation in humanen Spleißdonor-Stellen bei erblich bedingten Erkrankungen das Guanosin-Nukleotid innerhalb des hochkonservierten GT-Dinukleotides betrifft, wurde eine in *FANCC* Exon 2 gefundene +1G>T Spleißdonor-Mutation auf ihre Pathogenität untersucht. Obwohl bisher angenommen wurde, dass jeder Basenaustausch an dieser Position eine normale mRNA-Prozessierung vollständig verhindert, zeigten die Ergebnisse dieser Arbeit unerwarteterweise, dass die Spleißdonor-Stelle trotz der +1G>T Mutation mit stark reduzierter Effizienz in primären Fibroblasten von an Fanconi Anämie erkrankten Patienten genutzt wurde. Die systematische Mutation und vergleichende Analyse der *FGB* Exon 7 und *FANCC* Exon 2 +1G>T Spleißdonor-Stelle in Minigenkonstrukten machte deutlich, dass sowohl das nicht-kanonische TT-Dinukleotid als auch der genomische Kontext von *FANCC* für das Spleißen an der mutanten Spleißdonor-Stelle erforderlich waren.

Die lentiviral vermittelte stabile Expression von an die pathogene *FANCC* TT Spleißdonor-Stelle adaptierten U1 snRNA Molekülen verbesserte spezifisch die Erkennung von *FANCC* Exon 2 und konnte die normale Mono-Ubiquitinierung des *FANCD2* Proteins wiederherstellen. Darüber hinaus komplementierte der lentivirale Transfer der TT-adaptierten U1 snRNA Moleküle den für Zellen von Fanconi Anämie Patienten typischen G2-Zellzyklusarrest nach Stimulation mit DNA-schädigen Substanzen. Damit wurde im Rahmen der vorliegenden Arbeit erstmalig ein neuer RNA-basierter Gentherapieansatz für die Therapie von Spleißmutationen aufgezeigt.

SUMMARY

Human gene mutations interfering with precise precursor messenger RNA (mRNA) splicing are increasingly recognized as an important mechanism through which gene mutations cause human disease, since accurate exon recognition is a mandatory prerequisite for generation of intact and functional proteins. Although *in silico* tools predict the probability for aberrant splicing with considerable success reliable diagnosis of the splicing phenotype of a splice site mutation still requires functional splicing assays due to the complex interplay of splice site-defining sequence elements.

In this thesis a functional splicing assay based on a splicing reporter minigene construct was used to assess the requirements for exon recognition in a heterologous context. The results of this thesis showed that recognition of human *ATM* exon 9 and 54 was not only dependent on the exon sequence and the strength of its flanking splice sites but also on the strength of the splice sites of the splicing reporter construct. It was demonstrated that the natural intron sequence immediately downstream of the splice donor site of both exons can have a crucial influence on exon recognition in the minigene construct. Using the splicing reporter minigene it was validated that two putative pathogenic splice donor mutations found in the *RAD51C* gene cause loss of exon recognition or activation of cryptic splice donor sites and therefore were clearly associated with an elevated risk of breast and ovarian cancer. Likewise it was shown that a patient-derived homozygous micro-deletion within *BRCA2* exon 6 interfered with splicing pattern of the *BRCA2* transcript. RNA affinity chromatography visualized three cellular proteins bound only to the wild type *BRCA2* exon 6 but not to the mutant one, which were subsequently identified by mass spectrometry to be the heterogenous ribonucleoproteins (hnRNPs) H1, A1 and M4, which were previously shown to be involved in the regulation of splicing. Further the mechanism of cryptic splice donor activation was investigated exemplified by a +1G>T splice donor mutation described in *FGB* exon 7 suggesting that the local density of splicing enhancer elements and splice donor strength might be decisive whether a splice donor mutation results in skipping of the affected exon or in activation of cryptic splice sites.

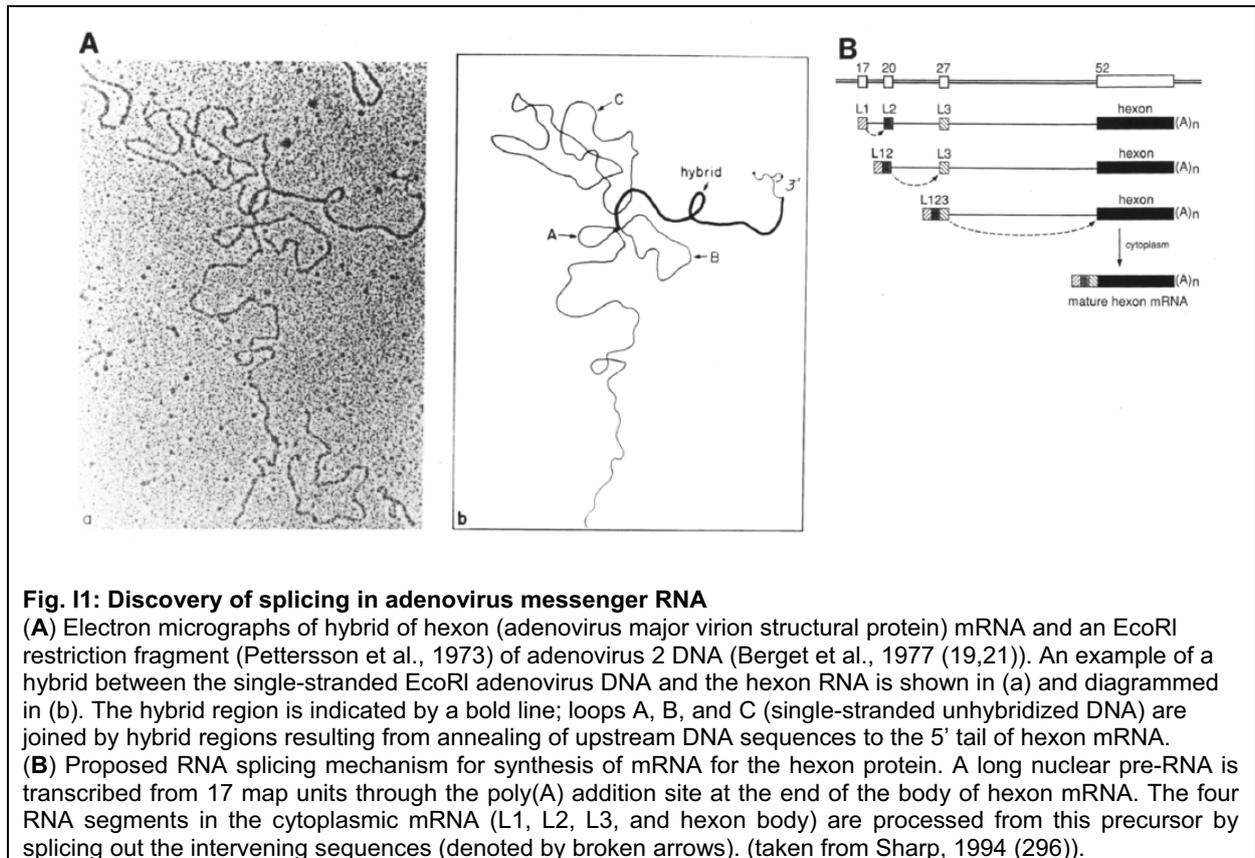
Since the most frequent base-pair mutation in human splice donor sites in inherited diseases comprises the guanosine within the highly conserved GT dinucleotide, a +1G>T splice donor mutation found in *FANCC* exon 2 was investigated for its pathogenicity within the context of this thesis. Although any base-pair substitution at this position was thought to completely abrogate normal mRNA processing the results of this thesis demonstrated in primary fibroblasts from Fanconi anemia patients that the mutation unexpectedly allowed correct splicing, albeit with decreased efficiency. Systematic mutation and comparative analysis of the *FGB* exon 7 and the *FANCC* exon 2 +1G>T splice donor within minigene constructs revealed that both the noncanonical TT dinucleotide and the genomic context of *FANCC* were required for the residual correct splicing at the mutant splice donor. Lentivirus-mediated expression of U1 snRNA molecules adapted to the mutant TT splice donor site specifically improved *FANCC* exon 2 inclusion and restored normal FANCD2-monoubiquitination in the patient-derived fibroblasts. Finally, lentiviral expression of the TT adapted U1 snRNA molecules corrected the DNA damage-induced G2 cell cycle arrest of primary patient derived fibroblasts. These data indicated that stably lentivirus-mediated expression of the TT-adapted U1 snRNA molecules can lead to the production of sufficient amounts of endogenous functional *FANCC* transcript for correction of the cellular phenotype of the disease, thus opening an alternative transcript-targeting approach for gene therapy of inherited splice site mutations.

1. INTRODUCTION

1.1. The human genome

1.1.1 The split nature of human genes

After the discovery of the structure of the DNA in the 1950's – it was generally thought that a gene is a contiguous string of base pairs, containing the information for the synthesis of a protein. The first indication that eukaryotic genes are not continuous like prokaryotic genes came when new methods allowing an accurate comparison of adenovirus DNA and the messenger RNA (mRNA) intermediate – delivering information from genes to ribosomes for protein synthesis – were applied to messenger RNAs produced by human adenovirus. When hybrids of the messenger RNA fraction coding for adenovirus major virion structural protein hexon and a single-stranded restriction endonuclease cleavage fragment of adenoviral DNA were visualized in the electron microscope, branched forms were observed which were not hydrogen bonded to the single-stranded DNA (Berget et al. 1977 (19,21) (**Fig. 11A**).



DNA sequences complementary to the messenger RNA sequences were found by electron microscopy to be located at 17, 20, and 27 units on the same strand indicating that the four segments of viral RNA may be joined together during the synthesis of the mature hexon messenger RNA (**Fig. 11B**). Thus, a model was suggested for adenovirus messenger RNA synthesis in which the initial transcript is processed into a mature messenger RNA by splicing out internal sequences (Berget et al., 1977 (19,21)).

Shortly after discovery of RNA splicing and split genes in adenovirus, a number of mammalian cellular genes were also shown to have intervening sequences. For example, it was discovered that the globin genes contain two intervening sequences (Jeffreys and Flavell, 1977 (145); Tilghman et al., 1978 (338)), that the ovalbumin gene is split into eight sets of sequences (Breathnach et al., 1977 (37)) and that the immunoglobulin genes contain both short and long intervening sequences (Tonegawa et al., 1978 (340)). In yeast, some tRNA (transfer RNA genes) were found to be interrupted by very short insertions and the sequences of these insertions were shown to be transcribed as a part of a precursor molecule (Goodman et al., 1977 (103); Valenzuela et al., 1978 (344); Hopper et al., 1978 (134)).

These observations suggested that in general in higher organisms the coding sequences on DNA, the regions that will ultimately be translated into an amino acid sequence, are not continuous but are interrupted by internal regions that are excised during maturation of the final messenger RNA being a spliced product. An alternative terminology, used by Gilbert and his colleagues referred to the intervening sequences as *introns*; those base sequences on the DNA which end up in the mRNA were referred as *exons* since they are the ones which are expressed (Gilbert, 1978 (99)).

1.1.2 The complex organization of the human genome

The human genome constitutes \approx 3 billion nucleotide base pairs. Of the 3.2 Gb (gigabases) that comprise the human genome 2.95 Gb are euchromatic (Lander et al., 2001 (182), Venter et al., 2001 (346)). Although genes represent the major biological function of the genome, genes – or at least their coding regions – constitute only a tiny fraction of human DNA. Only 1.1% to 1.4% is sequence that actually encodes protein; that is just 5% of the 28% of the sequence that is transcribed into RNA (Baltimore, 2001 (16)).

The public International Human Genome Sequencing project estimated that there were 31,000 protein-encoding genes in the human genome, of which they could provide a list of

Introduction

22,000 (Lander et al., 2001 (182)). Celera Genomics found about 26,000 (Venter et al., 2001 (346)).

Apart from the protein-encoding genes, thousands of human genes produce noncoding RNAs (ncRNAs) as their ultimate product (Eddy, 1999 (76)). Major classes of ncRNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs) that play an important role precursor messenger RNA splicing.

Human genes express highly complex precursor messenger RNAs, containing an average of eight exons, with introns making up 90% of the transcription unit. For each pre-mRNA that is expressed, its exons – which are separated by introns of up to hundreds of thousands of nucleotides in length – must be precisely joined together to generate the reading frame for translation (Wang & Cooper, 2007(351)). In the human genome, the overall gene size and intron size varies considerably. Most internal exons fall within a common peak between 50 and 200 bp (Lander et al., 2001 (182)). Intron size is much more variable in humans, with a peak at 87 bp but a very long tail resulting in a mean of more than 3,300 bp. The variation in intron size results in great variation in gene size. The variation in gene size and intron size could partly be explained by the fact that GC-rich regions in the human genome tend to be gene-dense with many compact genes, whereas AT-rich regions tend to be gene-poor with many sprawling genes containing large introns (Lander et al., 2001 (182)).

The number of coding genes in the human sequence compares with 6,000 for a yeast cell, 13,000 for a fly, 18,000 for a worm and 26,000 for a plant (Lander et al., 2001 (182)). The proteomic complexity of humans is achieved among other things by alternative splicing events allowing the production of many protein isoforms from a single gene. Assessment of the prevalence of alternative splicing within the framework of the human genome project found two or more alternatively spliced transcripts in 59% of the genes of chromosome 22 (Lander et al., 2001 (182)). More recent data generated by exon junction microarrays from 10,000 human genes probed using RNA from 52 different human tissues demonstrated that at least 74% of human multi-exon genes are alternatively spliced (Johnson et al., 2003 (146)).

Furthermore, the human genome project identified 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the genome, 60,000 of which fall in within exons – in coding and in untranslated regions (Sachidanandam et al., 2001 (272)). The study estimated that individual humans differ from one another by about one base pair per thousand. Thus, it appeared that SNPs are the main source of genetic and phenotypic variation. Moreover, genome-scanning technologies uncovered an unexpectedly large extent

of structural variation in the human genome (for a review see Feuk et al., 2006 (83)). Beyond the SNPs, copy-number variants (CNVs) of larger contiguous blocks of DNA sequence usually exceeding 1000 bp contribute an additional 0.4% difference in DNA sequence between any two individuals (Sebat et al., 2004 (287)). Classes of CNVs include insertions, deletions and duplications that can encompass genes leading to dosage imbalances. Variations in the genome sequence make an important contribution to human disease susceptibility and protection. They may also provide information about our personal responses to medicines.

1.1.3 Genetic factors in human disease

The ability to clone and sequence DNA made it possible to localize genes underlying the phenotypes of human disease (for a list see <http://ncbi.nlm.nih.gov/OMIM> (Online Mendelian Inheritance in Man)). As soon as the responsible locus has been identified, sequencing of the region in cases and controls may define the causal mutation, resulting in the study of the molecular and cellular functions of these genes (Altshuler et al., 2008 (6)). Most of them were rare diseases in which a mutation of a single gene is necessary and sufficient to cause disease. Common forms of human diseases often show complex inheritance and result from the combined action of alleles in many genes with modest contribution of each locus to the disease (for a review see Badano & Katsanis 2002 (15)).

1.1.3.1 Cancer susceptibility genes

The study of potential associations between specific genetic loci and various cancers has a long history in cancer epidemiology. The discovery of oncogenes and proto-oncogenes provided a simple and powerful explanation of how the proliferation of cells is driven (Huebner & Todaro, 1969 (139)). Bishop and Varmus used an oncogenic retrovirus to identify the growth-controlling oncogenes in normal cells (Stehelin et al., 1976 (321)). They draw the remarkable conclusion that the oncogene in the virus did not represent a true viral gene but instead was a normal cellular gene which the virus had acquired during replication in the host cell. Therefore, normal cellular genes that can become an oncogene due to mutations or an increased expression were termed proto-oncogenes (Bishop, 1981 (28); Weinberg, 1983 (356)).

The proteins encoded by proto-oncogenes participate in various ways in receiving and processing growth-stimulatory signals that originate in the extracellular environment. When these genes suffer mutation, the flow of growth-promoting signals released by these proteins becomes deregulated. Instead of emitting them in carefully controlled bursts, the oncoproteins release a steady stream of growth-stimulating signals, resulting in the

Introduction

unrelenting proliferation association with cancer cells (Sporn & Roberts, 1985 (316); Weinberg, 2007 (357)).

The logic underlying well-designed control systems dictates, however, that the components promoting a process must be counterbalanced by others that oppose that process. Growth-promoting genes provide only part of the story of growth control. Experiments involving somatic cell fusion and chromosome segregation had pointed to the existence of genes that could suppress tumorigenicity (Stanbridge, 1976 (319)). The antigrowth genes came to be called tumor suppressor genes (Murphree & Benedict, 1984 (216)). Their involvement in tumor formation seemed to happen when these genes were inactivated or lost. The inactivation of tumor suppressor genes plays a role in cancer pathogenesis that is as important to cancer as the activation of oncogenes. If mutant, inactive alleles of the tumor suppressor really did play a role in enabling the growth of cancer cells. Mitotic recombination that leads to homozygosity at the tumor suppressor locus is termed loss of heterozygosity, or simply LOH (Dracopoli & Fogh, 1983 (74)). LOH is responsible for the elimination of most of the second, surviving wild-type copies of tumor suppressor genes (Koufos et al., 1985 (172); Ponder, 1988 (245); Weinberg, 2007 (357)).

There are two types of tumor suppressor gene: gatekeepers and caretakers. Tumor suppressor genes that function to directly regulate the growth of tumors by inhibiting growth or promoting cell death are called gatekeepers. Inactivation of these genes is rate-limiting for the initiation of a tumor, and both the maternal and the paternal copies must be altered for tumor development. Predisposed individuals inherit one mutant copy of the gatekeeper gene, so they need only one additional somatic mutation to initiate neoplasia. Sporadic tumors form in people who do not have germline mutations when both copies of the relevant gatekeeper gene become mutated somatically. Because the probability of acquiring a single somatic mutation is exponentially greater than the probability of acquiring two such mutations, people with a hereditary mutation of a gatekeeper gene are at a much greater risk of developing tumors than the general population (Knudson, 1996 (167); Kinzler & Vogelstein, 1997 (165)).

Unlike gatekeeper genes, caretakers do not regulate directly cell proliferation but act to prevent genetic instability (Levitt & Hickson, 2002 (188)). Rather, tumor initiation occurs indirectly – inactivation leads to genetic instabilities which result in increased mutation of all genes, including gatekeepers. Once such a tumor is initiated by inactivation of a gatekeeper gene, it may progress rapidly due to an accelerated rate of mutation in other genes that directly control cell birth or death. In dominantly inherited cancer-predisposition syndromes of the caretaker type, patients inherit a single mutant caretaker gene from an affected parent. Three subsequent somatic mutations are usually required to initiate cancer: mutation of the normal caretaker allele inherited from the unaffected parent, followed by mutation of both

alleles of a gatekeeper gene. Because three mutations are needed, the risk of cancer in affected families is generally only 5-50-fold greater than in the general population – much less than the risk in families with inherited defects in a gatekeeper gene (Kinzler & Vogelstein, 1997 (165)).

The human genome is typically so stable that the many genetic alterations required for cancer to develop cannot accumulate unless the rate of mutation is increased - that is, it becomes genetically unstable. The fact that human tumor formation is a complex, multi-step process (Fearon & Vogelstein, 1991 (82)) reflects the multiple lines of defense against cancer that have been established within our cells (for a review see Hoeijmakers, 2001 (132)).

1.1.4 Splicing in human disease

Clinically identified sequence variants in disease specific genes are classified as either known deleterious (often protein-truncating) mutations, recognized polymorphisms (assumed to be) neutral with respect to disease risk, or variants of unknown significance (VUS).

From a protein coding viewpoint, sequence variations in the coding region are classified as either frame-shift, nonsense, missense or synonymous. Frame-shift or nonsense mutations produce truncated protein isoforms, whereas missense mutations affect amino acids that may be important for structure and function of a protein. Translationally synonymous mutations – allelic polymorphisms or so-called nucleotide variations – are considered to be neutral. From a transcript viewpoint, translationally neutral DNA alterations might very well affect RNA processing by altering an RNA stability element, or by affecting the splice site consensus sequence at the exon-intron border or an auxiliary splicing regulatory sequence element (for a review see Hartmann et al., 2008 (123)). Pathogenic splicing alterations are increasingly recognized as a widespread mechanism through which gene mutations cause disease. In the Human Gene Mutation Database (www.hgmd.org, currently operated by BIOBASE (Stenson et al., 2009 (322)), single base-pair substitutions within exon/intron boundaries constitute ~10% of the total number of listed mutations causing human inherited diseases.

1.2. Splicing

1.2.1 Sequences of human splice sites

The high fidelity of splicing is critically dependent on the recognition of the signals that mark the exon–intron boundaries. In particular, Breathnach et al. (1978) (37) noticed that the dinucleotide at the 5' end of introns is defined by an invariant GT (position +1 and +2),

Introduction

whereas it is always an AG dinucleotide at their 3' end (position -1 and -2). As the number of available splice junction sequences grew it became clear the vast majority of pre-mRNA introns obeyed this GT-AG rule and a somewhat longer consensus sequence could be written.

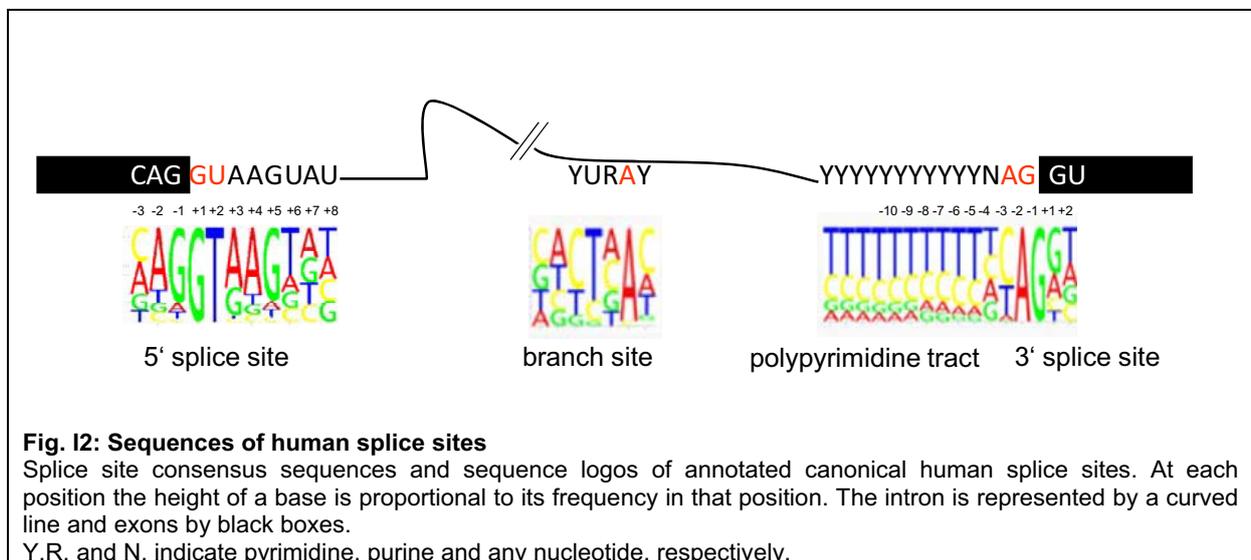
A statistical description of annotated human 5' splice sites or 3' splice sites can be obtained by aligning a large number of those, yielding a splice site motif specific for any given data set (for review see Hartmann et al., 2008 (123)). In such a motif, sequence conservation in fixed positions is indicated by one or two predominant nucleotide(s), while outside the conserved region the nucleotides are statistically distributed (with "background probability" of approx. 25% for G, T, A and C). Correspondingly, the splice site's *consensus sequence* is determined by picking the most frequent nucleotide in each conserved position.

For human 5' splice sites (5'ss) or splice donor sites, the consensus sequence MAG/GURAGU (where R = purine, M = C or A, and / denotes the exon-intron-border) includes positions -3 to +6 (i.e., the last 3 nucleotides (nt) of the upstream exon and the first 6 nt of the intron)). The sequence (Y)_nNYAG/G (Y = pyrimidine, N = any nucleotide, n ≥ 11) was found to be a consensus of human 3' splice sites (3'ss) or splice acceptor sites (Mount, 1981, see **Fig. I2**). In addition, the dinucleotide AG has never been seen in the -15 through -5 intronic region of a splice acceptor site (Seif and Dhar, 1979 (289)). In yeast, sequencing of introns revealed splice site sequences at their termini that conform to the well conserved consensus sequences derived from higher eukaryotes. However, the heptanucleotide TACTAAC occurred between 20 and 55 nucleotides upstream of the 3'ss in all known yeast intron sequences and was shown to be required for production of the spliced mRNA *in vivo* (Langford et al., 1983 (183)). It turned out that the TACTAAC sequence is the site of the intramolecular branch in intron lariat RNAs generated during splicing *in vivo* (Domdey et al., 1984 (73), Rodriguez et al. 1984 (266)). Within the yeast TACTAAC box the last adenosine serves as a branch point in splicing (Ruskin et al., 1985 (269)). In contrast to the strictly conserved branch point sequence (BPS) in yeast, the human BPSs are degenerative and have been recently described simply as yUnAy, where the underlined is the branch point at position zero and the lowercase pyrimidines ('y') are not as well conserved as the uppercase U and A (Gao et al., 2008 (96)). Thus, the mammalian 3' splice site consensus can be broken down into two parts: the highly conserved AG at position -1 and -2 relative to the intron-exon-border, and a stretch of pyrimidines (known as the polypyrimidine tract or PPT) extending 10 or more nucleotides back into the intron. The branch site is located upstream of the PPT, generally 11-40 nucleotides from the 3'ss AG (Senapathy et al., 1990 (292), Burge et al., 1999 (43), Moore 2000 (215)).

Introduction

Apart from the major GT-AG introns, a minor class of introns that possess AT and AC at their 5' and 3' ends has been identified in both vertebrate and invertebrate genomes (Jackson 1991 (144); Hall and Padgett, 1994 (115)). These minor class introns also exhibit longer, highly conserved but non-canonical sequences at their 5' and 3' splice sites. The 5' ss consensus for the minor class introns is /ATATCCTT and the 3'ss is CCAC/, while a third conserved intron element (TCCTTAAC) appeared upstream of the 3'ss.

Of 53,295 confirmed introns in the human genome project, 98.12% used the canonical dinucleotides GT at the 5'ss and AG at the 3'ss, another 0.76% used the related GC-AG and about 0.10% used the rare alternative AT-AC splice sites. The remaining 1% belonged to 177 types, some of which undoubtedly reflected sequencing or alignment errors (Lander et al., 2001 (182)).



1.2.2 The mechanism of splicing

In combination with *in vivo* studies, *in vitro* approaches have led to and refined a two-step model for the splicing reaction (Domdey et al. 1984 (73) ; Padgett et al. 1984 (234); Rodriguez et al. 1984 (266); Ruskin et al. 1984 (269); Zeitlin and Efstratiadis, 1984 (379)) (**Fig. I3**). In the first trans-esterification reaction, the 2' hydroxyl group of the conserved adenosine within the branch point sequence attacks the 5' phosphate of the conserved guanine at position +1 of the 5'ss at the 5' exon-intron junction. The reaction results in cleavage at the 5'ss producing a metastable free 5' exon intermediate and a second RNA with the 5' end of the intervening sequence joined through a 2'-5' phosphodiester bond producing a trinucleotide, which constitutes the branch point of a lariat structure. In the

second step, the 3' hydroxyl group from the free 5' exon attacks the phosphate group of the conserved guanine at position -1 of the 3' splice site at the 3' intron-exon-border in a transesterification reaction to produce the spliced exons and the excised intervening sequence.

Several lines of evidence suggested that the 3'ss is positioned for 3' cleavage and exon ligation, at least in part, through a non-Watson-Crick interaction between the guanosines at the 5'ss and 3'ss (Parker & Siliciano, 1993 (236); Chanfreau et al., 1994 (54); Deirdre et al., 1995 (70)). A possible non-Watson-Crick interaction between the 5'- and the 3' terminal nucleotide of the intron has also been suggested for the minor class introns (Dietrich et al., 1997 (72)).

1.2.3 The spliceosome

In vitro systems using a precursor RNA derived from the major late transcription unit of adenovirus 2 as substrate and a whole cell extract of HeLa cells showed that splicing requires Mg^{2+} and ATP (adenosine-triphosphat) (Hardy et al., 1984 (120)) and that the reaction is inhibited by antisera that recognize small nuclear ribonucleoprotein particles containing U1 snRNA (small nuclear RNA) (Padgett et al., 1983 (235)).

In analysis of *in vitro* splicing reactions of pre-messenger RNA (pre-mRNA) in yeast extract by glycerol gradient centrifugation labeled pre-mRNA appeared in a 40S peak only if the pre-mRNA was subjected to the first splicing reaction. Lariat form intermediates were found almost exclusively in this 40S complex and the cut 5' exon RNA was concentrated in this complex. This complex termed "spliceosome" was thought to contain components necessary for splicing (Brody and Abelson, 1985 (39)). In mammalian cells, a similar, but larger complex, sedimenting at 60S, was identified. The 60S RNA-protein complex formed only under conditions that permitted splicing: both ATP (adenosine-triphosphate) and a precursor RNA were required for its formation, while antiserum specific for U1 snRNP (U1 small nuclear ribonucleoprotein particle) inhibited its formation (Grabowski et al. 1985 (107)).

Since the 5' terminal region of U1 snRNA is highly complementary to the consensus 5'ss sequence it has been suggested that the U1 snRNP may be responsible for the recognition of the 5'ss sequence by intermolecular base-pairing between these regions. Indeed, it has been shown that the 5' terminus of the U1 snRNP particle which is complementary to the 5'ss is single stranded in the intact particle and is not protected by snRNP proteins (Rinke et al., 1984 (260)).

Introduction

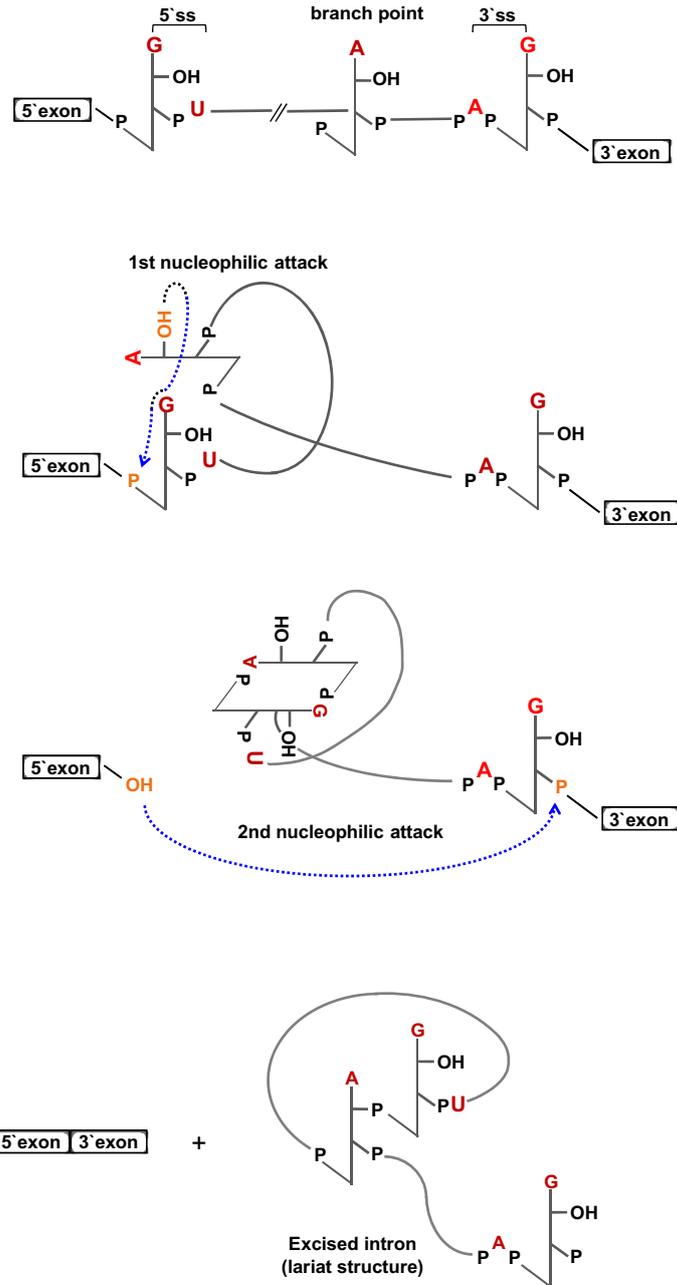


Fig. 13: The splicing reaction itself ensues by two consecutive trans-esterification reactions

The branch point A residue plays a critical role in the enzymatic reaction. The first step is a hydrophilic attack. The 2' hydroxyl group of the conserved adenosine within the branch point sequence attacks the 5' phosphate of the conserved guanine within the 5' splice site at the 5' exon-intron junction. An unusual 2'-5' phosphodiester bond is formed between both residues and the 5' exon-intron junction is cleaved. The products are a 2'-5' phosphodiester RNA lariat structure and a free 3'-OH (leaving group) that arises from the upstream exon. A rearrangement of spliceosomal components must follow to permit the second transesterification reactions. The second step is an additional hydrophilic attack. The 3'-OH end of the released 5' exon then attacks the scissile phosphodiester bond of the conserved guanine of the 3' splice site. This reaction liberates the 3'-OH of the intron resulting in a free lariat and spliced exons. The two exon sequences are joined together, while the intron sequence is released as a lariat structure.

Introduction

Selective degradation of U2 snRNP in a nuclear extract using ribonuclease H revealed that U2 as well as U1 snRNA are involved in pre-messenger splicing. Immunoprecipitated fragments protected from T1 RNase digestion included the branch point sequence, suggesting that the U2 snRNP is responsible for branch-point recognition during pre-mRNA splicing (Black et al., 1985 (29)). Biochemical complementation experiments identified a micrococcal nuclease-resistant factor, U2AF (U2 snRNP auxiliary factor), that is necessary for the U2 snRNP/branch point interaction and splicing complex assembly, promoting sequence-specific RNA-binding activity of the U2 snRNP despite the variability of mammalian branch-point sequences (Ruskin et al., 1988 (270), Nelson & Green, 1989 (221)). Binding of U2AF required an RNA substrate in which the polypyrimidine tract and the AG dinucleotide of the 3' splice site consensus sequence were present (Ruskin et al., 1988 (270)).

Apart from the U1 and U2 snRNA, the set of metabolically stable small RNAs in the 4-10S range within the cell nucleus includes U4, U5 and U6 snRNAs (Deimel et al., 1977 (69); Guimont-Ducamp et al., 1977 (111); Northemann et al., 1977 (227); Gallinaro and Jacob, 1979 (95)) present in greater than 10^5 copies per mammalian cell.

The U5 snRNP was suggested to be involved in recognizing the 3' ends of introns and to participate in pre-mRNA splicing in addition to the U1 and U2 snRNPs (Chabot et al., 1985 (51)). Selective cleavage of U4 and U6 RNA in HeLa cell nuclear extract showed that splicing *in vitro* required intact U4/U6 small nuclear ribonucleoproteins (Black et al., 1986 (30); Berget et al., 1986 (20)). Upon degradation of U4/U6 the block in the splicing pathway seemed to occur before the first cleavage and ligation step, just as in extracts where U1 or U2 snRNAs were specifically degraded (Black et al., 1985 (29); Krainer & Maniatis, 1985 (173); Black et al., 1986 (30)). Mutational analysis in yeast demonstrated a conserved base-pairing interaction between the U6 and U2 snRNAs that is mutually exclusive with the U4-U6 interaction (Madhani & Guthrie, 1992 (196)). Formation of U2-U6 snRNA intermolecular helices has been shown to be necessary for catalytic activation of the spliceosome (Sun and Manley, 1995 (328)). Biochemical and structural studies in a conserved stem-loop in U6 have shown specific metal ion binding (Yean et al., 2000 (372); Sigel et al., 2000; (302)). The ability of the protein-free RNA stem-loop domain of U6 to bind a divalent cation in the internal loop provided additional support for the competence of the spliceosomal snRNAs to form the active site of the spliceosome (Huppler et al., 2002 (140)).

1.2.4 Spliceosome assembly and catalysis

Biochemical data based on *in vitro* studies using native gel electrophoresis, affinity selection and glycerol gradient centrifugation indicated that the spliceosome assembles stepwise allowing the isolation of landmark assembly intermediates defined by sequential association and release of the spliceosomal snRNPs. Assembly intermediates of the human spliceosome that have been observed include the E, A, B, B*, and C complexes (**Fig. I4**) (for review, see Wahl et al., 2009 (349)).

Assembly of the spliceosome is initiated by recognition of the 5'ss by the U1 snRNP through base-pairing interactions of the free 5' end of the U1 snRNA and the 5'ss (Zhuang & Weiner, 1986 (384)). In addition to the U1-5'ss interaction, the earliest assembly phase of the spliceosome - although not in all cases - involves the cooperative binding of the splicing factor SF1/mammalian branch point binding protein (mBBP) to the branch point sequence (BPS) and of the 65 kDa subunit of the U2 auxiliary factor (U2AF) to the polypyrimidintract (PPT) (Guth & Valcarcel, 2000 (114)). In addition, the 35kDa subunit of U2AF, which is tightly bound to the U2AF65 in the U2AF heterodimer, binds the AG dinucleotide of the 3'ss (Zamore & Green, 1989 (376), Wu et al., 1999 (368)). Together, these molecular interactions yield the spliceosomal E complex and play crucial roles in initial recognition of the 5'ss and 3'ss of an intron.

Studies using a directed hydroxyl radical probe tethered to pre-mRNA substrates to map the structure of the pre-mRNA substrate during the spliceosome assembly process suggested that pre-mRNAs are organized at an early stage of spliceosome assembly such that the 5'ss and the branch region are directly proximal to one another (Kent & MacMillan, 2002 (163)). Binding of SF1/mBBP and U2AF is required for recruitment for U2 snRNP to the branch point sequence. U2AF65 recruits the U2 snRNP via binding to the U2 snRNA associated protein SF3b155 and promotes the base-pairing interaction by its arginine-serine-rich domain (Valcarcel et al., 1996 (342)). Moreover, the 56 kDa U2AF65 associated protein UAP56, which is a DEAD (Asp-Glu-Ala-Asp) box protein with ATP-dependent RNA helicase activity, is recruited to the pre-mRNA dependent on U2AF65 and necessary for U2 snRNP binding at the branchpoint region (Fleckner et al., 1997 (86)).

In the A complex, the U2 snRNA engages in an ATP-dependent manner in a base-pairing interaction with the branch point sequence. The RNA-dependent ATPase Prp5 (Pre-mRNA processing) of the DExD/H family (where x can be any amino acid) is suggested to hydrolyze ATP to promote stable association of U2 in the pre-spliceosome (Kosowski et al., 2009 (171)). Prp5 can physically associate with the U2 snRNP (Will et al., 2002 (364)) and seems to bridge U1 and U2 snRNPs at the time of pre-spliceosome formation (Xu et al., 2004 (371)).

Introduction

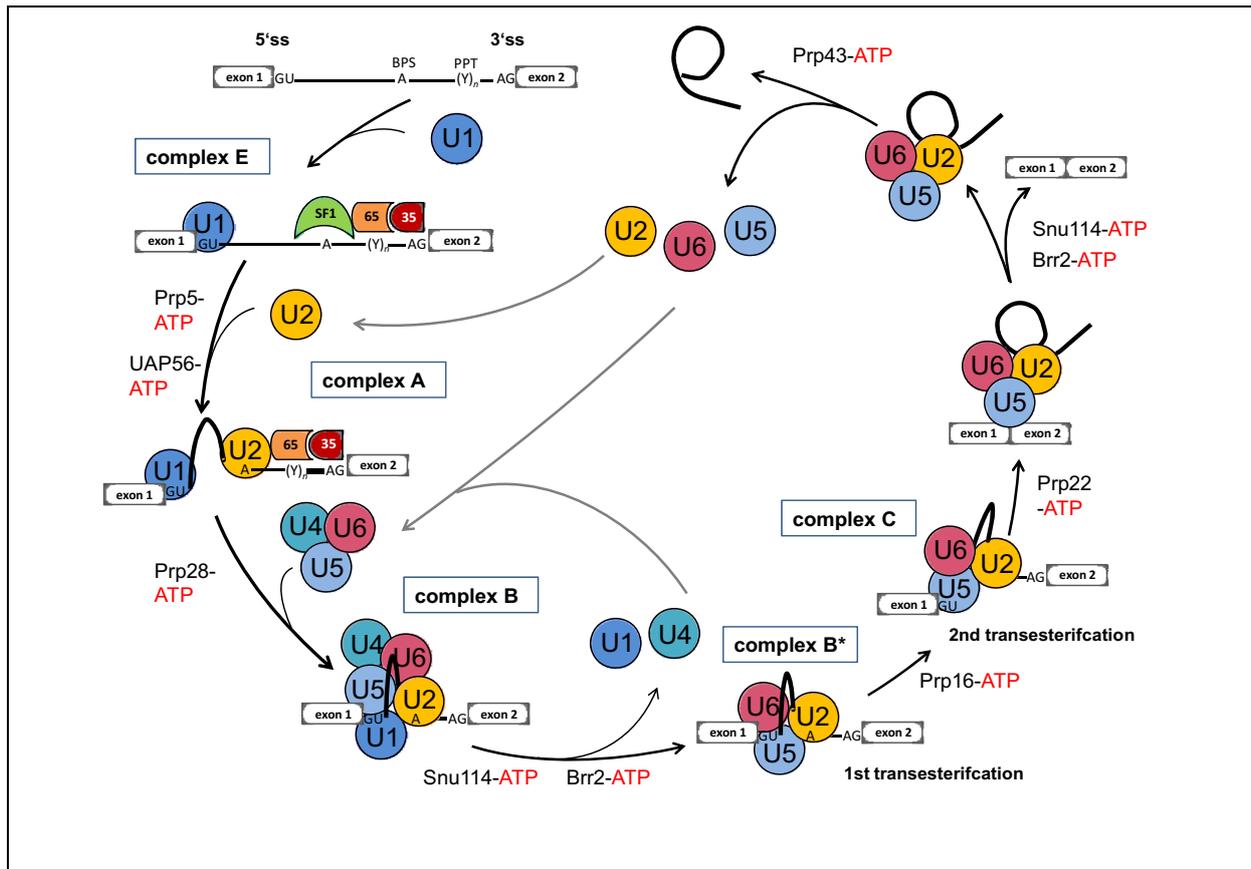


Fig. 14: Assembly cycle of the human major spliceosome

Assembly intermediates of the human spliceosome that can be resolved in mammalian splicing extracts by biochemical methods include the E, A, B, B*, and C complexes. The stepwise interaction of spliceosomal snRNPs (colored circles) in the removal of an intron from a pre-mRNA containing two exons is depicted.

Assembly of the spliceosome is initiated by recognition of the 5'ss by the U1 snRNP. The earliest assembly phase of the spliceosome (E-complex) involves the cooperative binding of the splicing factor SF1 to the branch point sequence (BPS) and of the 65 kDa subunit of the U2 auxiliary factor (U2AF) to the polypyrimidintract (PPT), whereas the 35 kDa subunit of U2AF binds the AG dinucleotide of the 3'ss. The 56 kDa U2AF65 associated protein UAP56, which is DEAD box protein with ATP-dependent RNA helicase activity, is recruited to the pre-mRNA dependent on U2AF65 and necessary for U2 snRNP binding at the branchpoint region.

In the A complex, the U2 snRNA engages in an ATP-dependent manner in a base-pairing interaction with the branch point sequence replacing SF1. The RNA-dependent ATPase Prp5 of the DExD/H family is suggested to hydrolyze ATP to promote stable association of U2 in the pre-spliceosome. As soon as both splice sites are recognized, the U4/U6.U5 tri-snRNP joins the spliceosome upon phosphorylation of the U5 snRNP associated RNA helicase hPrp28 of the DExD/H family - generating the B complex. The B-complex has no catalytic centre and must be activated to a catalytically competent state.

During this activation, the interaction of the U1 snRNP with the 5'ss is disrupted upon unwinding of the U1 RNA/5'ss duplex through the 100-kDa U5 snRNP associated DExD/H ATPase Prp28 which closely cooperates with the 220-kDa U5 snRNP associated protein Prp8. The U1 snRNP at the 5'ss is replaced by both the U5 snRNP and the U6 snRNP. Recognition of the 5'ss by U6 snRNP is a prerequisite for unwinding of the U6/U4 snRNPs by the U5 snRNP associated ATP dependent DExD/H box RNA helicase Brr2. The U5 snRNP associated GTPase Snu114 regulates the activity of Brr2, Snu114 is coordinated by the U5 snRNP Prp8 protein.

Release of the U1 snRNP and the U4 snRNP gives rise to the catalytically activated B* complex in which only U2, U5 and U6 snRNP are present and in which the first transesterification step of splicing takes place. In the C complex, the first of the two catalytic steps of splicing has occurred. Prior to the second transesterification step, the U2/U6 complex appears to be reformed by the DEAH-box ATPase Prp16. After the second transesterification step, the DEAH-box RNA helicase Prp22 catalyzes the release of the mRNA product from the spliceosome and thereby initiates disassembly of the spliceosome. Disassembly of the postsplicing U2/U6/U5 intron complex requires again the activity of the GTPase Snu114p and the ATPase Brr2p which are resident subunits of spliceosome. After taking part in splicing, the U5 and U4/U6 snRNPs reassemble. The RNA-dependent ATPase Prp43 is required for release of the lariat-intron from the spliceosome and promotes spliceosome disassembly after exon ligation (modified from Wahl et al., 2009 (349)).

Introduction

The U2 snRNP base pairs with the branch point region while the nucleophilic branch site adenosine does not base pair with the U2 snRNA, but rather bulges out of the recognition helix (Zhang & Weiner, 1989 (385); Query et al., 1994 (248); Berglund et al., 2001 (24)). Binding of mBBP/SF1 is mutually exclusive with the U2 snRNP, thus the U2 snRNP replaces SF1/mBBP (Berglund et al., 1998 (23); Rutz & Seraphin, 1999 (271)). U2-specific proteins, including the multimeric splicing factors SF3a and SF3b, anchor the U2 snRNP to the branch point sequence by binding to flanking sequences primarily upstream of the branch point (Gozani et al., 1996 and 1998 (105,106)). Upon stable integration of the U2 snRNP into the spliceosome, one SF3b subunit, p14, interacts directly with the branch adenosine (Will et al., 2001 (362)).

As soon as both splice sites are recognized, the U4/U6.U5 tri-snRNP joins the spliceosome upon phosphorylation of the U5 snRNP associated RNA helicase hPrp28 of the DExD/H family - generating the B complex (Mathew et al., 2008 (204)). Within the tri-snRNP the U4 and U6 snRNAs are base-paired with one another and the U4/U6-specific hPrp31 protein binds specifically to the U5-specific protein hPrp6, connecting U5 to U4/U6 (Makarov et al., 2002 (199); Schaffert et al., 2004 (279)). The B-complex has no catalytic centre and must be activated to a catalytically competent state and is the substrate for the major RNA remodelling events that lead to catalytic activation of the splicing machinery (Wolf et al., 2009 (366)).

During this activation, the interaction of the U1 snRNA with the 5'ss is disrupted by unwinding of the U1 RNA/5'ss duplex through the 100-kDa U5 snRNP associated DExD/H ATPase Prp28 which counteracts the stabilizing effect of the U1-C protein (Staley & Guthrie, 1999 (318); Ismaili et al., 2001 (142); Chen et al., 2001 (56)). Prp28 closely cooperates with the 220-kDa U5 snRNP associated protein Prp8 (Strauss & Guthrie, 1991 (325); Pena et al., 2008 (241)) which crosslinks to the highly conserved GU dinucleotide of the 5'ss with its C-terminal RNase H domain (Reyes et al., 1996 and 1999 (258,259)). The U1 snRNP at the 5'ss is replaced by both the U5 snRNP and the U6 snRNP: The ACA sequence within the phylogenetically conserved ACAGAG sequence of U6 snRNA can form complementary Watson-Crick base-pairs with the intron positions +4 to +6 of the 5'ss whereas the intron positions +1 to +3 of the 5'ss seem to interact through non-Watson-Crick interactions with the

Introduction

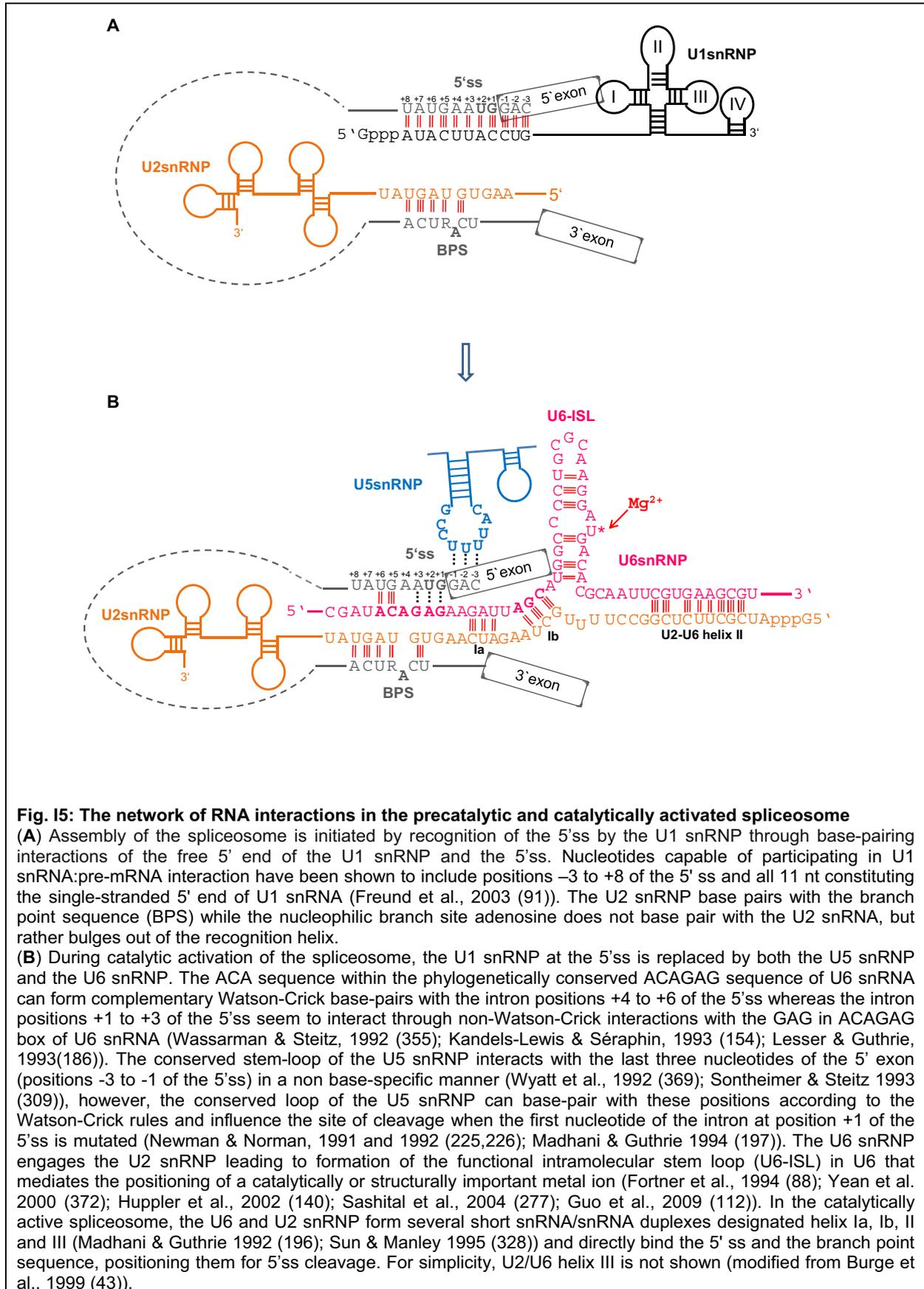


Fig. 15: The network of RNA interactions in the pre-catalytic and catalytically activated spliceosome

(A) Assembly of the spliceosome is initiated by recognition of the 5'ss by the U1 snRNP through base-pairing interactions of the free 5' end of the U1 snRNP and the 5'ss. Nucleotides capable of participating in U1 snRNA:pre-mRNA interaction have been shown to include positions -3 to +8 of the 5' ss and all 11 nt constituting the single-stranded 5' end of U1 snRNA (Freund et al., 2003 (91)). The U2 snRNP base pairs with the branch point sequence (BPS) while the nucleophilic branch site adenosine does not base pair with the U2 snRNA, but rather bulges out of the recognition helix.

(B) During catalytic activation of the spliceosome, the U1 snRNP at the 5'ss is replaced by both the U5 snRNP and the U6 snRNP. The ACA sequence within the phylogenetically conserved ACAGAG sequence of U6 snRNA can form complementary Watson-Crick base-pairs with the intron positions +4 to +6 of the 5'ss whereas the intron positions +1 to +3 of the 5'ss seem to interact through non-Watson-Crick interactions with the GAG in ACAGAG box of U6 snRNA (Wassarman & Steitz, 1992 (355); Kandels-Lewis & Séraphin, 1993 (154); Lesser & Guthrie, 1993(186)). The conserved stem-loop of the U5 snRNP interacts with the last three nucleotides of the 5' exon (positions -3 to -1 of the 5'ss) in a non base-specific manner (Wyatt et al., 1992 (369); Sontheimer & Steitz 1993 (309)), however, the conserved loop of the U5 snRNP can base-pair with these positions according to the Watson-Crick rules and influence the site of cleavage when the first nucleotide of the intron at position +1 of the 5'ss is mutated (Newman & Norman, 1991 and 1992 (225,226); Madhani & Guthrie 1994 (197)). The U6 snRNP engages the U2 snRNP leading to formation of the functional intramolecular stem loop (U6-ISL) in U6 that mediates the positioning of a catalytically or structurally important metal ion (Fortner et al., 1994 (88); Yean et al. 2000 (372); Huppler et al., 2002 (140); Sashital et al., 2004 (277); Guo et al., 2009 (112)). In the catalytically active spliceosome, the U6 and U2 snRNP form several short snRNA/snRNA duplexes designated helix Ia, Ib, II and III (Madhani & Guthrie 1992 (196); Sun & Manley 1995 (328)) and directly bind the 5' ss and the branch point sequence, positioning them for 5'ss cleavage. For simplicity, U2/U6 helix III is not shown (modified from Burge et al., 1999 (43)).

Introduction

GAG in ACAGAG box of U6 snRNA (Wassarman & Steitz, 1992 (355); Kandels-Lewis & Séraphin, 1993 (154); Lesser & Guthrie, 1993 (186)). The conserved stem-loop of the U5 snRNP interacts with the last three nucleotides of the 5' exon (positions -3 to -1 of the 5'ss) in a non base-specific manner (**Fig. I5**) (Wyatt et al., 1992 (369); Sontheimer & Steitz, 1993 (309)). However, the conserved loop of the U5 snRNP can base-pair with these positions according to the Watson-Crick rules and influence the site of cleavage when the first nucleotide of the intron at position +1 of the 5'ss is mutated (Newman & Norman, 1991 and 1992 (225,226); Madhani & Guthrie 1994 (197)). Recognition of the 5'ss by the ACAGAG box of the U6 snRNA is a prerequisite for unwinding of the U6/U4 snRNA helix by the U5 snRNP associated ATP dependent DExD/H box RNA helicase Brr2 (Laggerbauer et al., 1998 (181); Raghunathan & Guthrie, 1998 (249,250)).

The C-terminus of the U5 snRNP Prp8 protein stimulates the helicase activity of Brr2 and coordinates the U5 snRNP associated GTPase Snu114 that regulates the activity of Brr2 (Small et al., 2006 (306); Maeder et al., 2009 (198)). The U5-associated proteins Prp8, the ATPase dependent helicase Brr2, and the GTPase Snu114 together form a highly stable trimer with a putative 'molecular motor' function in U4/U6 unwinding (Achsel et al. 1998 (2)). The U4 snRNP is released from the spliceosome probably fulfilling a chaperone function protecting the U6 snRNA prior to activation of the spliceosome (Staley & Guthrie 1998 (317)). Disruption of the U4/U6 base-pairing interaction frees the U6 snRNA to engage the U2 snRNA leading to formation of functional intramolecular stem loop (U6-ISL) in U6 that mediates the positioning of a catalytically or structurally important metal ion and to formation of several U2/U6 helices (Fortner et al., 1994 (88); Yean et al. 2000 (372); Huppler et al., 2002 (140); Sashital et al., 2004 (277); Guo et al., 2009 (112)).

Release of the U1 snRNP and the U4 snRNP gives rise to the catalytically activated B* complex in which only U2, U5 and U6 snRNP are present and in which the first transesterification step of splicing takes place. In the catalytically active spliceosome, the U6 and U2 snRNP form several short snRNA/snRNA duplexes designated helix Ia, Ib, II and III (Madhani & Guthrie 1992 (196); Sun & Manley 1995 (328)) and directly bind the 5' ss and the branch point sequence, positioning them for 5'ss cleavage. This function is served in part by U2/U6 helix I, which is composed of two short helices, Ia and Ib.

Within the 3-helix structure, helix Ia is located proximal to both the U6/5'ss and U2/branch site interactions, helix Ib contains the highly conserved AGC triad of U6 that forms three base pairs with U2. The U6-ISL intramolecular stem-loop structure within the central region of the U6 snRNA plays a crucial role in the catalysis of the splicing reaction. In an alternative conformation of U2/U6 helix I the U6 strand of helix Ib extends the U6-ISL, while the U2

Introduction

strand of helix Ib is involved in formation of U2 snRNP stem I, forming a four-helix junction together with U2/U6 helix II and U2/U6 helix Ia (Sashital et al. 2004 (277)). It seems that the 4-helix junction in which the catalytically important AGC triad base-pairs only within U6 and not with U2 is relevant to the first, but not to the second step of splicing; helping to position the U6 ISL adjacent to the 5'ss.

Although mutational analysis of the invariant U5 snRNA loop sequence implied a potential role for U5 in influencing 5'ss cleavage site specificity (Newman & Norman, 1991 and 1992 (225,226)), activation of aberrant cleavage sites by U5 snRNA loop mutations proved to be strictly dependent on the presence of a mutation of the 5' terminal G residue of the intron. 5'ss cleavage can be uncoupled from the invariant U5 snRNA loop sequence without affecting accuracy or specificity, suggesting that U5 snRNA does not make an essential contribution to 5'ss definition (O'Keefe et al., 1996 (230); Segault et al., 1999 (288)). Instead, the U5 snRNP protein Prp 8 (p220) seems to target the highly GU nucleotide at the 5' exon-intron-border for attack by the the 2'-OH of the branchpoint adenosine. Prp8 is the largest and most highly conserved spliceosomal protein and is considered to be a master regulator of the spliceosome (Collins and Guthrie, 2000 (62); Grainger and Beggs, 2005 (108)). The RNase H domain of Prp8 seems to be the ideal device that needs to control coordinated structural rearrangements involving multiple RNA duplexes (Pena et al., 2008 (241)). The RNase H domain of Prp8 is placed in close proximity of the U6 ISL that is expected to constitute an active site component of the catalytically active spliceosome suggesting that Prp8 functions as a cofactor to an RNA enzyme and that the spliceosome may function as a ribonucleoprotein enzyme ('RNPzyme') (Collins & Guthrie 2000 (62); Abelson, 2008 (1); Pena et al., 2008 (241)).

In the C complex, the first of the two catalytic steps of splicing has occurred. Because the leaving group of the first reaction becomes the attacking group of the second reaction, the spliceosome must rearrange the substrate after 5'ss cleavage and undergoes substantial remodeling to generate the active sites responsible for exon ligation and intron excision. Prior to the second step of splicing, the U2/U6 complex appears to rearrange (Sashital et al., 2004 (277); Guo et al., 2009 (112)). The U2/U6 helix I which has configured the substrate during 5'ss cleavage seems to be destabilized and unwound by the DEAH-box ATPase Prp16 (Schwer & Guthrie, 1991 (286); Mefford & Staley 2009 (210)). Unwinding and reformation of the U2/U6 helix I promote repositioning of the substrate for exon ligation. Prp16 has also been implicated in remodeling of the U2 snRNP stem IIa and IIc driving the transition of the spliceosome from one step to another (Perriman et al., 2007 (242)). After U2 snRNP stem IIc has contributed to the first catalytic step, U2 snRNP stem IIa forms again between the first and second step transition and switches back to stem IIc before exon ligation (Hilliker et al.,

Introduction

2007 (130)). The structural changes enable recognition of the 3'ss by U5 snRNA and Prp8 after the first step of splicing. The conserved loop of the U5 snRNP also contacts exon nucleotides just downstream from the 3'ss and is thought not only to tether the 5' exon to the spliceosome after the first step, but also to align both exons for the second catalytic step (O'Keefe and Newman, 1998 (229)). It is thought that Prp8 recognizes an RNA tertiary structure element comprising nucleotides near the 5' and 3' splice sites of the pre-mRNA and U6 snRNA (Siatecka et al., 1999 (300); Collins & Guthrie, 2000 (62)). The RNA tertiary structure includes an interaction between the first and last Gs of the intron. Crosslinks of the human Prp8 protein span a region from the branch point to the 3'ss plus 13 bases into the 3'exon (MacMillan et al. 1994 (194); Teigelkamp et al. 1995 (335,336); McPheeters and Muhlenkamp 2003 (209); Grainger & Beggs 2005 (108)). Catalysis of the second trans-esterification reaction requires the ordered recruitment of Slu7, Prp18, and Prp22 to the spliceosome (Aronova et al., 2007 (8)). The Prp18 protein plays a role in fortifying U5/exon contacts prior to exon joining (Bacikova and Horowitz 2005 (14)). The human Slu7 protein associates with the spliceosome at a stage prior to recognition of the 3'ss for the second trans-esterification reaction and is required for structural rearrangement of the spliceosome prior to the establishment of the catalytically active spliceosome for the second step and for selection of the correct AG dinucleotide as 3'ss (Chua & Reed, 1999 (59,60)).

The DEAH-box RNA helicase Prp22 catalyzes the release of the mRNA product from the spliceosome by remodeling contacts within the spliceosome that involve the U5 snRNP (Schwer 2008 (285)). It has been proposed that the interactions between the U5 snRNP and exon bases persist after exon-joining and that Prp22 helicase breaks these contacts to release mRNA from the spliceosome and thereby initiates disassembly of the spliceosome (Aronova et al., 2007 (8)). The RNA-dependent ATPase Prp43 is required for release of the lariat-intron from the spliceosome and promotes spliceosome disassembly after exon ligation (Arenas & Abelson, 1997 (7)). Disassembly of the postsplicing U2/U6/U5 intron complex requires again the activity of the GTPase Snu114p and the ATPase Brr2p which are resident subunits of spliceosome (Small et al., 2006 (306)). After taking part in splicing, the U5 and U4/U6 snRNPs reassemble in Cajal bodies, away from the sites where splicing takes place (Stanek et al., 2008 (320)).

Recent observations suggest, in contrast to this stepwise snRNP recruitment model, the existence of a preassembled "holospliceosome" complex (Stevens et al., 2002 (324)). As the assembly intermediates were detected under stringent conditions, they most likely reflect different stabilization/destabilization states suggesting that the existence of both distinct assembly intermediates and a preassembled spliceosome is not necessarily contradictory (Will & Lührmann, 1999 (363)).

In contrast to other complex RNP enzymes like the ribosome, protein constitutes the majority of the spliceosome's mass. Improved methods for the isolation of spliceosomal complexes coupled with highly sensitive mass spectrometry (MS) techniques indicated that the spliceosome's proteome is much more complex than previously thought. Human spliceosomes contain ≈ 45 distinct snRNP associated proteins, which contributes ≈ 2.7 MDa of molecular mass (Wahl et al., 2009 (349)). Mass spectrometric analysis of affinity-purified spliceosomal complexes identified between 145 (Zhou et al., 2002 (382,383)) and 311 (Rappsilber et al., 2002 (251)) distinct spliceosomal proteins that copurify with splicing complexes.

The major U2 dependent spliceosome catalyzes the removal of the GT-AG and GC-AG introns. The less abundant U12-dependent minor spliceosome composed of the snRNPs U11, U12, U4atac/U6atac and U5, on the other hand, splices the rare alternative AT-AC splice sites (Hall & Padgett, 1996 (116); Tarn & Steitz 1996 (333,334); Patel & Steitz, 2003 (237); Will & Lührmann, 2005 (360)).

1.2.5 Splice site and exon recognition

1.2.5.1 5'ss recognition

Recognition of the 5'splice site (5' ss) in mRNA precursors by RNA duplex formation between U1 snRNA and a 5'ss initiates assembly of the spliceosome that catalyzes splicing. This RNA duplex formation is necessary for splicing and binding of U1 snRNP, and at least in some instances, also protects pre-mRNA against nuclear degradation (Kammler et al., 2001 (152)), as evident from human 5'ss mutations leading to RNA degradation rather than to aberrant splicing (Kirschner et al., 2000 (166); Wijk et al., 2004 (359)).

Sequence compilation of thousands of human 5' splice sites reveal a so-called consensus sequence, i.e., AG/GURAGU (where R=purine, and / indicates the exon-intron border) that reflects the frequency of a nucleotide in a given position of such a compilation. Although the human 5' ss consensus sequence includes positions -3 to $+6$ (i.e., the last 3 nucleotides [nt] of the upstream exon and the first 6 nt of the intron), nucleotides capable of participating in U1 snRNA:pre-mRNA interaction have been shown to include positions -3 to $+8$ of the 5' ss and all 11 nt constituting the single-stranded 5' end of U1 snRNA (Kammler et al., 2001 (152); Freund et al., 2003 (91)). Indeed, an alignment of 46,308 annotated canonical human 5'ss does not display a significant bias towards position $+7$ and $+8$. However, further classification of the annotated human canonical 5' ss into subsets showing either exonic, centred or intronic complementarities demonstrated that the subset of 3,830 5'ss sequences with *no*

Introduction

complementarity in the *exonic* positions of the 5'ss motif clearly displays a bias towards complementary bases even in position +7 and +8 (Hartmann et al., 2008 (123)).

In an experimentally approach to determine the intrinsic 5' ss strength, a hydrogen bond model for the complementarity between the free 5' end of the U1 snRNA and the 5' ss has been established (Kammler et al., 2001 (152); Freund et al., 2003 (91)). The hydrogen bond weight model translates the hydrogen bond pattern between the 5'ss and all 11 nt of the free 5' end of the U1 snRNA into a numerical HBond score (available at the web-interface <http://www.uni-duesseldorf.de/rna>).

Furthermore, the stability of the RNA duplex is not exclusively determined by its complementarity to U1 snRNA, but also by additional interactions of protein components with the pre-mRNA in the vicinity of the 5'ss, including the U1-specific proteins U1-A, U1-C and U1 70K (Surowy et al., 1989 (329); Nagai et al., 1990 (218); Heinrichs et al., 1990 (125)). The mammalian U1 snRNP consists of the 165-nt U1 snRNA, which forms four stem-loops (Sturchler et al., 1992 (326)), the three U1-specific proteins U1-A, U1-C, U1 70K and seven Sm proteins (B/B', D1, D2, D3, E, F and G). The U1-70K protein binds stem-loop I of the U1 snRNA whereas the U1-A protein interacts with stem-loop II (Hamm & Mattaj, 1987 (118); Patton & Pederson, 1988 (238); Scherly et al., 1989 (282); Will et al., 1996 (361)). The seven Sm proteins assemble around the Sm site nucleotides, located between stem loop III and stem loop IV (Bringmann et al., 1986 (38)). The Sm proteins are arranged in the order E-G-D3-B-D1-D2-F, in agreement with a heptameric ring model (Kambach et al., 1999 (149,150)). Moreover, the group of metabolically stable RNAs known as U snRNAs to which the U1 snRNA belongs is marked by a 5' terminal cap which contains the unusual nucleoside 2,2,7-trimethylguanosine (m_3G) at its 5' end. The association of the snRNP Sm proteins results in the hypermethylation of the snRNA's monomethylguanosine (m^7G) structure 2,2,7-trimethylguanosine (m_3G) form (Mattaj, 1986 (207)). The m_3G cap, together with the snRNP Sm proteins, forms a karyophilic signal required for the nuclear import of the spliceosomal snRNPs (Fischer & Lührmann, 1990 (84); Hamm & Mattaj, 1990 (118)).

The U1-70K protein binds directly to stem-loop I of U1 snRNA, the U1-A protein directly interacts with stem loop II, whereas the U1C protein interacts via protein-protein interactions (Hamm & Mattaj, 1987 (118); Patton & Pederson, 1988 (238), Scherly et al., 1989 (282); Will et al., 1996 (282); Varani & Nagai, 1998 (345)). Protein-protein contacts also appear to contribute to the association of U1-70K and with the U1 particle. The U1-70K protein contains a central RRM (RNA recognition motif) which binds the end of stem-loop I U1 snRNP (Ritchie et al., 2009 (261)). The first RRM of U1A is bound to the U1 snRNA stem-loop II (Oubridge et al., 1994 (233)). The human U1C protein is known to contain a zinc finger structure (Muto et al., 2004 (217)), and yeast U1C has been proposed to directly interact with

Introduction

the 5'ss (Du et al., 2002 (75)). Integration of U1C into the U1 snRNP particle is known on the N-terminal region of U1-70K and the Sm core domain (Nelissen, 1994 (220)). The zinc finger of U1C interacts the U1 snRNA-5'ss duplex. The U1C protein is positioned along the minor groove of the RNA duplex, including the location corresponding to the base pairs with the invariant GU dinucleotide which defines the 5'ss (Ritchie et al., 2009 (261)). It is needed for efficient complex formation of U1 snRNP with the 5'ss (Heinrichs et al., 1990 (125)). Complementation studies with U1 snRNPs lacking subsets of U1-specific proteins demonstrated a role for the U1C, but not U1A, in the formation of early splicing complexes (Will et al., 1996 (361)).

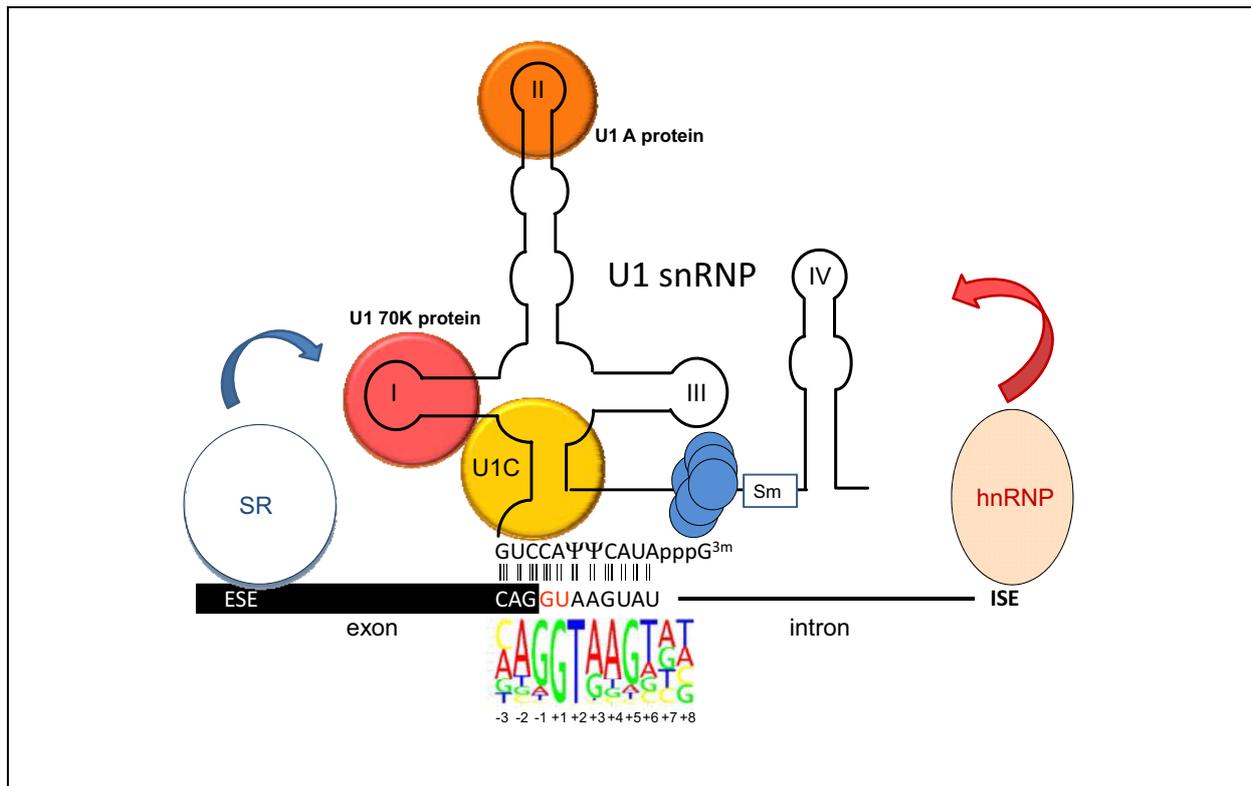


Fig. 16: Recognition of the 5' ss by RNA duplex formation between U1 snRNA

Assembly of the spliceosome is initiated by recognition of the 5'ss by the U1 snRNP through base-pairing interactions of the free 5' end of the U1 snRNP and the 5'ss. Nucleotides capable of participating in the U1 snRNA:pre-mRNA interaction have been shown to include positions -3 to +8 of the 5' ss and all 11 nt constituting the single-stranded 5' end of the U1 snRNA (Freund et al., 2003 (91)).

In an experimentally approach to determine the intrinsic 5' ss strength, a hydrogen bond model for the complementarity between the free 5' end of the U1 snRNP and the 5' ss has been established (Kammler et al., 2001(152); Freund et al., 2003 (91)). The hydrogen bond weight model translates the hydrogen bond pattern between the 5'ss and all 11 nt of the free 5' end of the U1 snRNA into a numerical HBond score (available at the web-interface http://www.uni-duesseldorf.de/rna/html/hbond_score.php). The stability of the RNA duplex is not exclusively determined by its complementarity to U1 snRNA, but also by additional interactions of protein components with the pre-mRNA in the vicinity of the 5'ss, including the U1-specific proteins U1-A, U1-C and U1 70K. The seven Sm proteins assemble around the Sm site nucleotides, located between stem loop III and stem loop IV. The group of metabolically stable RNAs known as U snRNAs to which the U1 snRNA belongs is marked by a 5' terminal cap which contains the unusual nucleoside 2,2,7-trimethylguanosine (m₃G) at its 5' end.

Many exonic splicing enhancer (ESE) sequences contain binding sites for members of the SR (serine/arginine-rich) family of proteins, which upon binding to *cis*-acting ESE sequences, enhance the interaction of the U1 snRNP with the 5'ss. Moreover, intronic splicing enhancer (ISE) or intronic splicing silencers (ISS) sequences can enhance or repress the use of nearby 5' ss.

Introduction

In addition to U1 snRNP-associated proteins, other splicing factors are involved in recruitment of the U1 snRNP to the 5'ss and in stabilizing binding of the U1 snRNP to the 5'ss. Several *cis*-acting elements and *trans*-acting factors have been identified preferentially, but not exclusively, upstream and downstream of 5'ss with low complementarity to the 5'end of U1 snRNA.

It appears that the number of complementary bases required for U1 snRNA binding is modulated by neighbouring exonic and intronic sequences [F.-J. Grosseloh diploma thesis, 2006 (110)]. Many exonic splicing enhancer (ESE) sequences contain binding sites for members of the SR (serine/arginine-rich) family of proteins.

The best studied example is the SR (serine/arginine-rich) protein ASF/SF2, which, upon binding to *cis*-acting ESE sequences, enhances the interaction of the U1 snRNP with the 5'ss, probably through a direct interaction between the arginine-serine-rich domain (referred to as the RS domain) of ASF/SF2 and the RS domain of the U1 snRNP component U1-70K (Wu & Maniatis, 1993 (367); Kohtz et al., 1994 (169)). However, most recent data demonstrate that the RNA recognition motif (RRM) of ASF/SF2 bridges the RRM of U1-70K (Cho et al., 2011 (57,58)). It appears that the hypo-phosphorylated RS domain of ASF/SF2 interacts with its own RRM, whereas the hyper-phosphorylated RS domain permits formation of a ternary complex containing an exonic splicing enhancer sequence, the SR protein ASF/SF2, and the U1 snRNP.

Moreover, intronic splicing enhancer (ISE) or intronic splicing silencers (ISS) sequences can enhance or repress the use of nearby 5' ss. Binding of the protein TIA-1 to uridine-rich sequences immediately downstream from the 5'ss helps to stabilize U1 snRNP recruitment via direct interaction with U1-C (Förch et al., 2002 (87)). Also, G triplets can function as intronic splicing enhancers, in most cases by binding of hnRNP (heterogeneous ribonucleoprotein) H/F (Hastings et al., 2001 (124); Caputi & Zahler, 2002 (46)). On the other hand, binding of hnRNP proteins to exonic splicing silencer (ESS) sequences and intronic splicing silencer (ISS) sequences can contribute to the repression of 5'ss recognition (Blanchette et al., 1999 (31); Eperon et al., 2000 (78)).

During catalytic activation of the spliceosome, the U1 snRNP at the 5'ss is replaced by both the U6 snRNP and the U5 snRNP (see also **Fig. I5**). This switch is thought to act as a sequential inspection mechanism of the 5'ss to ensure the fidelity of 5'ss cleavage (Staley & Guthrie, 1999 (318)).

1.2.5.2 3'ss recognition

The 3' ss is a multipart signal comprising a less conserved branchpoint consensus YNYURAY (Y = pyrimidine, R = purine, N = any nucleotide, branch point is underlined), and a stretch of pyrimidines (known as the polypyrimidine tract or PPT) adjacent to the invariant 3' ss AG (Moore, 2000 (215)). The distances between 3' ss signals are highly variable. The branch point sequence (BPS) is usually located 18-40 nucleotides upstream of the 3' ss AG, but may also reside up to several hundred nucleotides further upstream (Helfman & Ricci 1989 (126); Reed, 1989 (255); Gooding et al., 2006). Accordingly, polypyrimidine tracts vary in length and sequence composition. In particular, those polypyrimidine tracts composed of long uridine stretches promote the use of adjacent 3' ss (Reed, 1989 (255); Coolidge et al., 1997 (63)). However, natural polypyrimidine tracts are frequently interrupted by cytosines or purines (Senapathy et al., 1990 (292)).

This is reflected by the essential pre-mRNA splicing factor U2AF65, which coordinates the initial steps of 3'ss recognition by recognizing the polypyrimidine tract. U2AF (U2 auxiliary factor) is a heterodimer comprising a large subunit, U2AF65, and a small subunit, U2AF35 (Zamore & Green, 1989 (377); Zamore, 1992 (378)). U2AF65 contains three C-terminal RNA recognition motif (RRM) domains as well as an N-terminal region rich in basic residues and containing seven arginine-serine dipeptide repeats (RS domain) (Zamore, 1992 (378)). Association of U2AF65 with the polypyrimidine tract is most likely mediated by two of the three RRM domains. U2AF65 might distinguish purines (adenine and guanine) from pyrimidines (uracil and cytosine) on the basis of their size, but more likely on the basis of their unique patterns of hydrogen bond donors and acceptors (Sickmier et al., 2006 (301)). Since U2AF65 preferentially binds uridine-rich RNA segments, polypyrimidine tracts with long uridine stretches are stronger than those with interruptions of other nucleotides (Singh et al., 1995 (304)). These weak polypyrimidine tracts require an additional U2AF35-3' ss AG interaction for their recognition (Reed et al., 1989 (255); Wu et al., 1999 (368)). If the polypyrimidine tract is sufficiently long, the AG sequence will not be required until the second of the two chemical steps involved in splicing (AG independent introns). In contrast, so-called AG-dependent introns, which mostly have short or interrupted polypyrimidine tracts (Reed et al., 1989 (255)) require U2AF35 binding to promote or stabilize the binding of U2AF65 to the weak polypyrimidine tract (Guth et al., 2001(113)).

The primary structure of U2AF35 comprises a central RRM that is flanked by two zinc fingers in the N-terminus (Birney et al. 1993 (27)) and a glycine tract at the C-Terminus (Zhang et al. 1992; Kellenberger et al. 2002 (162)). U2AF35 binds both U2AF65 and the pre-mRNA through its RRM domain. Mutational analysis and *in vitro* genetic selection indicate that U2AF35 has a sequence specific RNA-binding activity that recognizes the 3' ss consensus,

Introduction

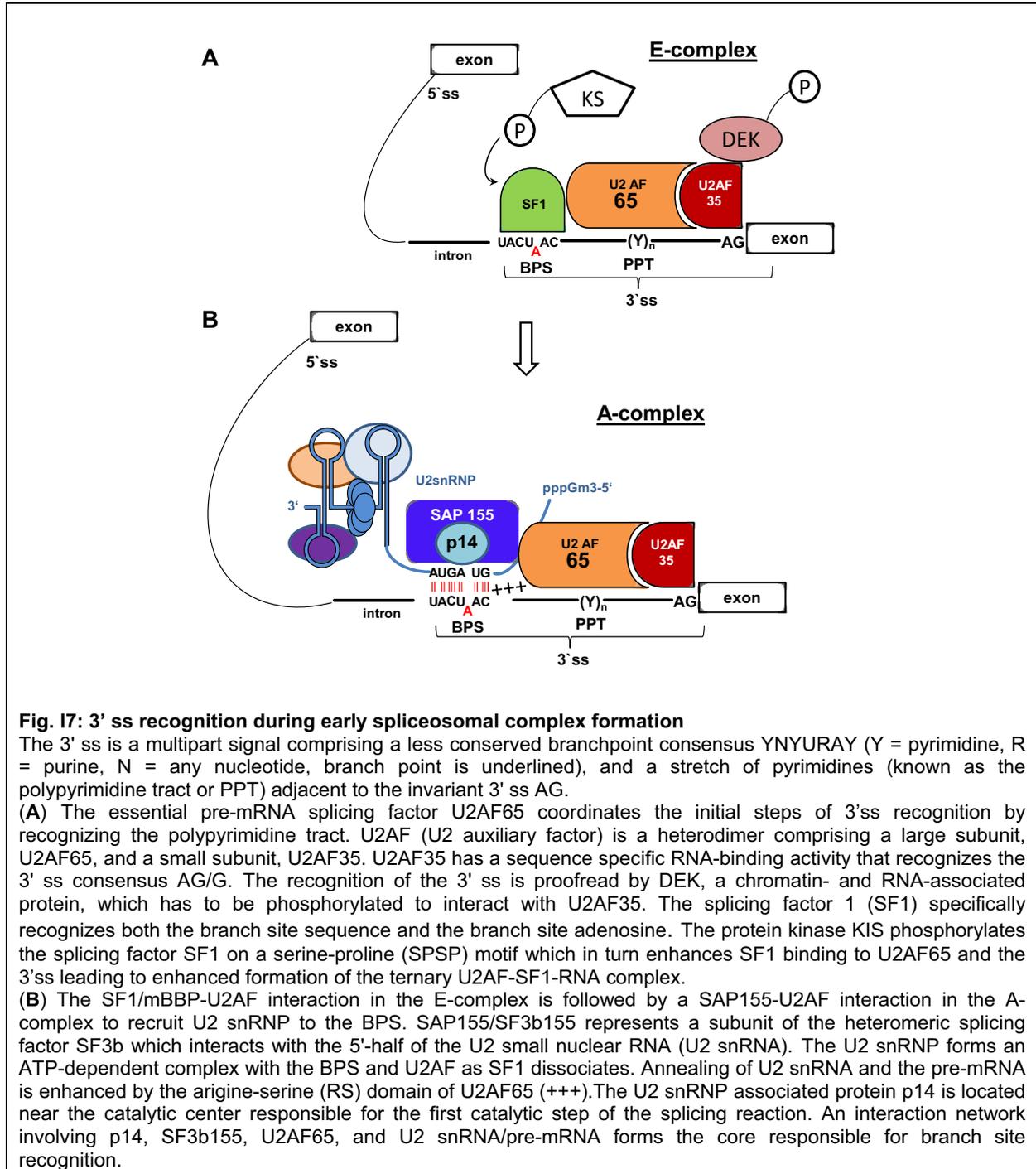
AG/G (Merendino et al., 1999 (212); Wu et al., 1999 (368)). It has been demonstrated that mutation of the 3'ss site AG/G to AG/C resulted in the loss of the stimulatory effect of U2AF35 on U2AF65 cross-linking (Guth et al., 2001 (113)). Moreover, minigene expression studies and RT-PCR analysis revealed that the nucleotide immediately downstream of the highly conserved AG dinucleotide appears to affect splice site recognition only in the presence of a suboptimal polypyrimidine tract with guanosine strongly promoting splicing compared to cytosine (L. Hartmann, diploma thesis (121)).

The branch point, which often bears little resemblance to the consensus motif, appears to be specified independently of the 3' ss AG by its immediate sequence context and by its proximity to the polypyrimidine tract (Smith et al., 1993 (307)). The splicing factor 1 (SF1, or mammalian branch point binding protein, mBBP) specifically recognizes both the branch site sequence and the branch site adenosine through its KH (hnRNP K homology) domain (Kramer et al., 1996 (175); Berglund et al., 1997 and 1998 (22,23); Peled-Zehavi et al., 2001 (240)). Binding of SF1 to the BPS, however, is weak ($K_s \approx 1 \mu\text{M}$) (Lui et al., 2001), but its affinity is significantly increased by simultaneous interaction with the third RNA recognition motif (RRM 3) of U2AF65 (Berglund et al. 1998 (23); Selenko et al., 2003 (290)). The protein kinase KIS phosphorylates the splicing factor SF1 on a Serine-Proline motif (SPSP) motif which in turn enhances SF1 binding to U2AF65 and the 3'ss leading to enhanced formation of the ternary U2AF-SF1-RNA complex (Manceau et al., 2006 and 2008 (200,201)).

The SF1/mBBP-U2AF interaction in the E-complex is followed by a SAP155-U2AF interaction in the A-complex to recruit U2 snRNP to the BPS (Gozani et al., 1998 (106)). SAP155/SF3b155 represents a subunit of the heteromeric splicing factor SF3b (Golas et al., 2003 (101), Spadaccini et al., 2006 (313)), which interacts with the 5'-half of the U2 small nuclear RNA (U2 snRNA), whereas SF3a associates with the 3'-portion of U2 snRNA (Kramer et al., 1996 (175)). The U2 snRNP forms an ATP-dependent complex with the BPS and U2AF as well as SF1 dissociates. U2AF65 stabilizes the interaction of U2 snRNP with the branch point (BP) by contacting the branch region through its N-terminal RS domain, promoting base pair interactions between U2 snRNA and the BP (Gaur et al. 1995 (98); Valcarcel et al., 1996 (342); Kent et al., 2003 (164); Shen and Green, 2004 (298)). U2AF65 at the 3'ss structures the PPT to juxtapose the branch point sequence and the 3'ss positioning the RS domain of U2AF65 in the vicinity of the branch point sequence and U2AF35 at the 3'ss (Kent al. 2003 (164)).

The U2 snRNP base pairs with the branch point region while the nucleophilic branch site adenosine does not base pair with the U2 snRNA, but rather bulges out of the recognition helix (Zhuang & Weiner, 1989 (385); Query et al., 1994 (248); Berglund et al., 2001(25)). There is also evidence that sequence-independent binding of the highly conserved

SF3a/SF3b subunits upstream of the branch site is essential for anchoring U2 snRNP to the pre-mRNA (Gozani et al., 1996 (105)). In particular, SAP155 was shown to crosslink to pre-mRNA on both sides of the BPS in the A complex (Gozani et al., 1998 (106)).



The p14 subunit of the essential splicing factor 3b (SF3b) which comprises a canonical RNA recognition motif (RRM) can be cross-linked to the branch-point adenosine and stably interacts with the SF3b subunit SF3b155. Therefore, an interaction network involving p14, SF3b155, U2AF65, and U2 snRNA/pre-mRNA forms the core responsible for branch site recognition (Spadaccini et al., 2006 (313)). The U2 snRNP associated protein p14 is located near the catalytic center responsible for the first catalytic step of the splicing reaction. A phylogenetically conserved pseudouridine in the U2 snRNA, located opposite of the branch point adenosine, may induce a unique conformation of the branch-point adenosine that primes for attack at the 5' splice site (Newby & Greenbaum, 2002 (224)).

The 3' splice site itself seems to be recognized in a scanning process for the first AG dinucleotide downstream of the branchpoint/polypyrimidine tract. Interestingly, CAG, UAG and AAG triplets were efficient 3' splice sites whereas GAG was not used at all (Smith et al., 1989 (307), Lev-Maor et al., 2003 (187)). This was also shown for 'tandem' (NAGNAG) 3' splice sites that effectively compete with each other (Hiller et al., 2006 (128,129)). Exceptions of the scanning process occurred, if the AG resides very close to the BPS and then can be bypassed (Chua & Reed, 2001 (61); Gooding et al., 2006 (102)). Moreover, the recognition of the 3' splice site is proofread by DEK, a chromatin- and RNA-associated protein. It has been demonstrated that depletion of DEK from nuclear extract reduced the ability of endogenous U2AF to discriminate between CG and AG dinucleotides and this activity was substantially restored by the addition of recombinant DEK protein. An interaction between *in vitro* synthesized U2AF35 and recombinant DEK was observed and phosphorylation of DEK was required for this interaction. Mutation of two known serine phosphorylation sites in DEK both abolished DEK phosphorylation and inhibited the interaction with U2AF35 (Soares et al., 2006 (308)) suggesting that phosphorylation of DEK promotes its association with U2AF35, which in turn enhances AG dinucleotide discrimination by the U2AF heterodimer.

1.2.5.3 Cis-active regulatory elements

Accurate splice site recognition further depends on *cis*-regulatory elements in the pre-mRNA that modulate splice site selection and allow to discriminate between real and pseudo splice sites (Sun & Chasin, 2000 (327); Sironi et al., 2004 (305)) (Fig. 1). Most exons contain exonic splicing enhancers (ESEs), which define them as recognition units promoting the use of their splice sites (Selvakumar et al., 1999 (291); Cartegni et al. 2003 (50); Fairbrother et al., 2004(81)). In addition, exons also contain functional splicing suppression units known as exonic splicing silencers (ESSs) (Wang et al., 2004 and 2006 (353,354)). Moreover, intronic splicing enhancers (ISEs) or intronic splicing silencers (ISSs) enhance or repress the use of

Introduction

nearby 5' or 3' ss (Carlo et al., 1996 (48); Ponthier et al., 2006 (246); Tange et al., 2001 (331); Modafferi & Black, 1997 (214); Kashima et al., 2007 (157,158)). These *cis*-acting splicing regulators are short degenerate RNA sequences, which occur frequently in the genome.

Enhancer motifs are frequently bound by the group of serine/arginine rich (SR) proteins, which mostly exerts a positive effect on splice site recognition and stimulates spliceosome assembly (Fu et al., 1992 (93); Zahler et al., 1993 (375); Berget, 1995 (18); Manley & Tacke, 1996 (202); Liu et al., 1998 and 2000 (191,192); Carlo et al., 2000 (47); Caputi et al., 2004 (44)). These positive effects can be antagonized by heterogeneous nuclear ribonucleoproteins (hnRNPs) that usually bind to silencer elements (Caputi et al., 1999 and 2002 (46); Crawford & Patton, 2006 (65); Hallay et al., 2006 (117); House & Lynch, 2006 (135)). However, it should be noted that the same sequence motif sometimes can act as an enhancer or silencer, depending on its position with respect to the splice sites (Goren et al., 2006 (104); Ule et al., 2006 (341)). The activities of *cis*-acting elements were shown to be context specific and there is compelling evidence that SR proteins can suppress splicing when bound to sequences located within the intron, and there are also examples of members of the hnRNPs exhibiting stimulating effects on splicing (Kanopka et al., 1996 (155); Chen et al., 1999 (55); Dauksaite & Akusjarvi, 2002 (67); Ibrahim et al., 2005 (141); Schaub et al., 2007 (280)). HnRNPs recognize the RNA via their KH (K homology) and RRM RNA-binding domains and RGG and glycine-patch domains. The multiple α -helices and antiparallel β -strands bind short motifs of 4-7 nucleotides in single-stranded DNA or RNA.

Moreover, the β -sheet surface on the RRM domain of many SR proteins recognizes specific RNA sequences through base stacking, hydrophobic, polar and electrostatic interactions (Jokan et al., 1997 (147); Lewis et al., 1999 (189); Braddock et al., 2002 (36); Auweter et al., 2006 (11-13)). The majority of KH and RRM proteins contain more than one copy of each RNA recognition domain engaging a range of different motifs leading to 'fuzzy' identity of *cis*-active regulatory elements (Chandler et al., 1997 (52)).

Specific splice site regulation, despite frequent occurrence of the degenerate target motifs, is achieved by clusters of degenerate RNA motifs bound by several different activator and repressor proteins. In addition, competition between SR proteins and hnRNPs or between these proteins and general splicing factors modulate splice site selection (Singh & Valcarcel, 2005 (303)). Furthermore, the activity of SR proteins as splicing factors depends on the phosphorylation lead to a movement into a different subcellular localization (such as from the nucleus to the cytoplasm), where they are unable to affect splicing (Tacke et al., 1997 (330); Kanopka et al., 1998 (156); Singh & Valcarcel, 2005 (303); Tacke et al., 1997 (330)). The RS domains of SR proteins engage in protein-protein interactions promoting interactions

between the components of the spliceosome to define exons or interactions across the intron during spliceosome assembly (Graveley, 2000 (109)). Binding of RS domains to RNA presumably shields negative charges facilitating annealing of complementary RNA strands during numerous base-pairing rearrangements required for spliceosome assembly and catalysis (Lee et al., 1993 (185); Singh & Valcarcel, 2005 (303); Shen & Green, 2006 (297)).

1.2.5.4 Exon recognition

A typical human gene contains relatively short exons (typically, 50-250 base pairs) in length separated by much larger introns (typically, hundreds to thousands of base pairs) that on average account for > 90% of the primary transcript. This transcript geometry, and the predominant exon skipping phenotype of splice site mutations, are consistent with the idea that in mammals splice sites are predominantly recognized in pairs across the exon termed “exon definition” (Robberson et al., 1990 (262); Nakai & Sakamoto, 1994 (219); Sterner et al., 1996 (323)).

Exon definition involves initial interactions across the exon between factors recognizing the 5'ss and the upstream 3'ss, whereas in the alternative model, intron definition, interactions firstly occur across the intron between factors recognizing the 5'ss and the downstream 3'ss (for a review, see Berget 1995 (18)).

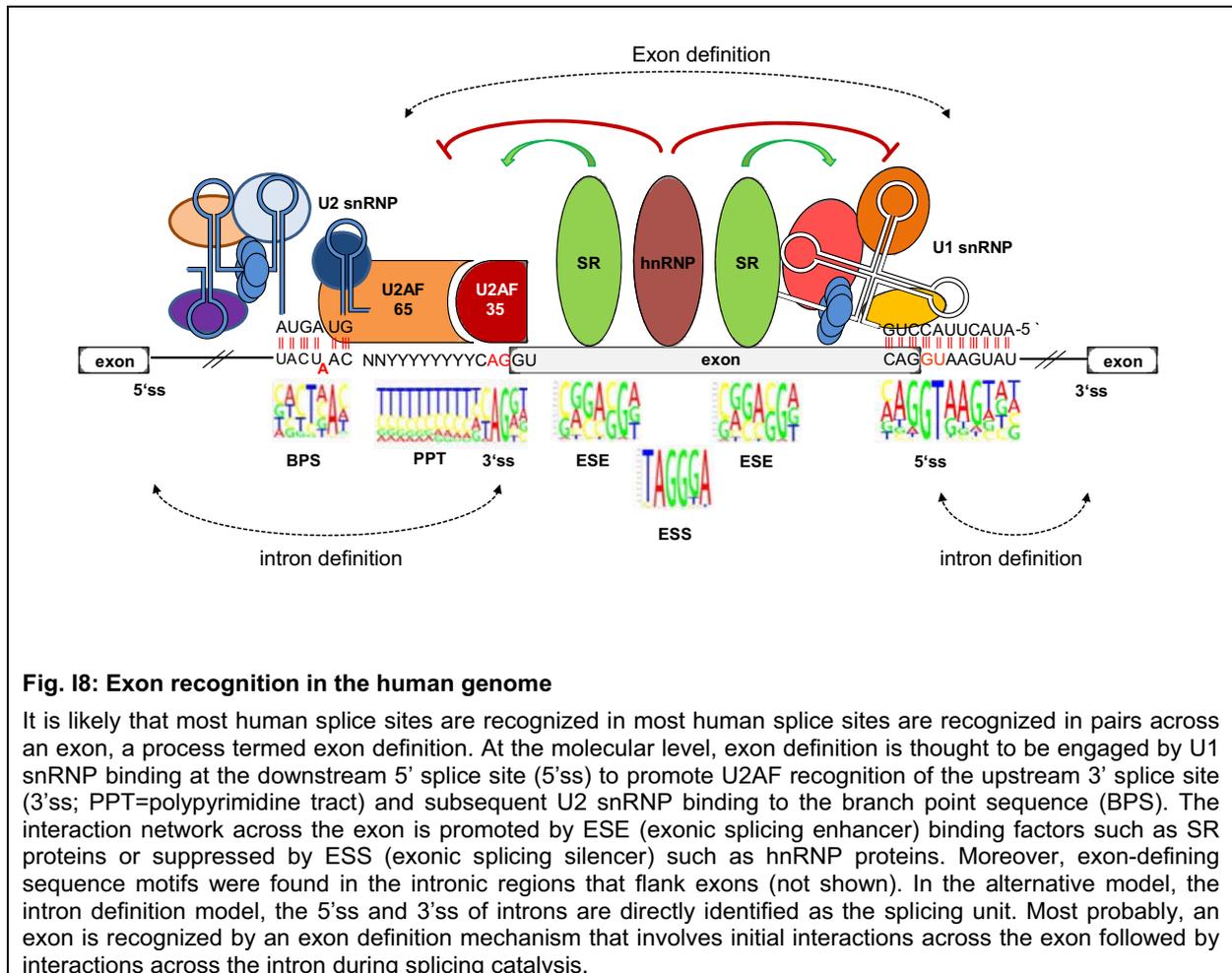
During exon definition, splicing enhancer sequences within the exon (ESEs) recruit SR proteins that establish a network of protein-protein interactions across the exon, thereby bridging U2 snRNP/U2AF at the 3'ss and U1 snRNP at the 5'ss and stabilizing the exon-defined complex (Hoffman & Grabowski, 1992 (133); Reed, 2000 (256)). By contrast, binding of hnRNP A1 can antagonize this activity of SR proteins. Recent data demonstrate that in addition to the U1 and U2 snRNP, cross-exon complexes contain U4, U5 and U6 snRNP, which form the tri-snRNP (Schneider et al., 2010 (283)). Moreover, exon-defining sequence motifs were found in the intronic regions that flank exons (Ke & Chasin, 2010 (160)). Many of these resemble binding sites the binding sites of hnRNPs.

After exon definition, splicing factors must form a complex across the upstream intron to allow splicing catalysis. It is thought that cross-exon interactions are disrupted and the cross-exon complex is converted into a cross-intron A complex, where a molecular bridge now forms between U2 snRNP and U1 snRNP bound to the upstream 5'ss (Reed, 2000 (256)). Alternatively, the switch from cross-exon to cross-intron complex can occur directly without prior formation of cross-intron A complex. Cross-exon complexes containing the tri-snRNP can directly engage an upstream 5'ss and thereby lead to pairing of splice sites across an intron (Schneider et al., 2010 (283)).

Introduction

Some data indicate that regulation of exon inclusion or skipping occurs during the switch from a cross-exon to a cross-intron complex (House & Lynch, 2006 (135); Bonnal et al., 2008 (33); Sharma et al., 2008 (295)). It seems that an irreversible and functional commitment to specific splice site pairing does not occur at E complex, but rather at A complex (Lim & Hertel, 2004 (190)).

Even though a pair of splice sites may be in close proximity during E complex, their association remains dynamic until an ATP-dependent lock of U2 snRNP on the pre-mRNA. Given the divergent sequence and architecture of genes, every exon has its specific set of elements that permits its recognition by the spliceosome. Each exon is flanked by a unique pair of splice site signals and contains a unique group of splicing enhancers and silencers and maybe secondary structures. The sum of contributions from each of these elements then defines the overall recognition potential of an exon (for a review see Hertel, 2008 (127)).



1.2.6 Splice site strength and identification of pathogenic splicing mutations

Mutations, even single nucleotide changes, can modify splicing in various ways: they can strengthen, weaken or even destroy an existing proper splice site or *cis*-regulatory element, or create a new one. Such splicing signal modifications may or may not lead to observable phenomena like exon skipping, activation of cryptic or *de novo* splice sites, or intron retention. Most patients, however, are *genotyped* only, and diagnostic RNA-level information about aberrant splicing is usually not available. Therefore, any computational prediction of DNA mutation effects on splicing (for an overview see Hartmann et al., 2008 (123)) can be beneficial for the human geneticist. Such predictions can be obtained from algorithms scoring the functionality of a given splice site and/or *cis*-regulatory element.

The “*splice site strength*” is a useful and central concept in judging the possible effect of a splicing signal mutation. Together with a “threshold” for splice site functionality, comparing strengths of wild type and mutant signal could yield reliable predictions of splicing effects (Sahashi et al., 2007 (275)). However, although widely used in the literature, the term “splice site strength” does not refer to a unique definition. In principle, any measure of “*functional* splicing signal strength” should quantitatively describe, why a given splice site is preferred over competing nearby potential (“*pseudo*”, “*mock*” or “*decoy*”) splice sites under cell specific conditions. It should take into account not only the proper 5' or 3' ss sequence, but also its context of *cis*-regulatory elements and pseudo splice sites, and even the cellular environment of SR proteins. In practice, this ambitious comprehensive concept (“*the splicing machinery itself*”) has not yet been implemented *in silico* and is approximated by more limited computational procedures. It comes natural that a wide variety of concepts from computational physics, artificial intelligence and machine learning have been applied to this problem.

In principle, two types of computational methods for splice site detection can be distinguished: those that are trained only by positive examples (*real* splice sites) – e.g. Weight Matrix/Array Models and Maximum Dependency Decomposition –, and those additionally requiring a training data set of negative examples (*decoy* splice sites). Locally, several different algorithms calculate a splice site’s *intrinsic* strength from a narrow region of nucleotides around the respective consensus dinucleotides (GT or AG), irrespective of its wider sequence context. A splice site’s *relative* strength then refers to the difference (or ratio) of its intrinsic strength to the neighboring pseudo sites, thus depending on the splice site context. The meaningful combination of *cis*-regulatory elements and *relative* splice site strength into a single *functional* strength measure still remains an open question, although a first step towards combining splice site scores and those of *cis*-regulatory elements has been taken by the splicing simulation software *ExonScan*, which independently adds up log-odds-

scores of individual components to obtain one overall score (<http://genes.mit.edu/exonscan/> (Wang et al. 2004 and 2006 (353,354))). However, all local primary sequence methods are bound to *misdiagnose* splice sites, due to the huge overlap of sites in the real and decoy data sets.

1.2.6.1 5'splice strength algorithms

The most widely-used *intrinsic* strength concept simply measures the 5' splice site's similarity with a consensus motif. Initially, Shapiro and Senapathy (S&S) developed a position-specific weight matrix (PSWM) for 5'ss, which reflects the degree of sequence conservation of the known 5' ss from position -3 (the third nucleotide from the 3' end of the upstream exon) to $+6$ (the sixth nucleotide in the intron) in an alignment of 1,446 5' ss. From this matrix they derived the S&S score in the range 0–100, with score 100 representing full coincidence with the consensus sequence, and score 0 obtained, if every position is occupied by the least likely nucleotide. All positions in the 5' ss are assumed independent by the S&S score, as with every weight matrix model.

Traditionally, splice sites with a high degree of resemblance to the consensus have been considered as strong splice sites, whereas non-consensus splice sites have been assumed to be intrinsically weak. Although this is still widely accepted, significance of such a consensus sequence remains arguable, because resemblance to frequency-based consensus matrices of independent nucleotides turned out to be insufficient for reliable prediction of 5' ss (Lear et al., 1990 (184)). Moreover, many matches to each consensus are present along pre-mRNAs, but the vast majority of these sequences are *pseudo* or *decoy* splice sites never selected for splicing (Sun & Chasin, 2000 (327)). Weight matrix models (WMM) represent an extension to the S&S score, indicating the relative importance of each base at every position: they quantify the relative likelihood of a given candidate splice site sequence with respect to the background nucleotide distribution from a training set of splice signals, but they still fail to incorporate nucleotide interdependencies.

An improvement for 5' ss prediction has been achieved by considering dependencies between bases of the 5' ss. Burge and colleagues developed three different algorithms that take into account dependencies between positions -3 to $+6$ of the 5' ss motif (Yeo & Burge, 2004 (373)): these algorithms apply probabilistic approaches to large datasets of known RNA splicing signals. The maximum dependence decomposition model (MDD) is an iterative decision-tree approach that captures the strongest dependencies – also between non-neighboring positions – in the early branches of the tree by WAM, and uses WMM for nearly independent positions. The maximum entropy model (MEM) performs better than previous

Introduction

models and is based on the maximum entropy distribution (MED). In statistical theory, this approach represents the least biased approximation for the distribution of sequence motifs, consistent with a set of constraints estimated from available data – known real and decoy signal sequences. It makes no further assumptions about the distribution than consistency with this empirical distribution, and different sets of constraints generate different models. The MEM incorporates local adjacent and nonadjacent position dependencies consistent with low-order marginal constraints for “few” nucleotides estimated from available data (MaxENTScan algorithm: <http://genes.mit.edu/burgelab/maxent/Xmaxentscanscoreseq.html>).

These algorithms use input sequences of constant length – a 9-mer in case of the 5' ss and 23-mer for the 3' ss –, and assign each sequence a numerical score reflecting the likelihood of the sequence being a true splice site.

While weight matrix/array models require the selection of relevant information features by hand, machine learning techniques automatically deduce a classification function (“rule”) that optimizes a given criterion in distinguishing training data sets of positive and negative sequences (real and decoy splice sites). For example, the neural network method (NN) is a machine learning approach that recognizes sequence patterns once it is trained with sets of DNA sequences encompassing authentic and decoy splice sites (http://www.fruitfly.org/seq_tools/splice.html) (Brunak et al., 1990 and 1991 (40,41), Reese et al., 1997 (257)). It employs a backpropagation feedforward neural network with one hidden layer, and produces an output score between 0 and 1 for each splice site candidate. Interestingly, decoy GT sites close to a real 5' ss have weaker neural network scores than those farther away, which seems consistent with the concept of *relative* splice site strength, comparing a real 5' ss with decoy sites in its neighborhood. Support vector machines (SVM) also belong to the category of machine learning systems that infer a classification function from a training data set. By using an appropriate representation for features of real and decoy sequences, specific splice site patterns can be obtained from the discrimination function of such models. Typically, only a small fraction of the large number of features represented by a high-dimensional feature vector are relevant for the classification and are mutually independent. Genetic algorithms have been successfully applied in the selection of such a “minimal feature set” with best classification performance. Estimation of distribution (EDA) algorithms have been shown to improve on these, most importantly providing normalized “feature weights” as ranking criterion (Degroeve et al., 2002 (68); Saeys et al., 2003 and 2004 (273,274)). SVM algorithms were also applied to detect splicing features in the human genome: 2,200 real and 2,300 pseudo exons including flanking intronic sequences were divided into five non-overlapping sequence compartments. The strongest features searched in words of length 4–7 nucleotides were the presence or absence of 4-

Introduction

mers and 5-mers, consistent with motifs identified by other methods, and at comparable sensitivities and specificities (Zhang et al., 2003 (381)).

With a view to the biological function of the 5' ss as a recognition site for the U1 snRNP early in spliceosome assembly, it seems obvious to determine a 5' splice site's intrinsic strength regarding this interaction. Indeed, stable RNA duplex formation between the U1 snRNA and the 5'ss is a prerequisite for spliceosome formation, and it has been shown that the stability of the U1 snRNA duplex has strong influence on the selection between two nearby 5' ss (Eperon et al., 2003 (78); Freund et al., 2003 (91); Bi et al., 2005 (26)). From a thermodynamic viewpoint, the 5'ss:U1 snRNA duplex stability can be quantified by its free energy ΔG , using the nearest-neighbor RNA base-pairing parameters reported by the Turner laboratory (Serra & Turner, 1995 (293)). These empirically fitted formulae are based on measurements with synthetic oligoribonucleotides and reflect the contribution of hydrogen bonding, base stacking, mismatches, and Watson-Crick or G·U base pairs (Reddy et al., 1981 (254)).

The nearest-neighbor approximation works very well for Watson-Crick base pairs, satisfactorily well for G·U base pairs flanked by Watson-Crick base pairs, but is less reliable for mismatches. Moreover, undetermined energy corrections at the ends of a short RNA duplex may impose limits on the accuracy of the free energy calculations (Freund et al., 2003 (91), Sorek et al., 2004 (311,312); Roca et al., 2005 (265)). Therefore, approximate free energies, calculated e.g. by popular computational web tools like DynAlign (Mathews & Turner, 2002 (205,206)), HyTher (Bommarito et al., 2000 (32)) and Bindigo (Hodas & Aalberts, 2004 (131)), seem insufficient for a reliable description of U1 snRNA duplex contribution to 5' ss strength.

In a complementary approach to experimentally determine intrinsic 5' ss strength in a model system, U1 snRNA duplex formation has been monitored within a retroviral-derived model transcription unit (Freund et al., 2003 (91)). It is well known that stable U1 snRNA duplex formation with 5' ss can protect pre-mRNA against degradation prior to splicing, and also initiates formation of the spliceosome. In combination with functional splicing assays, this protection mechanism has been used to obtain biological evidence of duplex stability. This experimental evidence was supplemented with a computational hydrogen bond weight model, translating the hydrogen bond pattern between the 5'ss and all 11 nt of the free 5' end of the U1 snRNA into a numerical HBond score (available at the web-interface http://www.uni-duesseldorf.de/rna/html/hbond_score.php). Beyond hydrogen bond formation at individual positions, the HBond algorithm also partially models nucleotide interdependence beyond nearest neighbor relationships. Contrary to purely statistical approaches currently ignoring nucleotides beyond position +6 due to lack of information content, the HBond

Introduction

algorithm fully takes positions +7 or +8 into account, with experiments confirming the dependency of the U1 snRNA duplex on these nucleotides. This observation is consistent with *in vitro* selection experiments to isolate functional 5' ss from pools of random sequences, where those 5' ss with the best complementarity to U1 snRNA were selected most efficiently, even if base pairing to U1 snRNA extended to positions +7 and +8 (Lund & Kjems, 2002 (193)). Moreover, mutual relationships between nucleotide positions within the 5' ss motif have been confirmed by human-mouse comparative genomics, and the contribution of individual 5' ss nucleotides to the intrinsic strength of human 5' ss has been examined extensively by *in vitro* 5' ss competition assays of the human β -globin gene (Carmel et al., 2004 (49); Sorek et al., 2004 (311,312); Roca et al., 2005 (265)). Studies with this gene revealed that the authentic 5' ss of the first exon lies in the vicinity of a cryptic 5' ss located 16 nucleotides upstream, which is only activated when the authentic one is sufficiently weakened by mutation (Roca et al., 2003 and 2005 (264,265)).

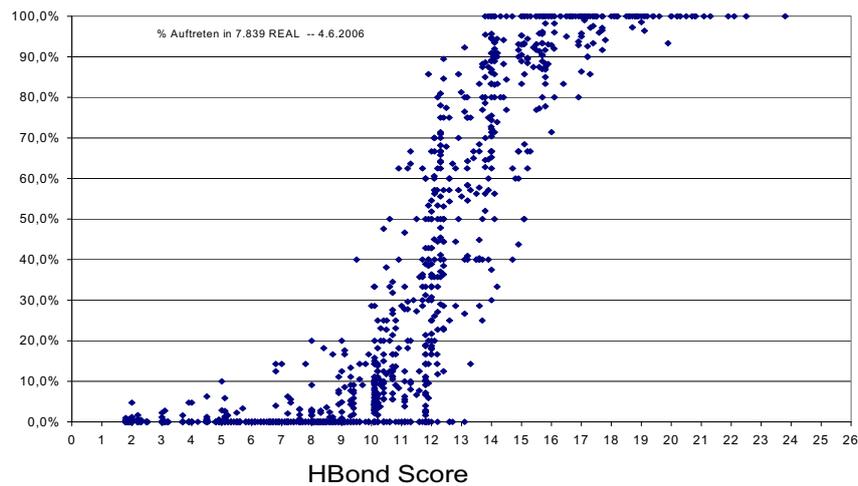


Fig. 19: HBond Score distribution of 7,849 real 5' ss

The computational hydrogen bond weight model, translating the hydrogen bond pattern between the 5' ss and all 11 nt of the free 5' end of the U1 snRNA into a numerical HBond score (available at the web-interface http://www.uni-duesseldorf.de/rna/html/hbond_score.php).

HBond score calculation of 7,849 real splice sites reveals that 5' ss with a HBond score beyond 12,3 are more frequent in the dataset than 5' ss with a HBond score lower than 12. This frequency distribution of the HBond score nicely correlates with splice strength. In other words, 5' ss with a HBond score lower than 12 can be considered as weak 5' ss.

In this case, the cryptic splice site can outweigh the mutant authentic one and be selected for splicing. Six 5' ss scores, including free energy ΔG , S&S, MM and MAXENT, were compared regarding their ability to explain these *in vitro* splicing analyses. However, no discriminating score threshold could be determined for any score that stringently separated activated from unused potential splice sites. Correlation (Pearson's r) between experimentally determined

percentage of splicing activation and scores was maximal for MAXENT, MM and ΔG in different competition schemes, suggesting mechanisms captured by different score algorithms. Indeed, both authentic and weakened 5' ss (reference sequences) have complementary nucleotides in positions +7 and +8, while the test sites do not. All examined 5' ss scores ignore these positions, which may be accountable for the lack of stringent differentiation. Interestingly, there was no correlation between the extent of complementarity of the 5' ss with U6 snRNA, which is in accordance with the observation that hyperstabilization of the 5' ss:U1 snRNA interaction does not inhibit replacement of the U1 snRNP by the U6 snRNP in higher eukaryotes (Freund et al., 2005 (92), Roca et al., 2005 (265)).

1.2.6.2 3'splice strength algorithms

The description of the inherent strength of 3' ss is more complicated due to sequence constraints of the 3' ss motif including the AG dinucleotide, the presence of the polypyrimidine tract (PPT) and the branch point sequence (BPS) upstream of the 3' ss. In addition, the distances between 3' ss signals are highly variable.

Algorithms that describe the intrinsic strength of 3' ss are based on nucleotide frequency matrices, machine learning approaches, neural networks, and on information contents of individual nucleotides, or apply probabilistic approaches considering dependencies between adjacent and non-adjacent positions (Shapiro & Senapathy 1987 (294); Brunak et al., 1990 (40); Reese et al., 1997 (257); Rogan & Schneider 1995 and 1998 (267,268); Senapathy et al., 1990 (292); Yeo & Burge, 2004 (373)).

The Shapiro and Senapathy matrix counts base frequencies at positions -14 to +1 of the 3' ss motif, whereas the first order Markov (MM) and maximum entropy model (MaxEntScore) use a wider sequence range of 3' ss positions from -20 to +3 (AG consensus at positions -1 and -2). Since the 3' ss sequence motif is much longer than the 5' ss, in a first step the maxent approach breaks up the 3' ss sequences into 3 consecutive non-overlapping fragments of length seven each, excluding the invariant AG dinucleotide.

This splitting, however, ignores the dependencies across fragment boundaries. To avoid that, six additional partially overlapping subfragments are introduced, and the final maxent likelihood is calculated from the appropriate ratio of individual segment distributions using second-order marginal constraints in each segment. While this second order Markov model is superior to a first-order model, performance is decreased again for third-order models.

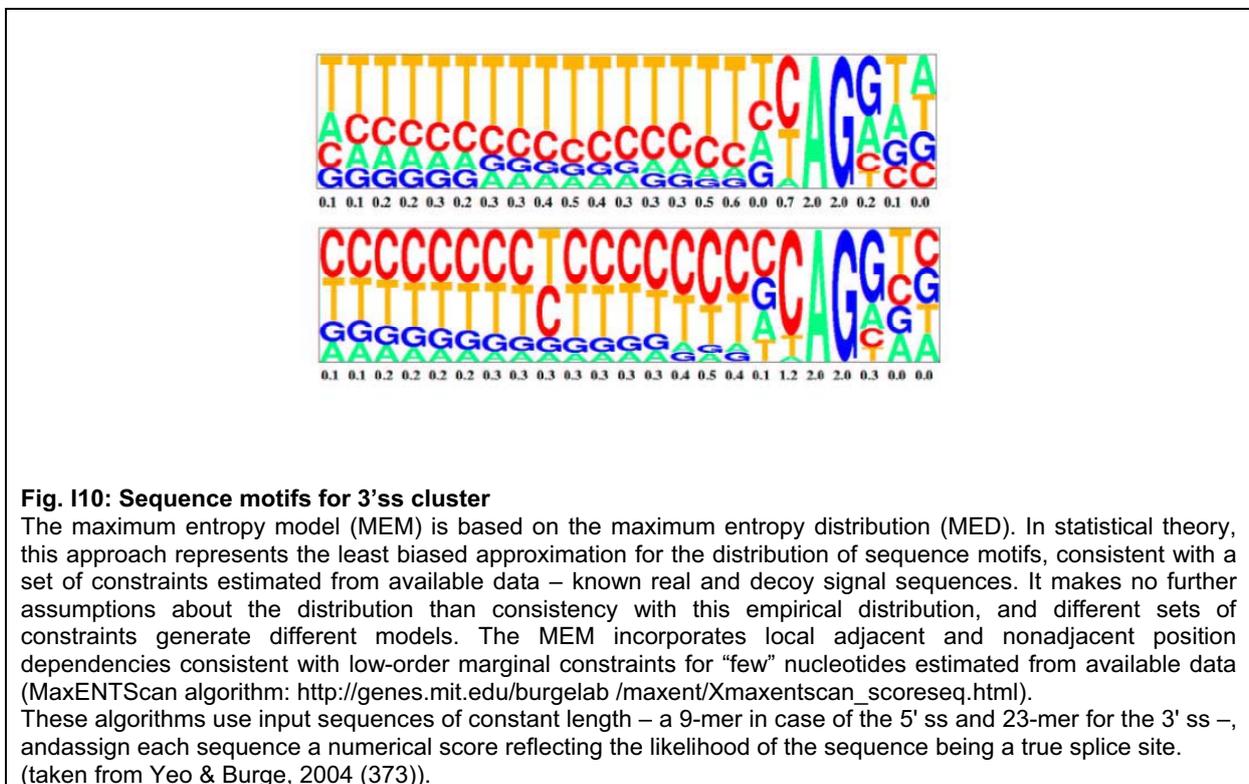
Introduction

Long-range dependencies across several “skipped” nucleotides are neglected in these models, but introducing additional dependencies does not significantly improve the performance beyond two-nucleotide-separation.

Comparison of the splice site strength using current prediction algorithms showed that the maximum entropy model class allowed the best discrimination between authentic and mutation induced aberrant 3' ss (Vorechovsky et al., 2006 (348)).

Ast and colleagues developed an algorithm which combines pairs of PPT and BPS to identify the location of functional BPS, since consensus scores alone are not sufficient to locate the BPS in introns due to frequent occurrence of high score motifs in exons and introns (<http://ast.bioinfo.tau.ac.il/>) (Kol et al., 2005 (170)). This algorithm is based on the BPS consensus calculated by Burge (Burge et al., 1999 (43)) and locates both the BPS and the PPT together by searching known combinations of BPS and PPT. The PPT borders are determined by a heuristic method based on experimental evidence (Coolidge et al., 1997 (63); Norton, 1994 (228)).

Their approach is contrasted by an algorithm which is primarily based on AG dinucleotide exclusion zones between the 3' ss AG and the BPS for branch point prediction (Gooding et al., 2006 (102)). This algorithm incorporates exons with distant BPS extending the usual search for probable branch points within a fixed distance of the 3' ss. Nevertheless, prediction of cryptic and *de novo* 3' ss is still a difficult task (Kralovicova et al., 2005 (174)).



2 RESULTS

Human gene mutations interfering with accurate exon recognition have a strong disease causing potential, since precise exon recognition in the precursor messenger RNA (mRNA) is a mandatory prerequisite for generation of intact proteins and correct cellular function. Although *in silico* tools predict the probability for aberrant splicing with considerable success reliable diagnosis of the splicing phenotype of a splice site mutation still requires functional splicing assays due to the complex interplay of splice site-defining sequence elements. If an RNA sample of a carrier of a putative pathogenic splice site mutation is not available, splicing minigene constructs provide a useful tool for analyzing such a splice site alteration.

In order to reliably test the effect of a mutation on exon recognition, most often a minimum of at least a three exon, two intron splicing minigene is necessary. However, in most human genes this would involve handling several thousand nucleotides due to the average large size of human introns (median size of 1458 base pairs for an internal intron (Scherer, 2008 (281)). In a heterologous splicing minigene only short DNA fragments need to be handled facilitating testing of putative pathogenic mutations and mutational analysis in general.

2.1. Requirements for the recognition of human exons with weak splice donor sites within a heterologous splicing minigene

2.1.1. Faithful *ATM* exon 54 recognition and intron removal in a heterologous splicing reporter minigene requires a strong terminal splice acceptor

Because the human *ATM* gene (ataxia telangiectasia mutated, Mendelian Inheritance in Man no. #607585, found at <http://www.ncbi.nlm.nih.gov/omim/>) was known to harbor a high number of exons and splicing mutations – approximately 50% of the ataxia telangiectasia (MIM #208900) patients were found to have disease due to mutations that resulted in aberrant splicing (Teraoka et al., 1999 (337)) – and as it has been suggested that this gene contains many exons with weak splice sites making this gene more susceptible to splicing mutations (Eng et al., 2004 (77)) – *ATM* exon 54 with its weak 5' splice site ((5'ss, HBond score = 12.3, calculated using the HBond Score algorithm (<http://www.uni-duesseldorf.de/rna>)) was chosen as a prototype human exon for establishment of a heterologous splicing reporter minigene.

In preparatory work in our group a heterologous transcription unit driven by the HIV-1 5' LTR (long terminal repeat) and terminated by the SV40 polyadenylation signal was generated. The 5' half of this construct comprised of the HIV-1 exon 1, the strong HIV-1 5' ss #1 - which

Results

is also called splice donor 1 or SD1/4 - with an HBond score of 20.8 and 68 base pairs of the HIV-1 intron 1. The 3'half of the construct was composed of intron 2 and an HIV-1 derived 3' splice site (3'ss) - which is also called splice acceptor (SA). Exon 3 in this splicing reporter was a hybrid of the CAT-ORF (chloramphenicol-acetyl-transferase-open reading frame) and the HIV-1 RRE (rev responsive element). Unique restriction sites within the reporter construct allowed both easy insertion of an internal test exon and splice site replacement (Neveling, K. diploma thesis, 2004 (222)) (**Fig. 1A**). Analysis of the influence of the intrinsic strength of the terminal 3'ss on the recognition of the central HIV-1 exon 2 within this reporter showed that exon recognition in the heterologous construct was affected by the strength of the terminal splice acceptor. The strength of the terminal splice acceptor appeared to be important if and only if one of the exon flanking splice sites was weak. This suggested that the recognition of the *ATM* exon 54 in the heterologous splicing minigene was also affected by the strength of the 3'ss within this reporter system (Neveling, K. diploma thesis, 2004 (222)).

To determine the impact of the strength of the terminal splice acceptor site on *ATM* exon 54 recognition within this heterologous minigene the human exon with its flanking splice sites (118 base pairs of the original upstream intron including the 3'ss and 11 base pairs of the original downstream 5'ss) was inserted into the reporter construct and subsequently the strength of the downstream splice acceptor was changed by mutagenesis (**Fig. 1B**): SA5 Py+ is a derivative of the HIV-1 SA5 in which the purine bases at position -4 and -5 within the polypyrimidine tract (PPT) close to the 3'ss AG dinucleotide were replaced by pyrimidine bases (pyrimidine content 60%). In SA5 Py++ all purine bases in the PPT except the AG dinucleotide of SA4b were substituted for pyrimidine bases (pyrimidine content 72%), whereas SA5 opt was further optimized by mutating the AG dinucleotides of SA4c, SA4a, SA4b and by perfect complementarity of BPS 2 (branch point sequence) and reduced complementarity of BPS 1 to U2 snRNA (pyrimidine content 80%). SA3 represents an efficient HIV-1 splice acceptor site without any modification (Kammler et al., 2006 (153)). The strength of these splice acceptor sites was calculated applying the MaxEntScore algorithm for 3'ss (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html).

HeLa cells were transiently transfected with these 3-exon-2-intron splicing minigenes harboring *ATM* exon 54 with its flanking splice sites as the middle exon and their splicing patterns were analyzed by RT-PCR. In the presence of SA5 Py+ (MaxEnt Score = 5.44) or SA5 Py++ (MaxEnt Score = 9.22) *ATM* exon 54 inclusion was unexpectedly low whereas splicing products with retention of the second intron were most abundant (65% and 61% respectively).

Results

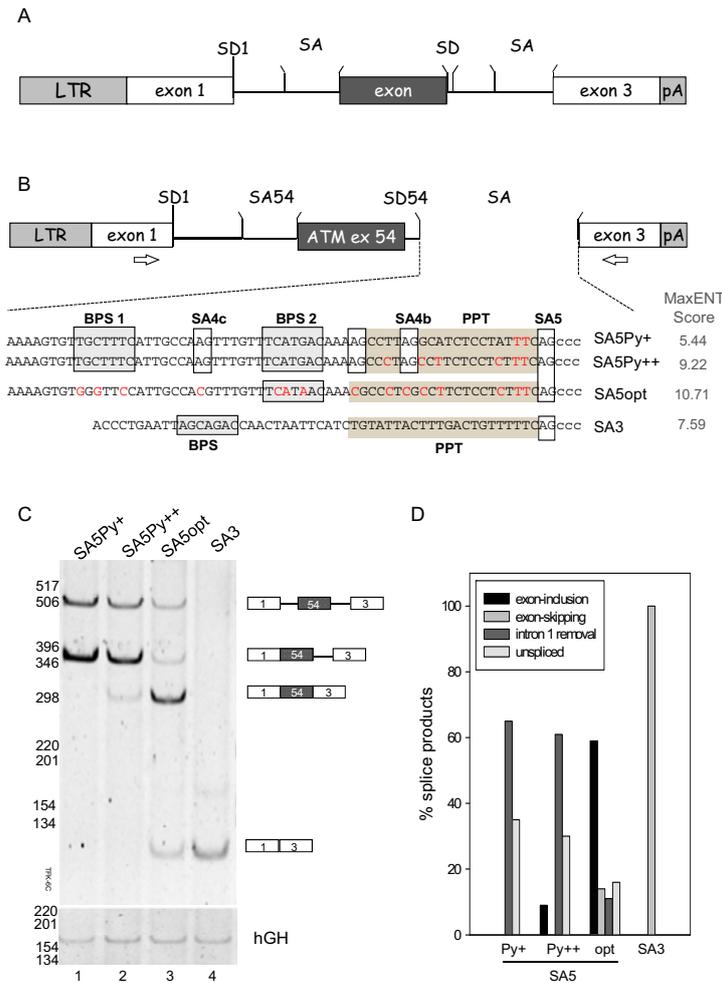


Fig. 1: Faithful *ATM* exon 54 exon recognition and intron removal in a heterologous splicing reporter minigene requires an optimized terminal splice acceptor

(A) Schematic drawing of the HIV-1 derived splicing reporter minigene driven by the HIV-1 5' LTR (long terminal repeat) and terminated by the SV40 polyadenylation signal. The 5' half of this construct comprised of the HIV-1 exon 1, the strong HIV-1 5' ss #1 - which is also called splice donor 1 or SD1/4 - with an HBond score of 20.8 and 68 base pairs of the HIV-1 intron 1 (The HBond was calculated using the HBond score algorithm (<http://www.uni-duesseldorf.de/rna>) - and 68 base pairs of the HIV-1 intron 1. The 3' half of the construct was composed of intron 2 and an HIV-1 derived 3' splice site (3'ss) - which is also called splice acceptor (SA). Exon 3 in this construct was a hybrid of the CAT-ORF (chloramphenicol-acetyl-transferase-open reading frame) and the HIV-1 RRE (rev responsive element). Unique restriction sites within the reporter construct allow both easy insertion of a test exon and splice site replacement.

(B) Schematic drawing of the reporter shown above harboring *ATM* exon 54 (NM_000051.2, exon numbering as reported in Platzer et al., 1997) including its flanking splice sites and either the HIV-1 splice acceptor 3 (SA3) or mutated variants of the HIV-1 SA5 as terminal splice acceptor. Base-pair substitutions within the polypyrimidine tract (PPT) or branchpoint sequence (BPS) in order to improve splice site efficiency are indicated in red. The strength of the splice acceptor sites was calculated applying the MaxEntScore algorithm for 3'ss (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html).

(C) RT-PCR analysis of HeLa-T4⁺ cells transiently transfected with the splicing reporter minigene harboring *ATM* exon 54 including its flanking splice sites and either the HIV-1 splice acceptor 3 (SA3) or mutated variants of the HIV-1 SA5 as terminal splice acceptor. 2.5 x 10⁵ cells were transfected with 1 µg of the splicing reporter, co-transfected with 0,1µg SVctat and 1 µg pXGH5 (hGH) expressing the human growth hormone mRNA to monitor equal transfection efficiency. Total RNA was isolated 30 h post transfection. mRNA was reverse transcribed using oligo(dT) as a primer. PCR was performed in the linear amplification range using vector specific primers as indicated. PCR products were separated on 6% polyacrylamide gels and stained with ethidium bromide. The splice products are schematically shown on the right.

(D) Quantification of the relative amount of the splicing products performed with the Lumi-Imager F1 (Roche Molecular Biochemicals) and the LumiAnalyst™ 3.1 software.

Results

Concomitantly, 35% and 31% of the reporter transcript remained unspliced (**Fig. 1C lanes 1 and 2; Fig.1D**). However, most efficient *ATM* 54 inclusion (59%) was achieved in the presence of the optimized acceptor SA5 opt (MaxEnt Score = 10.71) (**Fig. 1C lane 3 and Fig.1D**), although some unspliced transcripts (16%) as well as intron retaining transcripts (11%) could still be detected (**Fig. 1C lane 3 and Fig. 1D**). Interestingly, besides exon inclusion also exon skipping could be detected in the presence of the optimized 3'ss (**Fig. 1C lane 3 and Fig. 1D**). In contrast, in the presence of SA3 (MaxEnt Score = 7.59) as the terminal splice acceptor the reporter transcript was completely spliced and *ATM* exon 54 was skipped in 100% of the reporter transcripts (**Fig. 1C lane 4 and Fig. 1D**).

These results demonstrated that *ATM* exon 54 is not simply defined by its exon sequence and its flanking splice sites but additionally by the strength of the terminal 3' splice site. An optimized terminal splice acceptor was required for intron removal and inclusion of this human exon. However, low-level exon skipping and retention of intron 2 was still detectable suggesting that the natural sequence context of *ATM* exon 54 contains sequences that enhance recognition of this exon.

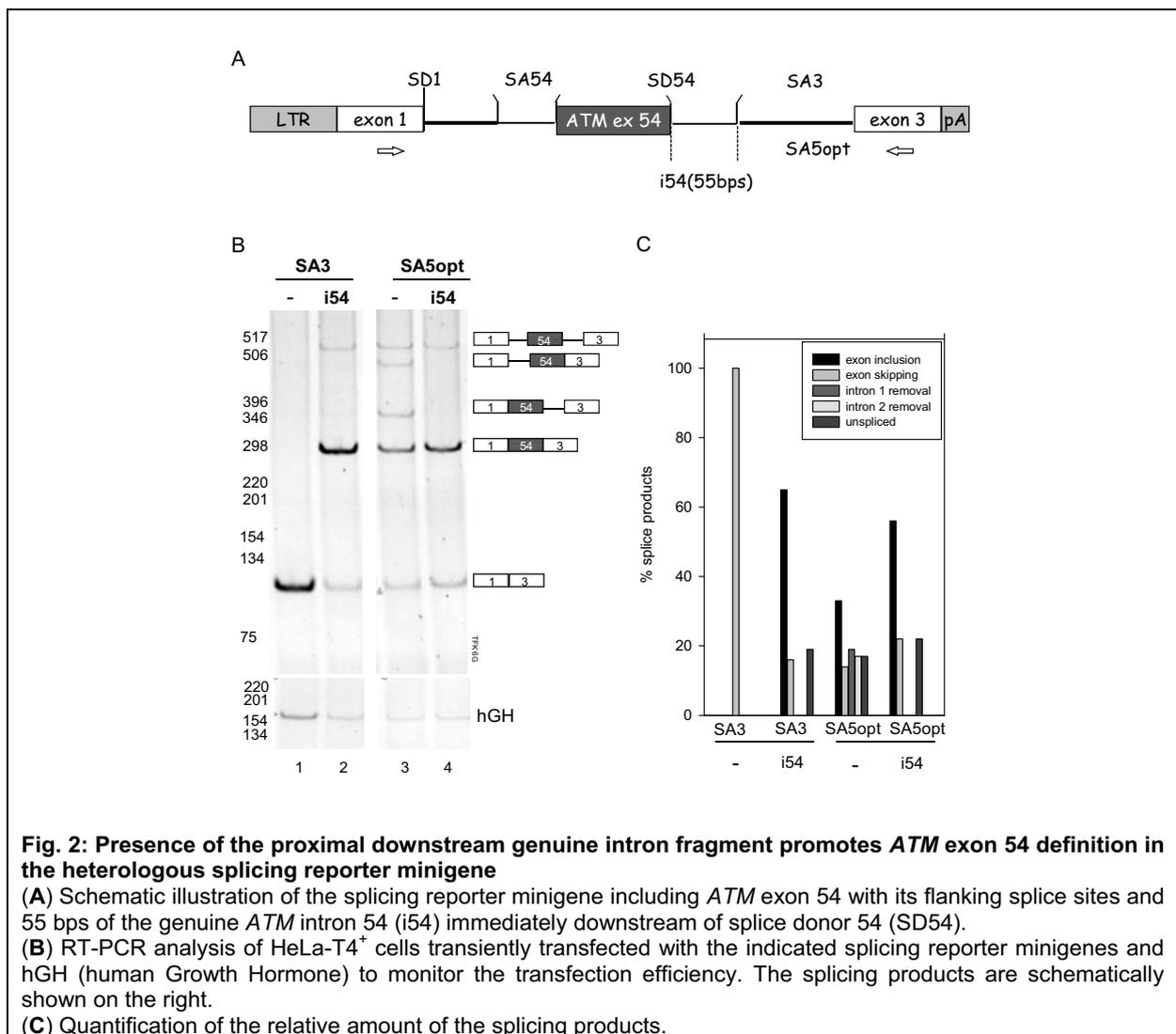
2.1.2. Adjacent genuine downstream intron segment promotes *ATM* exon 54 definition within the heterologous splicing reporter minigene

Given that the first intron was efficiently removed in the heterologous splicing reporter minigene whereas removal of the second intron was less efficient even in presence of an optimized terminal splice acceptor it seemed likely that recognition of *ATM* exon 54 with its weak splice donor (HBond score =12.3) is enhanced by the its natural downstream sequence missing in the splicing reporter. Because computational and experimental results have suggested that the intronic regions flanking constitutive exons might contain potential regulatory sequences with positional preference near the splice sites (Yeo et al., 2004 (374)) in an initial test experiment a short fragment (55 base pairs) of the adjacent genuine downstream intron sequence was included into the splicing reporter containing either SA5 opt or SA3 as terminal 3' splice site (**Fig. 2A**).

Remarkably, in the presence of its intronic sequence *ATM* exon 54 was efficiently recognized independently of the downstream splice acceptor (**Fig. 2B and C**). *ATM* exon 54 was efficiently included into the heterologous reporter construct harboring SA3 whereas in absence of the natural intron sequence adjacent to the splice donor of *ATM* exon 54 the human exon was skipped in this construct (**Fig. 2B lane 1 and 2**). Likewise, with insertion of the natural intronic sequence into the construct SA5 opt both intron 1 and intron 2 were more efficiently removed resulting in enhanced inclusion of *ATM* exon 54 (**Fig. 2B lane 3 and 4**).

Results

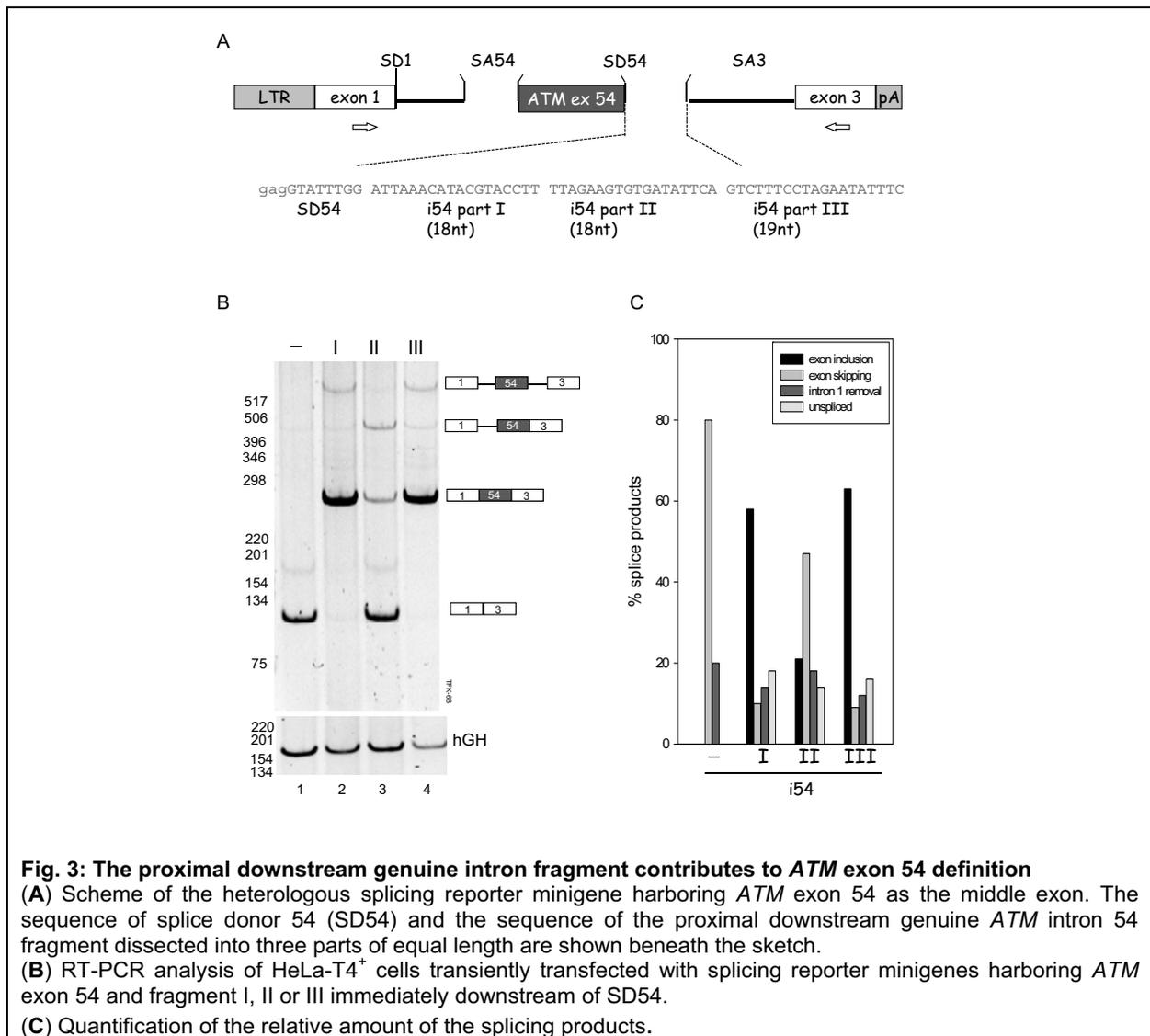
These data suggested that the genuine intron sequence adjacent to splice donor of *ATM* exon 54 contains a splicing regulatory element (SRE) enhancing recognition of the weak splice donor and promoting *ATM* exon 54 definition in the heterologous splicing minigene. Furthermore, these results demonstrated that *ATM* exon 54 definition initially occurs by cross-exon splicing complex formation in spite of the short introns in the heterologous minigene as evidenced by the fact that the strength of 3' splice site within the reporter was negligible for *ATM* exon 54 inclusion in the presence of the natural downstream exon-flanking intron sequence contributing to definition of this exon.



Results

2.1.3. The adjacent genuine downstream intron sequence contributes to *ATM* exon 54 definition in a sequence specific manner

To localize the putative splicing regulatory element (SRE) within the intronic sequence adjacent to the splice donor of *ATM* exon 54 the intron fragment was dissected into three parts of equal length (Fig. 3A). To determine the impact of each part on *ATM* exon 54 definition the respective sequence segment was inserted immediately adjacent to splice donor 54 (SD54) within the SA3 containing splicing reporter.



Results

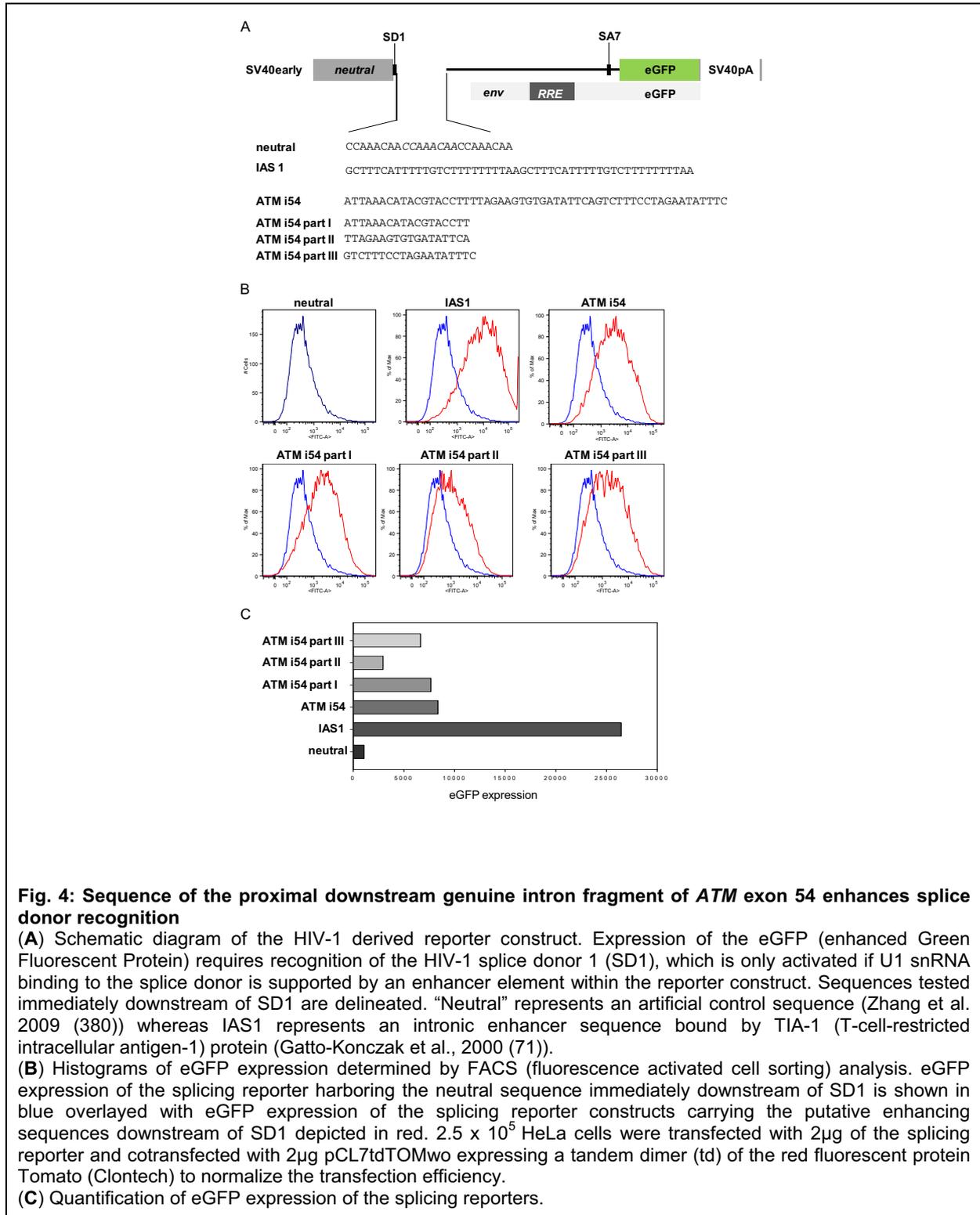
Analysis of *ATM* exon 54 recognition by RT-PCR in transiently transfected HeLa cells revealed efficient *ATM* 54 inclusion in the heterologous splicing reporter if either part I or part III of the *ATM* intron fragment was positioned adjacent to SD54 (76% and 68% of the reporter transcripts, respectively). In contrast, the insertion of part II at this position predominantly resulted in skipping of *ATM* exon 54 (55%) although exon inclusion was improved (24%) compared to the control construct containing *ATM* exon 54 with its flanking splice sites only (**Fig. 3B and C**).

These data revealed several interesting points. Firstly, the insertion of a genuine intron segment of only 18-19 base pairs caused a clear-cut shift from skipping of *ATM* exon 54 to inclusion of this exon into the reporter transcript. Secondly, although the sequences of part I and part III were entirely different both allowed efficient *ATM* exon 54 definition in the heterologous splicing reporter. Nevertheless, the presence of intron part II immediately downstream of SD54 mainly caused exon skipping indicating a sequence specific effect of the intron segments on *ATM* exon 54 inclusion. Thirdly, as the presence of part II within the complete *ATM* intron segment was compatible with efficient *ATM* exon 54 inclusion (Fig. 2B lane B) it appeared that the position of a specific sequence immediately adjacent to the splice donor was decisive for *ATM* 54 exon definition suggesting that these sequences affect U1 snRNP binding to this splice donor site.

2.1.4. The sequence of the proximal downstream genuine intron fragment of *ATM* exon 54 enhances splice donor recognition

U1 snRNP binding to a splice donor (SD) can be monitored using a sub-genomic HIV-1 glycoprotein (*Env*) expression vector whose unstable glycoprotein RNA can be protected from degradation by sufficient RNA duplex formation between U1 snRNA and the 5' splice site upstream of the *env* open reading frame (ORF) (Kammler et al., 2001 (152)). The published reporter construct has been further modified by replacement of the HIV-1 SD4 by SD1 and an *in frame* substitution of the region downstream of the HIV-1 splice acceptor (SA7) for the open reading frame of eGFP (enhanced green fluorescent protein). Thus, in this modified construct, expression of eGFP correlates with the recognition of the 5'ss by the spliceosomal U1 snRNP. Additionally, to be able to analyze intronic enhancer elements the enhancer sequence upstream of SD1 was replaced by three repeats of a *neutral* sequence predicted to have no effect on splicing (Zhang et al., 2009 (380)) (**Fig. 4A**).

Results



In order to functionally test the *ATM* intron segment and its fragments for their ability to support splice donor recognition the sequences were inserted immediately downstream of SD1. HeLa cells were transiently transfected with the resulting constructs and eGFP expression was measured by flow cytometry. To normalize eGFP expression to transfection

Results

efficiency a construct which constitutively expresses the red fluorescent protein Tomato was co-transfected and eGFP expression of 10,000 Tomato expressing cells was assessed. Using reporter constructs harboring three repeats of a *neutral* sequence upstream and downstream of SD1 the baseline eGFP expression level of the reporter constructs was determined. As a positive control a reporter harboring a known intronic splicing enhancer (ISE) (IAS1, TIA-1 binding site, Gatto-Konczak et al., 2000 (71)) immediately downstream of SD1 was used showing a 25-fold increase in eGFP expression compared to the *neutral* control (**Fig. 4B**).

The presence of the complete *ATM* intron segment (*ATM* i54) immediately downstream of SD1 increased eGFP expression 8-fold in comparison with the *neutral* sequence on the same position, achieving 32% of the induction by the known ISE. Insertion of *ATM* intron part I and part III enhanced eGFP expression 7-fold and 6-fold respectively, whereas positioning part II adjacent to SD1 resulted in a 3-fold increase in eGFP expression only.

These results demonstrated that in spite of the heterologous sequence context the *ATM* intron segment and its fragments supported splice donor recognition by U1 snRNA positioned immediately downstream of a splice donor site. However, the *ATM* intron sequences were less effective than the TIA-1 (IAS1) binding site. The observation that *ATM* i54 part II showed the slightest contribution to splice donor recognition by U1 snRNA whereas part I and III were as effective as the complete *ATM* intron segment was consistent with the analysis of the impact of the *ATM* intron fragments on *ATM* exon 54 recognition positioned immediately downstream of SD54 in the heterologous 3-exon-2-intron splicing reporter approving the assumption that that the sequences positioned immediately downstream of SD54 affect recognition of the splice donor site by U1 snRNA and thereby contribute to *ATM* exon 54 definition in the heterologous 3-exon-2-intron splicing reporter.

2.1.5. Identification of proteins bound to the intronic *ATM* fragments

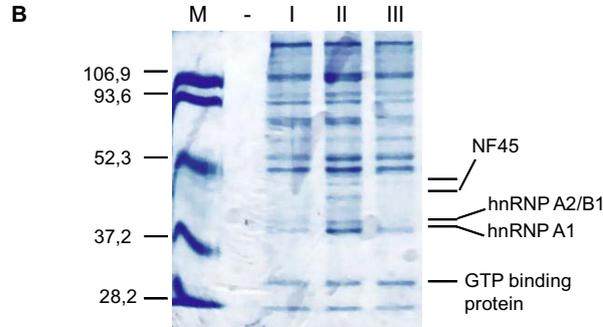
Because splicing regulatory sequences function by recruiting protein factors that activate or suppress splice site recognition by various mechanisms the question raised of whether the differential effects of the *ATM* intron sequence fragments on splice donor recognition were caused by a sequence specific prevalence or loss of distinct associated protein factors.

To this end, three different target RNA sequences for RNA affinity chromatography comprising *ATM* splice donor 54 directly followed by either *ATM* intron 54 part I, part II or part III were synthesized by *in vitro* transcription (**Fig. 5A**).

Results

A

SD 54	ATM i 54	
gagGTATTTGG	ATTAAACATACGTACCTT	part I
gagGTATTTGG	TTAGAAGTGTGATATTCA	part II
gagGTATTTGG	GTCTTTCCTAGAATATTTTC	part III



```
>gi|14043072|ref|NP_112533.1| heterogeneous nuclear ribonucleoproteins
A2/B1 isoform B1 [Homo sapiens]
HEKTLETVPLERKKREKEQFRKLF IGGLSFETTEESLRNYEQWQKLTDCVVHRDPASKRSRGFGVTFSSHAE
VDAAMAAARPHSIDGRVVEPKRAVAREESGKPGAHVTVKKLFVGGIKEDTEHHLRDYFEEYKIDTIEIITDRQ
SGKKRGFGVTFDDHDPVDKIVLQKYHTINGHNAEVRKALSRQEMQEVQSSRSRGGNFGFGDSRGGGGNFGPG
PGSNFRGGSDGYGSGRGRFGDGYNGYGGGPGGGNFGGSPGYGGGRGGYGGGGPGYGNQGGGYGGYDNYGGGNYG
SGNYNDFGNYNQQPSNYGPMKSGNFGGSRNMGGPYGGGNYGPGSGSGSGGYGGRSRY
```

```
>gi|47939618|gb|AAH71945.1| Heterogeneous nuclear ribonucleoprotein A1
[Homo sapiens]
MSKSESPKEPEQLRKLFIGGLSFETTESLRSHFQMGTLTDCVVHRDPNTKRSRGFGVTFYATVEEVDAAAMNA
RPHKVDGRVVEPKRAVAREESQRPGAHLTVKKIFVGGIKEDTEHHLRDYFEEYKIDTIEIITDRQSGKKRGF
AFVTFDDHDSVDRIVIQKYHTVNGHNCVVRKALSQEMASASSQRGRSGSGNFGGGGCGGNDNFGRGGNF
SGRGGFGGSRGGGGYGGSDGYNGFGNDGNSNFGGGGSYDFGNYNQSSNFGPHKGGNFGGRSSGPGYGGGQYF
AKPRNQGGYGGSSSSSSYGSRRF
```

```
gi|532313|gb|AAA20993.1| NF45 protein [Homo sapiens]
HRGDRGRGRGGFRGSRGGPGGFRPFVPHIPDFYLCENAFPRVKPAPDETSEALLKRNQDLAPNSAE
QASILSLVTKINNVIDNLIVAPGTFEVQIEEVRQVGSYKKGTHHTTGHNVADLVVILKILPTLEAVAALGN
KVVESLRAQDPSEVLTHLTNETGFEISSDQATVKILITVPPNLRKLDPELHLDIKVLQSAALAAIRHARU
FEENASQSTVKVLRLLKDLRIRFPGFEPFTPWILDLLGHYAVHNNPTRQPLALNVAYRCLQILAAGLF
LPGSVGITDPCESGNFRVHTVMTLEQQDHVCYTAQTLVRLSHGGFRKILGQEGDASYLASEISTWDGVI
VTPSEKAYEKPEKKEGEEEEENTERTTSRRGRKKGNSGVTFPSLLFLPKGKTGA
```

```
>gi|4092054|gb|AAC99400.1| GTP binding protein [Homo sapiens]
HAAQGEPEQVQFKLVLVGDGGTGKTFVKKRHLTGEFEKKYVATLGVVHPLVVFHTNRGPIKFNVDVDTAGQE
KFGGLRDGYIQAQCAIIFDVTSRVTKNVPNWHRDVLRVCENIPVLCGNKVDIKDRKVKAKSIVFHR
KKNLQYYDISAKSNYNFEKPFLLWARKLIGDPNLEFVAMPCLAPPEVVDPALAAQYEHDLVAQTALP
DEDDDL
```

Fig.5: HnRNP A2/B1 and hnRNP A1 bind ATM intron fragment II

(A) RNA affinity chromatography using RNA target sequences shown above. RNA sequences were synthesized by *in vitro* transcription with T7 polymerase. Agarose beads covalently linked to 2000pmol RNA were incubated with HeLa cell nuclear extract. Bound proteins were eluted with SDS-sample buffer and analyzed by 10% SDS-PAGE with Coomassie-blue staining.

(B) Labeled bands were isolated from the gel and proteins digested by trypsin. Resulting peptides were analyzed by mass spectrometry (BMFZ, HHU). Amino acid sequences of peptides leading to the identification of hnRNPA2/B1, hnRNPA1, NF45 and GTP binding protein are shown in red.

Results

The RNA molecules were covalently linked to dihydrazide agarose beads and incubated with HeLa nuclear extracts. Proteins that remained tightly bound to the RNA targets after washing were separated by SDS-PAGE and stained with Coomassie-blue.

The Coomassie stained gel revealed differences in the abundance of distinct protein bands between the RNA targets. In particular, two strong bands were observed at a size about 35 kDa to be more abundantly associated with the RNA harboring *ATM* intron 54 part II immediately downstream of SD54 whereas two bands in the range of about 25 to 35 kDa bound in equal magnitude to each RNA sequence (**Fig. 5B**). To identify these proteins bands containing proteins were excised from the SDS polyacrylamide gel, digested with trypsin and sequenced by mass spectrometry (BMFZ, HHU). Sequenced peptides were compared with the MASCOT database (MASCOT MS/MS ions search; www.matrixscience.com). Finally, several independent sequence hits led to the identification of heterogenous nuclear protein (hnRNP) A2/B1 [gi14043072] and hnRNPA1 [gi47939618], NF45 [gi47939618], and GTP binding protein [gi532313].

hnRNPA1, hnRNPA2/B1 and NF45 strongly bound to the RNA sequence containing *ATM* intron 54 part II adjacent to SD54, whereas these proteins bound in a considerably lower amount to the RNA sequences representing part I or part III. In contrast, the GTP binding nuclear protein Ran bound in equal amounts to each RNA sequence (**Fig. 5B**).

hnRNP A1 is known to act as a regulator of exon recognition (Mayeda et al., 1999 (208)) by interfering with U1 snRNP binding to a splice donor site (Eperon et al., 2000 (78)). hnRNP A1 and A2 have been reported to bind cooperatively to an intron splicing silencer (ISS) immediately downstream of the splice donor of *SMN2* (Survival Motor Neuron 2, MIM #601627) exon 7 causing skipping of this exon (Hua et al., 2008 (138)).

Therefore, it seemed likely that binding of hnRNP A1 and A2/B1 to *ATM* intron 54 part II affects binding of U1 snRNP to the adjacent splice donor if this intron segment was artificially positioned immediately downstream of a 5' splice site. This might provide an explanation for predominant skipping of *ATM* exon 54 in the heterologous 3-exon-2-intron splicing reporter when *ATM* intron 54 part II was positioned adjacent to SD54 (**Fig. 3B lane 3**). Interestingly, placing of part II immediately downstream of the splice donor was crucial for the observed exon skipping because insertion of *ATM* intron 54 part I or part III between the splice donor and part II resulted in efficient exon inclusion (data not shown) indicating that the natural arrangement of the flanking intron sequence buffers this exon against skipping.

2.1.6. Results from *ATM* exon 54 are applicable to *ATM* exon 9

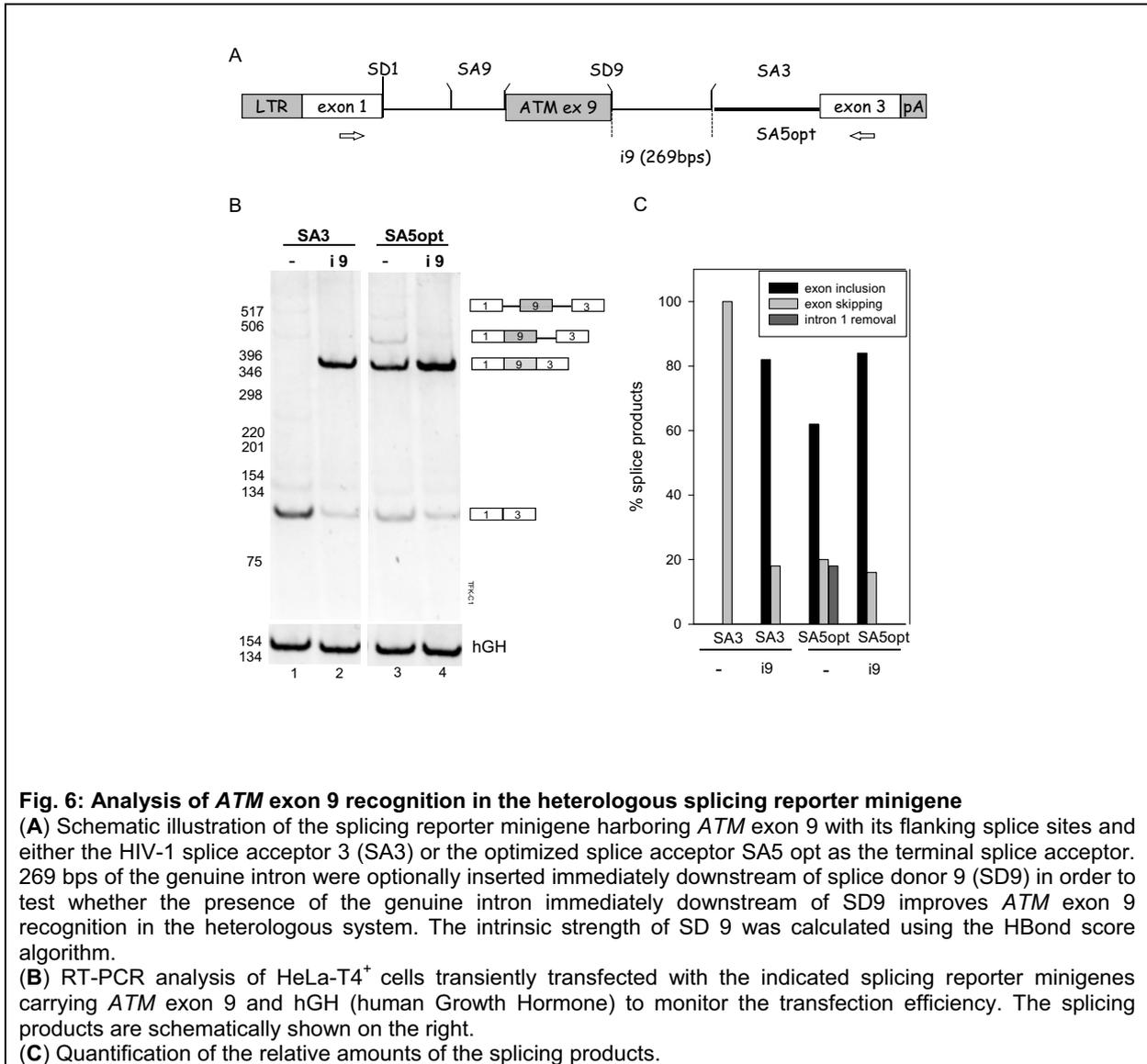
To exclude that the observed dependency of central exon recognition on the terminal splice acceptor and on the downstream flanking intron sequence in the heterologous splicing reporter was solely valid for *ATM* exon 54 this exon was replaced by *ATM* exon 9 also harboring a weak splice donor (HBond score = 12.3).

Again, this exon with its flanking splice sites (98 base pairs of the original upstream 3'ss and 11 base pairs of the original downstream 5'ss) was inserted as the central exon into the heterologous splicing minigene containing either the optimized acceptor SA5 opt or the less efficient acceptor SA3 as the terminal 3' splice site. To simultaneously test the effect of the genuine downstream intron sequence on *ATM* exon 9 inclusion in the heterologous minigene about 200 base pairs of the genuine downstream intron sequence were introduced into both minigenes (**Fig. 6A**). This number of base pairs was chosen because in computational searches for single motifs distinctive to the flanks of exons these could still be detected within a 200-nucleotide range (Xiao et al., 2007 (370), Ke et al. 2010 and 2011 (160,161)).

Transfection of HeLa cells followed by RT-PCR revealed that in the presence of the less efficient acceptor SA3 the genuine intronic sequence immediately downstream of SD9 was necessary for productive *ATM* exon 9 recognition within the heterologous splicing reporter minigene (**Fig. 6B, lanes 1 and 2**). However, within the heterologous splicing reporter harboring the optimized splice acceptor SA5 opt *ATM* exon 9 was efficiently recognized (61% of the reporter transcripts) even in the absence of the genuine downstream intron sequence (**Fig. 6B lane 3 and 6C**). Nevertheless, the presence of the genuine downstream intronic sequence enhanced exon recognition as evident by the efficient removal of the second intron (85%) (**Fig. 6B lane 4 and 6C**).

Thus, the results from *ATM* exon 9 and 54 were consistent and demonstrated efficient middle exon inclusion in the absence of additional downstream intronic splicing regulatory sequences only if the optimized 3' splice site SA5 opt was present, providing an adequate system for functional testing of splice site mutations on exon recognition. The presence of the genuine intron sequence immediately downstream of the exon associated splice donor further improved definition of both exons independently of the 3' splice site within the reporter system recommending the insertion of a test exon with its flanking downstream intron sequence of at least 200 nucleotides wherever possible.

Results

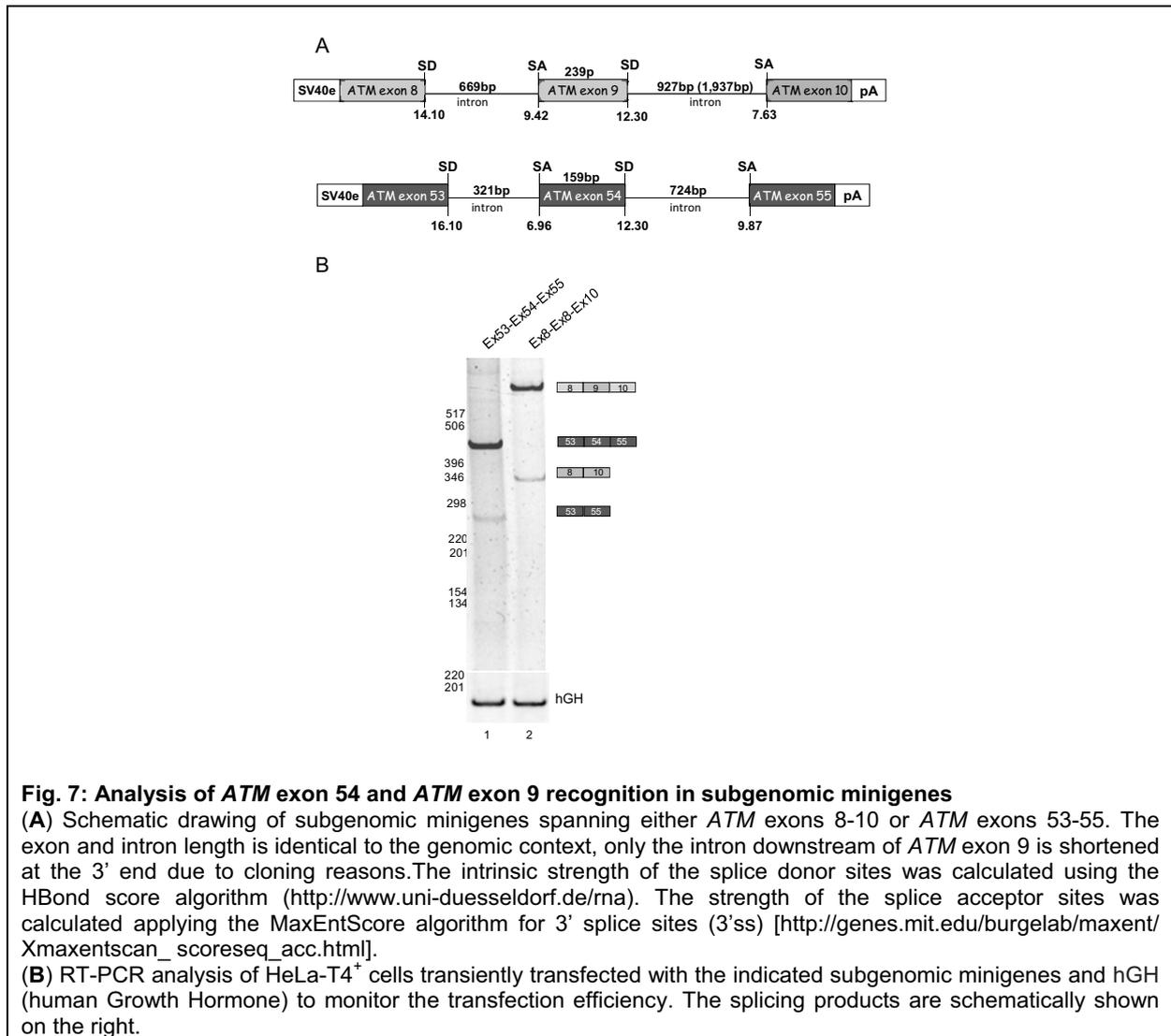


2.1.7. An extended genomic context is negligible for *ATM* exon 54 and *ATM* exon 9 recognition

Nonetheless, despite the presence of the genuine downstream intronic sequence residual exon skipping was observable for both exons within the heterologous minigenes raising the question of whether an extended genomic context comprising of the natural flanking exons and the entire flanking intronic sequence would culminate in perfect recognition of the *ATM* exons.

To clarify whether *ATM* exon 9 and exon 54 would be recognized more efficiently in their natural, i.e. extended genomic context, minigenes spanning *ATM* exons 8-10 and exons 53-55 were constructed (Figure 7A).

Results



Analysis of the splicing pattern by RT-PCR revealed that residual skipping of both exons was still detectable as seen in the heterologous setting (confer **Fig. 7B with 6B**). These results pointed to a complex splicing regulation of the *ATM* gene which has also been observed for the 5' UTR (untranslated region) undergoing extensive alternative splicing (Savitsky et al., 1997 (278)).

Nevertheless, an almost identical splicing outcome of these exons embedded either in their subgenomic context or in the heterologous splicing reporter indicated that definition of these exons predominantly relies, in addition to their intrinsic properties on their splice sites and flanking intron regions rather than on the wider genomic context. This indicates that insertion of these exons with part of their intronic flanking regions into heterologous splicing minigene provides a reliable model for investigating the effect of a splicing mutation on exon recognition.

2.2. Functional characterization of putative pathogenic splice donor mutations

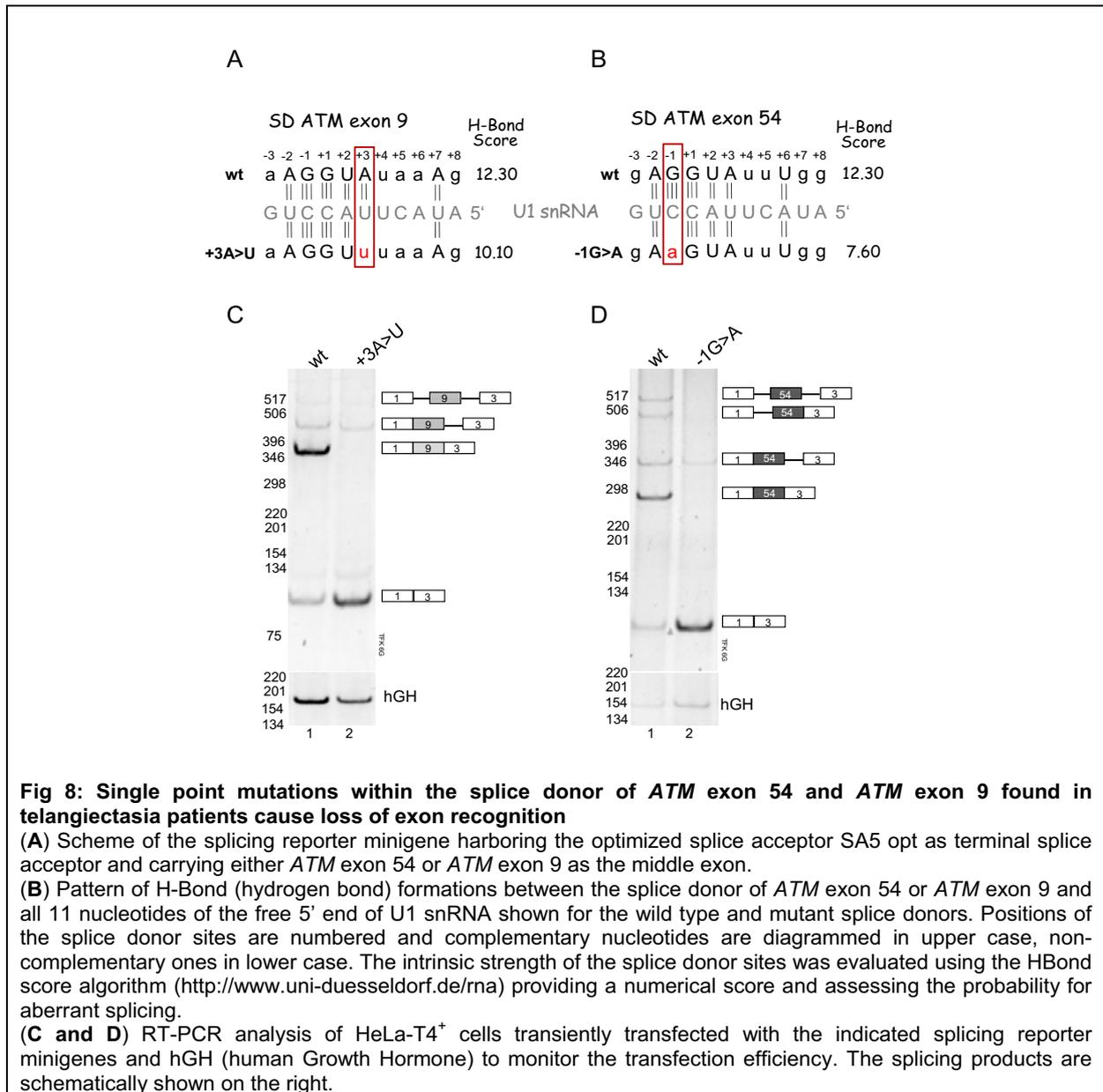
Single base-pair substitutions in human splice donor sites weakening RNA duplex formation between U1 snRNA and the splice donor commonly cause exon skipping or activation of cryptic splice sites resulting in loss of information for the encoded protein or causing a frameshift in the open reading frame usually generating non-functional transcripts with premature translation termination. Following up previous work in our group establishing a hydrogen bond model for the complementarity between the free 5' end of U1 snRNA and 5' splice sites predicting the probability of aberrant splicing for human splice donor mutations functional testing in heterologous minigenes allows to validate whether a splicing mutation causes skipping of the affected exon.

2.2.1. Single point mutations within the splice donor of *ATM* exon 54 and *ATM* exon 9 found in ataxia telangiectasia patients cause loss of exon recognition

Concerning the *ATM* exons 54 and 9, in patients suffering from ataxia telangiectasia (MIM # 208900) genomic sequencing identified single base-pair substitutions within the splice donor site of these exons (Prof. D.Schindler, Würzburg): in the splice donor sequence of *ATM* exon 54 a G>A mutation at position -1 severely decreasing the complementarity to U1 snRNA lowering the HBond score from 12.30 to 7.60 was found (**Fig. 8B**) whereas in the splice donor sequence of *ATM* exon 9 an A>G mutation at position +3 reducing the HBond score from 12.30 to 10.10 was detected (**Fig. 8A**).

Because the severe decrease in the HBond score suggested a high probability for aberrant splicing for both mutations although the consensus sequence of mammalian 5'ss (AG/GTRAGT (R=purin; A oder G)) allows any purin base at position +3 both mutations were introduced into the heterologous splicing reporter minigene harboring the affected *ATM* exon and the optimized terminal splice acceptor SA5 opt. RT-PCR analysis of transfected HeLa cells demonstrated that both mutations cause skipping of the affected *ATM* exon (**Fig. 8C and D**) providing evidence for the pathogenicity of these mutations.

Results



2.2.2. The *RAD51C* c.904+5G>T mutation in familial breast and ovarian cancer pedigree causes loss of *RAD51C* exon 6 recognition

In order to be able to reliably advise patients about their health risks evaluation of the expressivity and penetrance of a splicing mutation is necessary which could be done by studying segregation of the mutation in patients' families or through larger population studies.

In a collaborative project (Meindl et al., 2010 (211)) screening the *RAD51C* gene in 1.100 unrelated affected individuals from pedigrees with gynecological cancers that were negative for mutations in the breast cancer susceptibility genes *BRCA1* and *BRCA2* (MIM #604370 and #612555) 14 germline sequence alterations in *RAD51C* including 2 splice donor mutations were detected.

Results

The first splice donor mutation identified in the 5' splice site of *RAD51C* exon 6 (c.904+5G>T) affected an evolutionarily conserved position and was predicted to severely reduce the complementarity between the U1 snRNA and this splice donor as indicated by a decrease in the HBond score from 15.8 to 10.1 (**Fig. 9A**). An extended family tree with individuals depicted being at least 30 years of age showed high frequency of this mutation in the first degree relatives and siblings with both breast and ovarian cancers (**Fig. 9B**).

To validate that the *RAD51C* c.904+5G>T mutation causes aberrant splicing the *RAD51C* exon 6 with flanking 225- and 158-bp intronic sequences was amplified from genomic control DNA and inserted into the heterologous splicing reporter construct minigene (**Fig. 9C**). The c.904+5G>T mutation was introduced by PCR mutagenesis. Transfection of HeLa cells followed by RT-PCR demonstrated that the c.904+5G>T mutation resulted in loss of *RAD51C* exon 6 recognition (**Fig. 9D**).

As cells carrying the germline mutation were not available RT-PCR was performed on mRNA isolated from paraffin-embedded tumor samples from two carriers of this mutation. The amplified RT-PCR product showed that *RAD51C* exon 6 was skipped (Meindl et al., 2010 (211)). Finally, the pathogenic nature of this splice donor mutation was demonstrated by sequencing of DNA extracted from paraffin-embedded samples revealing that the loss of the wild-type allele had occurred independently in the breast and the ovarian cancer tissues in two affected individuals (**Fig. 9E** taken from Meindl et al., 2010 (211)).

Thus, in case of the *RAD51C* c.904+5G>T mutation the meaningful combination of *in silico* prediction, functional testing within the heterologous splicing reporter minigene, segregation analysis and the availability of tumor samples clearly confirmed the pathogenicity of this mutation.

2.2.3. The c.145+1G>T mutation within the splice donor site of *RAD51C* exon 1 resulted in enhanced production of non-functional *RAD51C* transcripts

The second splice donor mutation disrupted the canonical GT dinucleotide within the splice donor of *RAD51C* exon 1 (c.145+1G>T) and was found in a family with three sisters affected by breast or ovarian cancers (**Fig. 10A and B**) (Meindl et al., 2010 (211)). Direct analysis of the *RAD51C* splicing pattern in peripheral blood leukocytes from two heterozygous mutation carriers and comparison with normal controls by RT-PCR with primers located in *RAD51C* exon 1 and 3 amplified three transcripts which were identified as *RAD51C*-001, *RAD51C*-008 and *RAD51C*-009 by sequencing (**Fig. 10C and D** and **Fig. 11**) (transcripts and nomenclature according to the Ensembl genome browser).

Results

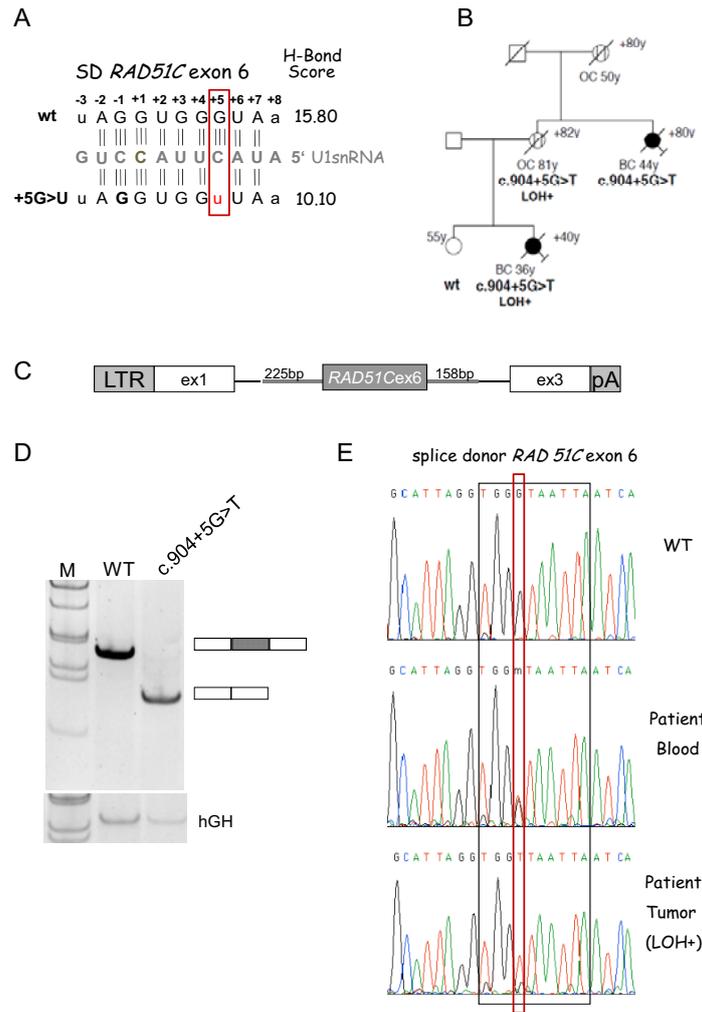


Fig.9: The *RAD51C* c.904+5G>T mutation in a familial breast and ovarian cancer pedigree causes loss of *RAD51C* exon 6 recognition

(A) Pattern of H-Bond (hydrogen bond) formations between the splice donor of *RAD51C* exon 6 harboring the c.904 5G>T mutation and all 11 nucleotides of the free 5' end of U1 snRNA shown for the wild type and mutant splice donor. Positions of the splice donor site are numbered and complementary nucleotides are diagrammed in upper case, non-complementary ones in lower case. The intrinsic strength of the splice donor sites was evaluated using the HBond Score algorithm (<http://www.uni-duesseldorf.de/rna>) providing a numerical score and assessing the probability for aberrant splicing.

(B) *RAD51C* c.904+5G>T mutations in a familial breast and ovarian cancer pedigree. Individuals with breast cancer (BC) are shown as filled circles, females with ovarian cancer (OC) as streaked circles; OP, surgery. Disease and age in years (y) at first diagnosis is given underneath the symbol, current age or age at death (+) above it. All affected individuals with breast or ovarian cancer not tested for germline mutations in *RAD51C* were deceased or refused testing. Carriers of *RAD51C* mutations are shown with their specific *RAD51C* mutation, whereas individuals who tested negative for the mutation in the specific pedigree are depicted as wild-type (WT). In addition, LOH data (+ for loss of the WT allele, - for a retained WT allele) is shown for the individuals for whom formalin-fixed, paraffin-embedded (FFPE) tissue samples of the tumor(s) could be analyzed (adapted from Meindl et al., 2010 (211)).

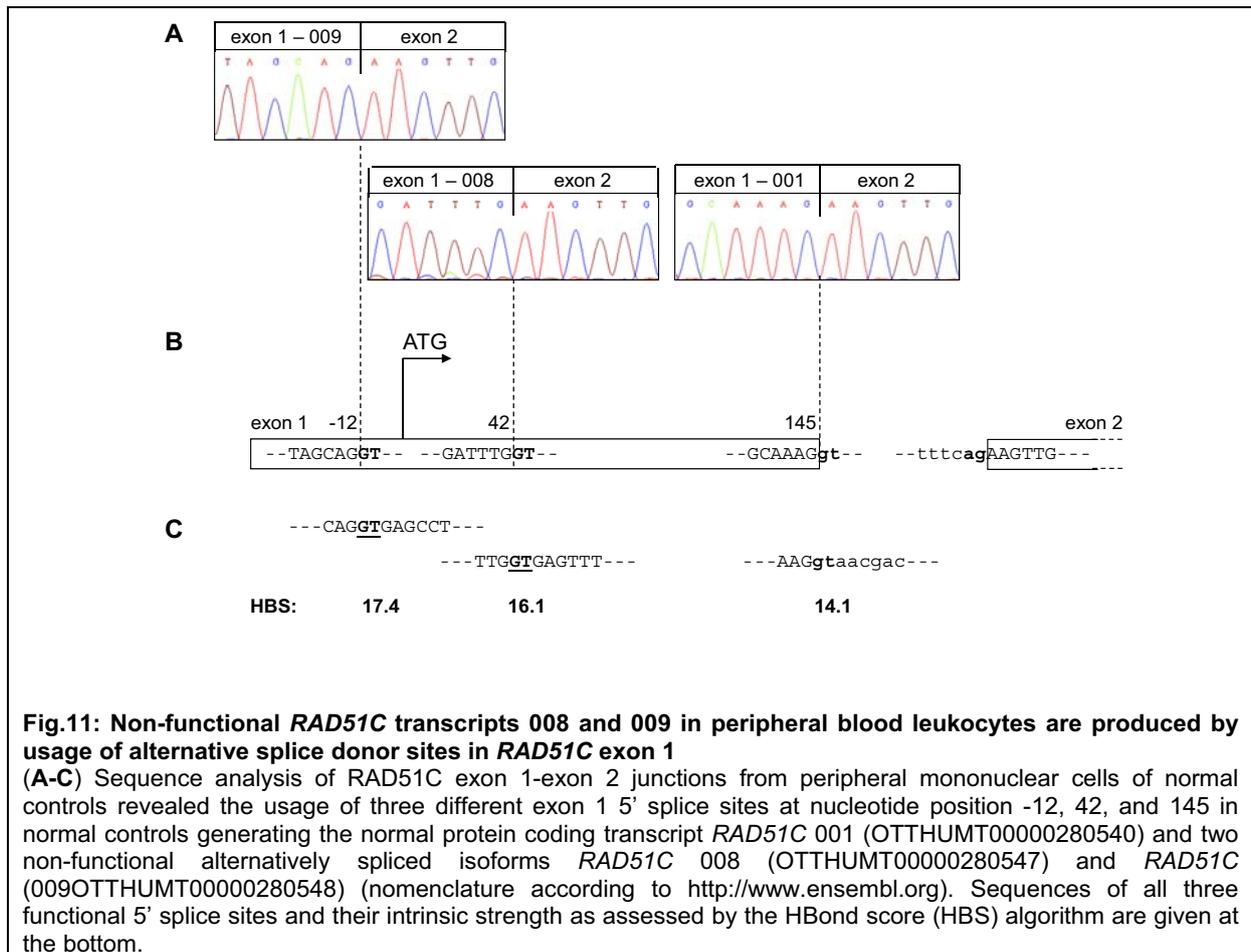
(C) Scheme of the heterologous splicing reporter minigene including *RAD51C* exon 6 and 225 bps of the genuine upstream and 158bps of the genuine downstream intron.

(D) RT-PCR analysis of HeLa-T4⁺ cells transiently transfected with the indicated wild type and mutant splicing reporter minigene and hGH (human Growth Hormone) to monitor the transfection efficiency. The splice products are schematically shown on the right

(E) LOH analyses in tumor samples. The wild type *RAD51C* sequence is shown in the upper row. The sequencing results from patients' germline DNA with the *RAD51C* c.904+5G>T and the corresponding tumor DNA with loss of the wild type allele is shown in the middle and lower row (cited from Meindl et al., 2010).

Results

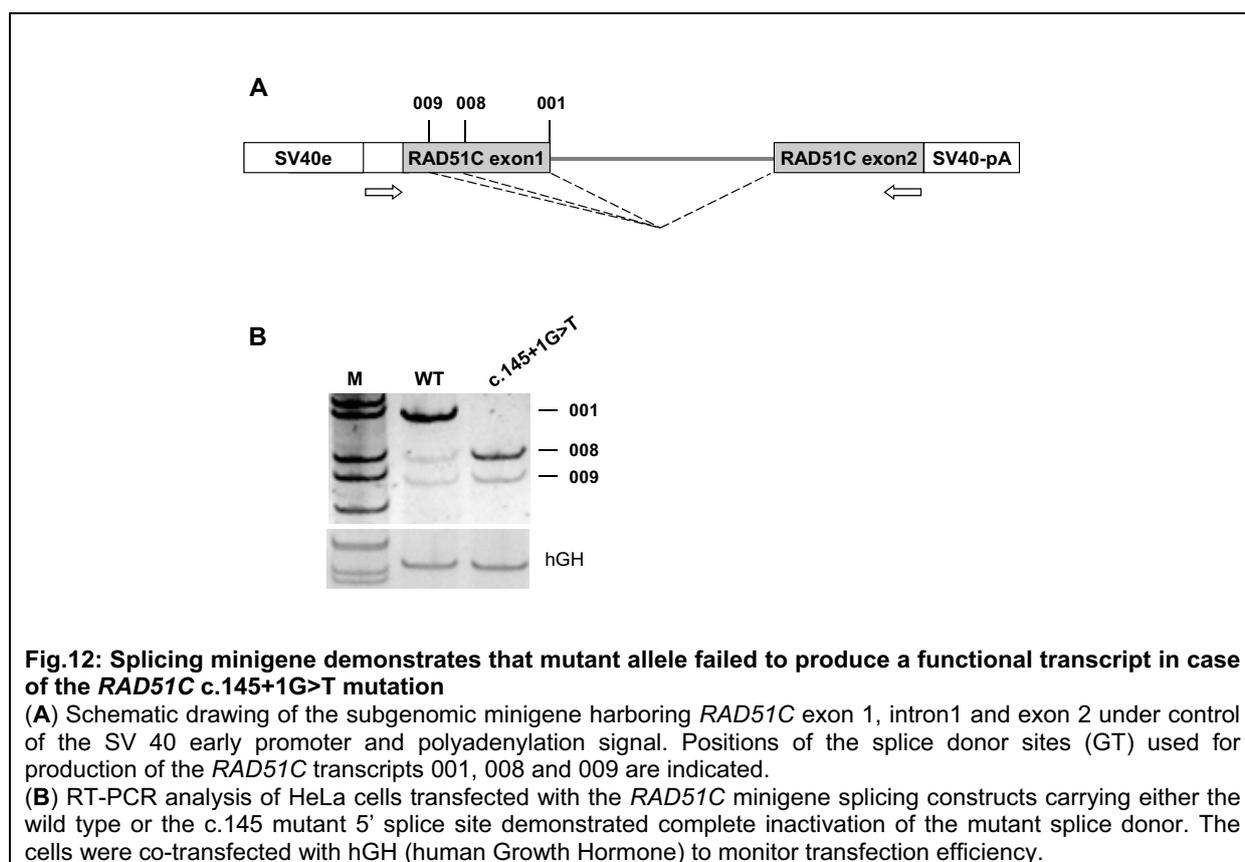
Relative quantification of these transcripts revealed reduced expression of the normal protein-coding *RAD51C*-001 and increased expression of the non-functional *RAD51C*-008 transcript in both mutation carriers, whereas the levels of the non-functional *RAD51C*-009 transcripts were unchanged compared to normal controls (**Fig. 10E and F**). The latter two transcripts were produced by usage of alternative splice donor sites within *RAD51C* exon 1 with intrinsic strengths of 17.4 and 16.1, respectively, as predicted by the HBond algorithm for 5' splice sites (**Fig. 11**). To prove that the normal *RAD51C* transcript was solely expressed from the wild-type allele in the heterozygous leukocytes, *RAD51C* exon 1, intron 1 and exon 2 was amplified from normal human control DNA and inserted into a splicing construct. The c.145+1G>T mutation was introduced by PCR mutagenesis. RT-PCR analysis after transfection of HeLa cells with the wild-type splicing minigene revealed that usage of the exon 1 splice donor was comparable to normal controls (**Fig. 12**). In contrast, the RT-PCR analysis of the c.145+1G>T splicing minigene showed complete inactivation of this mutant 5' splice site and increased transcript levels from the upstream proximal 5' splice site (transcript 008).



Results

Finally, the pathogenic nature of this 5' splice site mutation was demonstrated by the loss of the wild-type allele in the cancer tissue of the surviving subject with breast cancer (Meindl et al., 2010 (211)).

Therefore, the monoallelic *RAD51C* c.145+1G>T splice donor mutation was clearly associated with risk of breast and ovarian cancer. Because this germline mutation was present in one allele only, the minigene construct provided a valuable tool for characterizing its effect on splicing separately from the second allele.



2.3. The impact of a homozygous micro-deletion in *BRCA2* exon 6 on splicing

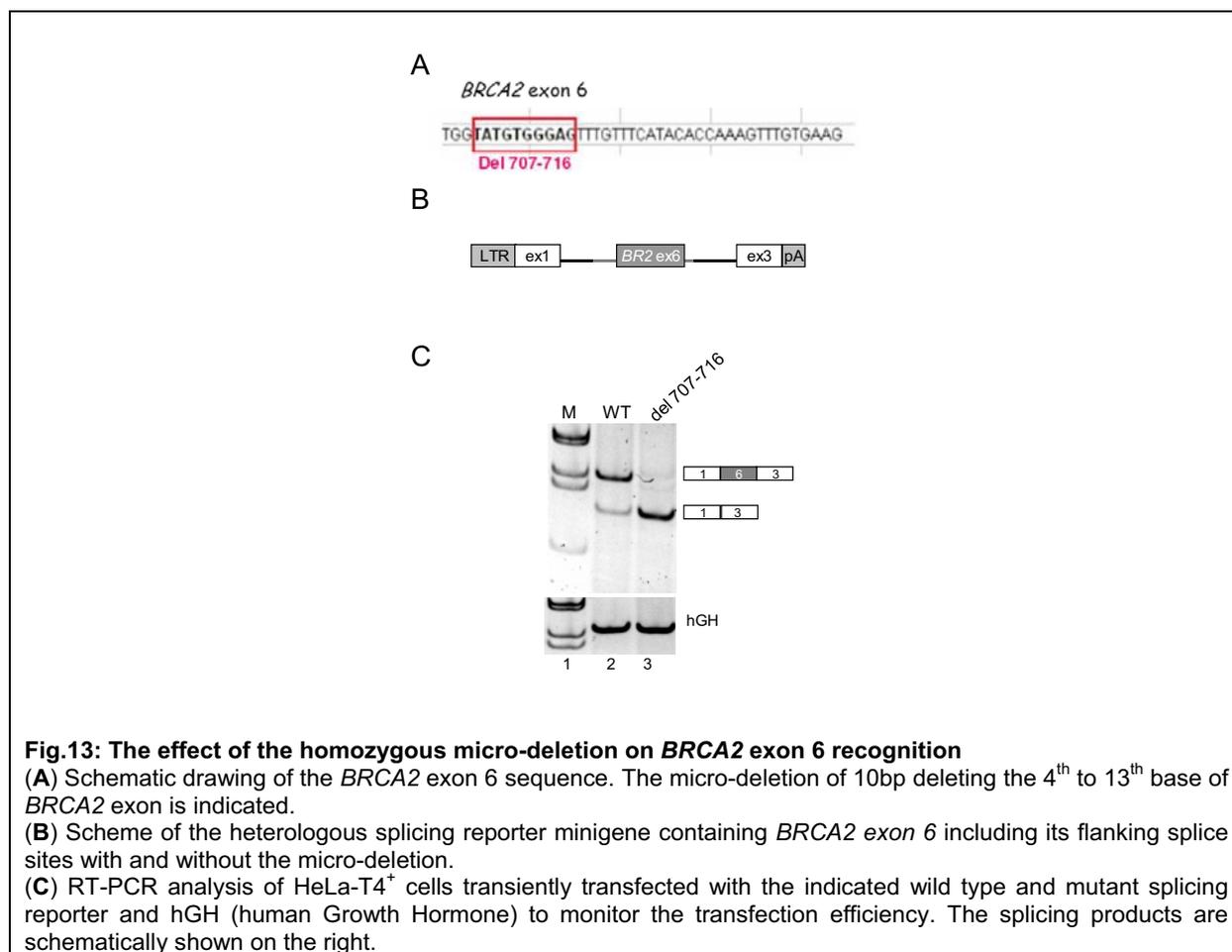
It has been well-established that heterozygous carriers of *BRCA2* mutations inherit a high risk of developing breast cancer (up to 85%) and other cancers such as ovarian and pancreatic. More recently, it has been discovered that germline inheritance of two defective copies of *BRCA2* can lead to Fanconi anemia (FA) (Howlett et al., 2002 (137)), a complex disorder characterized by congenital abnormalities, progressive bone marrow failure, and cancer susceptibility. Likewise, a homozygous micro-deletion of 10 bps in *BRCA2* exon 6 has

Results

been detected in siblings diagnosed with Fanconi Anemia in a genome wide linkage scan (Prof. R. Schneppenheim, Hamburg, unpublished data). This homozygous deletion was confirmed by direct sequencing of the *BRCA2* cDNA in both patients additionally identifying four different splice products surplus to the predicted RNA-product in the patients (Prof. R. Schneppenheim, Hamburg, unpublished data).

2.3.1. The effect of the homozygous micro-deletion on *BRCA2* exon 6 recognition

To analyze the effect of the identified 10bp deletion in *BRCA2* exon 6 on recognition of this exon *BRCA2* exon 6 with and without the 10bp deletion including its flanking splice sites was inserted into the heterologous splicing reporter minigene construct (**Fig. 13B**). Transfection of HeLa cells followed by RT-PCR demonstrated that the micro-deletion caused loss of *BRCA2* exon 6 recognition indicating that the sequence stretch affected by the deletion seemed to be necessary for recognition of this exon (**Fig. 13B**).



2.3.2. The biallelic micro-deletion in *BRCA2* exon 6 causes the generation of an additional in-frame transcript with unique skipping of exon 5 in the male patient

Establishment of fibroblast cultures and EBV (Epstein-Barr virus)-immortalized lymphoblastoid lines from both affected patients (Prof. R. Schneppenheim, Hamburg; Prof. H. Hanenberg, Düsseldorf) allowed direct analysis of the *BRCA2* transcript from these patients. To this end, total RNA was extracted from these cells and analyzed by RT-PCR using primers located in *BRCA2* exon 3 and 8 enabling the investigation of the consequences of the micro-deletion in *BRCA2* exon 6 on splicing within the endogenous transcript.

Analysis of the splicing pattern of the *BRCA2* transcript in fibroblasts grown from the affected boy and his affected sister in comparison to normal control fibroblasts predominantly showed a slightly shorter transcript in both patient derived fibroblasts as expected due to the 10bp deletion in *BRCA2* exon 6 (**Fig. 14B**). In addition to this transcript isoform, especially in the male patient a splice variant with skipping of the mutant *BRCA2* exon 6 and exon 5 was detectable in 15%. Additionally, splice variants lacking exon 5, 6 and 7 or exon 4, 5, 6 and 7 (5 % and 3 % respectively) were observable.

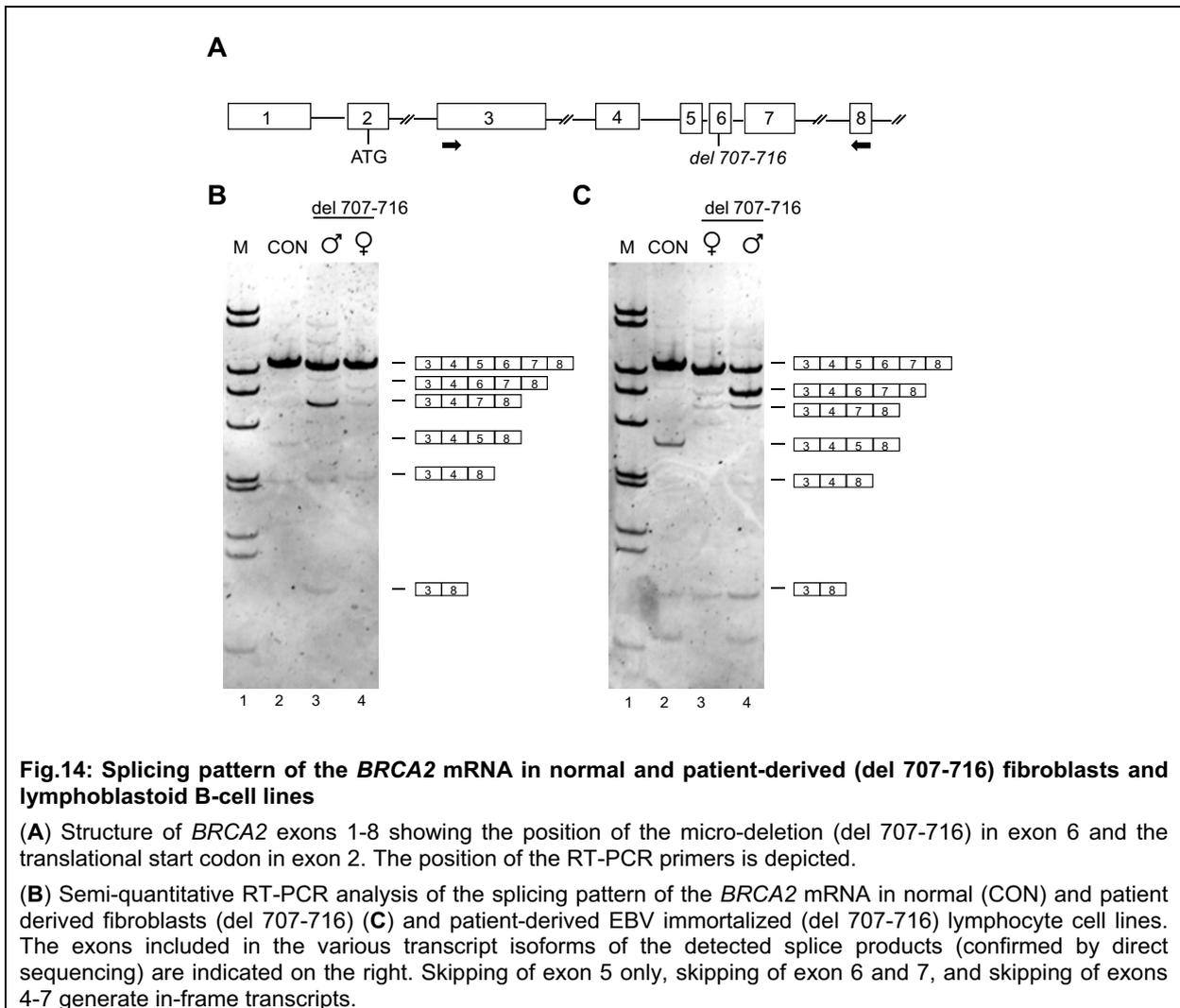
In the female patient, however, the overall level of splice variants additionally to the expected transcript was lower than in the male patient. In addition to low-level occurrence (4%) of the splice variant lacking the mutant *BRCA2* exon 6 and exon 5 a splice variant containing the mutant exon 6, but lacking exon 5 occurred (3%) in the female fibroblasts, whereas the variant lacking exon 4, 5, 6 and 7 was not detectable.

Within the EBV-immortalized normal control lymphocyte cell line a splice variant with skipping of exon 6 and 7 and the variant lacking exon 4, 5, 6 and 7 was found in 10% and in 4%, respectively (**Fig. 14C**). Of note, these splice variants generate in frame transcripts suggesting that low-level alternative splicing of the *BRCA2* transcript occurred naturally in these lymphocytes.

In the lymphocyte cell line derived from the male patient, remarkably, in about 40% of the detected *BRCA2* transcript exon 5 was skipped while the mutant exon 6 was retained in this transcript generating a mutant in-frame transcript in contrast to the expected transcript including all exons and thus, being out of frame due to the 10bp deletion in exon 6. In about 11% both, the mutant exon 6 and exon 5 was skipped and in 6% even a total of four exons, i.e. 4, 5, 6 and 7 were excluded from the transcript. Again, in the lymphocyte cell line derived from the female mutation carrier these splice variants were only faintly detectable.

Results

These results demonstrated several interesting points: Firstly, alternative splicing of the *BRCA2* transcript appeared to be more efficiently in the lymphocyte cell line compared to the fibroblasts pointing to a cell type dependent regulation of alternative splicing of *BRCA2*.

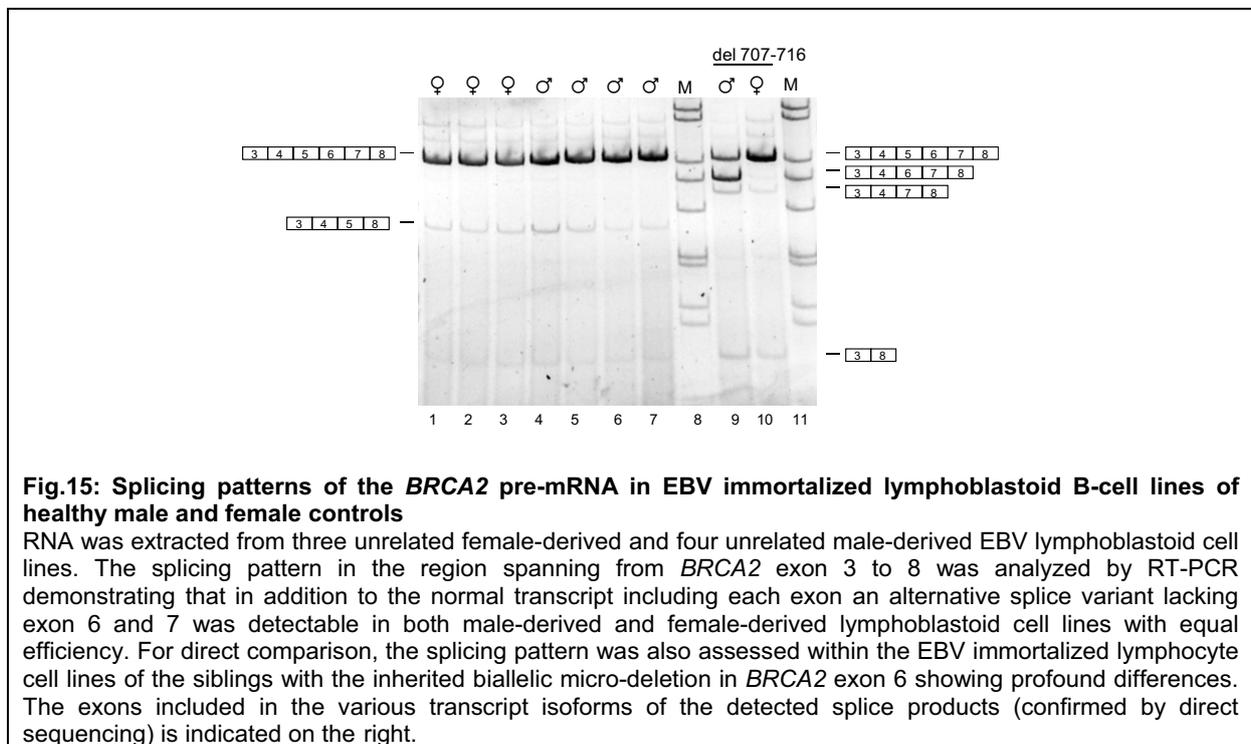


Secondly, the micro-deletion in *BRCA2* exon 6 had profound influence on alternative splicing of the *BRCA 2* transcript not only causing skipping of the affected exon 6 but also of exon 5 indicating that the definition of exon 5 is influenced by that of exon 6. Nonetheless, the occurrence of a transcript including the affected exon 6 and lacking only exon 5 is remarkable as skipping of this exon restored the open reading frame potentially retaining at least partial protein function. The micro-deletion in exon 6 on the other hand created a premature termination codon in exon 6 within the normal open reading frame and skipping of both exons generated a premature termination codon in exon 7. If the additional in-frame transcripts allowed residual protein function this would cause a proliferative advantage becoming operative in the fast proliferating tissue like the lymphocytes as opposed to the fibroblasts. This might provide an explanation for enhanced detection of in-frame splice variants in the lymphoblastoid line.

Results

Although it has reported that in-frame stop codons can cause skipping of the exon harboring the premature termination codon thereby maintaining the open reading frame (Valentine et al., 1998 (343), Wang et al., 2002 (352)) it appeared unlikely that an open reading frame preservation mechanism was underlying the alternative splicing of the *BRCA2* transcript upon the micro-deletion in *BRCA2* exon 6 because the most prominent in-frame transcript was generated by skipping of exon 5. The occurrence of several alternative splice variants caused by the micro-deletion in *BRCA2* exon 6 in both tissues rather provides evidence for a long-range interplay of splicing regulatory elements within the investigated exon cluster since the micro-deletion influenced not only recognition of exon 6 but also definition of the surrounding exons. And thirdly, more pronounced alternative splicing in the cells derived from the male patient compared to those derived from the female patient carrying the identical homozygous germline mutation suggested that gender specific differences may affect splicing of the *BRCA2* pre-mRNA.

To clarify whether gender specific differences influence splicing of the *BRCA2* pre-mRNA total RNA was extracted from three unrelated male-derived and four unrelated female-derived lymphoblastoid cell lines. Comparison of the splicing pattern within the region spanning from *BRCA2* exon 3 to 8 demonstrated that in addition to the normal transcript including each exon an alternative splice variant lacking exon 6 and 7 and one variant lacking exon 4, 5, 6 and 7 occurred in both male-derived and female derived lymphoblastoid cell lines with equal efficiency (**Fig. 15**).



Results

Therefore, gender specific differences did not affect splicing of the *BRCA2* pre-mRNA indicating that gender-independent genetic differences between both siblings may influence the splicing outcome upon the micro-deletion in *BRCA2* exon 6.

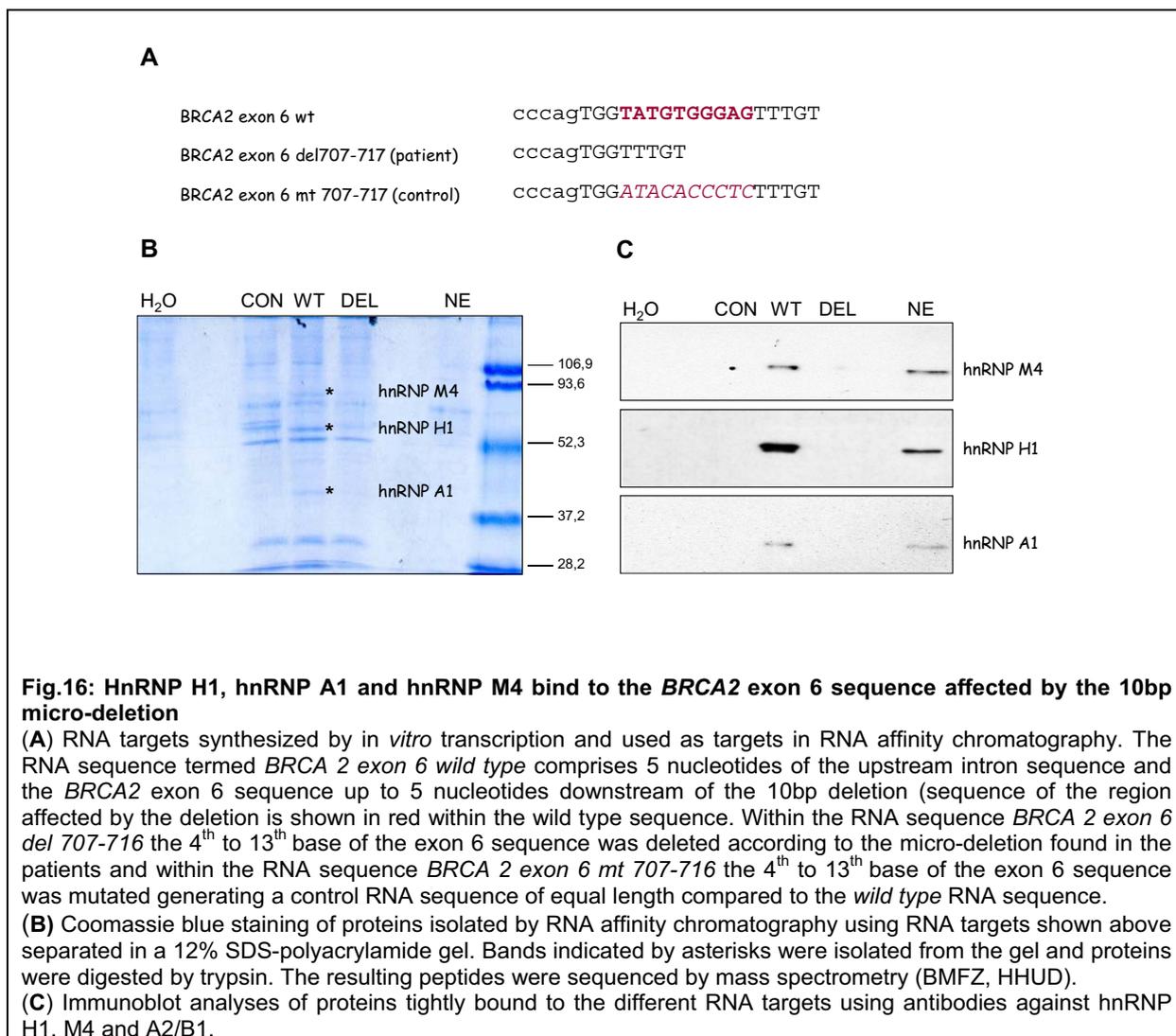
Taken together, the finding that the micro-deletion in *BRCA2* exon 6 causes alternative splicing of the *BRCA2* pre-mRNA implied that the micro-deletion disrupted a splicing regulatory element within exon 6 that seemed to influence not only recognition of exon 6 but also recognition of the surrounding exons within this cluster.

2.3.3. HnRNP H1, hnRNP A1 and hnRNP M4 bind to nucleotides deleted in the patient-derived *BRCA2* exon 6

To investigate whether the micro-deletion of 10 base pairs within *BRCA2* exon 6 interfered with binding of distinct protein factors RNA affinity chromatography experiments were performed with three different target RNA sequences (**Fig. 16A**): i) Because the micro-deletion was located close to the 5' end of *BRCA2* exon 6 the RNA sequence termed *BRCA2* exon 6 *wild type* was composed of 5 nucleotides of the upstream intron sequence and the exon 6 sequence up to 5 nucleotides downstream of the deletion. ii) Within the RNA sequence *BRCA2* exon 6 *del 707-716* the 4th to 13th base of the exon 6 sequence was deleted according to the micro-deletion found in the patients and iii) within the RNA sequence *BRCA2* exon 6 *mt 707-716* the 4th to 13th base of the exon 6 sequence was mutated generating a control RNA sequence of equal length compared to *wild type* RNA sequence.

These RNA sequences were generated by *in vitro* transcription and covalently linked to adipic acid dihydrazide-agarose beads. Following incubation with HeLa nuclear extract proteins that remained tightly bound to each RNA after washing were analyzed by SDS-PAGE. After Coomassie blue staining, two distinct protein bands in the separation range between 52 and 93 kDa and one distinct protein band in the range from 37 to 52 kDa (**Fig. 16B, asterisks**) were observed to bind to the *wild type* RNA sequence only and neither to the RNA sequence with the deletion nor to the mutant RNA sequence. These proteins bands were excised from the SDS polyacrylamide gel, digested with trypsin and sequenced by mass spectrometry (BMFZ, HHU). The following heterogenous ribonucleoproteins (hnRNPs) could be identified: H1 (www.uniprot.org/uniprot/P31943 49 kDa, A1 (P09651, 39 kDa) and M4 (P52272, 78 kDa). Immunoblotting confirmed strong binding of hnRNP H1 and moderate binding of hnRNP A1 and M4 to the *wild type* sequence whereas these proteins could not be detected on the RNA sequence harboring the deletion and also not on the control RNA.

Results



Furthermore, inspection of the *BRCA2* exon 6 sequence revealed the presence of the core-binding site GGA for hnRNP H1 (Caputi et al., 2001 (45)) within the region affected by the micro-deletion. Likewise, it has been reported that hnRNP M binds avidly to poly(G) homopolymers *in vitro* (Datar et al., 1992 (66)) indicating that both hnRNP H and M might specifically bind to the *wild type BRCA2* exon 6 sequence. As the 5'-end of *BRCA2* exon 6 does not contain an hnRNP A1 binding sites that exactly matches the consensus high-affinity hnRNP A1 binding site, UAGGGA/U (Burd et al., 1994 (42)), this might explain low-affinity binding of hnRNP A1 only.

It has been reported that hnRNP H1 and M are involved in the regulation of alternative splicing (Ohe et al., 2009 (232), Hovhannisyan et al., 2007 (136), Paul et al, 2006 (239)). Because it has been suggested that interactions between different hnRNP H1 and A1 proteins bound to distinct positions on a pre-mRNA can change its conformation to affect

splicing decisions (Fisette et al., 2010 (85)) it appeared likely that these proteins function as splicing regulators within the *BRCA2* transcript. Moreover, individual and cell-type specific expression levels of these proteins (Kamma et al., 1995 (151)) may contribute to the different splicing outcome upon the micro-deletion in *BRCA2* exon 6. Nevertheless, further studies including siRNA mediated knockdown of these proteins in different cell types and mutational analysis in extended minigenes will have to confirm potential direct mechanisms in control of *BRCA2* splicing.

2.4. Mechanisms of cryptic splice donor activation upon the *FGB* IVS7 +1G>T splice donor mutation

Even though exon skipping is by far the most frequent outcome of human splice donor mutations activation of cryptic splice donor sites located close to the authentic splice donor site is the second most frequent consequence of human splice donor mutations (Krawczak et al., 2007(176)).

The homozygous *FGB* IVS 7 +1G>T mutation affecting the highly conserved GT dinucleotide of the splice donor site of *FGB* exon 7 has been identified in a patient suffering from congenital afibrinogenemia (MIM #202400) by genomic sequencing (Spena et al., 2002(314)). This 5'ss mutation has been analyzed in a minigene construct composed of a portion of exon 6 (119 nucleotides), intron 6 (208 nucleotides), exon 7 (286 nucleotides), intron 7 (618 nucleotides), and a portion of exon 8 (273 nucleotides, comprising the first 41 nucleotides of the 3'UTR) by Spena and coworkers. Their analysis revealed that beside exon 7 skipping the main consequence of this mutation was the activation of three cryptic donor splice sites, localized in the *FGB* exon 7 at 106 nt (c1), 40 nt (c2), and 24 nt (c3) upstream from the physiological splice donor (Spena et al., 2006 (315)).

2.4.1 The *FGB* IVS7 +1G>T splice donor mutation causes activation of a putative splice donor in the downstream intron

Assessment of the intrinsic strength of GT sequences within exon 7 and its downstream intron applying the HBond algorithm calculated an HBond score (HBS) of 15.00 for the authentic wild type splice donor site of *FGB* exon 7. The HBond scores for the cryptic splice donor sites c1, c2 and c3 accounted for 12.20, 13.70 (calculated with GT instead of GC) and 10.80 respectively (**Fig. 17A**), demonstrating that the authentic splice donor had a significantly higher score value than the cryptic ones.

Results

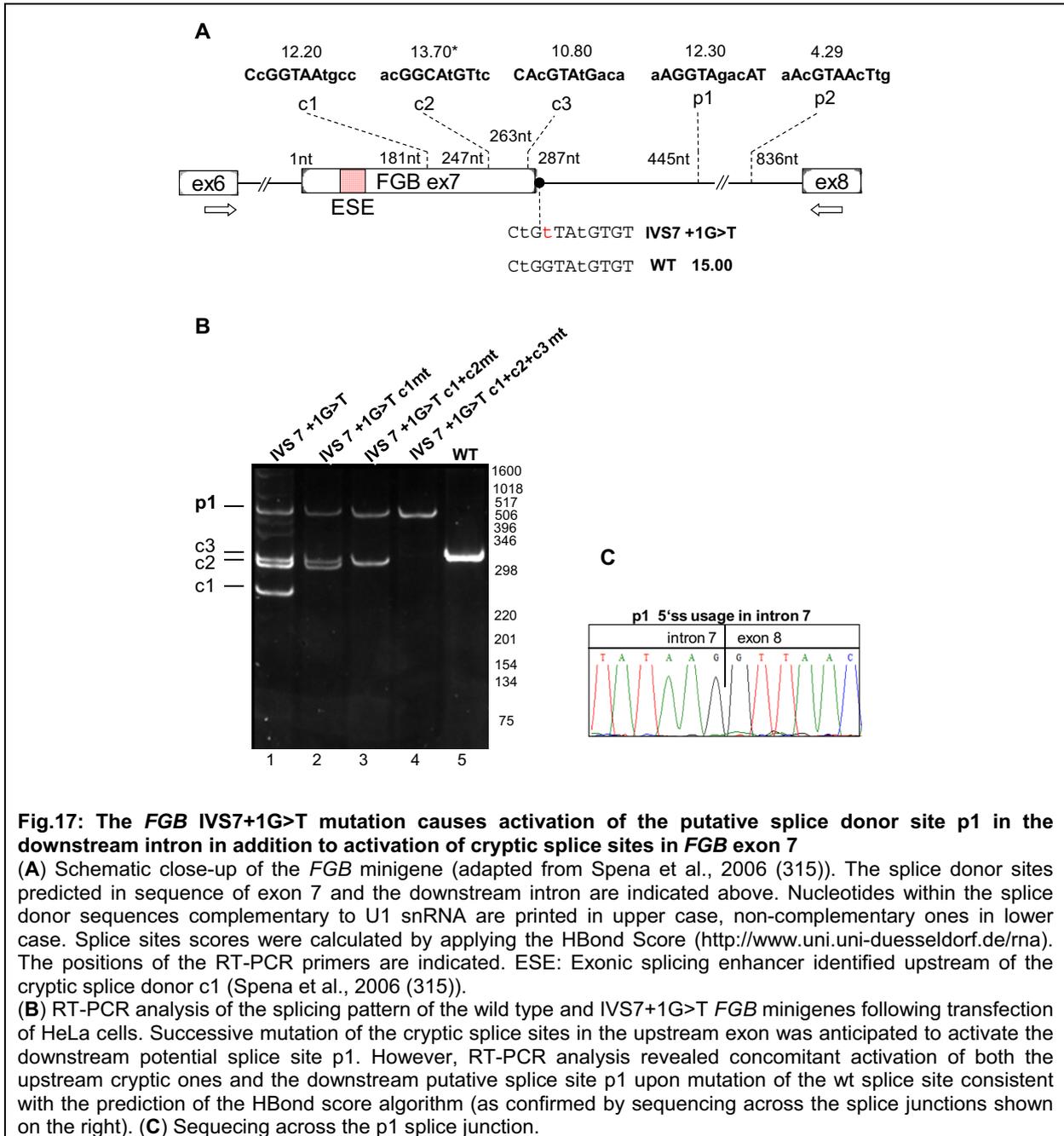


Fig.17: The *FGB* IVS7+1G>T mutation causes activation of the putative splice donor site p1 in the downstream intron in addition to activation of cryptic splice sites in *FGB* exon 7

(A) Schematic close-up of the *FGB* minigene (adapted from Spena et al., 2006 (315)). The splice donor sites predicted in sequence of exon 7 and the downstream intron are indicated above. Nucleotides within the splice donor sequences complementary to U1 snRNA are printed in upper case, non-complementary ones in lower case. Splice sites scores were calculated by applying the HBond Score (<http://www.uni.uni-duesseldorf.de/rna>). The positions of the RT-PCR primers are indicated. ESE: Exonic splicing enhancer identified upstream of the cryptic splice donor c1 (Spena et al., 2006 (315)).

(B) RT-PCR analysis of the splicing pattern of the wild type and IVS7+1G>T *FGB* minigenes following transfection of HeLa cells. Successive mutation of the cryptic splice sites in the upstream exon was anticipated to activate the downstream potential splice site p1. However, RT-PCR analysis revealed concomitant activation of both the upstream cryptic ones and the downstream putative splice site p1 upon mutation of the wt splice site consistent with the prediction of the HBond score algorithm (as confirmed by sequencing across the splice junctions shown on the right). (C) Sequencing across the p1 splice junction.

Apart from the cryptic splice donor sites within *FGB* exon 7, *in silico* tools suggested two additional putative intronic splice donor sites, located 158 nt (p1) and 549 nt (p2) downstream from the authentic splice donor site with an HBS HBond score of 12.30 for the putative splice site p1 and 9.4 for p2. Although the score of the putative splice site p1 was comparable high to the one of the cryptic splice site c1 usage of p1 had not been observed by Spena et al. (315).

To investigate whether the usage of p1 is outcompeted by the upstream cryptic splice sites the cryptic splice sites in *FGB* exon 7 were successively mutated within the minigene

Results

construct. After transfection of HeLa cells the splicing pattern was analyzed by performing an RT-PCR in the linear amplification range with primers located in *FGB7* exon 6 and exon 8 as done by Spena et al. Unexpectedly, in the presence of the 5'ss +1G>T mutation inactivation of the upstream cryptic splice donor sites was not necessary to activate p1. Instead, this analysis revealed concomitant activation of both the upstream cryptic sites (c1, c2 and c3) and the downstream putative splice site p1. Successive inactivation of each upstream cryptic splice donor site in exon 7 neither significantly increased the usage of the remaining cryptic sites nor the usage of the downstream potential site p1 (**Fig. 17 B and C**).

These results demonstrated that selection of the cryptic splice sites did not undergo a 5'>3' scanning process (Borensztajn et al., 2006 (35)) activating the most upstream splice donor site with sufficient complementarity to U1 snRNA. It appeared rather that they were activated independently from each other.

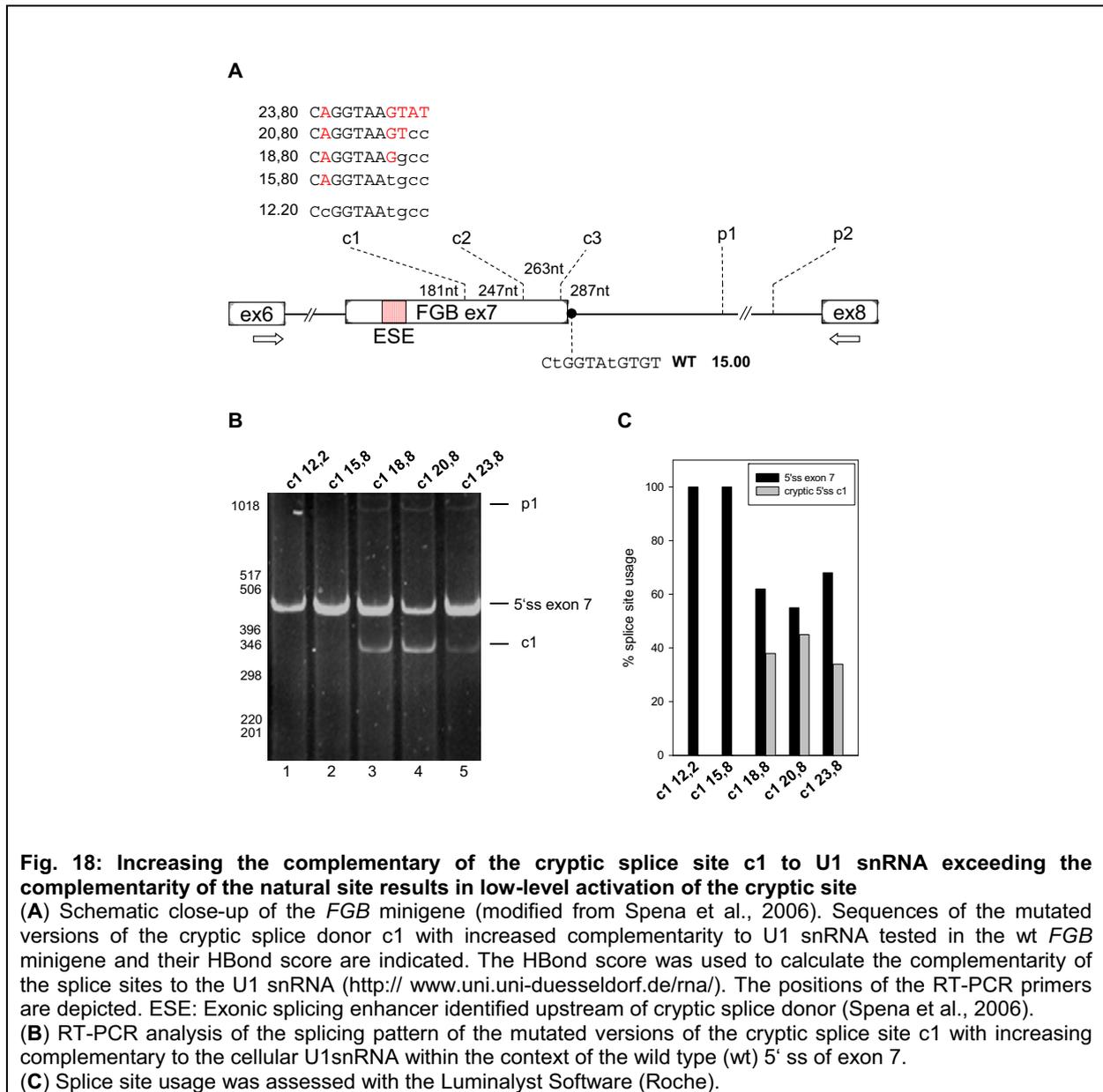
2.4.2. Increasing the complementary of the cryptic splice site c1 to U1 snRNA exceeding the natural site results in low-level activation of the cryptic site

As cryptic splice sites per definition are GT sequences that are not used as splice sites in wild type pre-mRNA, but are selected as a result of a mutation affecting the recognition of a wt 5'ss, the question remained whether a significant higher score value of the authentic site compared to the cryptic ones would account for the correct specification of the authentic site in the wild type pre-mRNA. In other words, would the reduction of the score difference between the cryptic splice sites and the authentic site by artificially increasing the complementarity of the cryptic splice donor sites to U1 snRNA result in activation of the cryptic sites despite the presence of the wild type splice donor?

To clarify this, the complementarity of the cryptic splice donor c1 to the U1 snRNA was successively increased in the *FGB* minigene harboring the wild type splice donor site of *FGB* exon 7 with an HBond score of 15.00. Within the cryptic splice donor site c1 (original HBond score of 12.20) non-complementary nucleotides were consecutively replaced by complementary nucleotides resulting in HBond score values of 15.80, 18.80, 20.80 and 23.80 for the cryptic splice donor c1 (**Fig. 18A**).

HeLa cells were transiently transfected with the corresponding *FGB* minigenes and the splicing pattern was analyzed by RT-PCR.

Results



Remarkably, if the intrinsic strength of the cryptic splice site c1 (HBS c1 = 15.8) was comparable to the intrinsic strength of the authentic splice donor site of *FGB* exon 7 (HBS = 15.0) the splicing machinery discriminated against the usage of the cryptic sites in favor of the natural site (**Fig. 18B lane 2**). However, increasing the intrinsic strength of c1 towards an HBond score value of 18.8 induced the usage of c1 instead of the wild type splice donor in 38% of the minigene transcripts (**Fig. 18B lane 3**). Further improvement of the cryptic splice donor c1 by increasing its complementarity to U1 snRNA towards an HBS of 20.8 resulted in activation of c1 in 45% of the minigene transcripts. However, the authentic splice donor of *FGB* exon 7 despite its significant lower complementarity was still preferred (55% of the

Results

minigene transcripts). Surprisingly, full complementarity of the cryptic site c1 to U1 snRNA achieving an HBond score value of 23.8 did not further increase the usage of the cryptic splice donor as in this case the cryptic splice site was activated in 34% of the minigene transcripts only (**Fig.18 B lane 5**).

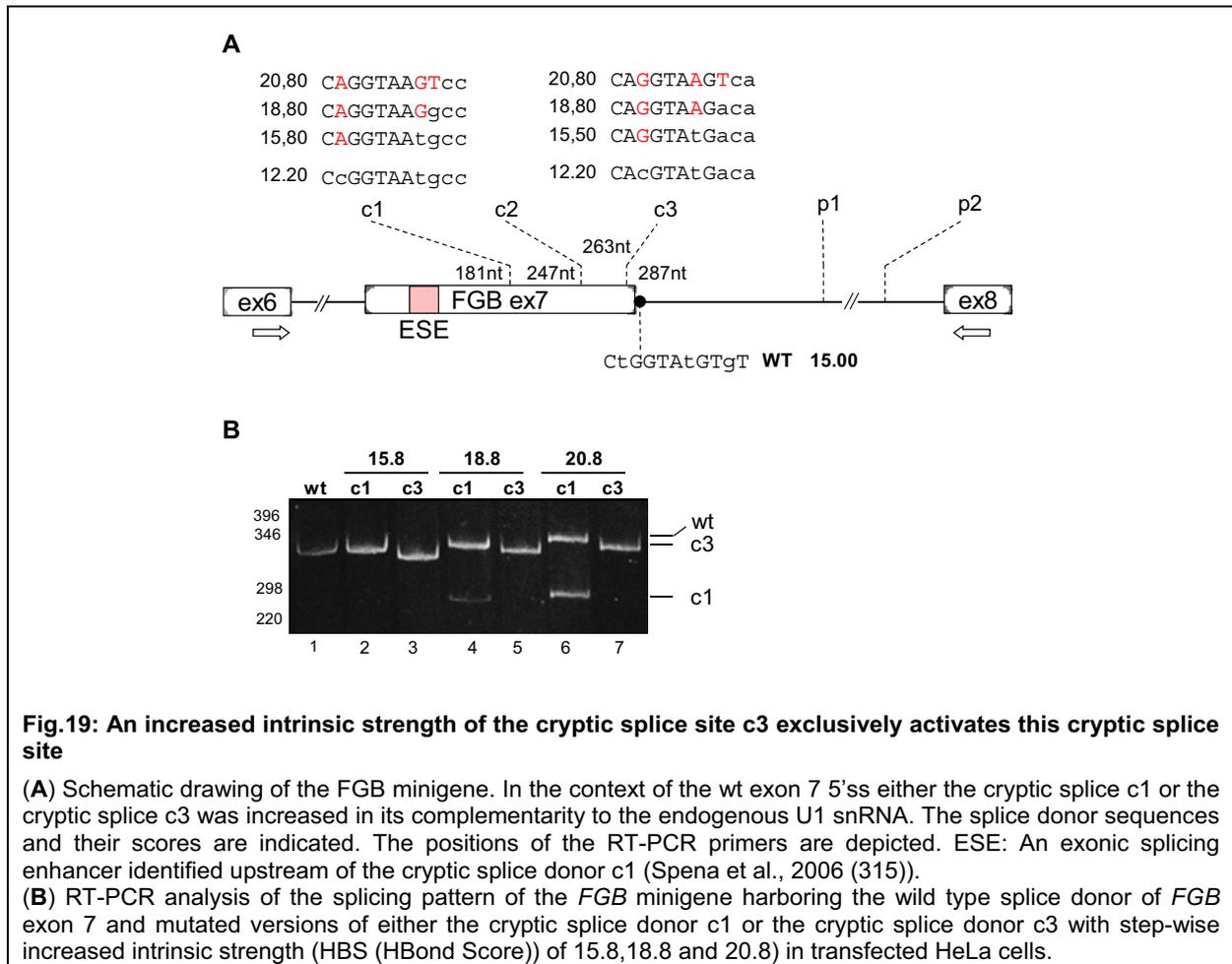
Taken together, an equal intrinsic strength of the cryptic splice donor c1 and the natural splice donor of *FGB* exon 7 discriminated the cryptic site and exclusively activated the natural splice site. This points to an enhanced functional strength of the natural splice donor taking into account its context of *cis*-regulatory elements allowing the splicing machinery to prefer the natural splice donor over competing nearby potential splice donor sites of comparable intrinsic strength. Moreover, less activation of c1 despite an intrinsic strength exceeding the intrinsic strength of the natural splice donor might be due to the weakness of the previously identified splicing enhancer upstream of c1 (Spena et al., 2002 and 2006 (314,315)).

2.4.3. Increased intrinsic strength of the cryptic splice site c3 exclusively activates this cryptic site

Since an enhancer element has been identified upstream of the cryptic splice donor c1 and has been shown to be crucial for activation of the cryptic site c1 (Spena et al., 2006 (315)) it seemed likely that activation of the cryptic splice site c3 upon disruption of the natural splice donor also was enhancer dependent in particular because the intrinsic strength of the cryptic splice donor c3 in *FGB* exon 7 accounting for an HBS of 10.8 was consistently lower compared to the intrinsic strength of the cryptic splice donor c1 with an HBS of 12.3. Noteworthy, the cryptic splice donor c1 was localized at 106 nt upstream from the physiologic splice donor whereas the cryptic splice donor c3 was identified only 24 nt upstream of the physiological one.

To clarify whether an increased intrinsic strength of the cryptic splice donor c3 permitted the cryptic splice donor c3 to outcompete the physiological wild type splice donor of *FGB* exon 7 the intrinsic strength of c3 was increased within the wild type *FGB* minigene by consecutively replacing non-complementary with complementary nucleotides to U1 snRNA achieving HBS values of 15.8, 18.8 and 20.8, respectively (**Fig. 19A**).

Results



To directly compare the splicing pattern of these minigenes to the one obtained by the wild type *FGB* minigene with the increased intrinsic strength of the cryptic splice donor c1 each of both constructs with almost identical scores of the respective cryptic splice donor were used to transiently transfect HeLa cells.

Analysis of the splicing pattern by semi-quantitative RT-PCR demonstrated that if the intrinsic strength of the cryptic splice donor c3 was comparable to the intrinsic strength of the physiological splice donor of *FGB* exon 7 (HBond score of 15.8 versus 15.0) the splicing machinery exclusively selected the cryptic splice donor c3 instead of the physiological splice donor (**Fig. 19B lane 3**). In contrast, in the case of identical intrinsic strength of the cryptic splice site c1 and the physiological splice donor, the splicing machinery discriminated against the cryptic splice site c1 and exclusively selected the physiological splice donor (**Fig. 19B, lane 2**). Therefore, the data provided evidence that the activation of the cryptic splice donor c3 as well as the authentic exon 7 splice donor was supported by an additional exonic enhancer element within *FGB* exon 7 that appeared to be much stronger than the previously identified splicing enhancer upstream of the cryptic splice donor c1.

2.4.4. *FGB* exon 7 contains multiple splicing enhancer elements

To provide experimental evidence for the presence of additional splicing enhancer elements within the *FGB* exon 7 and to localize such elements within *FGB* exon 7 the subgenomic HIV-1 glycoprotein (Env) expression vector was used (Kammler et al, 2001 (152)). Since it has been shown that stabilization of the unstable HIV-1 glycoprotein RNA (*env*) requires the presence of an enhancer element within the leader sequence immediately upstream of the SD4 supporting RNA duplex formation between the spliceosomal U1 snRNA and this splice donor, quantification of syncytium formation of HeLa T4⁺ cells after transient transfection with this reporter harboring a test sequence upstream of SD4 allows rapid identification of a putative enhancer sequence.

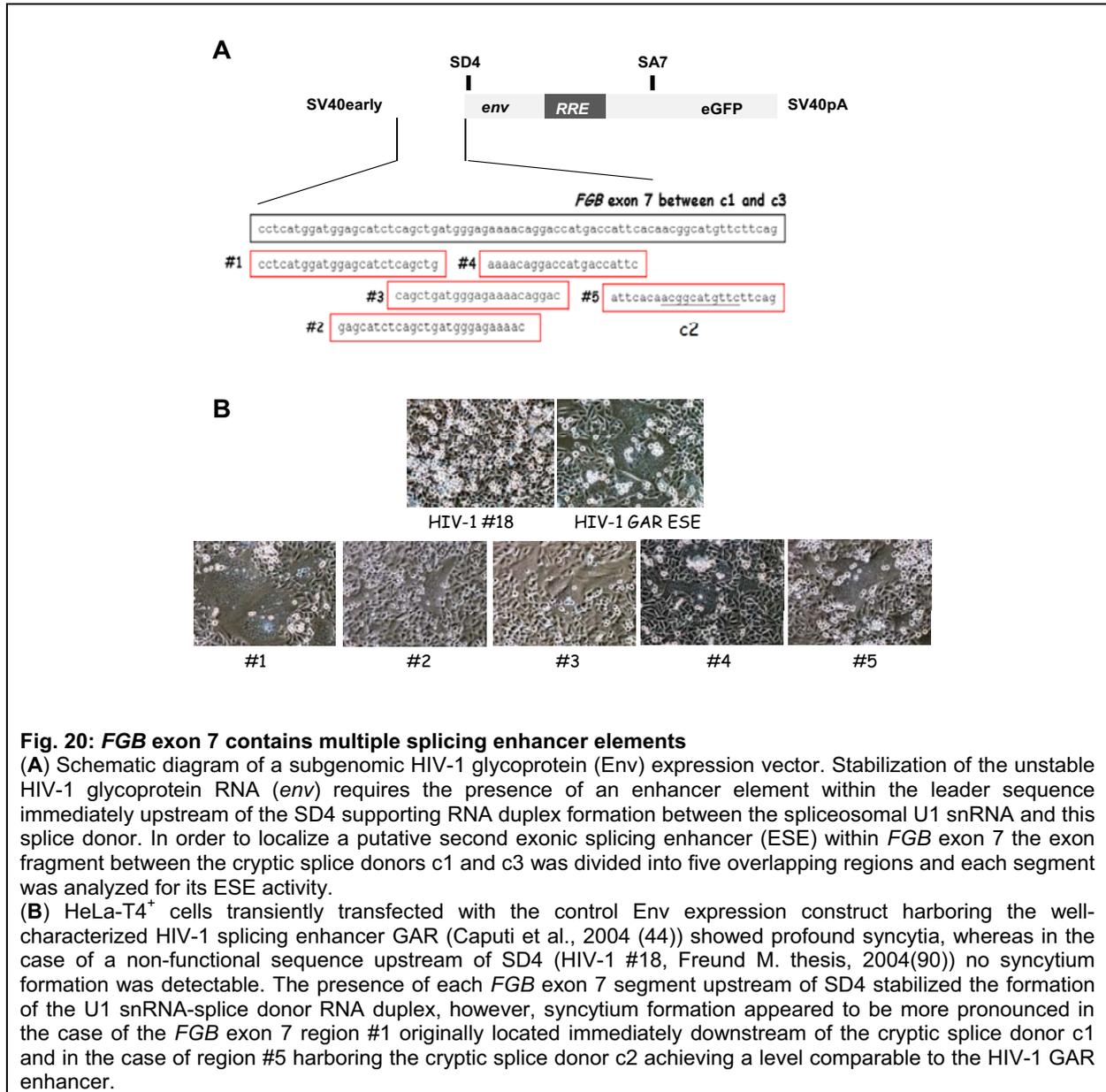
Therefore, the exon fragment between the cryptic splice donors c1 and c3 was divided into five overlapping regions and each segment was inserted immediately upstream of SD4 in the Env reporter construct (**Fig. 20A**). 48 hrs after transient transfection syncytium formation of HeLa-T4⁺ cells was assessed.

HeLa-T4⁺ cells transiently transfected with the control Env expression construct harboring the well-characterized HIV-1 splicing enhancer GAR (Caputi et al., 2004 (44); Asang et al., 2008 (10) and Asang C. thesis, 2010 (9)) showed profound syncytium formation, whereas in the case of the second control Env expression construct carrying a non-enhancer sequence upstream of SD4 (HIV-1 #18, Freund M. thesis, 2004 (90)) no syncytium formation was detectable (**Fig. 20B**).

Surprisingly, the presence of each *FGB* exon 7 segment upstream SD4 stabilized formation of the U1 snRNA-splice donor RNA duplex, albeit syncytium formation appeared to be more pronounced in the case of the *FGB* exon 7 region #1 originally located immediately downstream of the cryptic splice donor c1 within *FGB* exon 7. In the case of region #5 harboring the cryptic splice donor c2 a level of syncytium formation could be achieved which was comparable to the HIV-1 GAR enhancer-mediated syncytium formation (**Fig. 20B**).

Thus, the analysis of syncytium formation suggested that multiple enhancer elements within *FGB* exon 7 induce cryptic splice site activation upon disruption of the physiological splice donor. Continulative work in our group (Schöneweis K. diploma thesis, 2010 (284)) quantifying the enhancer activity of different regions of the *FGB* exon 7 using the Env-eGFP reporter construct and flow cytometry demonstrated that the enhancer activity of region #1 was even stronger than the one of the previously published enhancer sequence upstream of the cryptic splice donor c1 (data not shown).

Results



Moreover, additional work in our group demonstrated that disruption of the enhancer activity of region #1 (FGB 7D 5C8A mutation) allowed the preferential usage of the cryptic splice donor c1 in favor of the cryptic splice donor sites c2 and c3 and in favor of the natural splice donor. This was even more pronounced when the intrinsic strength of c1 was increased towards an HBS of 20.8 (K. Schöneweis diploma thesis (284) and S. Kübart bachelor thesis, 2010 (178)). In the presence of the IVS7+1G>T splice donor mutation disruption of both the previously published splicing enhancer and the newly identified enhancer resulted in increased *FGB* exon 7 exon skipping (Kübart S. bachelor thesis, 2010 (178)).

Results

Together, these results suggest that the density of enhancer elements and the intrinsic strength of GT sequences within human exons might be decisive whether a splice donor mutation results in skipping of the affected exon or in activation of cryptic splice sites.

2.5. Identification and characterization of a non-canonical TT splice donor

2.5.1. The *FANCC* c.165 +1G>T splice donor mutation in primary cells of FA-C patients allows correct splicing albeit at a reduced level

(cited from Hartmann et al., 2010 (122))

Genomic sequencing identified a single base-pair substitution in the 5'ss of *FANCC* exon 2, c.165 +1G>T, converting the highly conserved GT dinucleotide within the 5'ss to a TT dinucleotide (**Fig. 21B**) in three index Fanconi Anemia (FA) patients from two consanguineous families of Arabian ancestry and one mixed Arabian/British couple. These patients were assigned to the complementation group FA-C by transduction of primary skin fibroblasts of three index FA with gammaretroviral vectors expressing one of the following cDNAs: *FANCA*, *FANCC*, *FANCE*, *FANCF*, and *FANCG*. Transduced fibroblast cells were exposed to 33nM of the DNA crosslinker drug mitomycin (MMC) for three days and then harvested for cell cycle analysis by flow cytometry as described previously (Hanenberg et al. 2002 (119), Chandra et al. 2005 (53)). The cell cycle distribution of the fibroblasts revealed that overexpression of the *FANCC* cDNA specifically corrected the characteristic DNA cross-linker hypersensitivity of the patients' cells (data not shown).

To analyze the phenotypic consequence of the *FANCC* exon 2 c.165 +1G>T 5'ss mutation at the RNA level, RT-PCR analysis on mRNA from primary patient fibroblasts from pedigree 526 (**Table 1**) was performed. In contrast to the normal control, four distinct splice products were found contributing to 33, 27, 25 and 15% of the transcripts, respectively (**Fig. 21C**). Direct sequencing of the amplified products revealed that the three *FANCC* transcripts of aberrant size either lacked the translational start codon due to skipping of *FANCC* exon 2 (25%, skipping) or encoded mutant open reading frames with premature translation termination (33 and 15%, cryptic GC and GT). Remarkably, the fourth amplified product (27%, TT) was the normal wild-type *FANCC* transcript (**Fig. 21D**). Therefore, the c.165 +1G>T splice donor mutation still enabled normal *FANCC* transcript processing, albeit at lower efficiency compared to the wild-type canonical 5'ss.

Among the four male and five female patients, only the two patients from the family 640 with mixed ethnic background had typical severe congenital malformations as described for FA patients (Gillio et al., 1997 (100))(**Table 1**).

Results

Family/affected	Sex	Paternal mutation	Maternal mutation	Congenital abnormalities	Age at BMF	Diagnosis to SCT	SCT at age	Reason for SCT	Last follow-up
526/1	m	c.165+1G>T	c.165+1G>T	Café-au-lait spots	5.5	10.5	16	BMF	alive@20 years, 3.5 years post SCT
526/2	f	c.165+1G>T	c.165+1G>T	Café-au-lait spots	8	1	9	BMF	alive@29 years, 20 years post SCT
526/3	f	c.165+1G>T	c.165+1G>T	Café-au-lait spots	13	2	19	BMF	alive@19.5 years, 0.5 years post SCT
526/4	f	c.165+1G>T	c.165+1G>T	Café-au-lait spots	6	-	-		alive@11 years of age
526/5	f	c.165+1G>T	c.165+1G>T	Café-au-lait spots	7.5	-	-		alive@8 years of age
640/1	m	c.165+1G>T	c.1-250 del	Café-au-lait spots, malrotated kidney, ureter duplication, microphallus	5.5	-	-		died@13.5 years of age
640/2	m	c.165+1G>T	c.1-250 del	Café-au-lait spots, microphallus	7	8.5	15.5	BMF	died@16 years of age
1159/1	m	c.165+1G>T	c.165+1G>T	Café-au-lait spots	11.5	1	12.5	BMF	died@12.5 years of age
1159/2	f	c.165+1G>T	c.165+1G>T	none	none	-			alive@4.5 years of age

Table 1: Mild clinical manifestations of Fanconi anemia (FA) in the nine FA-C (FA subtype C) patients

Shown are the clinical characteristics of the nine FA-C patients from three pedigrees numbered 526, 640, and 1159, respectively. The gender, the paternal and maternal *FANCC* mutations, the café-au-laits spots and the major congenital abnormalities, the age at the onset of bone marrow failure (BMF), the time interval from diagnosis until stem cell transplantation (SCT), the age at and the indication for transplantation, and the last follow-up are shown. Cited from Hartmann et al., 2010 (122).

Results

Calculated for all nine affected individuals, the mean number of 0.45 malformations per patient was similar to that described for the 'European' c.67delC mutation in exon 2 (Gillio et al., 1997 (100)) but different to the c.456 +4A>T (IVS4 +4A>T) and c.1642 C>T (R548X) mutations described in Ashkenazi Jewish and North American patients (Kutler et al., 2003 (179)). Patients from the pedigrees 526 and 1159 were homozygous for this point mutation, while patients from the pedigree 640 carried a maternally inherited genomic deletion (Table 1), leading to the skipping of exons 2 and 3 in the *FANCC* mRNA (Fig. 22).

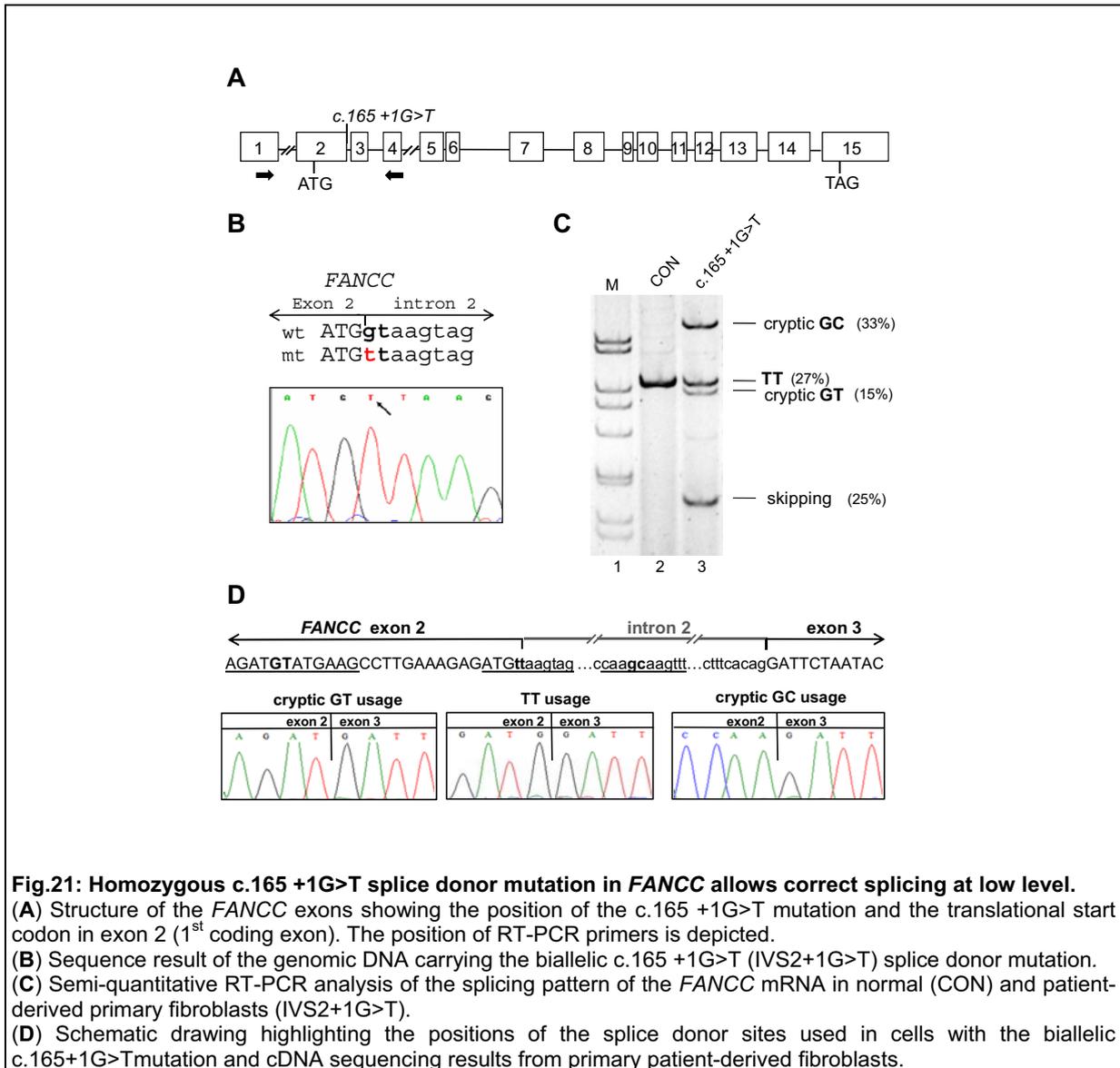
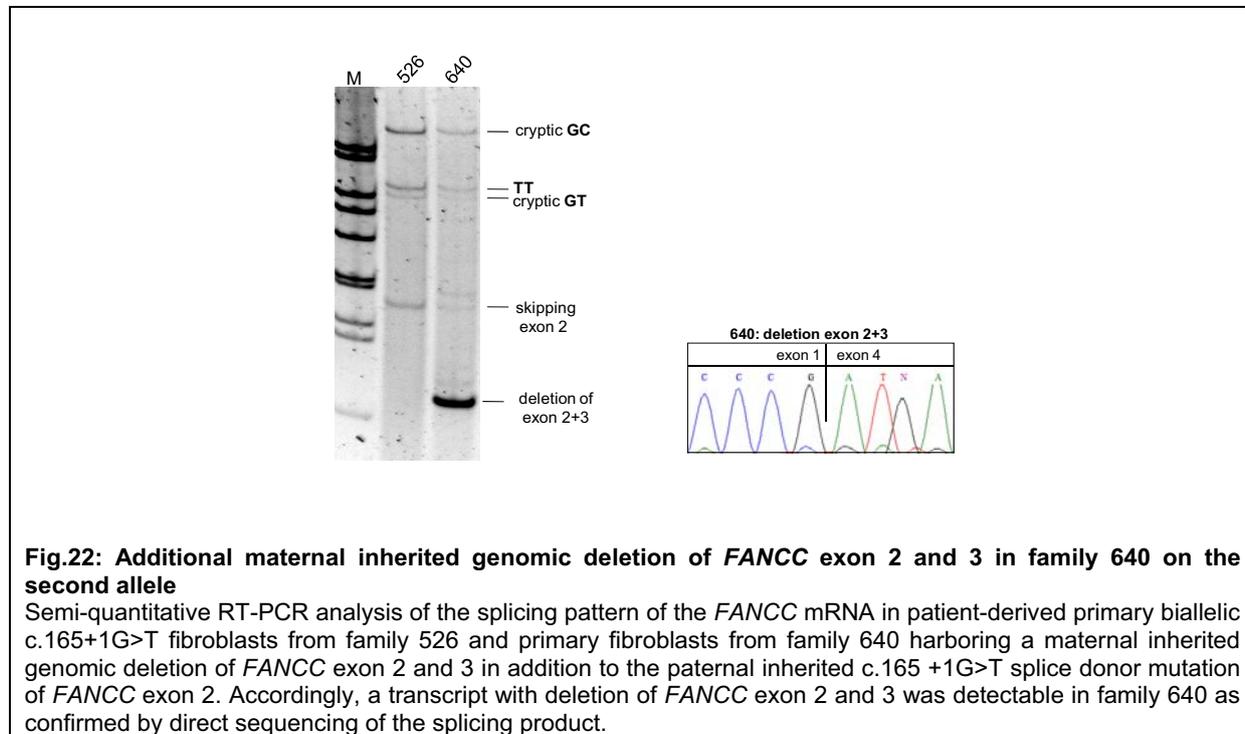


Fig.21: Homozygous c.165 +1G>T splice donor mutation in *FANCC* allows correct splicing at low level. (A) Structure of the *FANCC* exons showing the position of the c.165 +1G>T mutation and the translational start codon in exon 2 (1st coding exon). The position of RT-PCR primers is depicted. (B) Sequence result of the genomic DNA carrying the biallelic c.165 +1G>T (IVS2+1G>T) splice donor mutation. (C) Semi-quantitative RT-PCR analysis of the splicing pattern of the *FANCC* mRNA in normal (CON) and patient-derived primary fibroblasts (IVS2+1G>T). (D) Schematic drawing highlighting the positions of the splice donor sites used in cells with the biallelic c.165+1G>T mutation and cDNA sequencing results from primary patient-derived fibroblasts.

Thus, because the most frequent base-pair mutation in human splice donor sites in inherited diseases comprises the first intronic nucleotide which is a guanosine of the canonical GT dinucleotide (Krawczak et al., 2007 (176)) and until now, any base-pair substitution at this position has been thought to completely abrogate normal mRNA processing the finding that

Results

the *FANCC* c.165 +1G>T changing the canonical GT splice of *FANCC* exon 2 to TT allowed residual correct splicing was highly remarkable. Moreover, this phenomenon seemed to be the cause for a milder clinical phenotype of Fanconi anemia subtype C in these patients.



2.5.2. Increased complementarity to U1 snRNA specifically reconstitutes splicing at the TT dinucleotide in the heterologous splicing reporter minigene (cited Hartmann et al., 2010 (122))

To systematically analyze this unusual pathogenic *FANCC* splice donor, the *FANCC* exon 2 with flanking intronic nucleotides was inserted into the three-exon splicing reporter minigene (**Fig. 23**). HeLa cells were transfected with plasmids carrying either the wild-type GT or mutant TT *FANCC* exon 2 splice donor and analyzed for their splicing pattern by RT-PCR analysis. Although the intrinsic strength of the wild-type *FANCC* 5'ss is relatively high, due to the high degree of complementarity to the U1 snRNA (**Fig. 23C**), the analytical gel (**Fig. 23B**) demonstrated that the recognition of the wild-type *FANCC* exon 2 was not as effective as expected (**lane 2**) and that the mutant 5'ss was not recognized at all (**Fig. 23B, lane 3**). To rank the intrinsic strength of the wild-type *FANCC* exon 2 splice donor among human 5'ss, a representative group of 43.464 annotated 5'ss from constitutively spliced human exons were analyzed using the HBond algorithm. In this analysis, all annotated human splice donor sites

Results

had an average HBS of $15.001 \pm 2,591$ ($x \pm SD$), compared to the HBS of 18.7 of this *FANCC* 5'ss which thereby is ranked at the 92.3 percentile of all splice donor sites in this data set.

Because previous work in our group has demonstrated that an enhanced complementarity between a 5'ss and the U1 snRNA can improve the recognition of a 5'ss and compensate for the lack of supportive context missing in the heterologous splicing reporter minigene the nucleotides at positions -3 and -2 of the mutant TT 5'ss were replaced by nucleotides complementary to the 5'-end of the endogenous U1 snRNA (**Fig. 23C**).

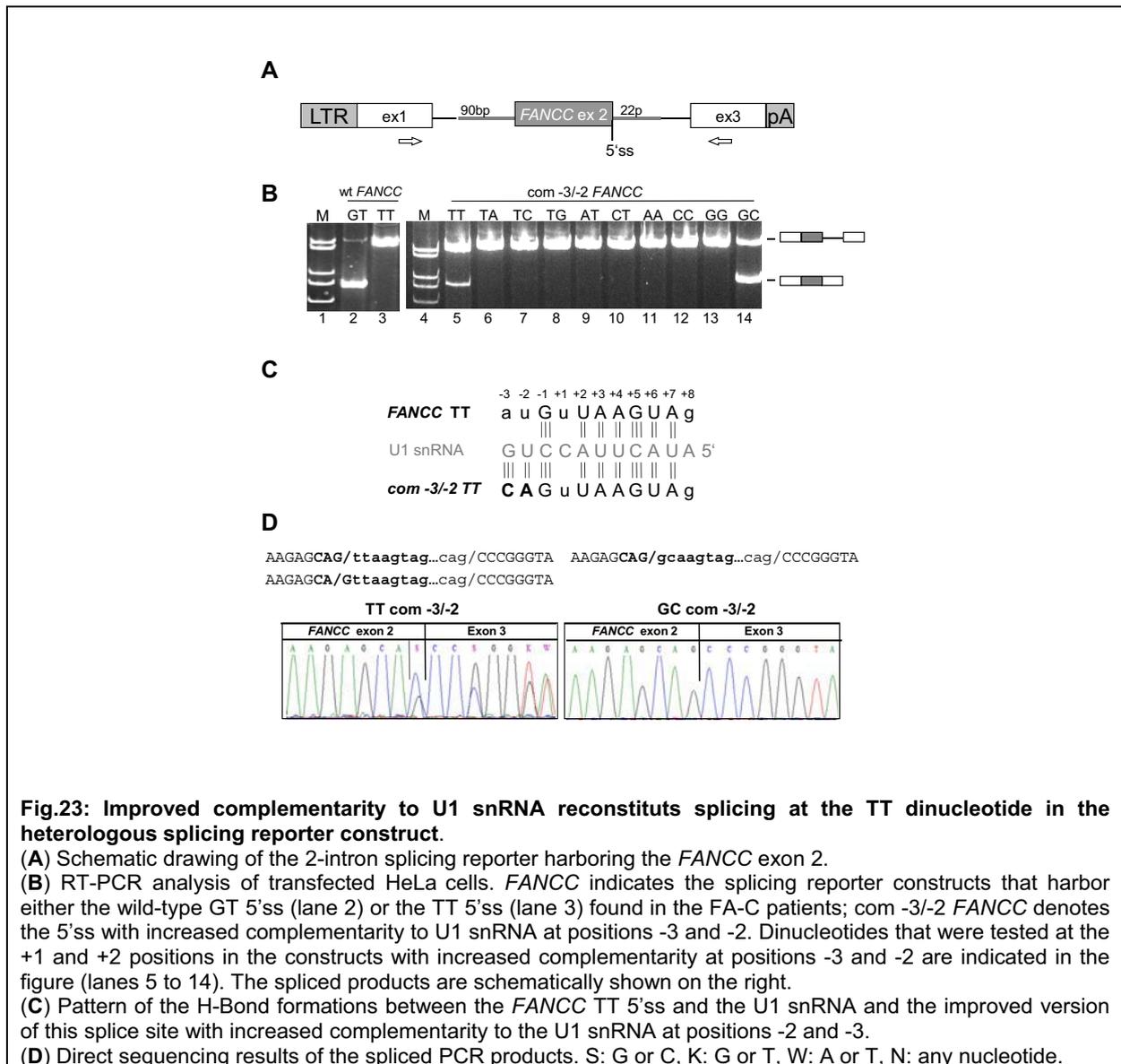


Fig.23: Improved complementarity to U1 snRNA reconstitutes splicing at the TT dinucleotide in the heterologous splicing reporter construct.

(A) Schematic drawing of the 2-intron splicing reporter harboring the *FANCC* exon 2.

(B) RT-PCR analysis of transfected HeLa cells. *FANCC* indicates the splicing reporter constructs that harbor either the wild-type GT 5'ss (lane 2) or the TT 5'ss (lane 3) found in the FA-C patients; com -3/-2 *FANCC* denotes the 5'ss with increased complementarity to U1 snRNA at positions -3 and -2. Dinucleotides that were tested at the +1 and +2 positions in the constructs with increased complementarity at positions -3 and -2 are indicated in the figure (lanes 5 to 14). The spliced products are schematically shown on the right.

(C) Pattern of the H-Bond formations between the *FANCC* TT 5'ss and the U1 snRNA and the improved version of this splice site with increased complementarity to the U1 snRNA at positions -2 and -3.

(D) Direct sequencing results of the spliced PCR products. S: G or C, K: G or T, W: A or T, N: any nucleotide.

As shown in Fig. 23B, lane 5 these two additional nucleotide adaptations facilitated inclusion of the *FANCC* exon 2 with the mutant TT splice donor. Direct sequencing of this splice product, however, revealed that splicing in this reporter transcript occurred not only at the TT

Results

dinucleotide at the authentic exon-intron border, but was also shifted to the GT dinucleotide one position upstream of TT (**Fig. 23D**, TT com -3/-2). The existence of this 2nd transcript was remarkable, because all available algorithms for splice donor sites unequivocally predicted that the intrinsic strength of the GT dinucleotide at -1 is very weak (e.g. the HBS is 2.3) due to the low complementarity to the U1 snRNA in this base-pairing frame.

To further clarify whether splicing at this artificially improved TT 5'ss was simply determined by the overall complementarity to U1 snRNA or also by the position of the GT in the -1 register, reporter constructs were generated that carried different dinucleotides at positions +1 and +2 (**Fig. 23B**). Noteworthy here was that the substitution of T at position +2 maintained the GT within the -1 register, however reduced the overall complementarity to the endogenous U1 snRNA in the original base-pairing frame. In contrast, substitution of the mismatching T at position +1 for A or C did not affect the overall complementarity in the original base-pairing frame, but specifically destroyed the GT in the -1 register. Thus, if the GT in the -1 register were important for recognition of the mutant TT 5'ss, the TA dinucleotide that specifically increased the U1 complementarity in the -1 register, should allow more efficient splice site recognition. As shown in **Fig. 23B**, lane 5 to 14 and confirmed by sequencing (**Fig. 23D**), splicing in this construct only occurred at the two physiological GT and GC splice donor sites or if a TT dinucleotide was present at position +1 and +2. These data demonstrated that a mutant TT splice donor site could be functional in a heterologous context if this site were highly complementary to the U1 snRNA. These results also suggested that the complementarity of the -1 GT register to the U1 snRNA is of less importance, since the TA dinucleotide despite higher complementarity did not allow splicing at this site (**Fig. 23B**, lane 6).

2.5.3 Artificial TT-adapted U1 snRNAs improve correct mRNA processing at the FANCC TT splice donor within the splicing reporter

(cited from Hartmann et al., 2010 (122))

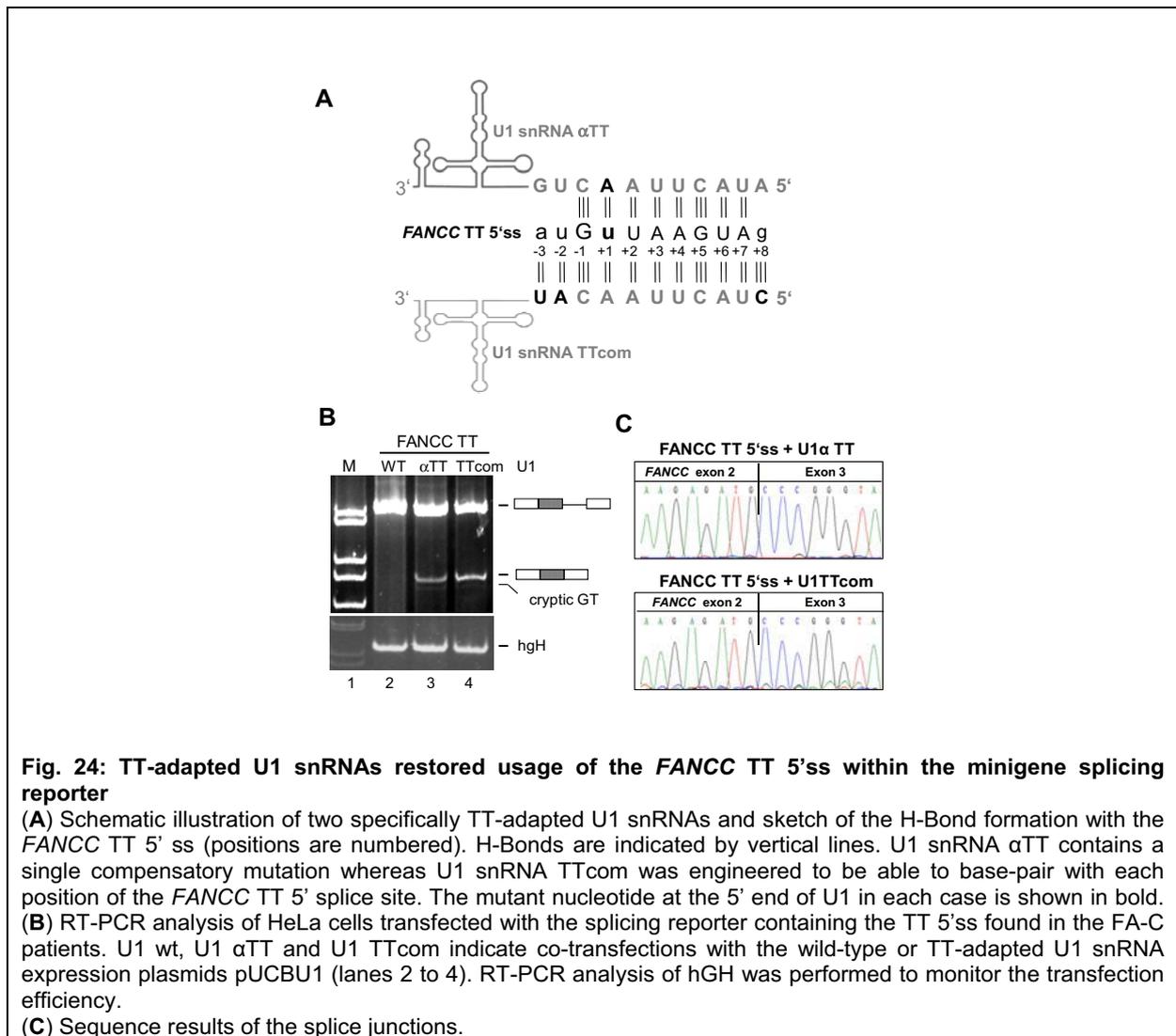
Since the mutant TT splice donor of *FANCC* exon 2 has been recognized in the heterologous splicing reporter only if its complementarity to the U1 snRNA has been increased, this raised the question of whether compensatory mutations within the 5' end of the U1 snRNA would also enable usage of the mutant *FANCC* TT 5'ss.

To this end, two artificial U1 snRNAs were constructed (**Fig. 24A**): The U1 snRNA α TT contained a single compensatory mutation complementary to the TT dinucleotide and the U1 snRNA TTcom matched each position of the mutant *FANCC* TT 5' splice site. While co-transfecting HeLa cells with the wild-type U1 snRNA and the minigene splicing reporter did

Results

not alter the splicing pattern of the construct (**Fig. 24B, lane 2**), co-transfection of either U1 snRNA α TT or TTcom partially restored recognition (8 and 12 %, respectively) of the mutant *FANCC* TT 5'ss (**Fig. 24B, lane 3, 4**). Here, sequence analysis of the splice products confirmed that splicing exclusively occurred at the correct exon-intron border (**Fig. 24C**).

Combining the results from the last two experimental settings of fully adapting either the 5'ss to the endogenous U1 snRNA or the U1 snRNA to the mutant splice site was striking: the exclusive use of the non-canonical TT as splice site was not simply determined by the free energy of the RNA duplex formed between the splice donor and the matching U1 snRNA (which was identical in both cases), but was predominantly dependent on the 5'ss sequence itself.



Results

However, RT-PCR analysis revealed that both tested U1A7 variants U1A7 TOPO 2C and 10T which differed in a single nucleotide polymorphism (SNP) did not affect recognition of the mutant *FANCC* TT 5'ss within the heterologous splicing minigene (**Fig. 26 C**). To rule out that this might be due to a lower expression level of U1A7 compared to U1A snRNA, the coding sequence U1A7 was cloned into the U1A snRNA expression plasmid, allowing expression of the U1A7 variant under the control of U1A promoter. Although RT-PCR confirmed expression of the U1A7 variant under the control of the U1A promoter in transfected HeLa cells (**Fig. 26E, right panel, lane 2**) the splicing pattern of the heterologous splicing minigene harboring *FANCC* TT splice donor was still unaffected by the U1A7 variant (**Fig. 26 D, lane 2 and 3**).

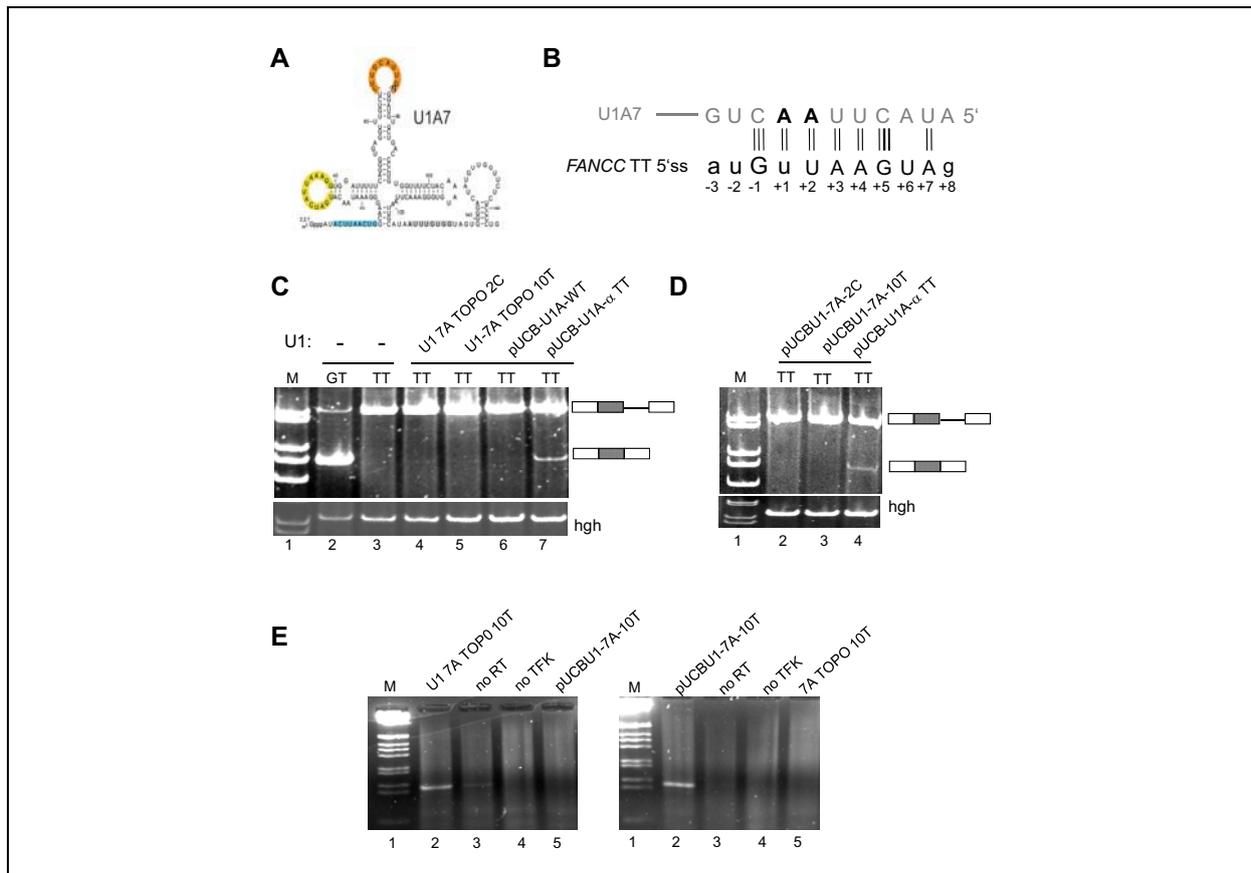


Fig. 26: The U1 snRNA A7 has no effect on usage of the *FANCC* TT splice donor within the splicing minigene.

(A) Predicted secondary structures for the U1A7 snRNA variant. Regions corresponding to important features within the U1A7 snRNA are color coded as specified above. Identified variable positions are highlighted (adapted from Kyriakopoulou et al., 2006 (180)).

(B) Sketch of the H-Bond formation between the free 5' end of U1A7 and the *FANCC* TT 5'ss (positions are numbered). H-Bonds are indicated by vertical lines.

(C) RT-PCR analysis of HeLa cells transfected with the splicing reporter containing the TT 5'ss found in the FA-C patients. U1 wt and U1 αTT indicate co-transfections with the wild-type or TT-adapted U1 snRNA expression plasmids pUCBU1 (lanes 6 to 7), whereas U1A7 TOPO 2C and 10T denote co-transfections with expression plasmids encoding the U1A7 variants differing in a single nucleotide polymorphism (lanes 4 and 5). RT-PCR analysis of hGH was performed to monitor the transfection efficiency.

(D) RT-PCR analysis of HeLa cells transfected with the splicing reporter containing the TT 5'ss found in the FA-C patients along with the U1 snRNA expression plasmids pUCBU1 in which the coding region of U1A snRNA was substituted for the coding region of the U1A7 variants.

(E) Confirmation of U1A7 TOPO 10T and pucBU1-7A10T expression in transfected HeLa cells by RT-PCR.

Results

Therefore, it remained unlikely that the U1A7 variant functions in splicing and that the *FANCC* TT splice donor is recognized by this variant.

2.5.4 Full complementary of the *FGB* TT splice donor to U1 snRNA results in activation of close-by GT dinucleotides

Because increasing the complementarity of the *FANCC* TT splice donor to U1 snRNA has reconstituted splicing at the TT dinucleotide in a heterologous splicing reporter construct, it seemed likely that the finding that the *FANCC* IVS2 +1G>T mutation allowed residual correct splicing at the mutant TT splice donor within its natural context whereas the *FGB* IVS7 +1G>T mutation resulted in complete disruption of the splice donor, was due to a higher intrinsic strength of the *FANCC* splice donor (HBS of 18.7) as opposed to the *FGB* splice donor (HBS of 15.0).

To investigate whether an increased complementarity of the mutant TT splice donor within the *FGB* minigene would allow splicing at the mutant TT splice donor all non-complementary nucleotides except the nucleotide at position +1 within the pathogenic *FGB* exon 7 TT 5'ss were replaced with complementary nucleotides (**Fig. 27A**).

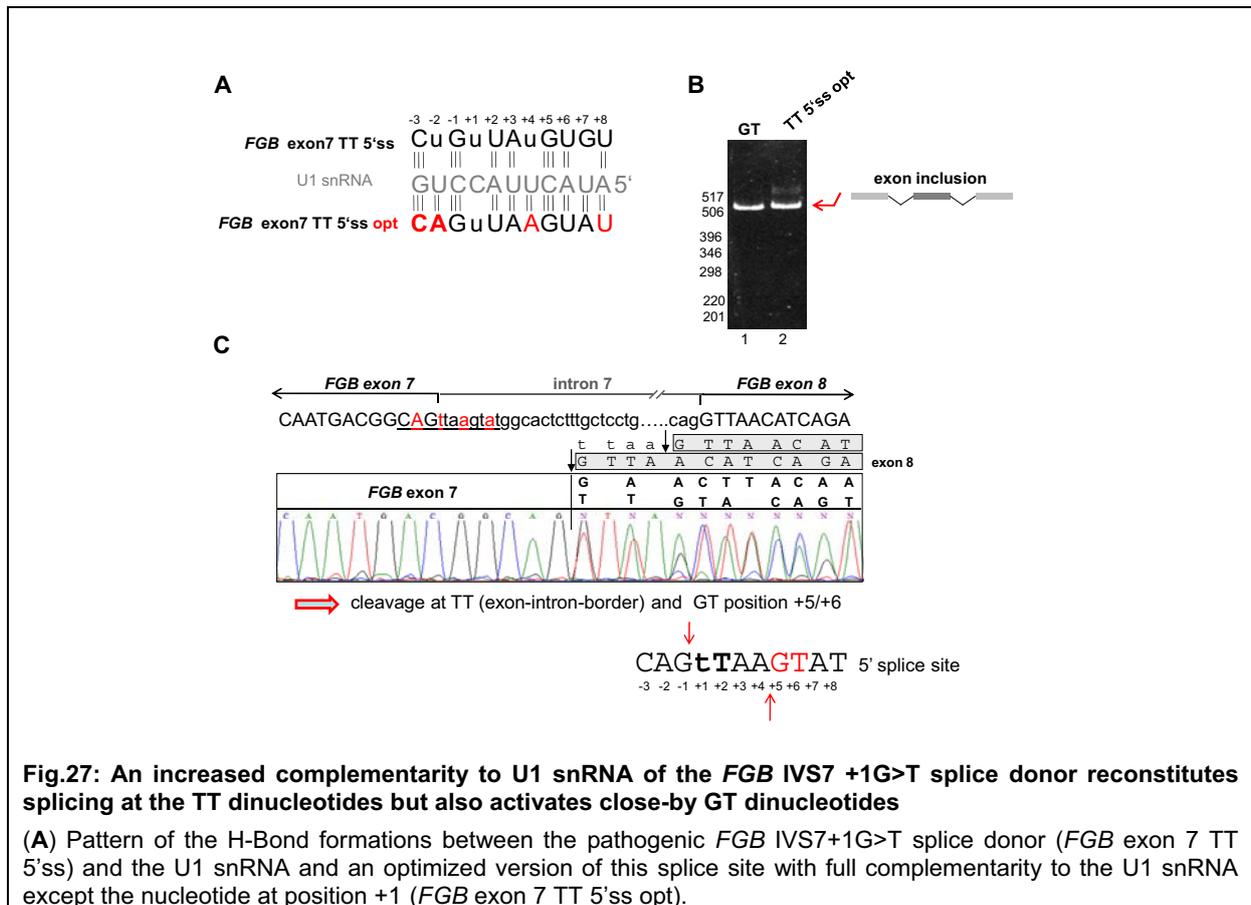


Fig.27: An increased complementarity to U1 snRNA of the *FGB* IVS7 +1G>T splice donor reconstitutes splicing at the TT dinucleotides but also activates close-by GT dinucleotides

(A) Pattern of the H-Bond formations between the pathogenic *FGB* IVS7+1G>T splice donor (*FGB* exon 7 TT 5'ss) and the U1 snRNA and an optimized version of this splice site with full complementarity to the U1 snRNA except the nucleotide at position +1 (*FGB* exon 7 TT 5'ss opt).

Results

(B) RT-PCR analysis of the splicing pattern of the *FGB* minigene harboring the wildtype (GT) or the optimized version of the pathogenic *FGB* IVS7+1G>T splice donor with full complementarity to U1 snRNA except the mutant nucleotide at position +1 (TT 5'ss opt, lane 2).

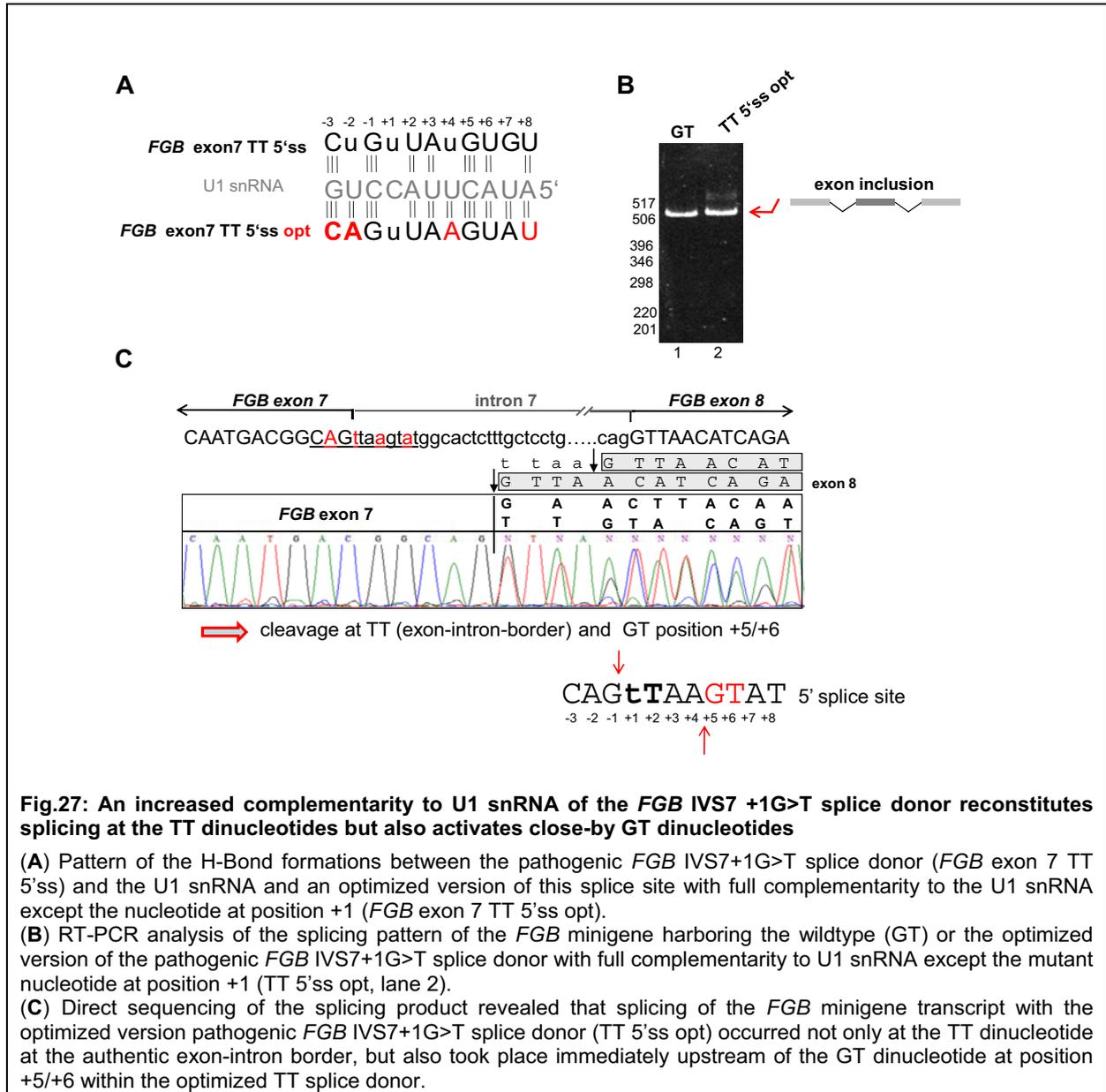
(C) Direct sequencing of the splicing product revealed that splicing of the *FGB* minigene transcript with the optimized version pathogenic *FGB* IVS7+1G>T splice donor (TT 5'ss opt) occurred not only at the TT dinucleotide at the authentic exon-intron border, but also took place immediately upstream of the GT dinucleotide at position +5/+6 within the optimized TT splice donor.

HeLa cells were transiently transfected with the mutated and the wild type *FGB* minigene and RT-PCR was performed to analyze the splicing pattern. As shown in **Figure 27B, lane 2**, the improved complementarity of the TT splice donor within the *FGB* minigene resulted in usage of TT splice donor instead of the cryptic splice donor sites within *FGB* exon 7. However, direct sequencing of the splicing product revealed that splicing of the *FGB* minigene transcript occurred not only at the TT dinucleotide at the authentic exon-intron border, but also took place immediately upstream of the GT dinucleotide at position +5/+6 within the optimized TT splice donor (**Fig. 27C**).

To clarify whether a reduction of the intrinsic strength of the GT at position +5/+6 would eliminate splicing at this site in favor of TT splicing at the exon-intron-border a G>C mutation was introduced at position +9 decreasing the HBS of the GT at +5/+6 from 9.1 to 7.6 without affecting the intrinsic strength of the TT splice donor (**Fig. 29A**). Transfection of the mutated minigene followed by RT-PCR again showed usage of the TT splice donor (**Fig. 28A lane 3**). Sequencing of the splice product nevertheless demonstrated that splicing did not exclusively occur at the TT dinucleotide at the exon-intron-border. Reduction of the intrinsic strength indeed eliminated usage of the GT at position +5/+6 but instead led to activation of the GT dinucleotide one position upstream (position -1) of the TT dinucleotide in addition to splicing at the TT dinucleotide as seen in the *FANCC* context (**Fig. 28B and Fig. 23**).

To clarify whether a reduction of the intrinsic strength of the GT at position +5/+6 would eliminate splicing at this site in favor of TT splicing at the exon-intron-border a G>C mutation was introduced at position +9 decreasing the HBS of the GT at +5/+6 from 9.1 to 7.6 without affecting the intrinsic strength of the TT splice donor (**Fig. 29A**). Transfection of the mutated minigene followed by RT-PCR again showed usage of the TT splice donor (**Fig. 28A lane 3**). Sequencing of the splice product nevertheless demonstrated that splicing did not exclusively occur at the TT dinucleotide at the exon-intron-border. Reduction of the intrinsic strength indeed eliminated usage of the GT at position +5/+6 but instead led to activation of the GT dinucleotide one position upstream (position -1) of the TT dinucleotide in addition to splicing at the TT dinucleotide as seen in the *FANCC* context (**Fig. 28B and Fig. 23**).

Results



Apart from that, along with optimization of the *FGB* TT splice splice donor in a low amount of the minigene transcripts the second intron remained unspliced (**Fig. 28A**), suggesting that in this splicing reporter an increased complementarity of the 5'ss to the U1 snRNA might lead to a hyperstabilization of the U1 snRNA which was shown before to have no influence on mammalian 5'ss recognition (Freund et al., 2005 (92)).

Inspection of the *FANCC* sequence revealed that the splice donor sequence of *FANCC* exon 2 also contained a GT dinucleotide at position +5/+6 and one at position+8/+9 (**Fig. 29B**). Assessment of the intrinsic strength of these splice sites, i.e., GT at position +5/+6 and at position +8/+9 revealed an HBS of 7.6 and 4.5, respectively.

Results

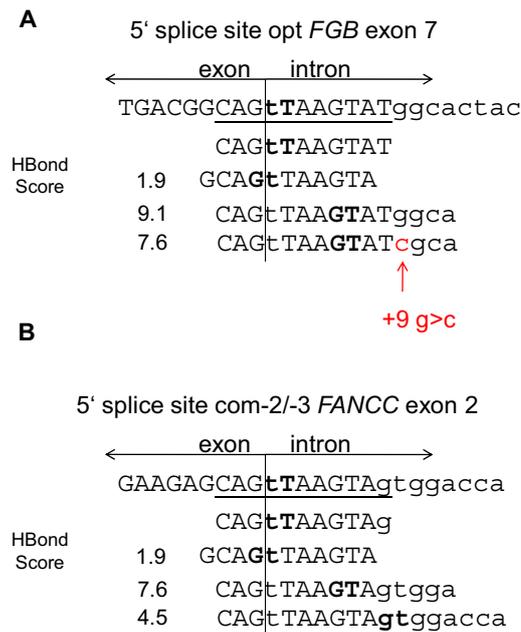


Fig.29: Assessment of the intrinsic strength of close-by GT dinucleotides within the *FGB* and *FANCC* TT splice donor

(A) Sequence of the optimized *FGB* exon 7 TT splice donor (underlined) and sequence immediately upstream and downstream of this splice site. The intrinsic strength of close-by GT dinucleotides was calculated using the HBond score algorithm. The +9 G>C mutation reduces the complementarity of the GT dinucleotide at position +5/+6 to U1 snRNA.

(B) Sequence of the optimized *FANCC* exon 2 TT splice donor (5' splice site com -2/-3 *FANCC* exon 2, underlined) and sequence immediately upstream and downstream of this splice site. The intrinsic strength of close-by GT splice sites was calculated using the HBond score algorithm.

Therefore, these results demonstrated that usage of a TT dinucleotide as a splice donor site upon disruption of the canonical GT is dependent on the overall complementarity of the splice donor site. However, similar to the phenomenon of cryptic splice site activation upon weakening a physiological splice donor by a mutation a high complementary TT splice donor remained vulnerable to close-by GT dinucleotides competing with the TT splice donor. Nevertheless, the results showed that the intrinsic strength of close-by GT splice sites is crucial for their competition with the TT dinucleotide as evident by the observation that a reduction of the HBS from 9.1 to 7.6 was sufficient to eliminate the usage of the GT dinucleotide at position +5/+6. Yet, low complementarity of the GT dinucleotide at position +5/+6 within the TT splice donor sequence of otherwise full complementarity in the three exonic positions (positions -3 to -1 of the 5'ss) contributing to an overall enhanced complementarity of the TT splice donor induced usage of the TT splice donor within both the heterologous splicing minigene harboring *FANCC* exon 2 and the *FGB* minigene - but also led to simultaneous activation of the GT at position -1/+1 despite its very low HBond score value of 1.9 (Fig. 29).

Results

As the TT splice donor sequence found in the FA-C patients with an HBS of 18.7 (calculated with GT instead of TT) with mismatches to U1 snRNA at positions – 3, –2 and +8 allowed exclusive splicing at the natural exon-intron-border necessary for generation of protein-encoding transcripts it appeared that a sufficient intrinsic strength was necessary for activation of a non-canonical TT splice donor yet full complementarity of the exonic positions was detrimental for exclusive splicing at the exon-intron-border.

2.5.5. Intrinsic features of the TT splice donor sequence determine the exclusive usage of the TT as splice donor

Given that co-transfection of an U1 snRNA expression plasmid adapted to the *FANCC* TT splice donor (HBS 18.7 calculated with GT instead of TT) reconstituted exclusive TT splicing at the correct exon-intron-border within the heterologous splicing minigene harboring *FANCC* exon 2 with its flanking splice sites this system allowed to determine sequence requirements of the splice donor allowing the usage of a non-canonical TT dinucleotide. For this purpose the sequence of the *FANCC* TT splice donor was replaced by the sequence of the *FGB* TT splice donor within the heterologous splicing minigene harboring *FANCC* exon 2. In order to narrow down requirements of the TT splice donor sequence necessary for TT splicing the sequence of the *FGB* TT splice donor was successively adapted to the sequence of the *FANCC* splice donor. As the *FANCC* TT splice donor was used in the heterologous construct only if an adapted U1 snRNA expression plasmid was co-transfected for each TT splice donor sequence an adapted U1 snRNA expression was designed and co-transfected with the respective heterologous splicing reporter minigene.

Analysis of the splicing pattern by RT-PCR revealed that along with the co-transfection of an adapted U1 snRNA molecule usage of the TT splice donor was already observable if the non-complementary nucleotide at position +4 within the *FGB* splice donor sequence was replaced by a complementary nucleotide (+4A, HBS of 18.30 calculated with GT) (**Fig. 30B, lane 7**). Upon co-transfection of an adapted U1 snRNA molecule, usage of the TT splice donor within the context of the *FANCC* reporter minigene therefore seemed to require complementary bases at position -1, +2, +3, +4, +5 and +6 (**see Fig. 30A**). Mismatches at positions -3 and -2 as in the original *FANCC* splice donor could be compensated by co-transfection of the adapted U1 snRNA molecule (-3A, +4A) as long as the nucleotides from position -1 up to position +6 were complementary ones (**Fig. 30B, lane 15**). In this case a complementary nucleotide at position +7 appeared to be negligible for usage of the TT splice donor (**Fig. 30B, lane 15**). However, direct sequencing of the splice products demonstrated that almost exclusive usage of the TT splice donor at the exon-intron-border required the

Results

Taken together, exclusive TT splicing at the exon-intron-border seemed to require non-complementary nucleotides in the two most upstream exonic positions (position -3, -2) and complementary nucleotides from position -1 up to +7. Therefore, the data evidently demonstrated that the exclusive splicing at the non-canonical TT at the exon-intron-border was not simply determined by the free energy of the RNA duplex formed between the splice donor and the matching U1 snRNA but was predominantly dependent on the 5'ss sequence itself.

2.5.6. The genomic context of FANCC exon 2 enhances splicing at the pathogenic TT splice donor

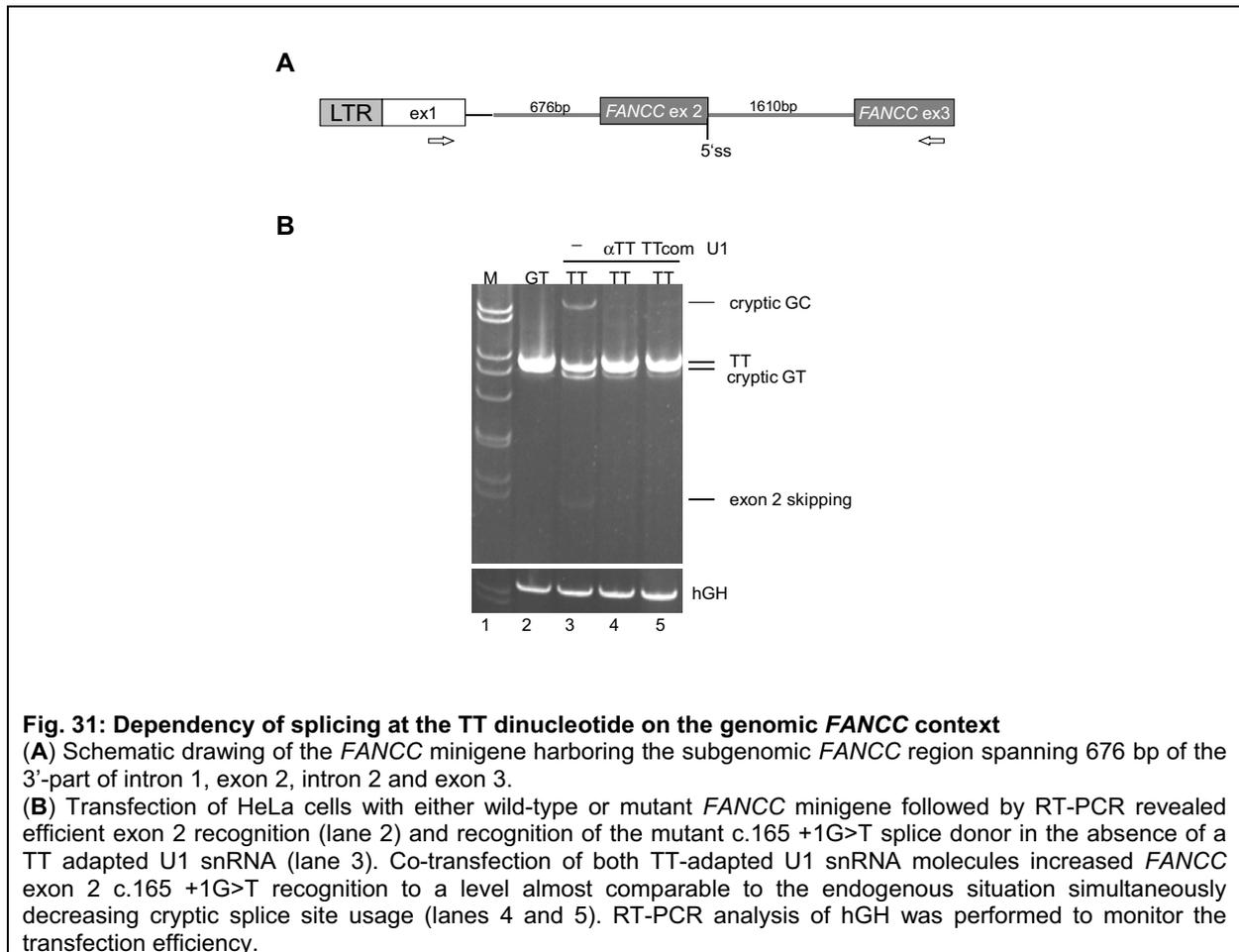
(cited from Hartmann et al., 2010 (122))

Because recognition of *FANCC* exon 2 in the heterologous splicing reporter minigene was not as effective as within the endogenous *FANCC* transcript and the mutant TT splice donor was only recognized upon an increased complementarity to U1 snRNA or upon co-transfection of a compensatory U1 snRNA molecule it seemed likely that recognition of the splice donor of *FANCC* exon 2 is enhanced by its genomic context.

To determine whether an extended subgenomic context improved the recognition of the *FANCC* exon 2 splice donor the genomic context of *FANCC* exon 2 within the minigene was extended to a region spanning 676 bp of the upstream intron, *FANCC* exon 2, intron 2 and exon 3 (**Fig. 31A**). Transfection of HeLa cells with this minigene harboring the wild-type GT or mutant TT *FANCC* exon 2 splice donor followed by RT-PCR revealed efficient *FANCC* exon 2 inclusion within the wild type minigene (**Fig. 31B, lane 2**). Remarkably, within the extended *FANCC* context the mutant TT splice donor was now efficiently recognized without co-transfection of the TT adapted U1 snRNA (**Fig. 31B, lane 3**), albeit recognition of the mutant TT splice donor was less efficient as opposed to the wild type GT splice donor (**Fig. 31B, compare lanes 2 and 3**). The GT to TT mutation of the splice donor site caused activation of a cryptic GT splice donor in exon 2 and a cryptic GC splice donor in intron 2 accompanied by low-level exon skipping as seen in the endogenous transcript (**Fig. 31B, lane 3 and Fig. 21C, lane 3**).

Nevertheless, recognition of the mutant TT splice donor site could be enhanced by co-transfection of both U1 snRNA α TT and TTcom achieving a level almost comparable to the wild type GT splice donor and simultaneously decreasing cryptic splice site usage (**Fig. 31B, lanes 4 and 5**).

Results



These results demonstrated that the genomic context of *FANCC* exon 2 enhanced *FANCC* exon 2 definition supporting recognition of the mutant TT splice donor site. Therefore, usage of the unusual TT splice donor site as seen in the Fanconi anemia patients was not only determined by the intrinsic features of the *FANCC* splice donor sequence but was also profoundly enhanced by enhancing sequences within the *FANCC* context not yet identified. Furthermore, they suggested that ectopic expression of the TT-adapted U1 snRNA molecules may also improve the recognition of the mutant *FANCC* TT splice donor within the endogenous *FANCC* transcript in the patient-derived fibroblasts.

2.6. An U1 snRNA based therapy approach for human splice donor mutations

2.6.1. Ectopic expression of the TT-adapted U1 snRNAs specifically enhances the amount of the endogenous in-frame transcript in FA patient-derived fibroblasts

(cited from Hartmann et al., 2010 (122))

As the TT-adapted U1 snRNAs improved the usage of the *FANCC* TT splice donor site within the minigenes, these results implied that transfection of the biallelic *FANCC* c.165 +1G>T patient-derived fibroblasts with the TT-adapted U1 snRNA molecules would improve recognition of the pathogenic TT splice donor and enhance the levels of endogenous TT-spliced in-frame transcripts. For this purpose, the primary fibroblasts from pedigree 526 were immortalized by a lentivirus expressing the SV40 large T cDNA and then transfected with the TT-adapted U1 snRNAs. By using RT-PCR primers in the 5' UTR and in exon 4 to distinguish the different endogenous *FANCC* transcripts in the biallelic *FANCC* c.165 +1G>T patients' cells (**Figure Fig. 32A**), it could be shown that transfection of patient-derived fibroblasts with the TT-adapted U1 snRNAs U1 α TT or the U1 TTcom specifically increased the amount of the TT-spliced in-frame transcript from 30% to 56% and 58%, respectively (**Fig. 32A, B**). Direct sequencing of both splice products confirmed that they were accurately spliced at the correct exon-intron border (**Fig. 32C**). Concomitantly, the amount of all aberrantly spliced transcripts decreased, indicating that the two mutant TT-adapted U1 snRNAs were capable of improving exon recognition and thereby facilitating production of the correctly spliced in-frame transcript. Thus, ectopic expression of the TT-adapted artificial U1 snRNAs significantly increased the usage of the pathogenic TT 5'ss in the patients' fibroblasts.

2.6.2. Phenotypic correction of *FANCC*-mutant fibroblasts by integrating lentivirus-mediated expression of TT-adapted U1 snRNAs

(cited from Hartmann et al., 2010 (122))

Permanent suppression of splice donor mutations in cells still actively dividing requires that the mutation-adapted U1 snRNA integrates into the genome of the mutant cells. As retroviruses are evolutionary optimized gene delivery systems for stably introducing foreign cDNA into the cellular DNA, the two TT-adapted U1 snRNA expression cassettes were transferred into a lentiviral vector (LV) which co-expressed the neomycin phosphotransferase (neoR) cDNA in opposite orientation (**Fig. 33A**).

Results

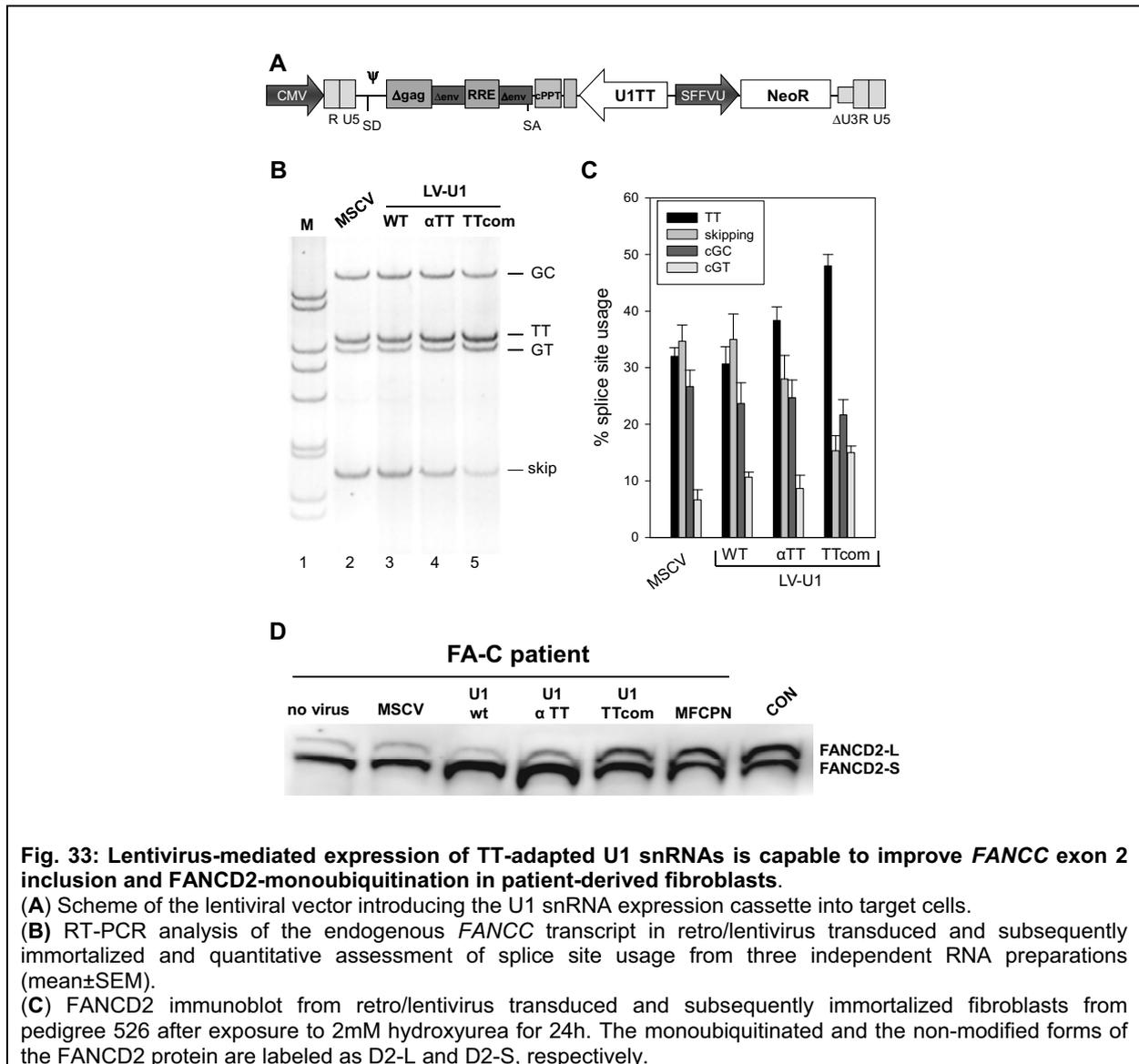


Fig. 33: Lentivirus-mediated expression of TT-adapted U1 snRNAs is capable to improve *FANCC* exon 2 inclusion and *FANCD2*-monoubiquitination in patient-derived fibroblasts.

(A) Scheme of the lentiviral vector introducing the U1 snRNA expression cassette into target cells.

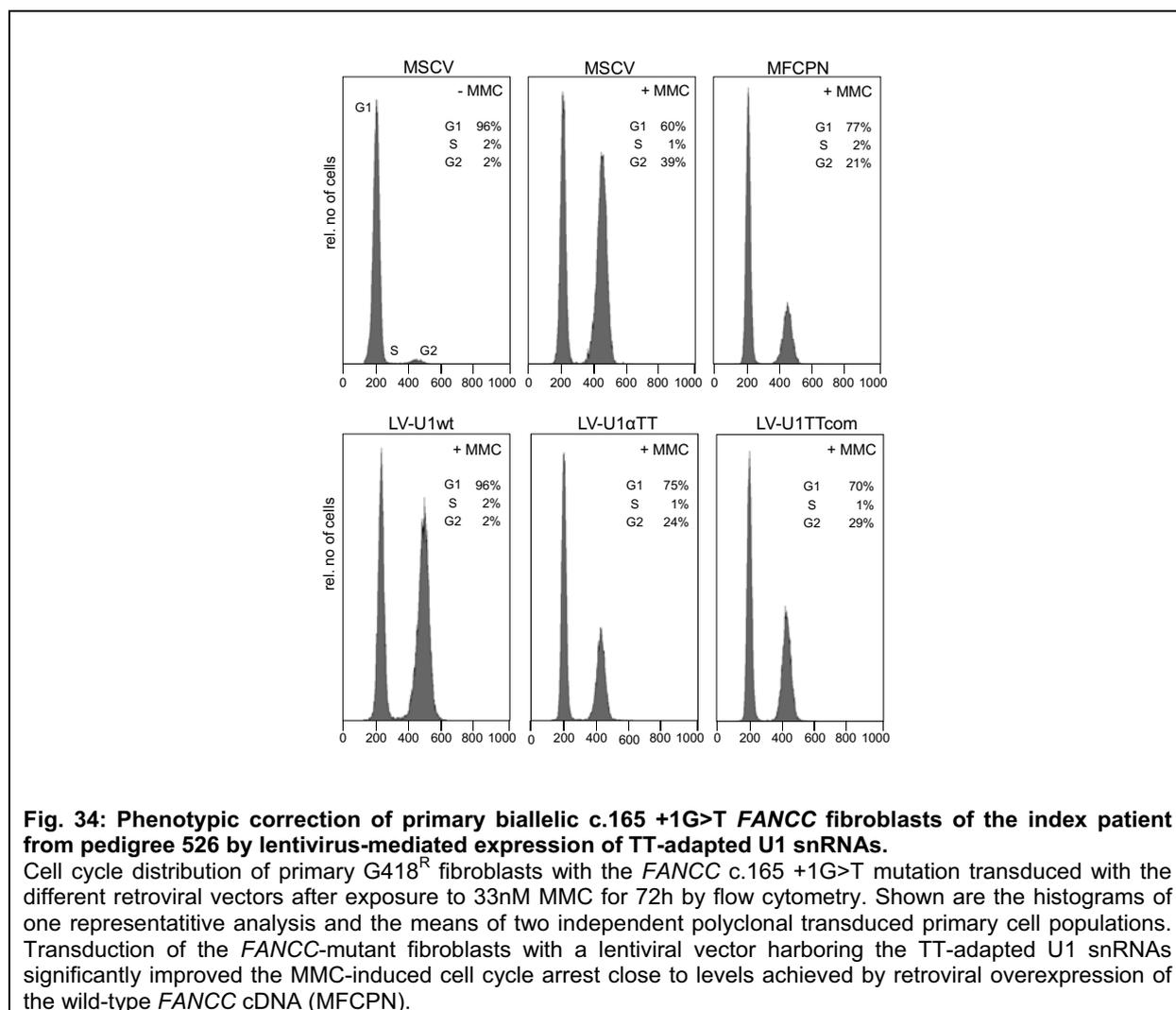
(B) RT-PCR analysis of the endogenous *FANCC* transcript in retro/lentivirus transduced and subsequently immortalized and quantitative assessment of splice site usage from three independent RNA preparations (mean±SEM).

(C) *FANCD2* immunoblot from retro/lentivirus transduced and subsequently immortalized fibroblasts from pedigree 526 after exposure to 2mM hydroxyurea for 24h. The monoubiquitinated and the non-modified forms of the *FANCD2* protein are labeled as D2-L and D2-S, respectively.

Transduction of the patient's fibroblasts carrying the pathogenic c.165 +1G>T mutation on both alleles with a retroviral vector containing the wild-type *FANCC* cDNA (MFCPN) corrected the MMC induced G2 arrest, whereas cells transduced with the mock vector (MSCV) exhibited a prominent G2 phase arrest typical for FA (Fig. 34). Expression of both TT-adapted U1 snRNAs (αTT, TTcom) in the LV vector significantly improved the MMC-induced cell cycle arrest while stable expression of the wild-type U1 snRNA did not influence the cell cycle distribution of the primary cells (Fig. 34). RT-PCR analysis of endogenous *FANCC* transcript in transduced and immortalized fibroblasts from pedigree 526 confirmed that the usage of the mutant TT splice donor and *FANCC* exon 2 recognition was clearly improved by lentivirus-mediated expression of both TT adapted U1 snRNAs (Fig. 33 B and C). An indication of a functional FA core complex with normal *FANCC* protein as an essential

Results

component is the mono-ubiquitination of the FANCD2 protein in response to exposure to DNA interstrand cross-linking agents (Kalb et al., 2007 (148)). As FANCD2 western blot analysis on primary fibroblasts is difficult due to the minute amounts of FANCD2 protein present, the primary fibroblasts from were immortalized with a lentivirus expressing the SV40 large T cDNA. Subsequently, the immortalized fibroblasts were exposed to 2mM hydroxyurea for 24h and then protein harvested as published (Garcia-Higuera et al. 2001 (97)). Results revealed that the immortalized fibroblasts with the biallelic *FANCC* c.165 +1G>T mutation already had minute levels of mono-ubiquitinated protein even in the absence of any U1 snRNA transfected (**Fig. 33D**). Residual activity of the FANCC protein could also be confirmed by formation of FANCD2 foci within the patient-derived fibroblasts upon challenge with MMC (**Fig. 35**).



Results

These results indicated that the lower levels of TT-spliced endogenous in-frame transcript encoded residual functional FANCC protein that was active in the FA core complex most likely accounting for the milder clinical phenotype of FA-C in these patients.

Finally, the results showed that transduction of the *FANCC* c.165 +1G>T fibroblasts with each of the two TT-adapted U1 snRNA expression constructs increased the levels of mono-ubiquitinated FANCD2 protein corresponding to the mRNA splicing pattern shown in **Figure 33C**. While expression of U1 snRNA TTcom resulted in FANCD2 monoubiquitination level comparable to the one induced by expression of the wild-type *FANCC* cDNA (MFCPN) and in the normal control (CON), expression of the U1 α TT was less effective on both mRNA and protein level (**Fig. 33C, D**).

These data indicated that stably lentivirus-mediated expression of the TT-adapted U1 snRNA molecules can lead to the production of sufficient amounts of endogenous functional *FANCC* transcript for restoration of the FA pathway and correction of the cellular FA phenotype of DNA cross-linker hypersensitivity, thus demonstrating the potential of lentivirus-mediated transfer of splice site mutation-adapted U1 snRNA genes as curative therapeutic strategy for splice donor site mutations in FA.

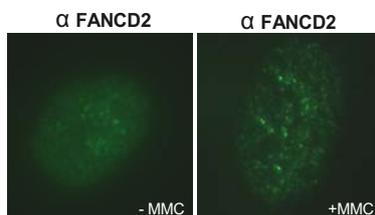


Fig. 35: Function of FANCD2 in biallelic *FANCC* c.165 +1G>T fibroblasts.

Formation of FANCD2 foci by immunofluorescent antibody staining of biallelic *FANCC* c.165 +1G>T fibroblasts after incubation for 24h in 150 nM mitomycin C (MMC). Shown is one representative analysis from three independent stainings.

2.6.3. Delivery of extended compensatory U1 snRNA molecules can improve exon recognition within patient-derived *FANCC* c.456+4A>T fibroblasts

Because the results from the *FANCC* c.165+1G>T splice donor mutation were very promising it was desirable to test the applicability of the approach to founder mutations in Fanconi anemia (FA) such as the highly frequent *FANCC* c.456+4A>T (IVS4 +4A>T) within the splice donor of *FANCC* exon 5 (Whitney et al., 1994 (358); Verlander et al., 1995 (347), Futaki et al., 2000 (94)). Due to a single point mutation causing a nucleotide substitution from

Results

A>T at position +4 within the splice donor of *FANCC* exon 5 the intrinsic strength of this splice donor site is severely decreased as evident from the reduction of the HBS from 15.5 to 10.1 (**Fig. 36A**).

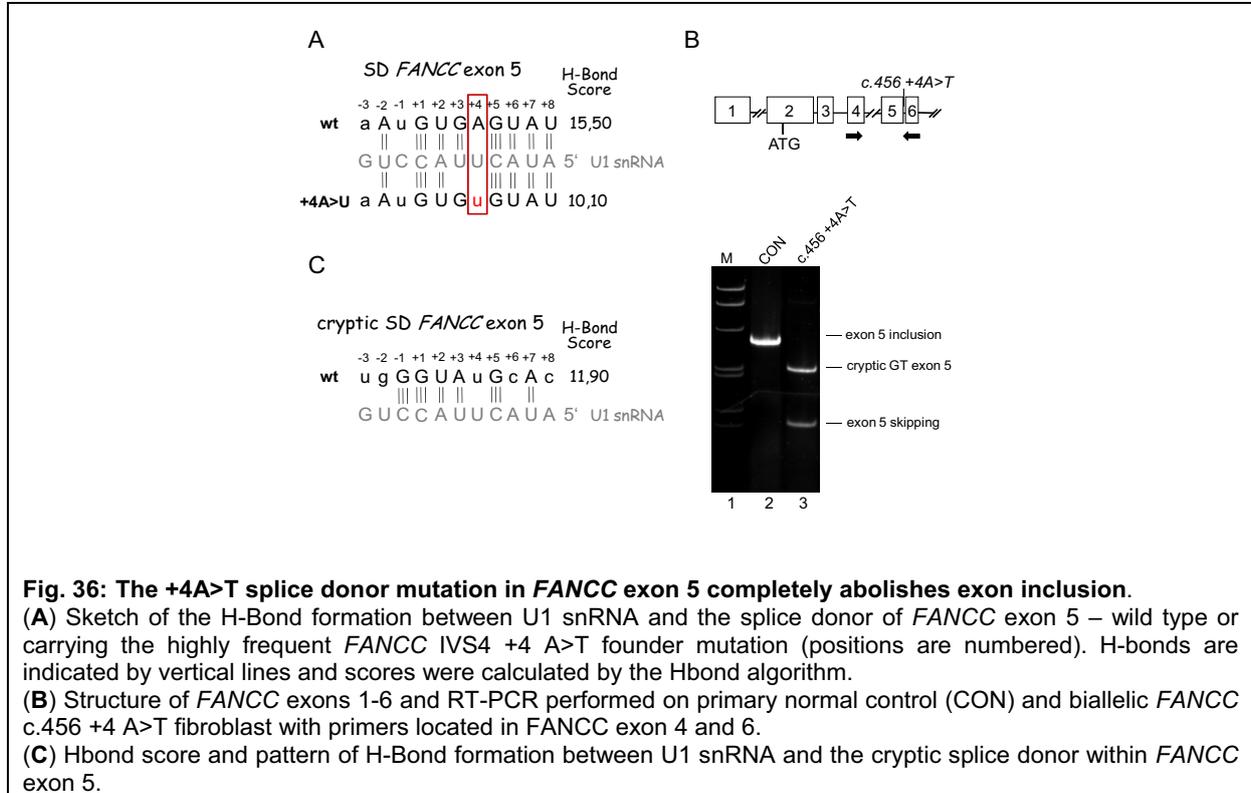


Fig. 36: The +4A>T splice donor mutation in *FANCC* exon 5 completely abolishes exon inclusion.

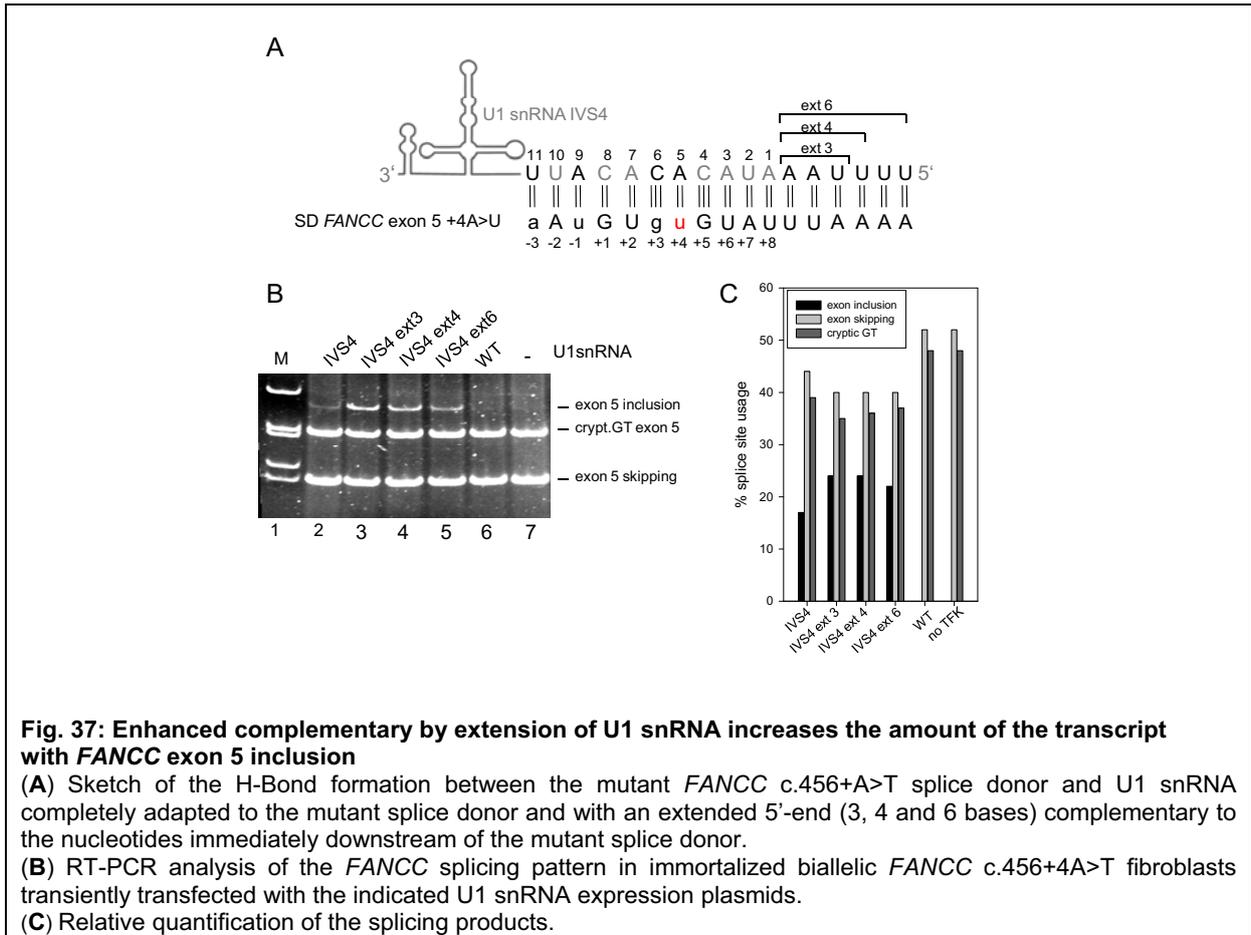
(A) Sketch of the H-Bond formation between U1 snRNA and the splice donor of *FANCC* exon 5 – wild type or carrying the highly frequent *FANCC* IVS4 +4 A>T founder mutation (positions are numbered). H-bonds are indicated by vertical lines and scores were calculated by the Hbond algorithm.

(B) Structure of *FANCC* exons 1-6 and RT-PCR performed on primary normal control (CON) and biallelic *FANCC* c.456 +4 A>T fibroblast with primers located in *FANCC* exon 4 and 6.

(C) Hbond score and pattern of H-Bond formation between U1 snRNA and the cryptic splice donor within *FANCC* exon 5.

Analysis of the splicing pattern of RNA of patient-derived fibroblasts (homozygous for the mutation) demonstrated that the mutation caused skipping of *FANCC* exon 5 and activation of a cryptic splice donor (HBS of 11.9) within this exon (**Fig. 36B and C**). To test whether transfection of an U1 snRNA molecule adapted to every nucleotide of the mutant exon 5 splice donor could restore exon inclusion SV40 large T-antigen immortalized c.456+4A>T fibroblasts were transiently transfected with the respective U1 snRNA expression plasmid (U1 IVS4, **Fig. 37A**). However, only very low-level *FANCC* exon 5 inclusion was detectable (17 % of the transcripts) while exon 5 skipping and usage of the cryptic splice donor were still predominant (44% and 39% respectively) (**Fig. 37B, lane2**).

Results



In order to improve exon inclusion mediated by U1 snRNA IVS4 the 5'-end of this U1 snRNA molecule was extended by 3, 4 and 6 additional nucleotides complementary to the *FANCC* sequence immediately downstream of the splice donor of *FANCC* exon 5 (**Fig. 37A**). These additional nucleotides at the 5' end of the U1 snRNA should further stabilize the RNA duplex formation between the mutant *FANCC* splice donor and the U1 snRNA molecules. Indeed, transfection of immortalized c.456+4A>T fibroblasts with the U1snRNA IVS ext 3, ext 4 and 6 resulted in enhanced *FANCC* exon 5 inclusion (24%, 24% and 22% respectively).

Therefore, extension of the 5' end of suppressor U1 snRNAs can improve RNA duplex formation between the U1 snRNA molecule and the mutant splice donor and thereby increasing the definition of *FANCC* exon 5 within the mutant pre-mRNA.

2.6.4. Delivery of an extended suppressor U1 snRNA does not produce sufficient amounts of endogenous functional *FANCC* transcript for phenotype correction of c.456 +4A>T fibroblasts

To determine whether the exon inclusion level achieved by the U1 snRNA IVS4 variants was sufficient to improve the phenotype of primary c.456+4A>T fibroblasts expression cassettes for U1 snRNA IVS4 and for two of the most efficient U1 snRNAs IVS4 ext 3 and 4 were transferred into lentiviral vectors (LV) which co-expressed the neomycin phosphotransferase (neoR) cDNA in opposite orientation. Primary c.456+4A>T fibroblasts were transduced with these lentiviral vectors and additional ones carrying the U1 snRNA wild type, U1 snRNA TT com and U1 snRNA 5A with a single compensatory mutation. G418 resistant fibroblasts were exposed for three days to 33 nM MMC and then analyzed by flow cytometry for their cell cycle distribution.

The histograms confirmed that transduction of the patient's fibroblasts with a retroviral vector containing the wild-type *FANCC* cDNA (MFCPN) corrected the MMC induced G2 arrest (38% of the cells in G2), whereas cells transduced with the mock vector (MSCV) exhibited a prominent G2 phase arrest typical for FA (65% of the cells in G2) (**Fig. 38**). Fibroblasts transduced with the control U1 snRNAs (WT, 5A, TTcom) had about 68% of cells in G2 phase after MMC challenge whereas in fibroblasts transduced with U1 snRNA IVS4, IVS4 ext 3 and ext 4 about 60% of the cells arrested in the G2 phase of the cell cycle (**Fig. 38**).

Therefore, transduction of the c.456 +4A>T fibroblasts with the mutation adapted U1 snRNA molecules U1 snRNA IVS4, IVS4 ext 3 and ext 4 seemed to slightly reduce the G2 arrest of c. 456+4A>T fibroblasts. Hence, the *FANCC* exon 5 inclusion level mediated by these U1 snRNA molecules was not sufficient for phenotypic correction of the fibroblasts.

These results indicated that a combination of the intrinsic strength of a given splice donor sequence, the position of the mutation within this sequence and features of the natural sequence context e.g. the presence of *cis*-regulatory enhancer sequences severely influence whether a splice donor mutation can be cured by the delivery of mutation adapted U1 snRNA molecules.

Results

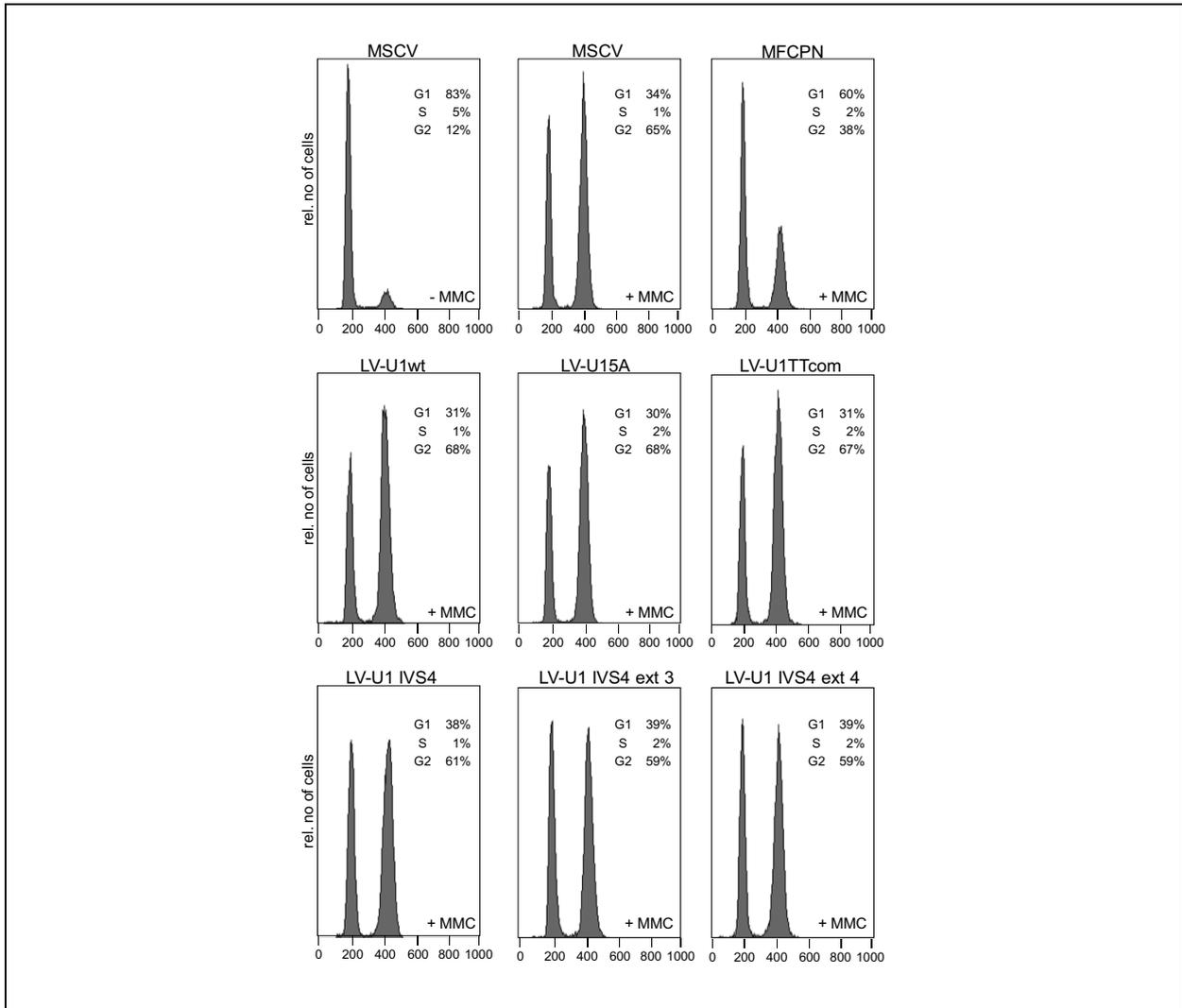


Fig. 38: Lentivirus-mediated expression of gene specific extended full complementary U1 snRNAs does not generate sufficient amounts of the functional transcript for phenotypic correction of the +4A>T splice donor mutation in *FANCC* exon 5.

Flow cytometry analysis of the cell cycle distribution of primary G418^R fibroblasts carrying the *FANCC* c.456 +4A>T mutation transduced with retroviral vectors after exposure to 33 nM MMC for 72 h. Shown are the histograms of one representative analysis.

3. Discussion

Human genes are discontinuous such that the exons whose sequence will be translated into an amino acid sequence for a protein are interrupted by intervening intron sequences, which are excised during the maturation of the final messenger RNA being a spliced product. In a typical human gene the exons are relatively short - about 50 - 250 base pairs in length - whereas the introns comprise hundreds to thousands base pairs (Fox-Walsh et al., 2005 (89)). This gene architecture, and the predominant exon skipping phenotype of splice site mutations, is consistent with the idea that in mammals splice sites are predominantly recognized in pairs across the exon through a process called exon definition (Robberson et al., 1990 (262); Sterner et al., 1996 (323)). Exon definition involves initial interaction across the exon between factors recognizing the 5' splice site (5'ss) or the upstream 3' splice site (3'ss), whereas in the alternative model intron definition, interactions occur first across the intron between factors recognizing the 5'ss and the downstream 3'ss (Berget, 1995 (18)). Exon recognition during pre-mRNA is mainly dependent on the strength of its flanking splice sites but further depends on *cis*-regulatory elements in the pre-mRNA that modulate splice site selection. Several intercommunicating layers of *cis*-acting elements that distinguish exons from introns appear to be particularly dense within and around exons (Fairbrother et al., 2002 (80), Sorek & Ast, 2003 (311)). These elements direct the spliceosome to the correct nucleotides for exon joining and intron removal as they serve as binding sites for *trans*-acting factors that regulate splicing. Most exons contain exonic splicing enhancers (ESEs), which define them as recognition units promoting the use of their splice sites (Cartegni et al., 2003; Fairbrother et al., 2004). In addition, exons also contain functional splicing suppression units known as exonic splicing silencers (ESSs) (Wang et al., 2004 and 2006 (353,354)). Moreover, intronic splicing enhancers (ISEs) or intronic splicing silencers (ISSs) enhance or repress the use of nearby 5' or 3' ss (Ponthier et al., 2006 (246); Kashima et al., 2007 (159)).

Genetic changes that interfere with pre-mRNA splicing are commonly associated with human genetic diseases. Splicing regulatory elements (SREs) are sensitive targets of nucleotide alterations: even single DNA mutations can strengthen, weaken or destroy a splice site or *cis*-regulatory element, or create a new one, and may thus lead to observable phenomena on RNA level like exon skipping, activation of cryptic or *de novo* splice sites, or intron retention. Although it has been recognized that splicing regulatory elements act in concert, and their interactions and dependencies play an important role in splice site functionality, *in silico* implementation of the comprehensive splicing machinery is still limited to a variety of independent algorithms scoring splice sites and/or *cis*-regulatory elements. These dedicated scores for 5' ss or 3' ss, as well as exonic or intronic splice enhancers or silencers, and have

been applied to the prediction with considerable success (for an overview, see Hartmann et al., 2008 (123)). However, the meaningful combination of *cis*-regulatory elements and splice site scores into a single functional measure still remains to be achieved. Reliable diagnosis of the splicing phenotype of a splice site mutation still requires functional splicing assays. This is of particular interest because most patients are genotyped only and through identification of disease-specific genes, genetic testing has found its way into clinical routine and supports a variety of clinical decisions in many common diseases and cancer syndromes.

The definitive test of whether a suspected disease causing mutation affects splicing ideally comes from RNA analysis of the affected tissue, as splicing mutations can have cell specific effects. If diagnostic RNA-level information is not available, the genomic segment comprising the mutation can be generated by PCR amplification directly from genomic DNA and inserted into an artificial minigene reporter construct in order to compare the splicing pattern of the mutant and wild type exon in a human cell line.

3.1. *In vivo* analysis of human exon recognition in a heterologous minigene

Analysis of human exon recognition within a minigene requires a system comprising a minimum of three exons and two introns (Baralle & Baralle, 2005 (17)). In most human genes this would require handling of several thousand nucleotides due to the large size of most human introns. In a heterologous splicing minigene the exon of interest is inserted along with its flanking splice sites and only a small amount of the natural flanking intron sequence. This has the advantage that only short DNA fragments need to be handled in order to validate a putative pathogenic splicing mutation or to identify an element that regulates exon recognition. In this thesis a heterologous splicing reporter minigene construct was used to validate putative pathogenic patient-derived splice donor mutations. The first exon of the heterologous splicing reporter construct was derived from HIV-1 and the third exon was a hybrid of the CAT (chloramphenicol-acetyl-transferase) open reading frame and the sequence of the HIV-1 RRE (rev responsive element). Immediately downstream of the first exon the heterologous construct contained the strong HIV-1 5' ss #1 - which is also called splice donor 1 or SD1/4 - with an HBond score of 20.8 , followed by 68 base pairs of the HIV-1 intron 1. The 3' half of the construct was composed of intron 2 and an HIV-1 derived 3' splice site (3'ss) - which is also called splice acceptor (SA).

Although the use heterologous splicing reporter minigene constructs for the analysis of putative pathogenic splicing mutations has been described elsewhere (Cooper, 2005 (64); Bonnet et al., 2008 (34)), the influence of the minigene sequences including the strength of its splice sites on exon recognition has never been investigated. Much attention has centered

on the exon sequence and the strength of its flanking splice sites. It could be shown that the level of internal exon inclusion is affected by the strength of both exon flanking splice sites (Neveling, K. diploma thesis, 2004 (222), Shepard et al., 2011 (299)). Nevertheless, results in our working group gave leads to the assumption that the strength of the 3'ss within the heterologous splicing reporter minigene can have a profound impact on exon recognition. In depth analysis in the context of this thesis demonstrated that the human *ATM* exon 54 was not simply defined by its exon sequence and its flanking splice sites. The level of *ATM* exon 54 inclusion was seen to be dependent on the strength of the 3'ss of the splicing reporter minigene and could be increased by improvement of the intrinsic strength of the 3'ss (**Fig.1**). This observation was consistent with the biochemistry of the splicing reaction. Regarding splicing catalysis, the cross-exon complex of splicing factors must be converted into a cross-intron complex allowing intron removal (Reed, 2000 (256)). Since recent data have suggest that the regulation of exon inclusion or skipping occurs during the switch from the cross-exon to a cross-intron complex (House & Lynch, 2006 (135); Bonnal et al., 2008 (33); Sharma et al., 2008 (295)), it seems natural that the splicing outcome of a minigene can be influenced by the strength of its splice sites. As a strong terminal splice acceptor was necessary for intron removal and recognition of *ATM* exon 54 in the context of the heterologous splicing reporter minigene a 3'ss with improved intrinsic strength was chosen for the establishment of a prototype heterologous splicing reporter minigene.

The *ATM* exon 54 - which was used as a prototype human exon for the analysis of exon recognition in the heterologous splicing reporter in this thesis - is characterized by a 3' ss with intermediate strength (MaxEnt score = 6.96) and by a weak 5'ss (HBond score = 12.3). It is known that when the 3' or 5' splice site is strong, most internal exons are efficiently recognized (Shepard et al., 2011 (299)). In contrast, it is assumed that exons with weak splice sites are not accurately spliced without the aid of additional enhancer elements. Accordingly, the RESCUE-ESE approach found a significant enrichment of hexamers with potential enhancing function in exons with weak splice sites (Fairbrother et al., 2002). Moreover, in a comparative analysis of human and mouse genomes, intronic sequences with a high degree of conservation were identified proximal to the enclosed exons (Sorek & Ast, 2003 (310)). It appeared therefore that the intronic sequence that flanks an exon might be crucial for exon definition especially for exons with weak splice sites. This should be taken in consideration for the analysis of exon recognition in a heterologous splicing reporter minigene.

The results of this thesis showed that the natural intron sequence immediately downstream of the weak splice donor site of *ATM* exon 54 had a profound influence on the recognition of this exon in the heterologous splicing reporter minigene (**Fig.2**). Since intronic regulatory

Discussion

elements seem to be preferentially located close to the splice sites in an initial test experiment a short fragment of only 55 base pairs of the natural intron sequence immediately downstream of the weak splice donor of *ATM* exon 54 was inserted together with this exon in the heterologous splicing reporter minigene. The results obtained here demonstrated that in the presence of the natural intron segment the strength of the 3'ss of splicing reporter minigene was negligible for the recognition of *ATM* exon 54. In the presence of the natural intron segment *ATM* exon 54 was also recognized when the strength of the 3' splice site within the reporter system was intermediate. Without the natural intron segment immediately downstream of the weak splice donor of *ATM* exon 54 the exon was only included in the reporter transcript if the 3'ss of the heterologous splicing reporter minigene was strong. In order to localize a putative splicing regulatory element within the intron sequence the sequence was dissected into three parts of equal length. Surprisingly, the presence of each part immediately downstream of the splice donor of *ATM* exon 54 improved the recognition of this exon in the heterologous minigene. However, the second segment was less efficient indicating a sequence specific effect of the intron sequence on exon recognition (**Fig.3**).

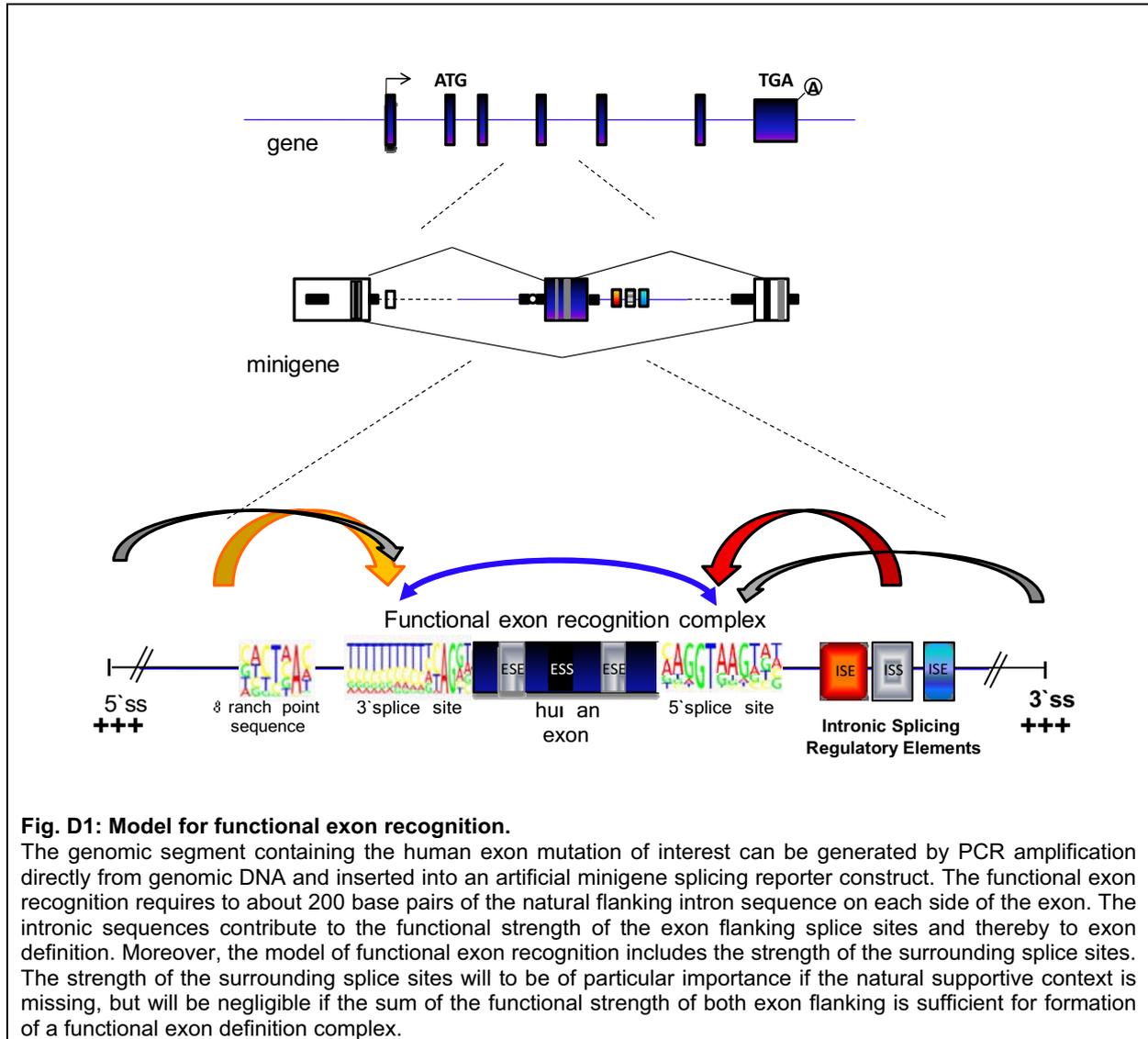
It appeared that the specific sequence immediately downstream of to the splice donor was decisive for *ATM* 54 exon definition which suggested that these sequences affect U1 snRNP binding to this splice donor site and thereby contribute to the functional strength of the exonic splice donor site. Therefore these sequences were tested in a different context for their ability to enhance U1 snRNP binding to a 5'ss from the downstream position. The results showed that the sequences were capable to enhance U1 snRNP binding to a 5'ss (**Fig.4**). Although splicing regulatory proteins that mediate this effect could not be clearly identified in the frame work of this thesis, it has been described for the hnRNP-like protein TIA-1 and for hnRNP proteins that these proteins promote U1 snRNP binding to a 5'ss from the downstream position (Förch et al., 2002 (87); Erkelenz, S., thesis, 2012 (79)).

To allow the drawing of general conclusions from these results and to exclude that the observed effects were specific for *ATM* exon 54 the experiments were repeated with *ATM* exon 9 instead of exon 54. Similar to exon 54, *ATM* exon 9 had a 3'ss of intermediate strength (MaxEnt score = 9.42) and a weak 5'ss (HBond score = 12.3). The obtained results demonstrated that in the absence of their natural intron sequence immediately downstream of the weak splice donor both exons were only recognized if the 3'ss of the heterologous splicing minigene was strong. If the 3'ss were less efficient, the presence of a small segment of authentic intron immediately downstream of the weak exonic splice donor would be necessary for exon recognition in the heterologous splicing reporter minigene.

These results were interesting because it has been described that if the 5'ss or 3'ss of an internal exon is of intermediate strength a strong compensating splice at the opposite end of

Discussion

the exon will be required to support exon recognition (Shepard et al., 2011 (299)). Therefore it seems plausible that not only the sum of cross-exon splice site strength but additionally the sum of cross-intron splice site strength is decisive for exon recognition. Apparently, the strong 3'ss within the heterologous the splicing reporter minigene was capable to compensate for the lack of the supportive intronic sequence. Nevertheless, the results were consistent with the model of exon definition because they demonstrate a clear priority of the functional strength of the exon flanking splice sites as a key to the decision to include or skip a particular exon. The functional splice site strength should take into account not only the intrinsic strength of the exon flanking splice site and the specific exon sequence but should also include the close-by intron sequence which can be a main modulator of the functional strength. Together, the results of this thesis suggest an extended and more flexible model for functional exon recognition. In this model, the functional exon definition complex is extended to about 200 base pairs of the natural flanking intron sequence on each side of the exon. Moreover, the model of functional exon recognition includes the strength of the surrounding splice sites. The strength of the surrounding splice sites will to be of particular importance if the natural supportive context is missing, but will be negligible if the sum of the functional strength of both exon flanking is sufficient for formation of a functional exon definition complex (**Fig.D1**). A more flexible model of exon definition is line with recent data showing the isolation of functional exon definition complexes containing the U4, U5 and U6 tri-snRNP in addition to U1 and U2 snRNP (Schneider et al., 2010 (283)) as well as many additional proteins (House et al., 2008 (135); Sharma et al., 2008 (295)). The group of Schneider et al. also demonstrated the existence of an alternative spliceosome assembly pathway in which the transition from the cross-exon complex to the cross-intron B-like complex can proceed directly without the preexistence of a cross-intron A complex. Other findings provide clear evidence that exon recognition includes multiple transitions along the spliceosome assembly and catalytic pathway (House et al., 2008 (135)), and is not limited to initial cross-exon splice site recognition (Lim et al., 2004 (190); Izquierdo et al., 2005 (143); Bonnal et al., 2008 (33)). Since it could be shown in the context of this thesis that the recognition of both *ATM* exon 54 and exon 9 was not improved in extended minigenes harboring the complete natural flanking introns and exons including the splice sites, the results allowed to draw the conclusion that a heterologous splicing reporter minigene is an adequate system for the investigation of exon recognition as long as the heterologous minigene contains strong splice sites.



3.2. Functional splicing assay contributes to establishment of *RAD51C* as cancer susceptibility gene

Genetic factors play a prominent role in common diseases and cancer syndromes including breast, colorectal, skin, prostate and ovarian cancer. Mutations in cancer susceptibility genes have been found in families with hereditary cancer and are believed to predispose carriers to breast, ovarian and other cancers. Reliable diagnosis of the pathogenicity of an inherited gene mutation is crucial for genetic counseling and the process of clinical decision making.

In particular, pathogenic splicing mutations require experimental validation due the complex interplay of splice site defining sequence elements. Moreover, for many cancer susceptibility genes mutation of a single allele confers an increased risk for cancer, e.g. women carrying heterozygous mutations in the cancer susceptibility genes *BRCA 1* or *BRCA 2* are estimated to have a lifetime breast cancer risk of up to 85% (Rebbeck & Domchek, 2008 (253)). These

Discussion

monoallelic pathogenic mutations may escape the identification or correct interpretation by *in vivo* splicing assays in peripheral blood leukocytes from heterozygous mutation carriers because the observable phenotype may be weakened due to the presence of the second normal wild type allele. In such cases, analysis of the putative pathogenic splicing mutation within a splicing minigene construct provides a useful model for comparing the expression from mutant allele opposed to the wild type allele.

In a collaborative project (Meindl et al., 2010 (211)) in the framework of this thesis, a splice donor mutation disrupting the canonical GT dinucleotide within the splice donor of *RAD51C* exon 1 (c.145+1G>T) and was found in a family with three sisters affected by breast or ovarian cancers. Direct analysis of the *RAD51C* splicing pattern in peripheral blood leukocytes from two heterozygous mutation carriers revealed reduced expression of the normal protein-coding *RAD51C* transcript and increased expression of a non-functional *RAD51C* transcript in both mutation carriers (**Fig. 10**), which appeared to be produced by the usage of an alternative splice donor within in *RAD51C* exon 1 which was predicted by the HBond algorithm (www.uni-duesseldorf.de/rna). However, to clearly confirm the pathogenicity of the observed c.145+1G>T splice donor mutation it was necessary to prove that the normal *RAD51C* transcript was solely expressed from the wild-type allele in the heterozygous leukocytes. This required a minigene based functional splicing assay which allowed to a monitor the splicing pattern of the mutated allele.

Here, *RAD51C* exon 1, intron 1 and exon 2 were amplified from normal human control DNA and inserted into a splicing construct. In a second construct the c.145+1G>T mutation was introduced by PCR mutagenesis. The RT-PCR analysis of the c.145+1G>T splicing minigene showed complete inactivation of this mutant 5' ss and increased transcript levels from the upstream proximal 5' splice site producing a non-functional transcript (**Fig. 12**), therefore the effect of the monoallelic *RAD51C* c.145+1G>T splice donor mutation could be clearly demonstrated. The minigene construct provided a valuable tool for characterizing the effect on splicing separately from the second allele.

Moreover, a second splice donor mutation identified in the 5'ss of *RAD51C* exon 6 (c.904+5G>T) affecting an evolutionarily conserved position was predicted to severely reduce the complementarity between the U1 snRNA and this 5'ss as indicated by a decrease in the HBond score from 15.8 to 10.1. This mutation showed a high frequency in the first degree relatives and siblings with both breast and ovarian cancers. In a heterologous splicing reporter minigene, in which *RAD51C* exon 6 was inserted along with its flanking splice sites, it could be demonstrated here that the c.904+5G>T mutation resulted in the loss of *RAD51C* exon 6 recognition (**Fig. 9**).

Thus, here the meaningful combination of *in silico* prediction and functional splicing assays, as well as segregation analysis and the availability of tumor samples, provided the first unambiguous evidence of highly penetrant mutations associated with human cancer in a *RAD51* paralog and supported the ‘common disease, rare allele’ hypothesis (Walsh & King, 2007 (350)) (published in Meindl. et al., 2010 (211)).

3.3. hnRNP H1, A1 and M4 seem to be involved in an exon definition net within the *BRCA 2* transcript

Inherited monoallelic mutations within the *BRCA2* gene have been shown to be associated with a high lifetime risk of developing breast cancer whereas germline inheritance of two defective copies of *BRCA2* causes a disease in childhood and adolescence called Fanconi anemia (Howlett et al., 2002 (137)). As described in the framework of this thesis, in siblings diagnosed with Fanconi anemia a biallelic micro-deletion of 10 base pairs was detected in *BRCA2* exon 6. This micro-deletion was shown here to cause loss of *BRCA2* exon recognition using a heterologous splicing reporter minigene (**Fig. 13**). However, when the splicing pattern of the *BRCA2* transcript was analyzed in cells derived from both patients it turned out that the effect of the micro-deletion on splicing was more complex. Analysis of the splicing pattern in fibroblasts and a lymphoblastoid cell line grown from the affected boy and his sister revealed the appearance of multiple splice variants. Apparently, the normal transcript including *BRCA2* exon 6 with the micro-deletion was still produced, albeit with decreased efficiency (**Fig. 14**). In addition, a splice variant with retention of the mutant *BRCA2* exon 6 but skipping of exon 5 was detectable especially in the male patient. Moreover, variants with skipping of the mutant *BRCA2* exon 6 and additional skipping of the surrounding exons were found in small amounts. Of note, the normal transcript including *BRCA2* exon 6 with the micro-deletion could not encode a functional protein since the micro-deletion caused a frameshift. However, retention of the mutant *BRCA2* exon 6 and skipping of exon 5 instead, restored the open reading frame. This additional transcript found in the patient-derived cells could therefore code for a *BRCA2* protein with residual activity (see also Ochman, T., diploma thesis, 2011 (231)).

The region affected by the micro-deletion within *BRCA2* exon 6 seemed to have a profound influence on alternative splicing of the *BRCA 2* transcript. The micro-deletion not only caused skipping of the affected exon 6 but also of exon 5 and other surrounding exons indicating that the definition of exon 5 and other exons in this cluster is influenced by that of exon 6. This pointed to a net regulation of these exons within the *BRCA2* transcript. Nonetheless, the occurrence of a transcript including the affected exon 6 and lacking only exon 5 is remarkable as skipping of this exon restored the open reading frame. The micro-deletion in

Discussion

exon 6 on the other hand created a premature termination codon in exon 6 within the normal open reading frame and skipping of both exons generated a premature termination codon in exon 7. Although it has been reported that in-frame stop codons can cause skipping of the exon harboring the premature termination codon thereby maintaining the open reading frame (Valentine et al., 1998 (343)) it appeared unlikely that an open reading frame preservation mechanism was underlying the alternative splicing of the *BRCA2* transcript upon the micro-deletion in *BRCA2* exon 6 because the most prominent in-frame transcript was generated by skipping of exon 5. The occurrence of several alternative splice variants induced by the micro-deletion in *BRCA2* exon 6 in both tissues rather provides evidence for a long-range interplay of splicing regulatory elements within the investigated exon cluster.

Further, more extensive alternative splicing of the *BRCA2* transcript in the lymphocyte cell line compared to the fibroblasts indicated a cell type dependent regulation of alternative splicing of *BRCA2*. More pronounced alternative splicing in the cells derived from the male patient compared to those derived from the female patient carrying the identical homozygous germline mutation initially suggested that gender specific differences may affect splicing of the *BRCA2* pre-mRNA. However, analysis of the splicing pattern of the *BRCA2* pre-mRNA in three unrelated male-derived and four unrelated female-derived lymphoblastoid cell lines demonstrated alternative splicing with equal efficiency in both male-derived and female derived lymphoblastoid cell lines. Therefore, gender-independent genetic differences between both siblings may influence the expressivity of the splicing outcome and disease phenotype upon the micro-deletion in *BRCA2* exon 6. General genetic differences between both siblings including structural variations in the genome might be a main modifier of the expressivity of a gene mutation.

The finding that the micro-deletion in *BRCA2* exon 6 caused alternative splicing of the *BRCA2* pre-mRNA implied that the micro-deletion disrupted a splicing regulatory region within exon 6 that seemed to influence not only recognition of exon 6 but also recognition of the surrounding exons within this cluster. RNA affinity chromatography in combination with mass spectrometry identified the binding of the proteins hnRNP H1, A1 and M4 to the wild type *BRCA2* exon 6 within the region affected by the micro-deletion. Immunoblotting confirmed strong binding of hnRNP H1 and moderate binding of hnRNP A1 and M4 to the wild type *BRCA2* exon 6 sequence whereas these proteins could not be detected on the RNA sequence harboring the deletion and also not on the control RNA (**Fig.16**). Inspection of the *BRCA2* exon 6 sequence revealed the presence of the core-binding site GGGA for hnRNP H1 (Caputi et al., 2001 (45)) within the region affected by the micro-deletion. Likewise, it has been reported that hnRNP M binds avidly to poly(G) homopolymers *in vitro* (Datar et al., 1992 (66)) indicating that both hnRNP H and M might specifically bind to the

wild type BRCA2 exon 6 sequence. As the 5'-end of *BRCA2* exon 6 does not contain an hnRNP A1 binding sites that exactly matches the consensus high-affinity hnRNP A1 binding site, UAGGGA/U (Burd et al., 1994 (42)), this might explain low-affinity binding of hnRNP A1 only.

It has been reported that hnRNP H1 and M are involved in the regulation of alternative splicing (Ohe et al., 2009 (232), Hovhannisyan et al., 2007 (136), Paul et al, 2006 (239)). Because it has been suggested that interactions between different hnRNP H1 and A1 proteins bound to distinct positions on a pre-mRNA can change its conformation to affect splicing decisions (Fisette et al., 2010 (85)) it appeared likely that these proteins function as splicing regulators within the *BRCA2* transcript. Moreover, this finding was consistent with a recent proteomic study of exon definition complexes in which hnRNPs were found within exon definition complexes (Sharma et al., 2008 (295)), suggesting that hnRNP H1, A1 and M4 are involved in an exon definition net within the *BRCA 2* transcript.

Individual and cell-type specific expression levels of these proteins (Kamma et al., 1995 (151)) may contribute to the different splicing outcome upon the micro-deletion in *BRCA2* exon 6. Nevertheless, further studies including siRNA mediated knockdown of these proteins in different cell types and mutational analysis in extended minigenes will have to confirm potential direct mechanisms in control of *BRCA2* splicing.

3.4. The local enhancer density and splice donor strength might bring about the decision between exon skipping or cryptic splice site activation

Even though exon skipping is by far the most frequent outcome of human splice donor mutations activation of cryptic splice donor sites located close to the authentic splice donor site is the second most frequent consequence of human splice donor mutations (Krawczak et al., 2007 (176)). Cryptic 5'ss per definition are GT sequences that are not used as splice sites in the wild type pre-mRNA, but are selected as a result of a mutation affecting the recognition of a wt 5'ss. Nevertheless, it remained a challenge to predict whether a splice donor mutation results in skipping of the affected exon or in activation of cryptic splice sites.

To gain insight into this question, a homozygous *FGB* IVS 7 +1G>T point mutation affecting the highly conserved GT dinucleotide of the splice donor site of *FGB* exon 7 identified by Spena and coworkers in a patient suffering from congenital afibrinogenemia provided here model for investigation of the mechanism of cryptic splice site activation. The *FGB* IVS 7 +1G>T splice donor mutation was analyzed in a minigene construct comprising a portion of *FGB* exon 6, intron 6, *FGB* exon 7, intron 7 and a portion of *FGB* exon 8 (Spena et al., 2002(314)). It has been described that the *FGB* IVS 7 +1G>T splice donor mutation resulted in the activation of three cryptic donor splice sites, localized in the *FGB* exon 7 at 106 nt (c1),

Discussion

40 nt (c2), and 24 nt (c3) upstream from the physiological splice donor (Spena et al., 2006 (315)). Assessment of the intrinsic strength of the GT sequences within exon 7 and its downstream intron applying the HBond algorithm calculated an HBond score (HBS) of 15.00 for the authentic wild type splice donor site of *FGB* exon 7. The HBond scores for the cryptic splice donor sites c1, c2 and c3 accounted for 12.20, 13.70 (calculated with GT instead of GC) and 10.80 respectively (**Fig. 17**), demonstrating that the authentic splice donor had a significantly higher score value than the cryptic ones.

This raised the question of whether a significant higher score value of the authentic site compared to the cryptic ones would account for the correct specification of the authentic site in the wild type pre-mRNA. If this were true, a reduction of the score difference between the cryptic splice sites and the authentic site by artificially increasing the complementarity of the cryptic splice donor sites to U1 snRNA should result in activation of the cryptic sites despite the presence of the wild type splice donor. Interestingly, the results of this thesis showed when the intrinsic strength of the cryptic splice site c1 (HBS c1 = 15.8) was comparable to the intrinsic strength of the authentic splice donor site of *FGB* exon 7 (HBS = 15.0) the splicing machinery discriminated against the usage of the cryptic sites in favor of the natural site (**Fig. 18**). However, if the intrinsic strength of the c1 was higher than the one of the authentic 5'ss this site was used in 38% instead of the wild type splice donor when the HBond score value of c1 was 18.8. Further improvement of the cryptic splice donor c1 by increasing its complementarity to U1 snRNA towards an HBS of 20.8 resulted in activation of c1 in 45% of the minigene transcripts. Nevertheless, the authentic splice donor of *FGB* exon 7 despite its significant lower complementarity was still preferred (55% of the minigene transcripts), even if the intrinsic strength of c1 was further increased towards a score value of 23.8. Therefore, it appeared that an enhanced functional strength of the natural splice donor taking into account its context of *cis*-regulatory elements allowed the splicing machinery to prefer the natural splice donor over competing nearby potential splice donor sites of comparable intrinsic strength. Moreover, less activation of c1 despite an intrinsic strength exceeding the intrinsic strength of the natural splice donor might be due to the weakness of the previously identified splicing enhancer upstream of c1 (Spena et al., 2006 (315)). This enhancer element has been described to be necessary for the activation of c1 which raised the question whether activation of the cryptic splice site c3 upon disruption of the natural splice donor was also enhancer dependent in particular because the intrinsic strength of the cryptic splice donor c3 accounted for an HBS of only 10.8. In order to clarify this it was tested here if an increased intrinsic strength of the cryptic splice donor c3 permitted the cryptic splice donor c3 to outcompete the physiological wild type splice donor of *FGB* exon 7. The results showed that if the intrinsic strength of the cryptic splice donor c3 was comparable to

Discussion

the intrinsic strength of the physiological splice donor of *FGB* exon 7 (HBond score of 15.8 versus 15.0) the splicing machinery exclusively selected the cryptic splice donor c3 instead of the physiological splice donor. This was in contrast to the cryptic splice c1 which was not selected when its intrinsic strength was identical with the physiological splice donor (**Fig. 19**). Therefore, it seemed that the activation of the cryptic splice donor c3 and maybe also the close-by authentic exon 7 splice donor was supported by an additional exonic enhancer element within *FGB* exon 7 that appeared to be much stronger than the previously identified splicing enhancer upstream of the cryptic splice donor c1.

Indeed, analysis of the enhancer activity within in *FGB* exon 7 in the region between the cryptic splice donors c1 and c3 suggested that this region contained multiple enhancer elements (**Fig. 20**) suggesting that multiple enhancer elements within *FGB* exon 7 induce cryptic splice site activation upon disruption of the physiological splice donor. This was confirmed by continuative work in our group demonstrating that the enhancer activity of the region immediately downstream of the cryptic splice donor c1 was stronger than the one of the previously published enhancer sequence upstream of c1 (Schöneweis K. diploma thesis, 2010(284)). Moreover, additional work in our group demonstrated that disruption of the additional identified enhancer activity (*FGB* 7D 5C8A mutation) allowed the preferential usage of the cryptic splice donor c1 in favor of the cryptic splice donor sites c2 and c3 and in favor of the natural splice donor. This was even more pronounced when the intrinsic strength of c1 was increased towards an HBS of 20.8 (K. Schöneweis diploma thesis (284) and S. Kübart bachelor thesis, 2010 (178)). Most importantly, disruption of both the previously published splicing enhancer and the newly identified enhancer in the presence of the IVS7+1G>T splice donor mutation caused increased *FGB* exon 7 exon skipping (Kübart S. bachelor thesis, 2010 (178)), demonstrating a switch from cryptic splice activation to exon skipping upon the *FBG* exon 7 splice donor mutation. This allowed to conclude that the density of enhancer elements and the intrinsic strength of GT sequences within human exons might be decisive whether a splice donor mutation results in skipping of the affected exon or in activation of cryptic splice sites. Moreover, the results demonstrated that not only the intrinsic strength but rather the functional splice site strength which quantitatively measures both the intrinsic strength and the context of *cis*-regulatory elements seemed to explain why a splice site is preferred over a nearby competing splice sites. Thus, the functional splice site strength appeared to be a useful concept in order to characterize differences between cryptic and authentic splice site. Finally, this information should be used to predict whether a putative pathogenic splice donor mutation results in activation of cryptic splice sites or exon skipping.

3.5. Intrinsic features of the 5'ss and the genomic context of *FANCC* exon 2 allow functional splicing at a mutant +1G>T splice donor

The most frequent base-pair mutation in human splice donor sites in inherited diseases comprises the first intronic nucleotide which is a guanosine of the canonical GT dinucleotide (Krawczak et al., 2007 (176)) and until now, it has been thought that any base-pair substitution at this position completely abrogates normal mRNA processing. In this thesis, however, it was demonstrated in primary fibroblasts from Fanconi anemia patients that a single base-pair mutation, changing the canonical GT splice donor of *FANCC* exon 2 to a TT splice donor, unexpectedly allowed correct splicing, albeit with decreased efficiency (**Fig. 21**). Moreover, this phenomenon seemed to be the cause for a milder clinical phenotype of Fanconi anemia subtype C in these patients.

Functional analysis in the patients's cells and within in the heterologous splicing reporter minigene in HeLa cells allowed to investigate the requirements that permit functional splicing at human mutant +1G>T splice donor sites. Although the intrinsic strength of the wild-type *FANCC* 5'ss was relatively high (HBS = 18.7), due to the high degree of complementarity to the U1 snRNA, recognition of the wild-type *FANCC* exon 2 in the heterologous splicing reporter minigene was not as effective as expected and the mutant TT 5'ss was not recognized at all (**Fig. 23**). Therefore the nucleotides at positions -3 and -2 of the mutant TT 5'ss were replaced by nucleotides complementary to the 5'-end of the endogenous U1 snRNA. This partially restored inclusion of *FANCC* exon 2 with the mutant TT splice donor. Direct sequencing of this splice product, however, revealed that splicing in this reporter transcript occurred not only at the TT dinucleotide at the authentic exon-intron border, but was also shifted to the GT dinucleotide one position upstream of TT. The existence of this 2nd transcript was remarkable, because all available algorithms for splice donor sites unequivocally predicted that the intrinsic strength of the GT dinucleotide at -1 was very weak (e.g. the HBS is 2.3) due to the low complementarity to the U1 snRNA in this base-pairing frame. This additional splicing at position -1 which has been recently characterized for a atypical 5'ss and has been assumed to precede base-pairing in a shifted register (Roca & Krainer, 2009). Nevertheless, further analysis showed that splicing in this construct only occurred at the two physiological GT and GC splice donor sites or if a TT dinucleotide was present at position +1 and +2. Therefore, additional splicing at -1 could not simply be explained by the increased complementarity to the U1 snRNA in the -1 register, as the TA dinucleotide in the splicing reporter that would otherwise have specifically increased base-pairing to U1 snRNA at -1 (gcaGTAAagta, HBond score 9.0 vs. gcaGTTAagta, HBond score 1.9) did not allow splicing, suggesting that U1 snRNA base-pairs with the mutant *FANCC* TT splice donor in the canonical register. Therefore a mutant TT splice donor site could be

Discussion

functional in a heterologous context if this site were highly complementary to the U1 snRNA. The complementarity of the -1 GT register to the U1 snRNA is seemed to be of less importance, since the TA dinucleotide despite higher complementarity did not allow splicing at this site.

In the natural context, however, as shown by the analysis in patient-derived fibroblasts, splicing of the mutant TT splice donor site exclusively occurred immediately upstream of the TT dinucleotide at the correct exon-intron-border, presumably due to additional sequences in the endogenous gene context.

Nevertheless, when U1 snRNA molecules specifically adapted to the mutant FANCC TT splice donor were co-transfected along with the splicing reporter construct containing the original mutant *FANCC* splice donor this restored recognition of the mutant *FANCC* TT 5'ss (**Fig. 24**). Here, sequence analysis of the splice products confirmed that splicing exclusively occurred at the correct exon-intron border. The combination of the results from both experimental settings of fully adapting either the 5'ss to the endogenous U1 snRNA or the U1 snRNA to the mutant splice site was striking: the exclusive use of the non-canonical TT as splice site was not simply determined by the free energy of the RNA duplex formed between the splice donor and the matching U1 snRNA (which was identical in both cases), but was predominantly dependent on the 5'ss sequence itself. Accordingly, when the TT adapted U1 snRNA was co-transfected with the reporter construct harboring the FANCC TT 5'ss with improved complementarity (**Fig. 25**) this significantly increased the efficiency of the splicing reaction but did not determine whether cleavage occurred at the TT at the exon-intron-border or at the -1 position. Since these results suggested that there might exist a not yet identified endogenous U1 snRNA that facilitates splicing at TT splice donor a recently published human U1 snRNA variant (U1A7) with complementarity to the *FANCC* TT 5'ss (Kyriakopolou et al., 2006 (180)) was co-transfected in HeLa cells along with the heterologous splicing minigene harboring *FANCC* exon 2 with the mutant TT splice donor. This U1A7 snRNA, however, did not enable TT splicing, neither the analyses here (**Fig.26**) nor in the work by Roca and Krainer (Roca & Krainer, 2009 (263)), most likely due to a nonfunctional snRNA body, thus suggesting that this U1A7 snRNA might be a transcript of a pseudogene. Alternatively, the U1A7 snRNA might be delayed in its biogenesis and thus its suppression capability could simply not been detected in the transient transfection assays. Therefore, it remained unlikely that the U1A7 variant functions in splicing and that the *FANCC* TT splice donor is recognized by this variant.

Moreover, replacement of the sequence of the mutant *FANCC* TT splice donor by sequence of the *FGB* IVS7 +1G>T splice donor within the heterologous splicing minigene harboring *FANCC* exon 2 allowed to determine intrinsic sequence requirements of a splice donor

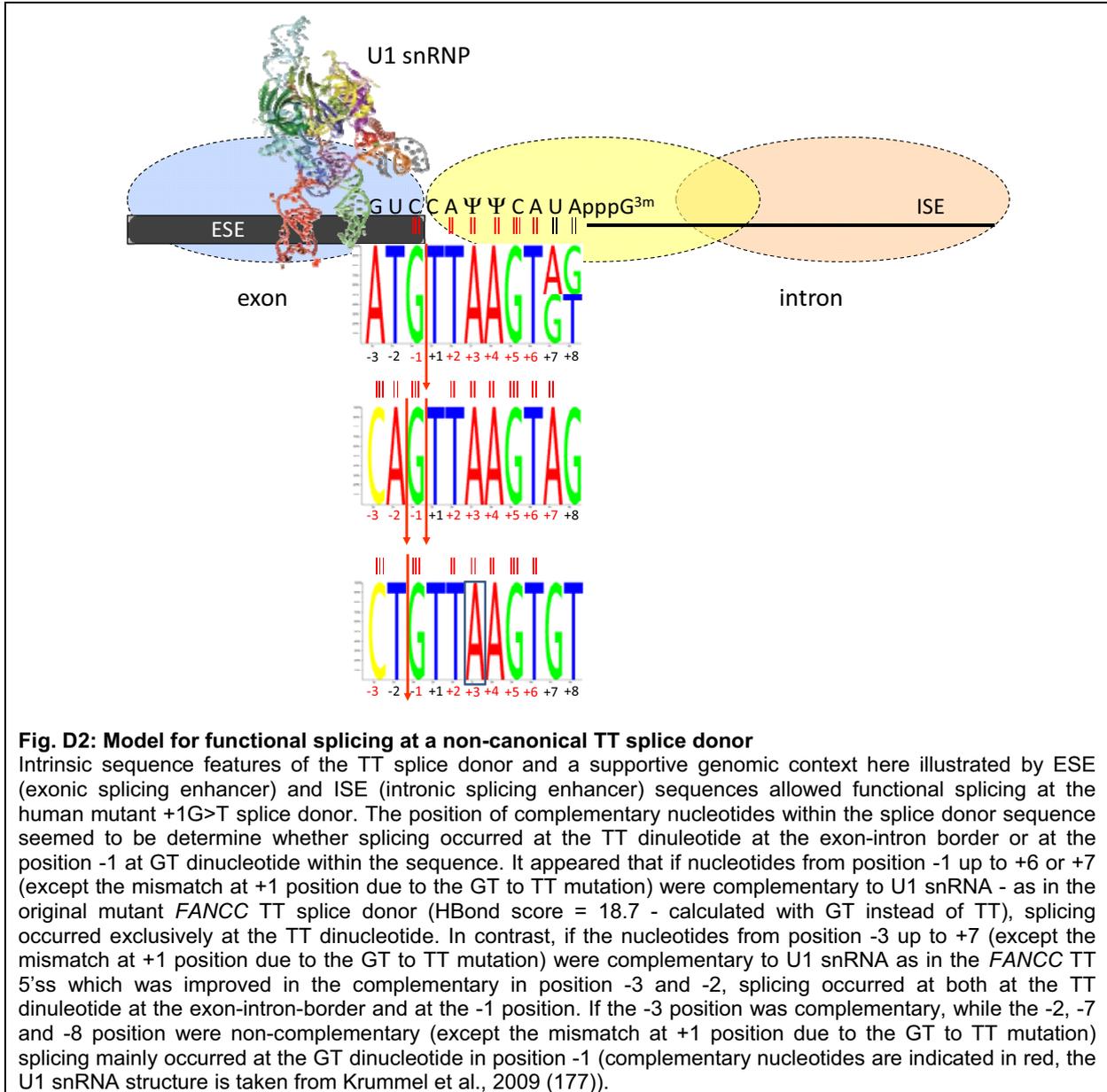
Discussion

allowing the usage of a non-canonical TT dinucleotide. Since the *FGB* IVS7 +1G>T splice donor represented also a TT 5'ss however which was not recognized in its natural context successive adaption of the sequence of the *FGB* TT splice donor towards the sequence of the *FANCC* TT splice donor permitted to narrow down intrinsic sequence requirements of the 5'ss for splicing at TT site. The results obtained here showed that along with the co-transfection of an adapted U1 snRNA molecule usage of the TT splice donor within the context of the *FANCC* reporter minigene seemed to require complementary bases at position -1, +2, +3, +4, +5 and +6 (**see also Fig. D2**). It turned out that in this sequence composition a complementary nucleotide at the +4 position was crucial for the recognition of the non-canonical TT as splice donor (**Fig.30**). Mismatches at positions -3 and -2 as in the original *FANCC* splice donor could be compensated by co-transfection of the adapted U1 snRNA molecule (-3A, +4A) as long as the nucleotides from position -1 up to position +6 were complementary ones. In this case a complementary nucleotide at position +7 appeared to be negligible for usage of the TT splice donor. But importantly, the results obtained by direct sequencing of the splice products demonstrated that almost exclusive usage of the TT splice donor at the exon-intron-border required the mismatches at positions -3 and -2 as in the original *FANCC* splice donor (**Fig. D2**). If the -3 position within the splice donor was complementary to U1 snRNA (and the nucleotide at position +7 was a non-complementary nucleotide) splicing mainly occurred at the GT dinucleotide in position -1. Exclusive usage of the TT splice donor however was detectable if the nucleotides from position -1 up to +7 (except the mismatch at +1 position due to the GT to TT mutation) were complementary to U1 snRNA - as in the original *FANCC* splice donor (HBond score = 18.7 - calculated with GT instead of TT). In contrast, if the nucleotides from position -3 up to +7 (except the mismatch at +1 position due to the GT to TT mutation) were complementary to U1 snRNA as in the *FANCC* TT 5'ss which was improved in complementary in positions -3 and -2 (**Fig.23**) splicing occurred at both at the TT dinucleotide at the exon-intron-border and at the -1 position (**Fig.D2**). Therefore, the data evidently demonstrated that here the position of complementary nucleotides within sequence of the TT splice donor determined the cleavage site and were crucial for correct splicing at the exon-intron-border.

Nevertheless, recognition of *FANCC* exon 2 in the heterologous splicing reporter minigene was not as effective as within the endogenous *FANCC* transcript and the mutant TT splice donor was only recognized upon an increased complementarity to U1 snRNA or upon co-transfection of a compensatory U1 snRNA molecule. Therefore it seemed likely that recognition of the splice donor of *FANCC* exon 2 is enhanced by its genomic context. Indeed, when the genomic context of *FANCC* exon 2 within the minigene was extended to a region spanning 676 bp of the upstream intron, *FANCC* exon 2, intron 2 and exon 3 (**Fig. 31**), the

Discussion

original mutant *FANCC* TT splice donor was efficiently recognized without co-transfection of the TT adapted U1 snRNA. Yet, recognition of the mutant TT splice donor site could be enhanced by co-transfection of both U1 snRNA α TT and TTcom achieving a exon recognition level almost comparable to the wild type GT splice donor.



These results demonstrated that the genomic context of *FANCC* exon 2 enhanced *FANCC* exon 2 definition supporting recognition of the mutant TT splice donor site. Therefore, usage of the unusual TT splice donor site as seen in the Fanconi anemia patients was not only determined by complementary nucleotides from position -1 up to +7 (except the mismatch at +1 position due to the GT to TT mutation; HBond score = 18.7 calculated with GT instead of

TT) but was also profoundly enhanced by not yet identified enhancing sequences within the *FANCC* context. As only 10% of annotated human 5'ss have an HBond of 18.7 (Theiss S. and Schaal H. unpublished data), this specific requirements provided a rationale why this phenomena has not been described earlier.

3.6. A novel U1 snRNA based therapy approach for human splice donor mutations

As the TT-adapted U1 snRNAs improved the usage of the *FANCC* TT splice donor site within the minigenes, this implied that transfection of biallelic *FANCC* c.165 +1G>T patient-derived fibroblasts with the TT-adapted U1 snRNA molecules would improve recognition of the pathogenic TT splice donor and enhance the levels of the endogenous TT-spliced in-frame transcripts. Indeed, the results obtained here showed that ectopic expression of the TT-adapted artificial U1 snRNAs significantly increased the usage of the pathogenic TT 5'ss in the patients' fibroblasts (**Fig. 32**). As permanent suppression of splice donor mutations in cells still actively dividing required that the mutation-adapted U1 snRNA integrates into the genome of the mutant cells the TT-adapted U1 snRNA expression cassettes were transferred into a lentiviral vector (LV) which co-expressed the neomycin phosphotransferase (neoR) cDNA in opposite orientation. Transduction of primary patient-derived fibroblasts harboring the biallelic *FANCC* c.165 +1G>T mutation with the lentiviral vector carrying the TT-adapted U1 snRNA significantly improved the MMC-induced cell cycle arrest and thereby the disease phenotype of this cells (**Fig. 33**). Moreover, RT-PCR analysis of the endogenous *FANCC* transcript in transduced and immortalized fibroblasts confirmed that the usage of the mutant TT splice donor and *FANCC* exon 2 recognition was clearly improved by lentivirus-mediated expression of both TT adapted U1 snRNAs (**Fig. 32**). Moreover, the results revealed that the immortalized fibroblasts with the biallelic *FANCC* c.165 +1G>T mutation already had minute levels of mono-ubiquitinated protein even in the absence of any U1 snRNA transfected (**Fig. 32**). This indicated that the lower levels of TT-spliced endogenous in-frame transcript encoded residual functional *FANCC* protein that was active in the FA core complex. This residual activity of the Fanconi anemia (FA) pathway might be the reason for the milder clinical phenotype as seen in these patients. This was in lines with findings that in certain FA complementation groups such as *FANCD2* and *FANCD1/BRCA2*, at least one hypomorphic allele with residual protein activity appears mandatory for the survival of patients with biallelic germ-line mutations (Popp et al., 2003 (247); Kalb et al., 2007 (148); Neveling et al., 2009 (223)).

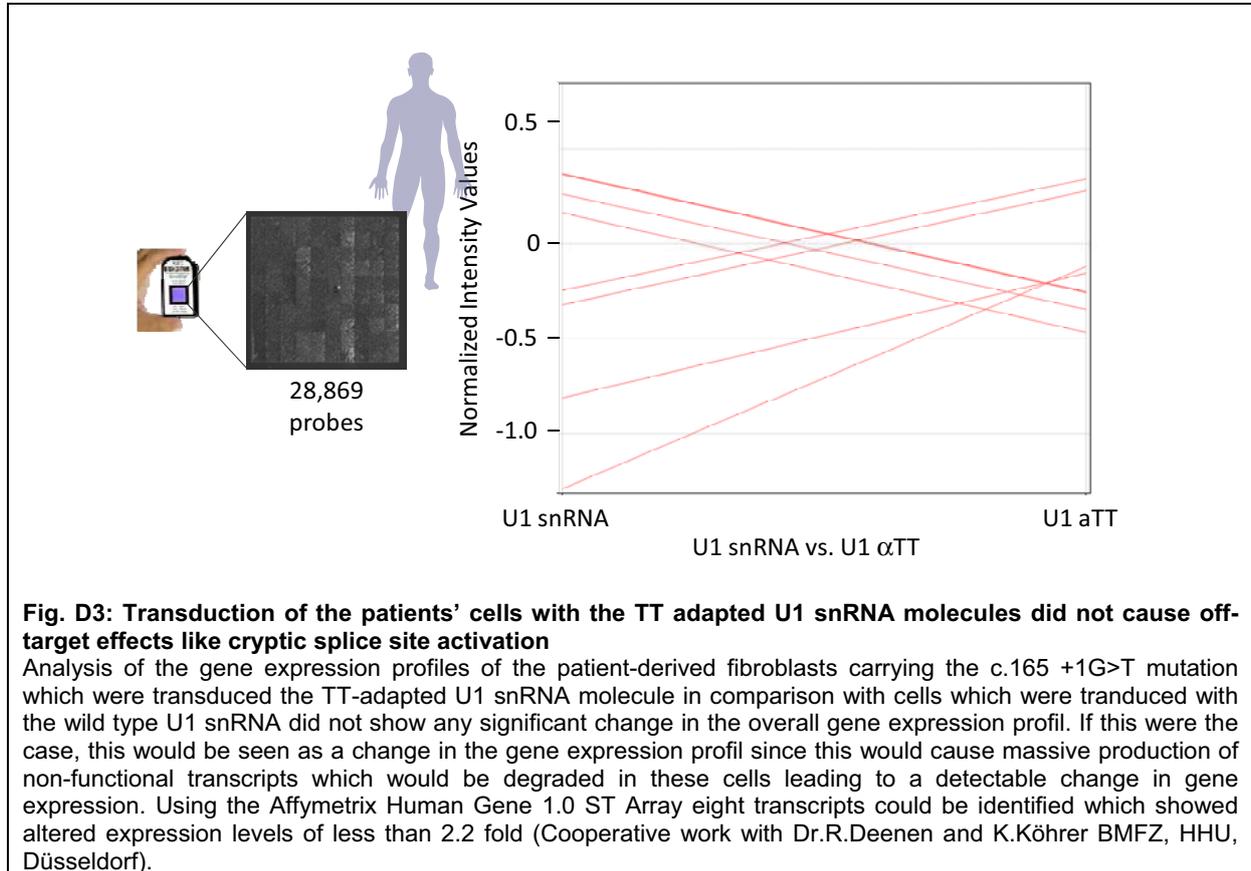
So far, genetic therapies aimed at correcting the underlying deficiency in hematopoietic stem cells utilized integrating retroviral vector systems to introduce a normal cDNA copy of the affected gene into the target cells. In the present study, we showed that understanding the

Discussion

phenotypic consequence of splice donor mutations at the mRNA level can be instrumental to develop novel therapeutic strategies to correct an aberrantly processed message. Since the initial report in 1986, compensatory mutations in U1 snRNA are known to have the capability to correct 5'ss mutations (Zhuang & Weiner, 1986 (384)). The suppressive efficiency of these altered U1 snRNAs however depends on the individual mutation and often can only be assessed by functional testing. Although correction of pre-mRNA processing in minigene constructs with the mutated splice sites have been reported by a few groups (Pinotti et al., 2008 and 2009 (243,244); Meyer et al., 2009 (213); Tanner et al, 2009 (332)) correction of the endogenous transcript and correction of the disease phenotype of primary human cells that are deficient in a cellular transcript, has not been reported so far. In a mouse model for spinal muscular atrophy, Meyer *et al.* elegantly showed in primary murine cells recently that the germ-line expression of an artificial U7 snRNA, that promoted inclusion of the mutant SMN2 exon 7, can efficiently complement the muscle tissue and significantly extend the limited life-span of these animals (Meyer et al., 2009 (213)). Here, it was shown for the first time in primary cells from patients with a monogenetic recessive disorder that stable expression of mutation-adapted U1 snRNAs can be utilized to rescue the pathological phenotype of these cells. Using lentivirus-based vectors as delivery systems for the U1 snRNA expression cassette allowed stable integration of the U1 snRNA expression cassette into the target cell genome in dividing and nondividing cells (Kohn & Candotti, 2009 (168)), e. g. hematopoietic stem cells and also retina cells as target cells for genetic correction (Baindrige et al., 2008; Aiuti & Roncarolo, 2009 (3-5)). Interestingly, the level of functional restoration of the FA/BRCA pathway in transduced cells differed between the two U1 snRNAs that were specifically adapted for the mutant *FANCC* exon 2 5'ss. The minimally adapted U1 snRNA α TT almost achieved a correction level of cells where the normal *FANCC* cDNA was overexpressed. Although it would appear likely that an increased complementarity of the TT-adapted U1 snRNA to the pathogenic 5'ss will more efficiently generate correct transcripts and also reduce the potential for deleterious off-target effects, surprisingly, the U1 snRNA TTcom with higher complementarity was less efficient in correcting the cell cycle arrest in the primary FA cells. It appeared that a complete match of the U1 snRNA to this TT splice donor might disturb consecutive steps during the splicing process, as the artificial TT U1 snRNA with unnaturally high-affinity might not be efficiently displaced by the U6 snRNA within sequence context. A more advanced strategy combining our U1 snRNA-based approach with efforts to support U1 snRNA binding by artificially recruited SR proteins (Marquis et al., 2007 (203)) should be further developed to achieve most efficient correction of a pathogenic 5'ss mutation on the RNA level. Correction of the endogenous transcript

would also obviate the inability to deliver large genes and ensure that the natural fine-tuning of the endogenous protein remains intact.

Moreover, analysis of the gene expression profiles of the patient-derived fibroblasts carrying the c.165 +1G>T mutation which were transduced the TT-adapted U1 snRNA molecule in comparison with cells which were transduced with the wild type U1 snRNA did not show any significant change in the overall gene expression profil (**Fig.D3**).



This indicated that transduction of the patients' cells with the TT adapted U1 snRNA molecules did not cause off-target effects like cryptic splice site activation. If this were the case, this would be seen as a change in the gene expression profil since this would cause massive production of non-functional transcripts which would be degraded in these cells leading to a detectable change in gene expression.

Therefore, correction of pathological mRNA processing at mutant splice sites might be an attractive gene therapy approach for certain FA complementation groups with either very large genes or toxicity of the overexpressed genes such as *BRCA2/FANCD1* (Howlett et al., 2002 (137)) or *FANCD2* (Timmers et al., 2001 (339)). This mutation specific approach might also be feasible in other genetic disorders with deficiencies in other genes such as *ATM* (Sandoval et al., 1999 (276)) and *NF1* (Wimmer et al., 2007 (365)) with a high percentage of 5'ss mutations.

4. MATERIALS AND METHODS

4.1. Material

Unless otherwise mentioned chemicals were supplied by Invitrogen, Merck, Riedel-de-Haen, Roth, Sigma and Serva. Preparation of growth media and solvents is described in the respective experimental protocols or derives from standard laboratory manuals (Ausubel et al. 1991, Sambrook et al. 1989).

4.1.1. Chemicals and Consumables

<u>Chemical / Consumable</u>	<u>Source</u>
Adipidic acid dihydrazide-Agarose	Sigma #A0802
Ampicillin	Roche #10835242001
Ampli-Taq [®] DNA Polymerase	AppliedBiosystems. #N0808-0166
BigDye [®] Terminator v1.1 Cycle Seq. Kit	Applied Biosystems #4336772
Complete Protease Inhibitors	Roche #1674498
DMEM	Gibco #41966
Dialysis membrane tubing (6-8 kDa cut off)	Spectrapor #132655
4',6 -Diamidino-2-phenylindole Dihydrochloride (DAPI)	Polysciences #09224
Dithiothreitol (DTT)	Serva #20710
DNase I recombinant, RNase-free	Roche #04716728001
dNTP Mix	Qiagen #201901
DPBS-CaCl ₂ -MgCl ₂	Gibco #14190
ECL [™] Western Blotting Detection Reagents	Amersham #RPN2106
FCS	Gibco #10270-106 / PAN Biotech
FastPlasmid [™] Mini	Eppendorf #0032007655
Formic acid (1M in H ₂ O)	Fluka #06473
FuGENE [®] 6 Transfection Reagent	Roche #1810575
High Fidelity Polymerase	Roche #1173265001
Hyperfilm [™] ECL	Amersham #RPN3103K
GeneAmp [®] dNTPs (2.5 mM each)	Applied Biosystems #N0808-0007
Gene Elute Mammalian Total RNA Kit	Sigma MRN70
Geneticin G-418 Sulphate	Gibco #11811-031
1 kb DNA Ladder	Invitrogen #15615-024
LB-Broth (Lennox)	Roth #X964.2
LB Agar (Lennox L Agar)	Invitrogen #22700-025
LE Agarose	Biozym #840004
Lysozym	AppliChem #A4972001
Leupeptin	Sigma #L8884
MetaPhor [®] Agarose	Biozym #50180E
Ni-NTA agarose	Invitrogen #R901-01
NP40/Igepal	Sigma #I3021
NuPAGE [®] Novex 7%Tris-Acetate Gels	Invitrogen

Materials and Methods

NuPAGE® Tris-Acetate SDS Running Buffer	Invitrogen #LA0041
NuPAGE® Antioxidant	Invitrogen
Nuclear extracts of 5x10 ⁹ HeLa cells	Cilbiotech S.A. #CC-01-20-25
PageRuler™ Prestained Protein Ladder Plus	Fermentas #SM1811
Polyethylenimine (PEI)	Aldrich
PenStrep	Gibco #15140
Pepstatin-A	Sigma #P4265
Plasmid Midi Kit	Qiagen #12145
Plasmid Maxi Kit	Qiagen #12163
Platinum® Taq DNA Polymerase	Invitrogen #10966-018
PMSF	Sigma #P7626
Primer p(dT) ₁₅ (Oligo-(dT))	Roche #10814270001
ProteinLoBind tubes (0.5ml)	Eppendorf #0030108.094
ProteinLoBind tubes (1.5ml)	Eppendorf #0030108.116
Pwo DNA Polymerase	Roche #1644955
QuickChange XL Site-Directed Mutagenesis Kit	Stratagene #200519
ReBlot Plus Stripping Solution	Chemicom #2509
Recombinant RNasin® Ribonuclease Inhibitor	Promega #N2511
RNA 6000 Nano Kit	Agilent #5067-1511
RPMI1640	Invitrogen
Rotiphorese Gel 30	Roth #3029.1
Sieve GP Agarose	Biozym #850050
SuperScript™ III One-Step RT-PCR system	Invitrogen #12574-026
SuperScript Reverse Transcriptase	Invitrogen #18080-085
T4-DNA Ligase	NEB #M0202S
TransIT®-LT1	Mirus Bio LLC #731-0028
0.05% Trypsin-EDTA	Gibco #25300
Trypsin (proteomics grade)	Sigma #T6567
Trypan Blue Stain 0.4%	Gibco #15250-061
Trypsin (proteomics grade)	Sigma #T6567

4.1.2. Enzymes

Restriction enzymes were supplied by New England Biolabs (NEB), Fermentas (MBI), Invitrogen, Promega and Roche. Enzymes were used as recommended by the manufacturer with the provided corresponding buffers. T7-RNA polymerase was purified in house.

4.1.3. Bacteria

Recombinant plasmids were transformed into *Escherichia coli* (E.coli) strain DH5αF'IQ (Invitrogen). For amplification of DNA plasmids subjected to cloning and restriction analyses with methylation-sensitive enzymes E.coli strain DM1 (Gibco) was used, which is deficient in functional methylases.

Genotypes of E.coli strains

DH5 α F'IQ: F ϕ lacZ Δ M15 Δ (lacZYA-argF) U169 *recA1 endA1 hsdR17* (*r_k⁻, m_k⁺*)
phoA supE44 λ ⁻ thi⁻1 gyrA96 relA1/F' proAB+ lacIqZ Δ M15 zzf::Tn5
[KmR]

DM1: F ϕ *dam⁻13::Tn9(Cm^r) dcm mcrB hsdR⁻M⁺ gal1 gal2 ara lac thr leu ton^r*
tsx^r Su^o λ ⁻

4.1.4. Cells

4.1.4.1 Human cell lines

HeLa cells

The human cervix carcinoma cell line HeLa and the derivative cell line HeLa-T4⁺ were used for transient transfection experiments. The cell line HeLa-T4⁺ is characterized by surface expression of the human CD4 receptor, which is stably integrated into the cells genome (Maddon et al. 1986 (195)). Cells containing the transgene were selected by addition of geneticin to the culture media.

HEK-293T or 293T cells

Human embryonic kidney (HEK) cell line derived from embryonic kidney tissue. Transformation with the adenoviral E1A gene product leads to production of SV40 the large T-antigen allowing the episomal replication of plasmids with SV40 origin of replication.

This cell line was used for virus particle production.

HT-1080 cells (CRL-12103; ATCC, Manassas, VA)

Human epithial fibrosarcom cell line. This line was used for titer estimation of viral supernatant (Rasheed et al. 1974 (252)).

4.1.4.2 Human primary cells

FANCC c.165+1G>T fibroblasts (IFAR526/1)

Primary patient-derived skin fibroblasts harboring the biallelic FANCC c.165+1G>T splice donor mutation. If necessary, the primary fibroblasts were immortalized with a lentivirus expressing the SV40 large T-antigen cDNA (performed by Prof. Helmut Hanenberg, Indianapolis, USA).

FANCC c.165+1G>T/c. 1-250del fibroblasts (IFAR640)

Primary patient-derived skin fibroblasts harboring a maternally inherited genomic deletion leading to FANCC exon 2 and 3 skipping and paternal the FANCC c.165+1G>T splice donor.

mutation. If necessary, the primary fibroblasts were immortalized with a lentivirus expressing the SV40 large T-antigen cDNA (performed by Prof. Helmut Hanenberg, Indianapolis, USA).

FANCC c.456 +4A>T (IVS4 +4A>T) fibroblasts

Primary patient-derived skin fibroblasts harboring the biallelic *FANCC* c.456 +4A>T (IVS4 +4A>T) splice donor mutation. If necessary, the primary fibroblasts were immortalized with a lentivirus expressing the SV40 large T-antigen cDNA (performed by Prof. Helmut Hanenberg, Indianapolis, USA).

BRCA 2 c.707-716del fibroblasts

Primary patient-derived skin fibroblasts harboring the biallelic *BRCA 2* c.707-716del genomic micro-deletion leading to the 4th to 13th base of the *BRCA2* exon 6. If necessary, the primary fibroblasts were immortalized with a lentivirus expressing the SV40 large T-antigen cDNA (performed by Prof. Helmut Hanenberg, Indianapolis, USA).

BRCA 2 c.707-716del LCL (lymphoblastoid B-cell line)

Patient-derived lymphoblastoid cell line harboring the biallelic *BRCA 2* c.707-716del genomic micro-deletion leading to the 4th to 13th base of the *BRCA2* exon 6. The cells were immortalized by EBV transformation (performed by Prof. Helmut Hanenberg, Indianapolis, USA).

4.1.5.1 Oligonucleotides for cloning

2-intron-3-exon splicing reporter minigenes:

ATM exon 54 or ATM exon 9

#2301: 5'-ATCGAATTCCACGCTCTACCC

5' primer for cloning of 2-intron-3-exon splicing reporter minigenes (EcoRI-Site)

#2302: 5'-ACCCTCGAGAAGGTACGTATGTTTAA

3' primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I-SA3 (XhoI-Site)

#2303: 5'-ACCCTCGAGTGAATATCACACTTCTAACCAAATACCTCATCAAGCTGAGAG

3' primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-II-SA3 (XhoI-Site)

#2304: 5'-ACCCTCGAGGAAATATTCTAGGAAAGACCCAAATACCTCATCAAGCTGAGAG

3' primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-III-SA3 (XhoI-Site)

#2305: 5'-ACCCTCGAGTGAATATCACACTTCTAA

3' primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I+II-SA3 (XhoI-Site)

Materials and Methods

#2306:5'-

ACCCTCGAGGAAATATTCTAGGAAAGACTGAATATCACACTTCTAACCAAATACCTCATC
AAGCTGAGAG

3'-primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-II+III-SA3 (XhoI-Site)

#2374:5'-

ACCCTCGAGTGAATATCACACTTCTAAGAAATATTCTAGGAAAGACCCAAATACCTCATC
AAGCTGAGAG

3'primer for cloning of LTR-SD1/4-ATM-exon54-3'intron-part-III+II-SA3 (XhoI-Site)

RAD51C exon 6

#3348: 5'-ATCGAATTCAGTGAAGTGGCAGCTCTTGGCTCACTGC

5' primer for cloning LTR-SD1-RAD51C-exon6-SA5opt using human gDNA as template
(EcoRI site)

#3349: 5'-CCTCGAGATCAGTATCTAACGGTACTGTGCTTAGTGC

3' primer for cloning LTR-SD1-RAD51C-exon6-SA5opt using human gDNA as template
(XhoRI site)

#3350: 5'-GCTTGTTCTGCATTAGGTGGTTAATTAATCAG

5' mutagenesis primer for cloning LTR-SD1-RAD51C-exon6-904+5G>T-SA5opt

#3351: 5'-CTGATTAATTAACCACCTAATGCAACAAGC

5' mutagenesis primer for cloning LTR-SD1-RAD51C-exon6-904+5G>T-SA5opt

BRCA2 exon 6

#197: 5'-TAATACGACTCACTATAGGG

5' primer for cloning of LTR-SD1-BRCA2-exon6-SA5opt (T7 Primer)

#2120: 5'-CTACTCGAGTTAATATTTACCTTC

3' primer for cloning of LTR-SD1-BRCA2-exon6-SA5opt (XhoI site)

FGB exon 6-8

#2619: 5'-ATCGGGACCCACAGAACTTTTGATAGAAATGGAG

5' primer for cloning of pT-Bbeta (PpuMI site)

#2620:5'-GATCCCGGGAAAGATTTGTTGTACATACAGAAG

5' primer for cloning of pT-Bbeta (PpuMI site)

#2621: 5'-ATCGAGGAACAGCCGCTAATGCCCTCATG

5' mutagenesis primer for cloning *FGB*-exon7-mt-c1

#2622: 5'-GATCATGAGGGCATTAGCGGCTGTTCCCTC

3' mutagenesis primer for cloning *FGB*-exon7-mt-c1

#2646: 5'-ATCCATTCACAACGCCATGTTCTTCAGC

5' mutagenesis primer for cloning *FGB*-exon7-mt-c2

Materials and Methods

#2647: 5'-GATGCTGAAGAACATGGCGTTGTGAATG
3' mutagenesis primer for cloning *FGB*-exon7-mt-c2

#2623: 5'-ATCCATTACAACGCCATGTTCTTCAGCACCTATGACAGAGAC
5' mutagenesis primer for cloning *FGB*-exon7-mt-c2/c3

#2624: 5'-GATGTCTCTGTCATAGGTGCTGAAGAACATGGCGTTGTGAATG
3' mutagenesis primer for cloning *FGB*-exon7-mt-c2/c3

#2764: 5'-CAGAGGAACAGCAGGTAATGCCCTC
5' mutagenesis primer for cloning pTbBeta-c1-15.8

#2765: 5'-GAGGGCATTACCTGCTGTTCTCTG
3' mutagenesis primer for cloning pTbBeta-c1-15.8

#2875: 5'-GAGGAACAGGTAAGTATCTCATGGATGG
5' mutagenesis primer for cloning pTbBeta-c1-23.8

#2876: 5'-CCATCCATGAGATACTTACCTGCTGTTCTCTC
5' mutagenesis primer for cloning pTbBeta-c1-23.8

#2924: 5'-GTTCTTCAGCAGGTAATACAGAGACAATGAC
5' mutagenesis primer for cloning pTbBeta-c3-15.8

#2925: 5'-GTCATTGTCTCTGTATTACCTGCTGAAGAAC
3' mutagenesis primer for cloning pTbBeta-c3-15.8

#2926: 5'-GTTCTTCAGCAGGTAAGACAGAGACAATGAC
5' mutagenesis primer for cloning pTbBeta-c3-18.8

#2927: 5'-GTCATTGTCTCTGTCTTACCTGCTGAAGAAC
3' mutagenesis primer for cloning pTbBeta-c3-18.8

#2928: 5'-GTTCTTCAGCAGGTAAGTCAGAGACAATGAC
5' mutagenesis primer for cloning pTbBeta-c3-20.8

#2930: 5'-GTTCTTCAGCAGGTAAGTATGAGACAATGAC
5' mutagenesis primer for cloning pTbBeta-c3-23.8

#2931: 5'-GTCATTGTCTCATACTTACCTGCTGAAGAAC
3' mutagenesis primer for cloning pTbBeta-c3-23.8

#2650: 5'-CAATGACGGCAGTTAAGTATGGCACTCTTTG
5' mutagenesis primer for cloning pTbBeta-IVS7+1G>T 5'ss-opt

#2651: 5'-CAAAGAGTGCCATACTTAACTGCCGTCATTG
3' mutagenesis primer for cloning pTbBeta-IVS7+1G>T 5'ss-opt

#2731: 5'-GGCAGTTAAGTATCGCACTCTTTGC
5' mutagenesis primer for cloning pTbBeta-IVS7+1G>T 5'ss-opt +9G>C

#2732: 5'-GCAAAGAGTGCGATACTTAACTGCC
3' mutagenesis primer for cloning pTbBeta-IVS7+1G>T 5'ss-opt +9G>C

FANCC exon 2

#2717: 5'-ATCGAATTCCAAAGATGGCTCCAGC

5' primer for cloning LTR-SD1-FANCC-Ex2-GT-SA5opt (EcoRI site)

#2718: 5-GGTAACCCTCGAGGGAGAC

3' primer for cloning LTR-SD1-FANCC-Ex2-GT-SA5opt (XhoI site)

#2723: 5'-CCTTGAAAGAGCAG**AAA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**AAI**/-2/-3opt-SA5opt

#2729: 5'-CCTTGAAAGAGCAG**ATA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**ATI**/-2/-3opt-SA5opt

#2730: 5'-GGTCCACTACTTATCTGCTCTTTCAAGG

3' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**ATI**/-2/-3opt-SA5opt

#2721: 5'-CCTTGAAAGAGCAG**CCA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**CCI**/-2/-3opt-SA5opt

#2722: 5'-GGTCCACTACTTGGCTGCTCTTTCAAGG

3' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**CCI**/-2/-3opt-SA5opt

#2727: 5'-CCTTGAAAGAGCAG**CTA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**CTI**/-2/-3opt-SA5opt

#2728: 5'-GGTCCACTACTTAGCTGCTCTTTCAAGG

3' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**CTI**/-2/-3opt-SA5opt

#2719: 5'-CCTTGAAAGAGCAG**GCA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**GCI**/-2/-3opt-SA5opt

#2720: 5'-GGTCCACTACTTGCCTGCTCTTTCAAGG

3' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**GCI**/-2/-3opt-SA5opt

#2725: 5'-CCTTGAAAGAGCAG**GGA**AGTAGTGGACC

5' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**GGI**/-2/-3opt-SA5opt

#2726: 5'-GGTCCACTACTTGCCTGCTCTTTCAAGG

3' mutagenesis primer for cloning LTR-SD1-FANCC-Ex2-**GGI**/-2/-3opt-SA5opt

FANCC exon 2 and exon 3

#3714: 5'-ATCGAATTCGTCAGGCTTATGAGATTTTATCTACTGTCCTGG

5' primer for amplification *FANCC* exon2-3 from gDNA

#3717: 5'-ATCCTCGAGCATATGCTAAAATAAAAGGATTCCAACAAGCTTTTGCCCAACA

3' primer for amplification *FANCC* exon2-3 from gDNA

#3718: 5'-GCCTTGAAAGAGATGTTAAGTAGTGGACCAG

5' mutagenesis primer for *FANCC* exon 2 +1G>T

#3719: 5'-CTGGTCCACTACTTAACATCTCTTTCAAGGC
3' mutagenesis primer for *FANCC* exon 2 +1G>T

RAD51C exon1-2

#3369: 5'-ATCGAATTCGTGCGGAGTTTGGCTGCTCCGGGG
5'-primer for amplification *RAD51C* exon1-2 gDNA

#3562: 5'-ATCCTCGAGCATAATTGTGTTTTTCCAACACCTGGTGC
3'-primer for amplification *RAD51C* exon1-2 gDNA

#3364: 5'-CCTCCGAGCTTAGCAAAGTTAACGACGACTCCTGATGGCT
5' mutagenesis primer for *RAD51C* exon 1 +1G>T

#3365: 5'-AGCCATCAGGAGTCGTTAACTTTGCTAAGCTCGGAGG
3' mutagenesis primer for *RAD51C* exon 1 +1G>T

4.1.6.2 Oligonucleotides for RT-PCR

#481: 5'-GCGCGCACGGCAAGA
5'-primer RT-PCR 2-intron-3-exon-reporter
#559: 5'-CTTTACGATGCCATTGGGA
5'-primer RT-PCR 2-intron- 3-exon-reporter

#1273: 5'-GATGCGGAGCAGCTCTAGGTTGGATTT
5'-primer hGH:
#1274: 5'-TTGACACCTACC AGGAGTTTGAAGAAG
3'-primer hGH

#2292: 5'-CCTGTTGTTCTACAATGTACACAT
5' primer for RT-PCR SVcATM Exon5-7
#2293: 5'-CTATGAGCACAGTAGAACTAAG
3' primer for RT-PCR SVcATM Exon5-7

#3034: 5'-GCCGCTGTACCAATCTCCTGTAAAAGAATTAG
5' primer for RT-PCR *BRCA2* exon 3
#3038: 5'-AGCAGTAGTATCATGAGGAAATACAGTTTCAG
3' primer for RT-PCR *BRCA2* exon 3

#3244: 5'-GAAGCAGCTCCCGCGAGGACCA
5'-primer for RT-PCR *FANCC* exon 1
#3245: 5'-CTGTGGTTCTTTGTTAATTAGACAACATAAGCACC
3'-primer for RT-PCR *FANCC* exon 4

Materials and Methods

#2911: 5'-CTGAGTGCTGAAAGTATATGAGATAATACACC

3'-primer for RT-PCR *FANCC* exon 5

#2922: 5'-GCTTATGTTGTCTAATTAACAAAGAAC

5'-primer for RT-PCR *FANCC* exon 4

#2923: 5'-CGCCTTTGAGTGTTAAATCC

3'-primer for RT-PCR *FANCC* exon 6

#3369: 5'-ATCGAATTCGTGCGGAGTTTGGCTGCTCCGGGG

5'-primer for RT-PCR *RAD51C* exon1

#3340: 5'-CCTCTCCCTTGTGTTTTTCTGCTATAAGC

3'-primer for RT-PCR *RAD51C* exon3

1.6.3 Oligonucleotides for RNA affinity chromatography

#2503: 5'-TAATACGACTCACTATAGG

T7 primer for *in vitro* RNA synthesis

5'-AAGGTACGTATGTTTAATCCAAATACCTCCCTATAGTGAGTCGTATTA

primer for RNA pulldown *ATM* exon 54 part I

5'-TGAATATCACACTTCTAACCAATACCTCCCTATAGTGAGTCGTATTA

primer for RNA pulldown *ATM* exon 54 part II

5'-GAATATTCTAGGAAAGACCCAAATACCTCCCTATAGTGAGTCGTATTA

primer for RNA pulldown *ATM* exon 54 part III

#2603: 5'-ACAAACTCCACATACCACTGGGCCTATAGTGAGTCGTATTA

primer for RNA pulldown *BRCA2* exon 6 wt

#2604: 5'-ACAAAGAGGGTGTATCCACTGGGCCTATAGTGAGTCGTATTA

primer for RNA pulldown *BRCA2* exon 6 del 707-717 (patient)

#2605:

5'GAAATATTCTAGGAAAGACAAGGTACGTATGTTTAATCCTATAGTGAGTCGTATTA

primer for RNA pulldown *BRCA2* exon 6 mt 707-717 (control)

4.1.7. Recombinant plasmids

Recombinant plasmids used here have an origin of replication (ori) for the amplification in *Escherichia coli* and contain the β -lactamase encoding the ampicillin resistance gene (*amp^r*) to allow selection in procaryotes. Additionally the plasmids carry the simian virus(SV40)

large T-antigen and the SV40early polyadenylation signal. HIV-1 sequences were derived from the vector pNLA-1 which is a cDNA derivative of NL4-3. The numeration system is based on the output sequence of NL4-3. Sequences of all generated recombinant plasmids were verified by DNA sequencing of the respective target regions.

4.1.7.1. Three-exon-two-intron splicing reporter minigenes

ATM exon 54 or ATM exon 9

The parental 2-intron-3-exon splicing reporter minigenes LTR-SD4-ATM-exon54-(minus'3intron)-SA5opt and LTR-SD4-ATM-exon54-(minus'3intron)-SA3 were kindly provided by Dr. K. Neveling (Neveling, K., diploma thesis, 2004). The 2-intron-3-exon splicing reporter minigene was driven by the HIV-1 5' LTR (long terminal repeat) and terminated by the SV40 polyadenylation signal. The 5' half of this construct comprised of the HIV-1 exon 1, a strong HIV-1 5' ss - which is a hybrid of the HIV-1 SD1 and SD4 (CtGGTAAGTAT) here referred as SD1-with an HBond score of 20.20 and 68 base pairs of the HIV-1 intron 1. The 3'half of the construct was composed of intron 2 and an HIV-1 derived 3' splice site (3'ss) - also called splice acceptor (SA). Exon 3 in this splicing reporter is a hybrid of the CAT-ORF (chloramphenicol-acetyl-transferase-open reading frame) and the HIV-1 RRE (rev responsive element). Unique restriction sites within the reporter construct allowed both easy insertion of an internal test exon and splice site replacement.

Based on LTR-SD4-ATM-exon54-(minus'3intron)-SA5opt the following constructs were cloned:

LTR-SD1/4-ATM-exon54-(minus'3intron)-SA5Py+: For cloning of LTR-SD1/4-ATM-exon54-(minus'3intron)-SA5Py+ the parental 2-intron-3-exon splicing reporter minigene LTR-SD4-ATM-exon54-(minus'3intron)-SA5opt was digested with the restriction enzymes *XhoI/MscI* and ligated with the *XhoI/MscI* fragment of LTR-SD4-Ex2-SD4-Py+ kindly provided by Dr. K. Neveling (Neveling, K. diploma thesis, 2004).

LTR-SD1/4-ATM-exon54-(minus'3intron)-SA5Py++: For cloning of LTR-SD1/4-ATM-exon54-(minus'3intron)-SA5Py+ the parental 2-intron-3-exon splicing reporter minigene LTR-SD4-ATM-exon54-(minus'3intron)-SA5opt was digested with the restriction enzymes *XhoI/MscI* and ligated with the *XhoI/MscI* fragment of LTR-SD4-Ex2-SD4-Py++ kindly provided by Dr. K. Neveling (Neveling, K. diploma thesis, 2004).

LTR-SD1/4-ATM-exon54-3'intron-part-I-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2302

and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-II-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-II-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2303 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-III-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-III-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2304 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-I+II-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I+II-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2305 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-II+III-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-II+III-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2306 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-III+II-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-III+II-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2374 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. the parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1/4-ATM-exon54-3'intron-part-I+III-SA3: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I+III-SA3 a polymerase-chain-reaction (PCR) using the primers #2301 and #2375 and the plasmid LTR-SD1/4-ATM-exon54-SA3 as template was performed. The parental vector LTR-SD1/4-ATM-exon54-(minus'3intron)-SA3 was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

The plasmids LTR-SD4-ATM-exon9-SA3, LTR-SD4-ATM-exon9-(minus'3intron)-SA3, LTR-SD4-ATM-exon9-SA5opt, LTR-SD4-ATMEx9 (mt, minus 3' intron)-SA5opt, SV-ATM-exon-

53-55 and SV-ATM-exon 8-10 were kindly provided by Dr. K. Neveling (Neveling, K., diploma thesis, 2004; Neveling, K., thesis, 2007).

RAD51C exon 6

LTR-SD1-RAD51C-exon6-SA5opt: For cloning of LTR-SD1-RAD51C-exon6-SA5opt a PCR using the primers #3348 and #3349 and human gDNA as template was performed. The parental vector LTR-SD1-FANCC-Ex1-GC-2/-3opt-SA5opt was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

LTR-SD1-RAD51C-exon6-904+5G>T-SA5opt: For cloning of LTR-SD1-RAD51C-exon6-904+5G>T-SA5opt a mutagenesis PCR of the wild type construct using primers #3348 and #3351 and primers #3349 and #3350 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-GC-2/-3opt-SA5opt.

BRCA2 exon 6

LTR-SD1-BRCA2-exon6-SA5opt: For cloning of LTR-SD1-BRCA2-exon6-SA5opt a PCR using the primers #197 and #2120 and human SVh-BRCA2-SA-exon6 as template was performed. The parental vector LTR-SD1-SA5opt-C-exon2 SD4-SA5opt was digested with *EcoRI/XhoI* and ligated the *EcoRI/XhoI* digested PCR product.

The plasmid LTR-SD1-BRCA2-exon6del-SA5opt was kindly provided by Dr. K. Neveling (Neveling, K., diploma thesis, 2004; Neveling, K., thesis, 2007).

FGB exon 6-8

The parental 2-intron-3-exon splicing reporter minigenes pT-Bbeta-wt and pT-beta-IVS7+1G>T were kindly provided by Dr. Silvia Spina and Dr. Emanuele Buratti (Spina et al., 2006).

pT-beta-IVS7+1G>T-mt-c1: For cloning of pT-beta-IVS7+1G>T-mt-c1 a mutagenesis PCR of pT-Bbeta-IVS7+1G>T using primers #2619 and #2622 and primers #2620 and #2621 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-IVS7+1G>T.

pT-beta-IVS7+1G>T-mt-c1/c2: For cloning of pT-beta-IVS7+1G>T-mt-c1/c2 a mutagenesis PCR of pT-Bbeta-IVS7+1G>T using primers #2619 and #2647 and primers #2620 and #2646 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-IVS7+1G>T.

Materials and Methods

pT-beta-IVS7+1G>T-mt-c1/c2/c3: For cloning of pT-beta-IVS7+1G>T-mt-c1/c2/c3 a mutagenesis PCR of pT-Bbeta-IVS7+1G>T using primers #2619 and #2624 and primers #2620 and #2623 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-IVS7+1G>T.

pT-beta-c1-15.8: For cloning of pT-beta-c1-15.8 a mutagenesis PCR of pT-Bbeta-wt using primers #2619 and #2765 and primers #2620 and #2764 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c1-18.8: For cloning of pT-beta-c1-18.8 a mutagenesis PCR of pT-Bbeta-c1-15.8 using primers #2619 and #2872 and primers #2620 and #2871 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c1-20.8: For cloning of pT-beta-c1-20.8 a mutagenesis PCR of pT-Bbeta-c1-15.8 using primers #2619 and #2874 and primers #2620 and #2873 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c1-23.8: For cloning of pT-beta-c1-23.8 a mutagenesis PCR of pT-Bbeta-c1-15.8 using primers #2619 and #2876 and primers #2620 and #2875 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c3-15.8: For cloning of pT-beta-c3-15.8 a mutagenesis PCR of pT-Bbeta-wt using primers #2619 and #2925 and primers #2620 and #2924 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c3-18.8: For cloning of pT-beta-c3-18.8 a mutagenesis PCR of pT-Bbeta-wt using primers #2619 and #2927 and primers #2620 and #2926 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c3-20.8: For cloning of pT-beta-c3-20.8 a mutagenesis PCR of pT-Bbeta-wt using primers #2619 and #2929 and primers #2620 and #2928 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-c3-23.8: For cloning of pT-beta-c3-23.8 a mutagenesis PCR of pT-Bbeta-wt using primers #2619 and #2931 and primers #2620 and #2930 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-wt.

pT-beta-IVS7+1G>T-5'ss-opt: For cloning of pT-beta-IVS7+1G>T-5'ss-opt a mutagenesis PCR of pT-Bbeta-IVS7+1G>T using primers #2619 and #2651 and primers #2620 and #2650 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-Bbeta-IVS7+1G>T.

pT-beta-IVS7+1G>T-5'ss-opt-+9G>C: For cloning of pT-beta-IVS7+1G>T-5'ss-opt-+9G>C a mutagenesis PCR of pT-beta-IVS7+1G>T-5'ss-opt using primers #2619 and #2731 and primers #2620 and #2732 was performed. The *PpuMI/XmaI* digested PCR product was ligated with the *PpuMI/XmaI* digested parental vector pT-beta-IVS7+1G>T-5'ss.

FANCC exon 2

The parental 2-intron-3-exon splicing reporter minigenes LTR-SD1-FANCC-Ex1-GT-SA5opt, LTR-SD1-FANCC-Ex1-TT-SA5opt and LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt were kindly provided by Dr. K. Neveling (Neveling, K., diploma thesis, 2004, Neveling, K., thesis, 2007).

LTR-SD1-FANCC-Ex1-AA-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-AA-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers #2717 and #2724 and primers #2718 and #2723 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/3opt-SA5opt.

LTR-SD1-FANCC-Ex1-AT-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-AT-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers #2717 and #2730 and primers #2718 and #2729 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/3opt-SA5opt

LTR-SD1-FANCC-Ex1-CC-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-CC-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers #2717 and #2722 and primers #2718 and #2721 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/3opt-SA5opt

LTR-SD1-FANCC-Ex1-CT-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-CT-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers #2717 and #2728 and primers #2718 and #2727 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/3opt-SA5opt.

LTR-SD1-FANCC-Ex1-GC-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-GC-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers

#2717 and #2720 and primers #2718 and #2719 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/-3opt-SA5opt.

LTR-SD1-FANCC-Ex1-GG-2/3-opt-SA5opt: For cloning of LTR-SD1-FANCC-Ex1-GG-2/3-opt-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-TT-2/3-opt-SA5opt using primers #2717 and #2726 and primers #2718 and #2715 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-TT-2/-3opt-SA5opt.

LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-SA5opt: For cloning of LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-GT-SA5opt using primers #2717 and #3457 and primers #2718 and #3456 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-GT-SA5opt.

LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-/-3A-SA5opt: For cloning of LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-GT-SA5opt using primers #2717 and #3465 and primers #2718 and #3464 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-GT-SA5opt.

LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-7,8AG-SA5opt: For cloning of LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-7,8AG SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-GT-SA5opt using primers #2717 and #3463 and primers #2718 and #3462 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-GT-SA5opt.

LTR-SD1-FANCC-Ex2-IVS7-1G>T-7,8AG-SA5opt: For cloning of LTR-SD1-FANCC-Ex2-IVS7-1G>T/+4A-SA5opt a mutagenesis PCR of LTR-SD1-FANCC-Ex1-GT-SA5opt using primers #2717 and #3459 and primers #2718 and #3458 was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector LTR-SD1-FANCC-Ex1-GT-SA5opt.

The plasmids LTR-SD1-FANCC-Ex2-IVS7-1G>T-SA5opt and its derivatives were cloned by K. Schöneweis (Schöneweis K., diploma thesis, 2010).

LTR-SD1-FANCC-Ex2-3: For cloning of LTR-SD1-FANCC-Ex2-3 a PCR using the primers #3714 and #3717 and human gDNA as template was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *XhoI/BamHI* fragment of SVcrev and with *XhoI/BamHI* digested parental vector LTR-SD1-FANCC-Ex1-TT-SA5opt.

LTR-SD1-FANCC-Ex2-3+1G>T: For cloning of LTR-SD1-FANCC-Ex2-3+1G>T mutagenesis PCR of the wild type construct LTR-SD1-FANCC-Ex2-3 using the mutagenesis primers #3718 and #3718 was performed (Stratagene mutagenesis kit.)

4.1.7.2. 1-intron-2-exon splicing reporter minigenes

SV-RAD51C-exon1-2: For cloning of SV-RAD51C-exon1-2 a PCR using the primers #3369 and #3562 and human gDNA as template was performed. The *EcoRI/XhoI* digested PCR product was ligated with the *EcoRI/XhoI* digested parental vector SVcrev.

SV-RAD51C-exon1-2-c.145+1G>T: For cloning of SV-RAD51C-exon1-2-c.145+1G>T mutagenesis PCR of the wild type construct SV-RAD51C-exon1-2 using the mutagenesis primers #3334 and #3335 was performed (Stratagene mutagenesis kit.)

4.1.7.3. SV-env/eGFP reporter plasmids

SV-GAR-SD1-del-vpu-env-D36G-eGFP

The parental HIV-1 glycoprotein/eGFP expression plasmid SV-GAR-SD1-del-vpu-env-D36G-eGFP was kindly provided by Dr. M. Freund (Freund, M., thesis, 2003; Caputi et al. 2004). The subgenomic SV-env/eGFP splicing reporter contains the coding sequence for the viral glycoprotein (*env*). The eGFP (enhanced green fluorescent protein) coding sequence was cloned into the plasmid by substitution of the 3'-terminal region for a PCR product amplified of pEF eGFP-neo (kindly provided by Prof. Dr. Dirk Lindemann) as a template. Thereby the cytoplasmic domain of the gp41 subunit of the viral glycoprotein was partially removed because it is dispensable for fusogenicity assays and syncytia formation in the context of Hela-T4⁺ cells stably expressing the viral entry receptor CD4.

The plasmids SV-neutral-r(CCAAACAA)₃-SD1-neutral-r(CCAAACAA)₃-delvpuenv-eGFP D36GpA and SV-neutral-r(CCAAACAA)₃-SD1-IAS-delvpuenv-eGFPD36GpA were kindly provided by S.Erkelenz (Erkelenz, S., thesis, 2012; Zhang et al., 2009).

Based on these plasmids the following splicing reporter constructs were cloned:

SV-neutral-r(CCAAACAA)₃-SD1-ATM-intron-54-delvpuenv-eGFP-D36GpA: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I+II-SA3 a PCR using the primers #3550 and #3561 and the plasmid SV-GAR-SD1-del-vpu-env-D36G-eGFP as template was performed. The *SacI/NdeI* digested PCR product was ligated with the *SacI/Clal* fragment of SV-neutral-r(CCAAACAA)₃-SD1delvpuenv-eGFP-D36GpA and with *Clal/NdeI* digested parental vector SV-neutral-r(CCAAACAA)₃-SD1delvpuenv-eGFP-D36GpA.

SV-neutral-r(CCAAACAA)3-SD1-ATM-intron-54-part-I-delvpuenv-eGFP-D36GpA: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-I-II-SA3 a PCR using the primers #3551 and #3561 and the plasmid SV-GAR-SD1-del-vpu-env-D36G-eGFP as template was performed. The *SacI/NdeI* digested PCR product was ligated with the *SacI/Clal* fragment of SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA and with *Clal/NdeI* digested parental vector SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA.

SV-neutral-r(CCAAACAA)3-SD1-ATM-intron-54-part-II-delvpuenv-eGFP-D36GpA: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-II-SA3 a PCR using the primers #3554 and #3561 and the plasmid SV-GAR-SD1-del-vpu-env-D36G-eGFP as template was performed. The *SacI/NdeI* digested PCR product was ligated with the *SacI/Clal* fragment of SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA and with *Clal/NdeI* digested parental vector SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA.

SV-neutral-r(CCAAACAA)3-SD1-ATM-intron-54-part-III-delvpuenv-eGFP-D36GpA: For cloning of LTR-SD1/4-ATM-exon54-3'intron-part-III-SA3 a PCR using the primers #3553 and #3561 and the plasmid SV-GAR-SD1-del-vpu-env-D36G-eGFP as template was performed. The *SacI/NdeI* digested PCR product was ligated with the *SacI/Clal* fragment of SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA and with *Clal/NdeI* digested parental vector SV-neutral-r(CCAAACAA)3-SD1delvpuenv-eGFP-D36GpA

The control plasmids SV-GAR-SD4-delvpuenvD36GeGFP and SV-HIV#18-SD4-delvpuenvD36GeGFP kindly provided by Dr. M. Freund; Freund, M., thesis, 2004).

4.1.7.4. U1 snRNA expression plasmids

pUCB-U1wt: The plasmid pUCB-U1wt was kindly provided by Prof. Dr. Alan Weiner, Seattle, USA.

pUCB-U1alphaTT: For cloning of pUCB-U1alphaTT a PCR using primers #2040 and #1131 and the plasmid pUCBU1wt as the template. The *XhoI/Bgl-II* digested PCR product was ligated with the *XhoI/Bgl-II* digested plasmid pUCB del U1(cloned by M.Freund).

pUCB-U1-TTcom: For cloning of pUCB-U1-TTcom a PCR using primers #2809 and #1131 and the plasmid pUCBU1alphaTT as the template. The *XhoI/Bgl-II* digested PCR product was ligated with the *XhoI/Bgl-II* digested plasmid pUCBU1alphaTT.

pUCB-U1-5A: For cloning of pUCB-U1-TTcom a PCR using primers #2811 and #1131 and the plasmid pUCBU1alphaTT as the template. The *XhoI/Bgl-II* digested PCR product was ligated with the *XhoI/Bgl-II* digested plasmid pUCBU1alphaTT.

pUCB-U1-IVS4: For cloning of pUCB-U1-IVS4 a PCR using primers #2810 and #1131 and the plasmid pUCBU1alphaTT as the template. The *XhoI/Bgl-II* digested PCR product was ligated with the *XhoI/Bgl-II* digested plasmid pUCBU1alphaTT.

U17A+TOPO 10T, U17A+TOPO2C: The plasmids U17A+TOPO 10T and U17A+TOPO2C were kindly provided by Dr. Anders Virtanen, Upsala, Sweden

pUCB-U1A7-10T: For cloning of pUCB-U1A7-10T a PCR using primers #2497 and #2564 and the plasmid pUCB-U1A7-10T as the template. The *Bgl-II/BsiWI* digested PCR product was ligated with the *Bgl-II/BsiWI* digested plasmid pUCBU1alphaTT.

pUCB-U1A7-2C: For cloning of pUCB-U1A7-2C a PCR using primers #2497 and #2564 and the plasmid pUCB-U1A7-10T as the template. The *Bgl-II/BsiWI* digested PCR product was ligated with the *Bgl-II/BsiWI* digested plasmid pUCBU1alphaTT.

The plasmids pUCB-U1-FGB7-IVS7-+1G>T and its derivatives were cloned by K. Schöneweis (Schöneweis K., diploma thesis, 2010).

4.1.7.5. Control plasmids

pXGH5 (Selden et al. 1986) was cotransfected to monitor transfection efficiency in quantitative and semi-quantitative RT-PCR analyses. The plasmid encodes the human growth hormone 1 (hGH1) under control of the mouse metallothionein-1 promoter.

4.1.7.6. Plasmids for protein expression

SVctat: The plasmid SVctat expresses the *tat*-cDNA derived from the HIV-1 isolate NL4-3 under control of the SV40 early promoter.

SVcrev: The plasmid SVcrev expresses the *rev*-cDNA derived from the HIV-1 isolate NL4-3 under control of the SV40 early promoter and was cloned by replacing the *EcoRI/XhoI*-fragment from pSVT7 with the *EcoRI/XhoI*-fragment from pUHcrev.

pQE-80L-T7RNAP: expression plasmid encoding recombinant T7 RNA polymerase. The gene encoding the RNA polymerase of the bacteriophage T7 was subcloned into the plasmid pQE-80L (Qiagen) generating plasmid pQE-80L-T7RNAP and is expressed with an Nterminal 6xHis-tag. The ampicillin resistance gene is used for selection. was a kindly provided by Prof. Dr. M. Caputi (Florida Atlantic University, USA).

4.1.5.7. Lentiviral vectors

The parental lentiviral vector pCL1-N-PB was kindly provided by Prof. Dr. Helmut Hanenberg, Indianapolis, USA. The lentiviral vector pCL1N-PB-WT-U1 E/X was cloned by

Materials and Methods

ligation of *EcoRI/XhoI* fragment of pCL1-N-MCS-PB with the *EcoRI/SalI* fragment of pUCBU1. This vector was kindly provided by Dipl.-Biologin Stephanie Borkens.

pCL1N-PB-U1-alphaTT: For cloning of the lentiviral vector pCL1N-PB-U1-alphaTT the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-alpha TT.

pCL1N-PB-U1-TTcom: For cloning of the lentiviral vector pCL1N-PB-U1-TTcom the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A-E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-TTcom.

pCL1N-PB-U1-5A: For cloning of the lentiviral vector pCL1N-PB-U1-5A the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A-E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-5A.

pCL1N-PB-U1-IVS4: For cloning of the lentiviral vector pCL1N-PB-U1-IVS4 the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A-E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-IVS4.

pCL1N-PB-U1-IVS4-ext3: For cloning of the lentiviral vector pCL1N-PB-U1-IVS4-ext3 the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A-E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-IVS4-ext3.

pCL1N-PB-U1-IVS4-ext4: For cloning of the lentiviral vector pCL1N-PB-U1-IVS4-ext4 the *EcoRI/PvuI* and *PvuI/XhoI* fragment of CL1-N-PB-U1-6A-E/X were ligated with the *EcoRI/XhoI* fragment of pUCBU1-IVS4-ext4.

pczVSV-G: The pczVSV-G expression plasmid coding for *vesicular stomatitis virus* G protein was kindly provided by Prof. Dr. Helmut Hanenberg, Indianapolis, USA (Mochizuki et al., 1998).

pCD/NL-BH: HIV-1 derived helper plasmid coding for *gag* and *pol* (*env* deletion) kindly provided by Prof. Dr. Helmut Hanenberg, Indianapolis, USA

puc2CL7-EGwo: lentiviral vector encoding the greenfluorescent protein eGFP,

puc2CL7-dTOMwo: lentiviral vector encoding the red fluorescent protein Tomato

Both vectors kindly provided by Prof. Dr. Helmut Hanenberg, Indianapolis, USA.

MFPCN: retroviral vector **expressing the FANCC** cDNA

MSCV: *murine stem cell virus* derived retroviral vector

Both vectors kindly provided by Prof. Dr. Helmut Hanenberg, Indianapolis, USA

4.1.8. Antibodies

4.1.8.1. Primary Antibodies

α hnRNP H: The rabbit-derived α hnRNP H polyclonal antibody (AN113) was kindly provided by Prof. Dr. Douglas Black, USA and used in 1: 5000 dilution.

α hnRNP A1: For detection of hnRNP A1 a polyclonal goat-derived α hnRNP A1 antibody (sc-10032, Santa Cruz) was used in 1: 200 dilution.

α hnRNP M4: For detection of hnRNP M4 a polyclonal mouse-derived α hnRNP M4 antibody (Santa Cruz) was used in 1: 200 dilution.

α FANCD2: For detection of FANCD2 a monoclonal mouse-derived α FANCD2 antibody (Santa Cruz) was used in 1: 800 dilution.

4.1.8.2. Secondary Antibodies

α rabbit: Sigma (A6154), horseradish-peroxidase-linked, 1 : 2.000 dilution

α goat: Dianova (705-035-147), horseradish-peroxidase-linked, 1 : 5.000 dilution

α mouse: Amersham Biosciences (NA9310), horseradish-peroxidase-linked, 1:2.000 dilution

4.2. Methods

4.2.1. Cloning

4.2.1.1. Polymerase Chain Reaction (PCR)

DNA fragments used for cloning of recombinant plasmids were amplified in a volume of 100 µL using 2.5 U Pwo DNA Polymerase (Roche) and 100 ng DNA template in a reaction containing 10 mM Tris-HCl, pH 8.85, 25 mM KCl, 5 mM (NH₄)₂SO₄, 2 mM MgSO₄, 200 µM desoxynucleosidetriphosphates (dNTP) (Applied Biosystems), 200 nM sense and antisense primer, respectively. DNA was amplified in a Robocycler Gradient 96 (Stratagene) using following conditions: denaturation 1 x 94°C 3 min; amplification 30 cycles à 94°C 30 sec, 60°C 1 min 72°C 1 min; final elongation 1 x 72°C 10 min. Long templates e.g. amplicons from human gDNA were amplified with High Fidelity polymerase (Roche) according to the recommendation of the manufacturer.

PCR products were purified from the reaction by adding 1 vol. phenol (Roth) and 1 vol. chloroform/isoamyl alcohol (24:1). After vortexing phases were separated by centrifugation (12.000 rpm, 5 min, Eppendorf microcentrifuge) and the supernatant again extracted with 1 vol. chloroform/isoamylalcohol (24:1). After separation (12.000 rpm, 5 min, Eppendorf microcentrifuge), DNA in the aqueous phase was precipitated with 0.1 vol. 4M LiCl and 2.5 vol ethanol (96%) at -80°C for 20 min. After centrifugation (12.000 rpm, 15 min, Eppendorf microcentrifuge), DNA was washed with 200 µL ethanol [70% (v/v)] (12.000 rpm, 10 min, Eppendorf microcentrifuge), air-dried and resuspended in 30 µL ddH₂O.

4.2.1.2 Restriction and purification of PCR products or plasmid fragments using agarose gel electrophoresis

The DNA restriction was performed with restriction enzymes according to the recommendation of the manufacturer (New England Biolabs, Roche, MBI Fermentas). The reaction was carried out with 1-3 µg DNA in a total volume of 20µl. Using a 1% agarose gel (Biozym) und 1 x TBE (10 x TBE: 890mM Tris-HCl, pH 8; 980 mM Boracid; 25 mM Na-EDTA, pH 8) as running buffer the restriction products were separated by their size. The desired product was excised using 370nm UV light.

4.2.1.3. Ligation

If necessary, the restricted targeted vector was dephosphorylated before ligation using Alkaline Phosphatase (NEB). The reaction was performed in at total volume with 5U Phosphatase, 1/10 Volumen recommended buffer and 17µl of the gel eluted DNA.

Target vectors and PCR product with complementary ends were ligated using T4-DNA ligase (NEB) in a 20µl reaction containing 1µl T4-DNA ligase (400U/ml), 2µl 10x T4 ligase buffer (100mM MgCl₂, 100mM DTT, 10mM ATP, 500mM Tris-HCL, pH 7.5). Relative amounts of fragments were adjusted to an approximately 3:1 insert:vector molar ratio. In generally, ligation reactions were performed over night at 16°C and afterwards transformed into competent E.coli cells.

4.2.1.4. Transformation

For amplification of plasmid DNA competent E.coli DH5αF'H (Invitrogen) cells were mixed with either 5 µl from the ligation reaction or 0.5 ng of highly pure plasmid DNA, respectively, and incubated for 20 min on ice. After heat-shocking for 90 sec at 42°C, the transformed bacteria cells were chilled on ice and 800 µl LB medium were added prior to incubation for 1h at 37°C and 220rpm. Subsequently, 200µl of the samples were streaked on ampicillin-containing LB agar plates (100µg/ml) and grown overnight at 37°C.

4.2.1.5. Analytical plasmid DNA isolation

Single colonies were picked from an agar plate and transferred into 5 ml LB medium plus ampicillin (100µg/ml). Bacteria cells were incubated overnight at 37°C meanwhile continuously rotating at 220 rpm. After centrifugation for 1 min at maximum speed (Eppendorf micocentrifuge), cell pellets were resuspended in 300 µl buffer 1 (50mM Tris-HCL pH7.5, 10mM EDTA, 400 µg/ml RNase A) and lysed at RT for 5 min by addition of buffer 2 (0.2M NaOH, 1% (w/v) SDS). Addition of buffer 3 (3M KAc, pH 5.5) neutralised the lysates and precipitated the proteins and bacterial debris, while the plasmid DNA remained in solution. Following centrifugation for 15 min at 12.000 rpm and 4°C the supernatant was removed and precipitated with 0.7 vol. isopropanol. Plasmid DNA was obtained by

centrifugation for 15-30 min at full speed and RT, washing with 120 µl ethanol (70% (v/v)) and resuspension in TE buffer (pH 8.0). Positive clones were confirmed by restriction analyses with respective DNA endonucleases and specific digestion patterns resolved by gel electrophoresis using 1% agarose (LE, Biozym).

For immediate sequencing plasmid DNA was purified using silica-columns (FastPlasmid™ Mini, Eppendorf).

4.2.1.6. Preparative plasmid DNA isolation

150 ml of ampicillin-containing LB medium (100µg/ml) were inoculated with 200 µl of a 5 ml bacterial pre-culture (see II.2.4) or a cryo-stock (prepared from 700 µl bacteria culture and 300 µl glycerol, stored at -80°C) and grown overnight at 37°C constantly shaking at 220 rpm. Bacteria cells were pelleted by centrifugation for 10 min at 5.000 rpm (Beckmann JS-21, JA14 Rotor) and resuspended in 4 ml buffer 1 (50mM Tris-HCL pH7.5, 10mM EDTA, 400 µg/ml RNase A). For alkaline lysis 4 ml buffer 2 (0.2M NaOH, 1% (w/v) SDS) was added, followed by incubation for 5 min at RT. Addition of 4 ml buffer 3 (3M KAc, pH 5.5) neutralised the lysate and after 15 min incubation on ice, proteins and bacterial debris was pelleted via centrifugation for 30 min at 10.000 rpm and 4°C (Beckmann JS-21, JA14 rotor). Supernatant was cleared through a folded filter (Schleicher & Schüll, 5951/2, Ø150 mm) and loaded on silica-based anion-exchange-columns (Plasmid DNA Midi Kit, Qiagen) for plasmid DNA purification according to the manufacturer's instructions. Plasmid DNA eluted from the columns was precipitated with 0.7 vol. isopropanol and centrifuged for 30 min at 10.000 rpm and RT (Beckman JS-21, JS13.1). After washing with 70% ethanol, plasmid-DNA was air-dried and resuspended in 50-300µl TE buffer (pH 8). DNA concentrations were quantitated by spectral photometry at 260 nm and 280nm (NanoDrop) and adjusted to ~1 µg/ µl with TE buffer (pH 8). Positive clones were controlled by sequencing reactions performed in the Analytical Core Facility of the Biological-Medical Research Centre (BMFZ, HHUD).

4.2.1.7. DNA sequencing

To confirm mutations introduced into plasmids or to identify RT-PCR products after gel electrophoretic separation, DNA was sequenced using the dideoxy method by Sanger et al. Sequencing reactions were performed in a volume of 20 µL containing 5 pmol primer, 4 µL BigDye v1.1 RR-24 reaction mix (includes DNA-polymerase, labelled and unlabelled dNTPs and reaction buffer) and template-dependent DNA amounts (200-500 ng DNA for plasmid and 50 ng for PCR product sequencing). Sequencing PCR was performed by 26 cycles of denaturation (94°C, 0:30 min), annealing (55°C, 0:30 min) and elongation (60°C, 4:00 min). Sequencing reactions were purified by ethanol/sodium acetate precipitation [0.3 mM sodium

acetate, pH 5.2, 78% (v/v) ethanol] for 15 min. After centrifugation at 14,000 rpm for 20 min the pellet was washed with 250 μ L 75% (v/v) ethanol. After a second centrifugation at 14,000 rpm for 5 min the supernatant was removed, the pellet was air-dried and dissolved in 10 μ L formamide. Sequencing reactions were protected from light and stored at 4°C till separation on an automated DNA sequencer (3130 Genetic Analyzer, Applied Biosystems). Additional sequencing reactions were performed by the Analytical Core Facility of the Biological-Medical Research Centre (BMFZ, HHUD).

4.2.2. Eukaryotic cell culture

4.2.2.1 Cell Culture and Transfection

For the splicing reporter assay, 2.5×10^5 HeLa cells were seeded per 6-well plate in Dulbecco's modified Eagle's medium (DMEM; Invitrogen) supplemented with 10% fetal bovine serum (Pan Biotech), 2 mM L-glutamine, and 50 U/ml penicillin and streptomycin (both Invitrogen), 24 hr before transfection. Cells were transfected with 1 μ g of the splice reporter constructs or their mutated derivatives with FuGENE 6 according to the manufacturer's protocol (Roche Molecular Biochemicals). For cotransfection experiments, cells were transfected with 1 mg of pXGH5 encoding human growth hormone and 2 μ g of the respective plasmid.

Primary fibroblast strains were established by standard cell culture procedures and maintained in complete DMEM in high humidity incubators in an atmosphere of 5% (v/v) CO₂ and 5% (v/v) O₂. For splicing analysis of the endogenous transcript, fibroblasts were seeded in T75 flasks and grown to approximately 80% confluency. For transfection of immortalized fibroblasts, cells were seeded 24 hr before transfection and transfected with 16 μ g of the respective plasmid and 8 μ g of pXGH5 using FuGENE 6 (Roche). For both assays RNA was isolated 30h after transfection.

The EBV immortalized lymphoblastoid B-cell line were cultured in RPMI1640 (Invitrogen) medium supplemented with 10% fetal bovine serum (Pan Biotech), 2 mM L-glutamine, and 50 U/ml penicillin and streptomycin (both Invitrogen).

4.2.4. Flow cytometrical analysis of transiently transfected HeLa cells

Cells samples were collected and washed with PBS. After trypsination for 5 min at 37°C and several washing steps in FACS buffer (PBS + 3% FCS), samples were resuspended and acquired on a FACS-CANTO II cytometer (Becton Dickinson). To quantify the mean fluorescence intensity data were exported to the FlowJo (Tree Star, Inc.) analysis software.

4.2.3 Lentiviral particle production

For the production of lentiviral particles, 6×10^6 293T cells were plated per 10-cm cell culture dish 24h prior to transfection with $6\mu\text{g}$ of pCD/NL-BH, $6\mu\text{g}$ of an expression plasmid coding for vesicular stomatitis virus G protein and of $6\mu\text{g}$ pCL1NPB-U1, using polyethylenimine (PEI, Aldrich). Supernatants were harvested 48h after transfection and filtered through a $0.45\text{-}\mu\text{m}$ filter. Functional Neomycin titers of the lentiviral vectors (LV) were determined in HT1080 cells, plated at 3.5×10^4 cells per well in 6-well plates the day before, and infected with different dilutions of either LV. Cells were washed and incubated for 7d with fresh medium containing 0.8mg/ml G418 (Invitrogen). Colonies were fixed with methanol and stained with methylenblue. Titters were calculated, usually obtaining 10^{6-7} infectious virus particles/mL.

4.2.5. Reverse transcriptase (RT)-PCR analysis

4.2.5.1. Isolation of total RNA using anionic exchange columns

RNA of adherent cells was isolated using microspin columns containing a silica-matrix (GenElute™ Mammalian Total RNA Kit, Sigma). Cells were washed twice with 2 mL PBS each and lysed by addition of 250 μL lysis buffer per 6-well. The lysate was centrifuged through filtration columns (14.000 rpm, 2 min, Eppendorf microcentrifuge). The flow-through was mixed with 1 vol. ethanol (70%) and loaded on RNA binding columns by centrifugation (14.000 rpm, 15 sec, Eppendorf microcentrifuge). Column-bound RNA was washed with 500 μL washbuffer 1, the column transferred to another reaction tube and subsequently washed with 500 μL wash buffer 2 (14.000 rpm, 15 sec, Eppendorf microcentrifuge). A second washing step with wash buffer 2 was performed for 2 min (14.000 rpm, Eppendorf microcentrifuge). RNA was eluted from the column with 50 μL elution buffer. After determining the concentration by photometry at 260nm and 280nm RNA was stored at -80°C until further analyses.

4.2.5.2. Reverse transcription and PCR analysis

RT-PCR was performed using SuperScript™ III RT-PCR System with Platinum Taq Polymerase (Invitrogen). For analysis of the splicing pattern, prior to RT, $4\mu\text{g}$ (endogenous transcript) or $1\mu\text{g}$ of total RNA (splicing reporter transcript) was subjected to DNase I digestion with 10U of DNase I at 70°C for 5min (Roche), $2\mu\text{L}$ of the DNase I-digested RNA samples were reverse transcribed with SuperScript™ III RT-PCR System with Platinum Taq Polymerase (55°C , 30 min) using $0.2\mu\text{M}$ of the transcript specific antisense primer. $2\mu\text{L}$ of the SuperScript III/Taq-DNA polymerase mixture (Invitrogen) and subsequently amplified with transcript specific primers according to the protocol provided by the manufacturer in a 20

μ L reaction (initial denaturation: 94°C 3 min; 30 cycles: 94°C 30 sec, 60°C 1 min, 68°C 1 min; final elongation : 68°C 10 min) (Robocycler Gradient 96 , Stratagene). To ensure a linear PCR amplification range allowing semiquantitative assessment of the spliced products, cycle number of the PCR reaction was adapted to the specific transcript conditions. As a control for transfection efficiency, human growth hormone (hGH) mRNA was detected with a specific primer pair. As negative control for remaining DNA contamination of each sample, a second assay was performed with Platinum Taq Polymerase (Invitrogen).

4.2.5.3. Native gel electrophoresis and EtBr staining to visualize RT-PCR products

10 μ l of the RT-PCR products were separated on 6-10% non-denaturing polyacrylamide (PAA) gels (Rotiphorese Gel 30, Roth) using 1xTBE running buffer. Gels were run at 200 V dependent on the percentage for 1h up to 2h, stained with ethidium bromide (EtBr, 4 μ g/ml in 1xTBE) for 5-10min and exposed to UV light excitation in the Lumi-Imager F1 (Roche).

4.2.5.4. Purification of RT-PCR products from native polyacrylamide gels (PAA)

RT-PCR products visualized by EtBr-staining were cut out from the gels using long wave UV light (320 nm) and diced into small pieces before transfer into a 1.5ml reaction tube. DNA was eluted from the gel by overnight incubation at 37°C in PAA elution buffer (0.5M sodium acetate, 0.1% (w/v) SDS, 1mM EDTA). After centrifugation for 1 min at full-speed and 4°C (Eppendorf microcentrifuge) supernatant was removed into a new 1.5 ml reaction tube and gel pieces once more mixed with 0.5 vol. PAA elution buffer. After centrifugation, both supernatants were pooled and purified from gel leftovers through filtration using glass fibre filters (GF/C filter, Whatman). DNA was precipitated by addition of 2 vol. ethanol (96%) on ice for 30 min followed by centrifugation for 10 min at full-speed and 4°C (Eppendorf microcentrifuge). Pellets were resuspended in 200 μ l TE (pH 8) and 25 μ l 3M sodium acetate (pH 5.2). Precipitation with 2 vol. ethanol (96%) was repeated and after a new centrifugation DNA pellets were washed with 120 μ l ethanol (70% (v/v)), air-dried and resuspended in 10 μ l ddH₂O. DNA was used as template for re-amplification by a PCR reaction using proofreading Pwo DNA polymerase (Roche). and purified by phenol/chloroform extraction. The DNA was purified by phenol/chloroform extraction or from a 1% agarose gel using the Qiagen gel extraction kit. After concentration was photometrically measured, 10 up to 50 ng of DNA were applied per 20 μ L sequencing.

4.2.6. RNA affinity chromatography

4.2.6.1. Purification of DNA oligos

Full-length oligonucleotides used for *in vitro*-transcription were purified by separating 100 μ L oligonucleotides (100 μ M) supplemented with 150 μ L 8M urea containing bromophenolblue in

15 % polyacrylamide gels (300 V, 2:30 h). DNA was detected by UV shadowing (320 nm) and full-length oligonucleotides cut from the gel. Gel pieces were further cut into smaller pieces and rotated in 600 μ L elution buffer [0.5 M NH_4Ac , 0.1% (w/v) SDS, 1 mM EDTA] at 4°C ON. Eluted DNA was purified by phenol-chloroform extraction. After addition of 0.1 vol 3 M NaAc (pH 5), 1 vol. phenol (pH 4) and 0.2 vol. chloroform/isoamyl alcohol (24:1) and centrifugation (13.000 rpm, 4°C, 5 min), DNA in the aqueous phase was precipitated with 1 mL ethanol (96%) at -20°C for 5 min. DNA was sedimented (13.000 rpm, 30 min) and air-dried. DNA was resolved in 52 μ L DMDC-ddH₂O and the concentration determined by photometry. For *in vitro*-transcription 500 pmol of the respective sequence-specific primer and the T7 primer were adjusted to a total volume of 500 μ L with DMDC-ddH₂O. Primers were denatured at 90°C for 5 min and subsequently annealed by cooling down at RT for 5 min.

4.2.6.2. Expression and purification of recombinant T7 RNA polymerase

A glycerol stock of the *E. coli* strain BL21(DE3) transformed with the T7 RNA polymerase expression plasmid pQE-80-L-T7RNAP was streaked out on LB agar plates containing ampicillin (100 μ g/mL). A single colony was transferred into 2 mL ampicillin-containing LB medium (100 μ g/mL) and incubated for 2 h at 220 rpm and 37°C. The 2 mL preparatory culture was transferred into 50 mL LB medium containing ampicillin (100 μ g/mL) and incubated ON at 220 rpm and 37°C. 10 mL of the overnight culture were transferred into 500 mL LB medium containing ampicillin (100 μ g/mL) and propagated for 2-3 h at 220 rpm and 37°C until the optical density at 600nm reached 0.6. Recombinant protein expression was induced by addition of IPTG in a final concentration of 1 mM to the bacterial culture. 1 mL aliquots were taken prior to and after IPTG supplementation to control the induction of T7 RNA polymerase expression. After another 4 h cultivation bacteria were harvested by centrifugation (4.000 g, 4°C, 4 min) and resuspended in 5 mL chilled LB medium. After centrifugation (4.000 g, 4°C, 4 min) cells were resuspended in binding buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 10 mM imidazole) (2-5 mL/g bacteria pellet). After addition of PMSF (Sigma) to a final concentration of 2 mM, 2 mg/mL lysozym (AppliChem) and 1000 U DNase I, RNase-free (Roche) bacteria were incubated for 30 min on ice and subsequently sonicated for 4 x 10 sec. The bacteria suspension was cleared from cell debris by centrifugation at 35.000 rpm and 4°C for 30 min (Ultracentrifuge; Beckmann). Recombinant T7 RNA polymerase was purified from the supernatant by affinity chromatography of the His-tagged protein using Ni-NTA agarose (Invitrogen) in a C 26/40 chromatography column (GE Healthcare). Unspecifically bound proteins were removed from the affinity column by washing with 10 vol. binding buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 10 mM imidazole)

followed by 8 vol. washing buffer (50 mM NaH₂PO₄ pH 8, 300 mM NaCl, 20 mM imidazole). T7 RNA polymerase was eluted from the affinity column with 3 vol. elution buffer (50 mM NaH₂PO₄ pH 8, 300 mM NaCl, 250 mM imidazole). Salt conditions of the eluted enzyme solution were adjusted by dialysing against dialysis buffer [20 mM NaH₂PO₄ pH 7.7, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol] for 12-22 h. The purified enzyme was stored at -20°C.

4.2.6.3. *In vitro* transcription

For *in vitro* transcription a 1 mL reaction containing 500 pmol pre-annealed oligos or 250 ng PCR products, 50 mM Tris-HCl, pH 7.5, 15 mM MgCl₂, 5 mM DTT, 5 mM NTPs (pH 8, Sigma), 2 mM spermidine and 60 µL T7 RNA polymerase (B.2.2), aliquoted into 500 µL and incubated at 37°C for 5 h. RNA was precipitated by addition of 1 mL ethanol (96%) to each aliquot and incubation at -80°C for 5 min. RNA was sedimented by centrifugation at 13.200 rpm for 7 min at 4°C. To purify full-length transcripts, the RNA pellet was resolved in 200 µL 8 M urea containing bromophenolblue and separated in 15% polyacrylamide gels (300 V, 2-3h). RNA was detected by UV shadowing and the slowest migrating bands cut from the gel. Gel fragments of both 500 µL aliquots were chopped into pieces, combined in a 15 mL falcon tube and eluted by rotating in 3 mL elution buffer [0.5 M NH₄Ac, 0.1% (w/v) SDS, 1 mM EDTA] at 4°C ON. RNA was isolated by addition of 0.1 vol 3 M NaAc (pH 5), 1 vol. phenol (pH 4) and 0.2 vol. chloroform/isoamyl alcohol (24:1). After centrifugation (4.000 rpm, 4°C, 7 min, Eppendorf 5810 R), RNA in the aqueous phase was precipitated by addition of 6 mL ethanol (96%) at -80°C for 5 min and subsequently seeded by centrifugation (4.000 rpm, 4°C, 45 min, Eppendorf 5810 R). RNA pellets were air-dried, resolved in 52-102 µL DMDC-ddH₂O depending on the pellet size and the RNA concentration photometrically determined. RNAs were stored at -80°C until RNA affinity chromatography.

4.2.6.4. Protein isolation by RNA affinity chromatography

RNA affinity chromatography was performed by modification of a published procedure (see also Asang C. thesis, 2010). 900-2000 pmol of *in vitro* transcribed RNA were chemically activated in the dark in Protein LoBind reaction tubes (Eppendorf) in a 400 µL reaction for 1 h (0.1 M NaAc, pH 5, 5 mM Na-m-JO₄), precipitated with 0.2 vol. NaAc (1 M, pH 5) and 2.5 vol. ethanol (96%) at -80°C for exactly 5 min and sedimented (13.200rpm, 4°C, 30 min). For each sample 125 µL Adipic acid dihydrazid-Agarose suspension (Sigma) were washed four times with 0.1 M NaAc (pH 5) (300 rpm, 4°C, 3 min) and after the last washing step adjusted to 1 mL with 0.1 M NaAc (pH 5). Washed Adipic acid dihydrazid-Agarose beads were added

given to the RNA precipitate and bound ON at 4°C. Unbound RNA was removed by two washing steps each with 1 mL 2 M NaCl (800 rpm, 2 min, Eppendorf micro centrifuge). Bound RNA was adjusted to the nuclear salt concentration by washing three times with 1 mL buffer D each [20 mM HEPES-KOH, pH 7.6, 5% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol (DTT)]. HeLa cell nuclear extract (Cilbiotech s.a., Belgium) was diluted with buffer D and rotated with the RNA-coupled agarose beads for 20-30 min at 30°C. Unbound proteins were removed from the reaction by washing five times with 1 mL buffer D each containing 4mM MgCl₂ (800 rpm, 2 min, Eppendorf microcentrifuge). After final washing, 12.5-60 µL 2 x protein sample buffer [0.75 M Tris-HCl, pH 6.8, 20% (v/v) glycerol, 10% (v/v) β-mercapto-ethanol, 4% (w/v) SDS] were added to the bead pellet depending on the amount of input RNA. Proteins were dissociated from the RNA by incubating at 95°C for 10 min. Agarose beads were pelleted by centrifugation, the supernatant transferred to another ProteinLoBind reaction tube (Eppendorf) and stored at -20°C until protein analyses.

4.2.6.7.1 Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis

Protein separation was performed under denaturing conditions as vertical flat bed gel electrophoresis in discontinuous 0.1% SDS-10% polyacrylamide gels (Rotiphorese Gel 30, Roth). Mini-gels were operated in 1 x SDS running buffer [0.8% (w/v) SDS, 0.2 M Tris-Base, 1.9 M glycine] for 1 h applying a current of 20 mA per gel. To monitor protein size and subsequent blotting efficiency molecular weight markers (PageRuler™ Prestained ProteinLadder Plus, Fermentas; Prestained SDS-PAGE Standard, Low Range, Biorad).

4.2.4.7.2 Immunoblotting

Proteins were transferred from SDS-polyacrylamide gels to PVDF membranes (Millipore, Immobilon-P) by electroblotting either in a tank blot system (Biorad) in transfer buffer [200 mM glycine, 25 mM Tris-Base, 20% (v/v) methanol] for 1 h using 150 mA and additional cooling or in a semi-dry system (Biometra) for 1:30 h applying 0.8 mA/cm² membrane. The membrane was blocked in TBS-T [20 mM Tris-HCl, pH7.5, 150 mM NaCl, 0.05% (v/v) Tween-20] containing 10% (w/v) dry milk for 1 h at RT or ON at 4°C. Binding of the primary antibody was performed for 1 h in TBS-T containing 5% dry milk. After washing the membrane three times for 10 min each in TBS-T, the membrane was incubated with appropriate secondary antibodies in TBS-T containing 5% dry milk for 30 min. The membrane was washed with TBS-T three times 10 min each and twice shortly with TBS. Antibody binding was visualised using the ECL system (Amersham) for peroxidase-conjugated secondary antibodies, whereas the CDP Star system (Roche) was employed to detect alkaline phosphatase-conjugated secondary antibodies. Both detection assays were

used according to the manufacturer's protocol. Chemiluminescence was measured by exposure to ECL hyperfilm (Amersham) or to the Lumi-Imager F1 operating a CCD camera (Roche). For immunoblot reprobing antibodies were removed by incubating the membrane in Antibody Stripping Solution (Chemicon) for 15 min. Membranes were washed twice 5 min each in blocking solution [10% (w/v) dry milk in TBS-T] and either reprobbed immediately or stored at 4°C.

4.2.8 Protein sequencing by mass spectrometry

4.2.8.1. In gel digestion and sample preparation

Bands containing proteins to be identified were cut from the SDS polyacrylamide gels, cutted into approximately 1 mm³ pieces and transferred into a 0.5 mL reaction tube (Protein-Low-Bind reaction tube, Eppendorf). To remove salts, which could interfere with peptide ionisation, gel pieces were agitated four times in 100 µL freshly prepared 25 mM ammonium hydrogen carbonate buffer/50% acetonitrile each, first for 10 min and then three times for 30 min at RT. Gel pieces were completely dehydrated by incubation in acetonitrile (100%) for 30 min and after removal of the acetonitrile dried in a vacuum centrifuge (DNA110 SpeedVac®, Thermo Scientific). Gel pieces were rehydrated in trypsin solution (0.1 µg/µL [Sigma] in 25 mM ammonium carbonate buffer, pH 8), excessive trypsin solution removed and overlaid with 25 mM ammonium carbonate buffer. Proteins were in gel digested for 12-16 h at 37°C.

The supernatant of the in gel digestion was collected in a second reaction tube (ProteinLoBind reaction tube, Eppendorf). Gel pieces were rocked in 2 vol. ddH₂O for 5 min. After sonication for 5 min the supernatant was removed and combined with the supernatant extracted before. Afterwards gel pieces were three times agitated in 1 vol. elution buffer (50% acetonitrile, 5%formic acid) each for 30 min at RT and all supernatants pooled with the supernatants collected before. In the final elution gel pieces were agitated with 1 vol. acetonitrile (100%) for 30 min and the supernatant was also combined with the protein supernatant eluted before. Eluted proteins were lyophilised in a vacuum centrifuge (DNA110 SpeedVac®, Thermo Scientific) and stored at -20°C until mass spectrometry analyses.

4.2.8.2. Mass spectrometry

Eluted proteins were dissolved in 5 µL 4% methanol/1% formic acid, desalted and concentrated by ZipTip_{C18} reversed-phase purification (Millipore). The C18-resin of the ZipTip_{C18} pipette tip was wetted three times with 60% methanol/1% formic acid and equilibrated three times with 4% methanol/1% formic acid. Proteins were loaded on the ZipTip resin by 10 x aspirating and dispensing the sample. Proteins bound to the C18-resin

were washed four times with a total volume of 30 μ L 4% methanol/1% formic acid and eluted in 5 μ L 60% methanol/1% formic acid. Mass spectrometry was performed by Dr. W. Bouschen using an ESIQuadrupol-TOF (QSTAR XL; Applied Biosystems) at the Analytical Core Facility of the Biological-Medical Research Centre (BMFZ, HHUD).

4.2.9. FANCD2 immunoblotting

FANCD2 immunoblotting was performed as described previously (Kalb et al., 2007, (148)) with minor modifications: immortalized fibroblasts were seeded in T75 flasks and grown to approximately 70% confluence. Fibroblasts were transfected with 16 μ g of the respective U1 snRNA expression plasmid using FuGENE 6 (Roche). 24h after transfection cells were exposed to 150nM MMC for 16h. After 16h the cells were harvested by trypsination and washed three with PBS to ensure that the Trypsin is removed. The cell pellet was resolved in a small amount of PBS and transferred into a Eppendorf tube. After removal of the supernatant and cell pellet was frozen at -20°C or -80°C.

Cells lysis was performed with lysis puffer (150mM Tris, pH 7.4, 50mM NaCl, 0.2% Triton-X 100, 0.2% Triton-X 1, 0.3% NP-40, 2mM EGTA, 2mM EDTA, 25mM glycerol-2-phosphate disodium salt pentahydrate, 40mM NaF, 0.1 mM Na₃VO₄; 1 pill Complete Protease Inhibitor Cocktail (Roche)). The proteinase inhibitors were added immediately before use, or alternatively to a small amount of lysis buffer (e.g. 20ml or 50ml) and was frozen in aliquots at -20°C after addition of inhibitors. The pelleted cells were treated with 20-60 μ l lysis buffer and incubated for 45min on ice. After lysis, cell lysates were cleared at 14.000rpm, 4°C, 10min. The supernatant was transferred in into a fresh Eppendorf tube. Any cell debris were avoided by insertion of a 50-200 μ l tip into a 1-10 μ l tip. Lysats were kepted on ice. The protein content was measured using Bradford reagent. For protein separation the NuPAGE® Tris-Acetate Systems (Invitrogen) for gel electrophoresis was used (NuPAGE® Novex 7%Tris-Acetate gels and NuPAGE® Tris-Acetate SDS Running Buffer, Invitrogen). 300ml buffer were used per chamber and 200 μ l NuPAGE® Antioxidant were added to the upper (cathode) buffer chamber to prevent reduced proteins from reoxidizing during electrophoresis (minimizes protein oxidation during electrophoresis and keeps reduced protein bands sharp and clear, better transfer efficiency by eliminating inter-protein disulfide formation). Each slot holded a maximum sample volume of 30 μ l. 50 μ g protein were loaded and samples were prepared like this: 7,5 μ l of NuPAGE® LDS Sample Preparation Buffer (pH 8.4, 4x) was provided and 3 μ l of NuPAGE® Reducing Agent was added in a Eppendorf Tube. 50 μ g lysat were added and add ddH₂O was added to 30 μ l. Samples were incubated at 95°C for 5 minutes. Samples were transferred on ice and spinned down. 15 μ l of an adequate protein

Materials and Methods

standard and 2x15 μ l of each sample were loaded in each slot. The gel was running at 4°C (cool room) for 6-8h at 120-130 Volt. For blotting NuPAGE® Transfer Buffer was used surplus 5% methanol. Blotting was performed ON at 20V, 4°C onto a Hybond-P PVDF membran. Add transfer buffer to the blotting core of the chamber only. 200 μ lNuPAGE® Antioxidant is added to the transfer buffer (upper buffer chamber) for enhanced blotting results with reduced proteins. Fill the chamber with pre-cooled water. The membrane was blocked with 5% (w/v) skim milk /0.05% Tween (PBS-T) for 1h at room Biosciences) dilution temperature (RT) and washed 3x 10min with PBS-T. The membrane was probed with the primary mouse monoclonal anti-FANCD2 antibody (Santa Cruz) at a concentration of 1:800 in 5% (w/v) skim milk /0.05% Tween (PBS-T) for 1,5-3h at RT. The membrane was washed again for 3x 10min with PBS-T. The membrane was then probed with the secondary antibody: antimouse IgG horseradish-peroxidase-linked F(ab)₂ from sheep (Amersham 1: 2000 in 5% (w/v) skim milk /0.05% Tween (PBS-T) for 1h at room temperatur. Finally, the membrane was washed 4x15min with PBS-T. For chemiluminescence detection, a standard ECL reagent (Amersham Biosciences, Little Chalfont, UK) was employed.

4.2.10 Foci assay

For indirect immunofluorescence staining of foci, we seeded cells onto coverslips (Nalgene NUNC) and incubated them the next day with 150 nM MMC (Medac) as described previously³⁵. After 24 h, cells were fixed with 3.7% paraformaldehyde (Sigma-Aldrich) for 15 min at 20–25 °C and permeabilized with 0.5% (vol/vol) Triton X-100 for 5 min. After 30 min in blocking buffer (10% (wt/vol) BSA (PAA), 0.1% (vol/vol) NP-40 (Sigma-Aldrich)), cells were incubated at 4 °C with mouse anti-FANCD2 (Santa Cruz) at 1:200 dilution for 45 min. Cells were washed three times in TBS (Invitrogen) and then incubated with a 1:500-diluted FITC–conjugated polyclonal anti-mouse. After 45 min, cells were washed three times with TBS and the slides were mounted in ProLong Gold antifade reagent (Invitrogen) with 4,6-diamidino-2-phenylindole (DAPI, Sigma-Aldrich). We viewed specimens with an inverted microscope (Axiovert 200M, Zeiss) and fluorescence imaging workstation and acquired images at 20–25 °C with a Plan-Apochromat \times 63, 1.4 numerical aperture oil immersion lens using a digital camera (AxioCam MRm, Zeiss).

4.2.11. Cell cycle analysis

Primary patient derived *FANCC* fibroblasts were transduced with either pCL1NPB-U1 wt or mutant U1 snRNA derivatives or as control, with MFCPN and murine stem cell virus (MSCV) at equivalent multiplicities of infection. G418 selected cells were cultured for 72h with or without

Materials and Methods

33nM of the interstrand cross-linker drug, mitomycin C (MMC; Sigma-Aldrich). Cells were harvested by trypsinization and washed with 1% (w/v) bovine serum albumine fraction V (BSA) in PBS (Invitrogen). The cell pellets were resuspended in PBS and fixed overnight in 98% ethanol at -20°C. After centrifugation (600 x g, 4°C), resuspended cell pellets were incubated with 100µg/ml RNase (Invitrogen) in PBS for 15min at 37°C. Cell pellets were resuspended in staining buffer containing 0.5% (w/v) BSA and 10µg/ml propidiumiodid (PI; Sigma) in PBS. DNA histograms were recorded using the flow cytometer FACSCalibur (Becton Dickinson, Heidelberg, Germany). Quantitative assessment was performed with ModFit® (Verity Software House).

REFERENCES

1. **Abelson, J.** 2008. Is the spliceosome a ribonucleoprotein enzyme? *Nat.Struct.Mol.Biol.* **15**:1235-1237. doi:nsmb1208-1235 [pii];10.1038/nsmb1208-1235 [doi].
2. **Achsel, T., K. Ahrens, H. Brahms, S. Teigelkamp, and R. Luhrmann.** 1998. The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. *Mol.Cell Biol.* **18**:6756-6766.
3. **Aiuti, A., I. Brigida, F. Ferrua, B. Cappelli, R. Chiesa, S. Markt, and M. G. Roncarolo.** 2009. Hematopoietic stem cell gene therapy for adenosine deaminase deficient-SCID. *Immunol.Res.* **44**:150-159. doi:10.1007/s12026-009-8107-8 [doi].
4. **Aiuti, A., F. Cattaneo, S. Galimberti, U. Benninghoff, B. Cassani, L. Callegaro, S. Scaramuzza, G. Andolfi, M. Mirolo, I. Brigida, A. Tabucchi, F. Carlucci, M. Eibl, M. Aker, S. Slavin, H. Al-Mousa, G. A. Al, A. Ferster, A. Duppenhaler, L. Notarangelo, U. Wintergerst, R. H. Buckley, M. Bregni, S. Markt, M. G. Valsecchi, P. Rossi, F. Ciceri, R. Miniero, C. Bordignon, and M. G. Roncarolo.** 2009. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N.Engl.J.Med.* **360**:447-458. doi:360/5/447 [pii];10.1056/NEJMoa0805817 [doi].
5. **Aiuti, A. and M. G. Roncarolo.** 2009. Ten years of gene therapy for primary immune deficiencies. *Hematology.Am.Soc.Hematol.Educ.Program.*682-689. doi:2009/1/682 [pii];10.1182/asheducation-2009.1.682 [doi].
6. **Altshuler, D., M. J. Daly, and E. S. Lander.** 2008. Genetic mapping in human disease. *Science* **322**:881-888. doi:322/5903/881 [pii];10.1126/science.1156409 [doi].
7. **Arenas, J. E. and J. N. Abelson.** 1997. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc.Natl.Acad.Sci.U.S.A* **94**:11798-11802.
8. **Aronova, A., D. Bacikova, L. B. Crotti, D. S. Horowitz, and B. Schwer.** 2007. Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. *RNA.* **13**:1437-1444. doi:rna.572807 [pii];10.1261/rna.572807 [doi].
9. **Asang, C.** 2010. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
10. **Asang, C., I. Hauber, and H. Schaal.** 2008. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic Acids Res.* **36**:1450-1463. doi:gkm1147 [pii];10.1093/nar/gkm1147 [doi].
11. **Auweter, S. D., R. Fasan, L. Reymond, J. G. Underwood, D. L. Black, S. Pitsch, and F. H. Allain.** 2006. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**:163-173. doi:7600918 [pii];10.1038/sj.emboj.7600918 [doi].

Referenzen

12. **Auweter, S. D., F. C. Oberstrass, and F. H. Allain.** 2006. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* **34**:4943-4959. doi:gkl620 [pii];10.1093/nar/gkl620 [doi].
13. **Auweter, S. D., F. C. Oberstrass, and F. H. Allain.** 2007. Solving the structure of PTB in complex with pyrimidine tracts: an NMR study of protein-RNA complexes of weak affinities. *J.Mol.Biol.* **367**:174-186. doi:S0022-2836(06)01744-X [pii];10.1016/j.jmb.2006.12.053 [doi].
14. **Bacikova, D. and D. S. Horowitz.** 2005. Genetic and functional interaction of evolutionarily conserved regions of the Prp18 protein and the U5 snRNA. *Mol.Cell Biol.* **25**:2107-2116. doi:25/6/2107 [pii];10.1128/MCB.25.6.2107-2116.2005 [doi].
15. **Badano, J. L. and N. Katsanis.** 2002. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat.Rev.Genet.* **3**:779-789. doi:10.1038/nrg910 [doi];nrg910 [pii].
16. **Baltimore, D.** 2001. Our genome unveiled. *Nature* **409**:814-816. doi:10.1038/35057267 [doi].
17. **Baralle, D. and M. Baralle.** 2005. Splicing in action: assessing disease causing sequence changes. *J.Med.Genet.* **42**:737-748. doi:42/10/737 [pii];10.1136/jmg.2004.029538 [doi].
18. **Berget, S. M.** 1995. Exon recognition in vertebrate splicing. *J.Biol.Chem.* **270**:2411-2414.
19. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc.Natl.Acad.Sci.U.S.A* **74**:3171-3175.
20. **Berget, S. M. and B. L. Robberson.** 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell* **46**:691-696. doi:0092-8674(86)90344-2 [pii].
21. **Berget, S. M. and P. A. Sharp.** 1977. A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Brookhaven.Symp.Biol.*332-344.
22. **Berglund, J. A., K. Chua, N. Abovich, R. Reed, and M. Rosbash.** 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**:781-787. doi:S0092-8674(00)80261-5 [pii].
23. **Berglund, J. A., M. L. Fleming, and M. Rosbash.** 1998. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA.* **4**:998-1006.
24. **Berglund, J. A., M. Rosbash, and S. C. Schultz.** 2001. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA.* **7**:682-691.
25. **Berglund, J. A., M. Rosbash, and S. C. Schultz.** 2001. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA.* **7**:682-691.
26. **Bi, J., H. Xia, F. Li, X. Zhang, and Y. Li.** 2005. The effect of U1 snRNA binding free energy on the selection of 5' splice sites. *Biochem.Biophys.Res.Commun.* **333**:64-69. doi:S0006-291X(05)01068-5 [pii];10.1016/j.bbrc.2005.05.078 [doi].

Referenzen

27. **Birney, E., S. Kumar, and A. R. Krainer.** 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**:5803-5816.
28. **Bishop, J. M.** 1981. Enemies within: the genesis of retrovirus oncogenes. *Cell* **23**:5-6. doi:0092-8674(81)90263-4 [pii].
29. **Black, D. L., B. Chabot, and J. A. Steitz.** 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* **42**:737-750. doi:0092-8674(85)90270-3 [pii].
30. **Black, D. L. and J. A. Steitz.** 1986. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell* **46**:697-704. doi:0092-8674(86)90345-4 [pii].
31. **Blanchette, M. and B. Chabot.** 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**:1939-1952. doi:10.1093/emboj/18.7.1939 [doi].
32. **Bommarito, S., N. Peyret, and J. SantaLucia, Jr.** 2000. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**:1929-1934. doi:gkd309 [pii].
33. **Bonnal, S., C. Martinez, P. Forch, A. Bachi, M. Wilm, and J. Valcarcel.** 2008. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol.Cell* **32**:81-95. doi:S1097-2765(08)00544-3 [pii];10.1016/j.molcel.2008.08.008 [doi].
34. **Bonnet, C., S. Krieger, M. Vezain, A. Rousselin, I. Tournier, A. Martins, P. Berthet, A. Chevrier, C. Dugast, V. Layet, A. Rossi, R. Lidereau, T. Frebourg, A. Hardouin, and M. Tosi.** 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J.Med.Genet.* **45**:438-446. doi:jmg.2007.056895 [pii];10.1136/jmg.2007.056895 [doi].
35. **Borensztajn, K., M. L. Sobrier, P. Duquesnoy, A. M. Fischer, J. Tapon-Breaudiere, and S. Amselem.** 2006. Oriented scanning is the leading mechanism underlying 5' splice site selection in mammals. *PLoS.Genet.* **2**:e138. doi:05-PLGE-RA-0315R2 [pii];10.1371/journal.pgen.0020138 [doi].
36. **Braddock, D. T., J. M. Louis, J. L. Baber, D. Levens, and G. M. Clore.** 2002. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415**:1051-1056. doi:10.1038/4151051a [doi];4151051a [pii].
37. **Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon.** 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc.Natl.Acad.Sci.U.S.A* **75**:4853-4857.
38. **Bringmann, P. and R. Luhrmann.** 1986. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**:3509-3516.
39. **Brody, E. and J. Abelson.** 1985. The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**:963-967.

Referenzen

40. **Brunak, S., J. Engelbrecht, and S. Knudsen.** 1990. Neural network detects errors in the assignment of mRNA splice sites. *Nucleic Acids Res.* **18**:4797-4801.
41. **Brunak, S., J. Engelbrecht, and S. Knudsen.** 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J.Mol.Biol.* **220**:49-65. doi:0022-2836(91)90380-O [pii].
42. **Burd, C. G. and G. Dreyfuss.** 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**:1197-1204.
43. **Burge, C. B., T. Tuschl, Sharp, and P.A.** 1999. Splicing of Precursors to mRNAs by the Spliceosomes, *In: The RNA World.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
44. **Caputi, M., M. Freund, S. Kammler, C. Asang, and H. Schaal.** 2004. A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *J.Virol.* **78**:6517-6526. doi:10.1128/JVI.78.12.6517-6526.2004 [doi];78/12/6517 [pii].
45. **Caputi, M. and A. M. Zahler.** 2001. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J.Biol.Chem.* **276**:43850-43859. doi:10.1074/jbc.M102861200 [doi];M102861200 [pii].
46. **Caputi, M. and A. M. Zahler.** 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J.* **21**:845-855. doi:10.1093/emboj/21.4.845 [doi].
47. **Carlo, T., R. Sierra, and S. M. Berget.** 2000. A 5' splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon. *Mol.Cell Biol.* **20**:3988-3995.
48. **Carlo, T., D. A. Sterner, and S. M. Berget.** 1996. An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA.* **2**:342-353.
49. **Carmel, I., S. Tal, I. Vig, and G. Ast.** 2004. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA.* **10**:828-840.
50. **Cartegni, L., J. Wang, Z. Zhu, M. Q. Zhang, and A. R. Krainer.** 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**:3568-3571.
51. **Chabot, B., D. L. Black, D. M. LeMaster, and J. A. Steitz.** 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science* **230**:1344-1349.
52. **Chandler, S. D., A. Mayeda, J. M. Yeakley, A. R. Krainer, and X. D. Fu.** 1997. RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc.Natl.Acad.Sci.U.S.A* **94**:3596-3601.
53. **Chandra, S., O. Levran, I. Jurickova, C. Maas, R. Kapur, D. Schindler, R. Henry, K. Milton, S. D. Batish, J. A. Cancelas, H. Hanenberg, A. D. Auerbach, and D. A. Williams.** 2005. A rapid method for retrovirus-mediated identification of complementation groups in Fanconi anemia patients. *Mol.Ther.* **12**:976-984. doi:S1525-0016(05)00199-1 [pii];10.1016/j.ymthe.2005.04.021 [doi].

Referenzen

54. **Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier.** 1994. Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3' splice site selection in *S. cerevisiae*. *Nucleic Acids Res.* **22**:1981-1987.
55. **Chen, C. D., R. Kobayashi, and D. M. Helfman.** 1999. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.* **13**:593-606.
56. **Chen, J. Y., L. Stands, J. P. Staley, R. R. Jackups, Jr., L. J. Latus, and T. H. Chang.** 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol.Cell* **7**:227-232. doi:S1097-2765(01)00170-8 [pii].
57. **Cho, S., A. Hoang, S. Chakrabarti, N. Huynh, D. B. Huang, and G. Ghosh.** 2011. The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRM. *Nucleic Acids Res.* **39**:9413-9421. doi:gkr663 [pii];10.1093/nar/gkr663 [doi].
58. **Cho, S., A. Hoang, R. Sinha, X. Y. Zhong, X. D. Fu, A. R. Krainer, and G. Ghosh.** 2011. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc.Natl.Acad.Sci.U.S.A* **108**:8233-8238. doi:1017700108 [pii];10.1073/pnas.1017700108 [doi].
59. **Chua, K. and R. Reed.** 1999. Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes Dev.* **13**:841-850.
60. **Chua, K. and R. Reed.** 1999. The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* **402**:207-210. doi:10.1038/46086 [doi].
61. **Chua, K. and R. Reed.** 2001. An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol.Cell Biol.* **21**:1509-1514. doi:10.1128/MCB.21.5.1509-1514.2001 [doi].
62. **Collins, C. A. and C. Guthrie.** 2000. The question remains: is the spliceosome a ribozyme? *Nat.Struct.Biol.* **7**:850-854. doi:10.1038/79598 [doi].
63. **Coolidge, C. J., R. J. Seely, and J. G. Patton.** 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**:888-896. doi:gka144 [pii].
64. **Cooper, T. A.** 2005. Use of minigene systems to dissect alternative splicing elements. *Methods* **37**:331-340. doi:S1046-2023(05)00173-8 [pii];10.1016/j.ymeth.2005.07.015 [doi].
65. **Crawford, J. B. and J. G. Patton.** 2006. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol.Cell Biol.* **26**:8791-8802. doi:MCB.01677-06 [pii];10.1128/MCB.01677-06 [doi].
66. **Datar, K. V., G. Dreyfuss, and M. S. Swanson.** 1993. The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. *Nucleic Acids Res.* **21**:439-446.

Referenzen

67. **Dauksaite, V. and G. Akusjarvi.** 2002. Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. *J.Biol.Chem.* **277**:12579-12586. doi:10.1074/jbc.M107867200 [doi];M107867200 [pii].
68. **Degroeve, S., B. B. De, Y. Van de Peer, and P. Rouze.** 2002. Feature subset selection for splice site prediction. *Bioinformatics.* **18 Suppl 2**:S75-S83.
69. **Deimel, B., C. H. Louis, and C. E. Sekeris.** 1977. The presence of small molecular weight RNAs in nuclear ribonucleoprotein particles carrying HnRNA. *FEBS Lett.* **73**:80-84.
70. **Deirdre, A., J. Scadden, and C. W. Smith.** 1995. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14**:3236-3246.
71. **Del Gatto-Konczak, F., C. F. Bourgeois, G. C. Le, L. Kister, M. C. Gesnel, J. Stevenin, and R. Breathnach.** 2000. The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Mol.Cell Biol.* **20**:6287-6299.
72. **Dietrich, R. C., R. Incorvaia, and R. A. Padgett.** 1997. Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol.Cell Biol.* **17**:151-160. doi:S1097-2765(00)80016-7 [pii].
73. **Domdey, H., B. Apostol, R. J. Lin, A. Newman, E. Brody, and J. Abelson.** 1984. Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* **39**:611-621. doi:0092-8674(84)90468-9 [pii].
74. **Dracopoli, N. C. and J. Fogh.** 1983. Loss of heterozygosity in cultured human tumor cell lines. *J.Natl.Cancer Inst.* **70**:83-87.
75. **Du, H. and M. Rosbash.** 2002. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**:86-90. doi:10.1038/nature00947 [doi];nature00947 [pii].
76. **Eddy, S. R.** 1999. Noncoding RNA genes. *Curr.Opin.Genet.Dev.* **9**:695-699. doi:S0959-437X(99)00022-2 [pii].
77. **Eng, L., G. Coutinho, S. Nahas, G. Yeo, R. Tanouye, M. Babaei, T. Dork, C. Burge, and R. A. Gatti.** 2004. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum.Mutat.* **23**:67-76. doi:10.1002/humu.10295 [doi].
78. **Eperon, I. C., O. V. Makarova, A. Mayeda, S. H. Munroe, J. F. Caceres, D. G. Hayward, and A. R. Krainer.** 2000. Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol.Cell Biol.* **20**:8303-8318.
79. **Erkelenz, S.** 2012. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
80. **Fairbrother, W. G., R. F. Yeh, P. A. Sharp, and C. B. Burge.** 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**:1007-1013. doi:10.1126/science.1073774 [doi];1073774 [pii].

Referenzen

81. **Fairbrother, W. G., G. W. Yeo, R. Yeh, P. Goldstein, M. Mawson, P. A. Sharp, and C. B. Burge.** 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res.* **32**:W187-W190. doi:10.1093/nar/gkh393 [doi];32/suppl_2/W187 [pii].
82. **Fearon, E. R.** 1991. A genetic basis for the multi-step pathway of colorectal tumorigenesis. *Princess Takamatsu Symp.* **22**:37-48.
83. **Feuk, L., C. R. Marshall, R. F. Wintle, and S. W. Scherer.** 2006. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum.Mol.Genet.* **15 Spec No 1**:R57-R66. doi:15/suppl_1/R57 [pii];10.1093/hmg/ddl057 [doi].
84. **Fischer, U. and R. Luhrmann.** 1990. An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science* **249**:786-790.
85. **Fisette, J. F., J. Toutant, S. Dugre-Brisson, L. Desgroseillers, and B. Chabot.** 2010. hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection. *RNA.* **16**:228-238. doi:rna.1890310 [pii];10.1261/rna.1890310 [doi].
86. **Fleckner, J., M. Zhang, J. Valcarcel, and M. R. Green.** 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**:1864-1872.
87. **Forch, P., O. Puig, C. Martinez, B. Seraphin, and J. Valcarcel.** 2002. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J.* **21**:6882-6892.
88. **Fortner, D. M., R. G. Troy, and D. A. Brow.** 1994. A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev.* **8**:221-233.
89. **Fox-Walsh, K. L., Y. Dou, B. J. Lam, S. P. Hung, P. F. Baldi, and K. J. Hertel.** 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc.Natl.Acad.Sci.U.S.A* **102**:16176-16181. doi:0508489102 [pii];10.1073/pnas.0508489102 [doi].
90. **Freund, M.** 2004. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
91. **Freund, M., C. Asang, S. Kammler, C. Konermann, J. Krummheuer, M. Hipp, I. Meyer, W. Gierling, S. Theiss, T. Preuss, D. Schindler, J. Kjems, and H. Schaal.** 2003. A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.* **31**:6963-6975.
92. **Freund, M., M. J. Hicks, C. Konermann, M. Otte, K. J. Hertel, and H. Schaal.** 2005. Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res.* **33**:5112-5119. doi:33/16/5112 [pii];10.1093/nar/gki824 [doi].
93. **Fu, X. D., A. Mayeda, T. Maniatis, and A. R. Krainer.** 1992. General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. *Proc.Natl.Acad.Sci.U.S.A* **89**:11224-11228.

Referenzen

94. **Futaki, M., T. Yamashita, H. Yagasaki, T. Toda, M. Yabe, S. Kato, S. Asano, and T. Nakahata.** 2000. The IVS4 + 4 A to T mutation of the fanconi anemia gene FANCC is not associated with a severe phenotype in Japanese patients. *Blood* **95**:1493-1498.
95. **Gallinaro, H. and M. Jacob.** 1979. An evaluation of small nuclear RNA in hnRNP. *FEBS Lett.* **104**:176-182. doi:0014-5793(79)81110-2 [pii].
96. **Gao, K., A. Masuda, T. Matsuura, and K. Ohno.** 2008. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* **36**:2257-2267. doi:gkn073 [pii];10.1093/nar/gkn073 [doi].
97. **Garcia-Higuera, I., T. Taniguchi, S. Ganesan, M. S. Meyn, C. Timmers, J. Hejna, M. Grompe, and A. D. D'Andrea.** 2001. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol.Cell* **7**:249-262. doi:S1097-2765(01)00173-3 [pii].
98. **Gaur, R. K., J. Valcarcel, and M. R. Green.** 1995. Sequential recognition of the pre-mRNA branch point by U2AF65 and a novel spliceosome-associated 28-kDa protein. *RNA.* **1**:407-417.
99. **Gilbert, S. F.** 1978. The embryological origins of the gene theory. *J.Hist Biol.* **11**:307-351.
100. **Gillio, A. P., P. C. Verlander, S. D. Batish, P. F. Giampietro, and A. D. Auerbach.** 1997. Phenotypic consequences of mutations in the Fanconi anemia FAC gene: an International Fanconi Anemia Registry study. *Blood* **90**:105-110.
101. **Golas, M. M., B. Sander, C. L. Will, R. Luhrmann, and H. Stark.** 2003. Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**:980-984. doi:10.1126/science.1084155 [doi];300/5621/980 [pii].
102. **Gooding, C., F. Clark, M. C. Wollerton, S. N. Grellscheid, H. Groom, and C. W. Smith.** 2006. A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol.* **7**:R1. doi:gb-2006-7-1-r1 [pii];10.1186/gb-2006-7-1-r1 [doi].
103. **Goodman, H. M., M. V. Olson, and B. D. Hall.** 1977. Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. *Proc.Natl.Acad.Sci.U.S.A* **74**:5453-5457.
104. **Goren, A., O. Ram, M. Amit, H. Keren, G. Lev-Maor, I. Vig, T. Pupko, and G. Ast.** 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol.Cell* **22**:769-781. doi:S1097-2765(06)00300-5 [pii];10.1016/j.molcel.2006.05.008 [doi].
105. **Gozani, O., R. Feld, and R. Reed.** 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10**:233-243.
106. **Gozani, O., J. Potashkin, and R. Reed.** 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol.Cell Biol.* **18**:4752-4760.
107. **Grabowski, P. J., S. R. Seiler, and P. A. Sharp.** 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**:345-353. doi:S0092-8674(85)80130-6 [pii].

Referenzen

108. **Grainger, R. J. and J. D. Beggs.** 2005. Prp8 protein: at the heart of the spliceosome. *RNA*. **11**:533-557. doi:11/5/533 [pii];10.1261/rna.2220705 [doi].
109. **Graveley, B. R.** 2000. Sorting out the complexity of SR protein functions. *RNA*. **6**:1197-1211.
110. **Grossloh and F.J.** 2006. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
111. **Guimont-Ducamp, C., J. Sri-Widada, and P. Jeanteur.** 1977. Occurrence of small molecular weight RNAs in Hela nuclear ribonucleoprotein particles containing HnRNA. *Biochimie* **59**:755-758.
112. **Guo, Z., K. S. Karunatilaka, and D. Rueda.** 2009. Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat.Struct.Mol.Biol.* **16**:1154-1159. doi:nsmb.1672 [pii];10.1038/nsmb.1672 [doi].
113. **Guth, S., T. O. Tange, E. Kellenberger, and J. Valcarcel.** 2001. Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol.Cell Biol.* **21**:7673-7681. doi:10.1128/MCB.21.22.7673-7681.2001 [doi].
114. **Guth, S. and J. Valcarcel.** 2000. Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *J.Biol.Chem.* **275**:38059-38066. doi:10.1074/jbc.M001483200 [doi];M001483200 [pii].
115. **Hall, S. L. and R. A. Padgett.** 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J.Mol.Biol.* **239**:357-365. doi:S0022-2836(84)71377-5 [pii];10.1006/jmbi.1994.1377 [doi].
116. **Hall, S. L. and R. A. Padgett.** 1996. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**:1716-1718.
117. **Hallay, H., N. Locker, L. Ayadi, D. Ropers, E. Guittet, and C. Branlant.** 2006. Biochemical and NMR study on the competition between proteins SC35, SRp40, and heterogeneous nuclear ribonucleoprotein A1 at the HIV-1 Tat exon 2 splicing site. *J.Biol.Chem.* **281**:37159-37174. doi:M603864200 [pii];10.1074/jbc.M603864200 [doi].
118. **Hamm, J., M. Kazmaier, and I. W. Mattaj.** 1987. In vitro assembly of U1 snRNPs. *EMBO J.* **6**:3479-3485.
119. **Hanenberg, H., S. D. Batish, K. E. Pollok, L. Vieten, P. C. Verlander, C. Leurs, R. J. Cooper, K. Gottsche, L. Haneline, D. W. Clapp, S. Lobitz, D. A. Williams, and A. D. Auerbach.** 2002. Phenotypic correction of primary Fanconi anemia T cells with retroviral vectors as a diagnostic tool. *Exp.Hematol.* **30**:410-420. doi:S0301472X02007828 [pii].
120. **Hardy, S. F., P. J. Grabowski, R. A. Padgett, and P. A. Sharp.** 1984. Cofactor requirements of splicing of purified messenger RNA precursors. *Nature* **308**:375-377.
121. **Hartmann, L.** 2005. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
122. **Hartmann, L., K. Neveling, S. Borkens, H. Schneider, M. Freund, E. Grassman, S. Theiss, A. Wawer, S. Burdach, A. D. Auerbach, D. Schindler, H. Hanenberg, and H. Schaal.** 2010. Correct mRNA processing at a mutant TT splice donor in FANCC ameliorates the clinical phenotype in patients and is enhanced by delivery of

Referenzen

- suppressor U1 snRNAs. *Am.J.Hum.Genet.* **87**:480-493. doi:S0002-9297(10)00465-9 [pii];10.1016/j.ajhg.2010.08.016 [doi].
123. **Hartmann, L., S. Theiss, D. Niederacher, and H. Schaal.** 2008. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci.* **13**:3252-3272. doi:2924 [pii].
 124. **Hastings, M. L. and A. R. Krainer.** 2001. Pre-mRNA splicing in the new millennium. *Curr.Opin.Cell Biol.* **13**:302-309. doi:S0955-0674(00)00212-X [pii].
 125. **Heinrichs, V., M. Bach, G. Winkelmann, and R. Luhrmann.** 1990. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247**:69-72.
 126. **Helfman, D. M. and W. M. Ricci.** 1989. Branch point selection in alternative splicing of tropomyosin pre-mRNAs. *Nucleic Acids Res.* **17**:5633-5650.
 127. **Hertel, K. J.** 2008. Combinatorial control of exon recognition. *J.Biol.Chem.* **283**:1211-1215. doi:R700035200 [pii];10.1074/jbc.R700035200 [doi].
 128. **Hiller, M., K. Huse, K. Szafranski, N. Jahn, J. Hampe, S. Schreiber, R. Backofen, and M. Platzer.** 2006. Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *Am.J.Hum.Genet.* **78**:291-302. doi:S0002-9297(07)62360-X [pii];10.1086/500151 [doi].
 129. **Hiller, M., K. Szafranski, R. Backofen, and M. Platzer.** 2006. Alternative splicing at NAGNAG acceptors: simply noise or noise and more? *PLoS.Genet.* **2**:e207. doi:06-PLGE-C-0307R2 [pii];10.1371/journal.pgen.0020207 [doi].
 130. **Hilliker, A. K., M. A. Mefford, and J. P. Staley.** 2007. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev.* **21**:821-834. doi:21/7/821 [pii];10.1101/gad.1536107 [doi].
 131. **Hodas, N. O. and D. P. Aalberts.** 2004. Efficient computation of optimal oligo-RNA binding. *Nucleic Acids Res.* **32**:6636-6642. doi:32/22/6636 [pii];10.1093/nar/gkh1008 [doi].
 132. **Hoeijmakers, J. H.** 2001. Genome maintenance mechanisms for preventing cancer. *Nature* **411**:366-374. doi:10.1038/35077232 [doi];35077232 [pii].
 133. **Hoffman, B. E. and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.* **6**:2554-2568.
 134. **Hopper, J. E. and L. B. Rowe.** 1978. Molecular expression and regulation of the galactose pathway genes in *Saccharomyces cerevisiae*. Distinct messenger RNAs specified by the *Gali* and *Gal7* genes in the *Gal7-Gal10-Gal1* cluster. *J.Biol.Chem.* **253**:7566-7569.
 135. **House, A. E. and K. W. Lynch.** 2006. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat.Struct.Mol.Biol.* **13**:937-944. doi:nsmb1149 [pii];10.1038/nsmb1149 [doi].
 136. **Hovhannisyan, R. H. and R. P. Carstens.** 2007. Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively

Referenzen

- spliced exons. *J.Biol.Chem.* **282**:36265-36274. doi:M704188200 [pii];10.1074/jbc.M704188200 [doi].
137. **Howlett, N. G., T. Taniguchi, S. Olson, B. Cox, Q. Waisfisz, C. De Die-Smulders, N. Persky, M. Grompe, H. Joenje, G. Pals, H. Ikeda, E. A. Fox, and A. D. D'Andrea.** 2002. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**:606-609. doi:10.1126/science.1073834 [doi];1073834 [pii].
 138. **Hua, Y., T. A. Vickers, H. L. Okunola, C. F. Bennett, and A. R. Krainer.** 2008. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am.J.Hum.Genet.* **82**:834-848. doi:S0002-9297(08)00163-8 [pii];10.1016/j.ajhg.2008.01.014 [doi].
 139. **Huebner, R. J. and G. J. Todaro.** 1969. Oncogenes of RNA tumor viruses as determinants of cancer. *Proc.Natl.Acad.Sci.U.S.A* **64**:1087-1094.
 140. **Huppler, A., L. J. Nikstad, A. M. Allmann, D. A. Brow, and S. E. Butcher.** 2002. Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat.Struct.Biol.* **9**:431-435. doi:10.1038/nsb800 [doi];nsb800 [pii].
 141. **Ibrahim, E. C., T. D. Schaal, K. J. Hertel, R. Reed, and T. Maniatis.** 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc.Natl.Acad.Sci.U.S.A* **102**:5002-5007. doi:0500543102 [pii];10.1073/pnas.0500543102 [doi].
 142. **Ismaili, N., M. Sha, E. H. Gustafson, and M. M. Konarska.** 2001. The 100-kda U5 snRNP protein (hPrp28p) contacts the 5' splice site through its ATPase site. *RNA.* **7**:182-193.
 143. **Izquierdo, J. M., N. Majos, S. Bonnal, C. Martinez, R. Castelo, R. Guigo, D. Bilbao, and J. Valcarcel.** 2005. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol.Cell* **19**:475-484. doi:S1097-2765(05)01418-8 [pii];10.1016/j.molcel.2005.06.015 [doi].
 144. **Jackson, I. J.** 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**:3795-3798.
 145. **Jeffreys, A. J. and R. A. Flavell.** 1977. The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell* **12**:1097-1108. doi:0092-8674(77)90172-6 [pii].
 146. **Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton, and D. D. Shoemaker.** 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**:2141-2144. doi:10.1126/science.1090100 [doi];302/5653/2141 [pii].
 147. **Jokan, L., A. P. Dong, A. Mayeda, A. R. Krainer, and R. M. Xu.** 1997. Crystallization and preliminary X-ray diffraction studies of UP1, the two-RRM domain of hnRNP A1. *Acta Crystallogr.D.Biol.Crystallogr.* **53**:615-618. doi:10.1107/S09074444997003326 [doi];S09074444997003326 [pii].
 148. **Kalb, R., K. Neveling, H. Hoehn, H. Schneider, Y. Linka, S. D. Batish, C. Hunt, M. Berwick, E. Callen, J. Surralles, J. A. Casado, J. Bueren, A. Dasi, J. Soulier, E. Gluckman, C. M. Zwaan, S. R. van, G. Pals, J. P. de Winter, H. Joenje, M. Grompe, A. D. Auerbach, H. Hanenberg, and D. Schindler.** 2007. Hypomorphic

Referenzen

- mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am.J.Hum.Genet.* **80**:895-910. doi:S0002-9297(07)60945-8 [pii];10.1086/517616 [doi].
149. **Kambach, C., S. Walke, and K. Nagai.** 1999. Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr.Opin.Struct.Biol.* **9**:222-230. doi:sb9212 [pii].
150. **Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li, and K. Nagai.** 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**:375-387. doi:S0092-8674(00)80550-4 [pii].
151. **Kamma, H., D. S. Portman, and G. Dreyfuss.** 1995. Cell type-specific expression of hnRNP proteins. *Exp.Cell Res.* **221**:187-196. doi:S0014-4827(85)71366-3 [pii];10.1006/excr.1995.1366 [doi].
152. **Kammler, S., C. Leurs, M. Freund, J. Krummheuer, K. Seidel, T. O. Tange, M. K. Lund, J. Kjems, A. Scheid, and H. Schaal.** 2001. The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable env pre-mRNA. *RNA.* **7**:421-434.
153. **Kammler, S., M. Otte, I. Hauber, J. Kjems, J. Hauber, and H. Schaal.** 2006. The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology.* **3**:89. doi:1742-4690-3-89 [pii];10.1186/1742-4690-3-89 [doi].
154. **Kandels-Lewis, S. and B. Seraphin.** 1993. Involvement of U6 snRNA in 5' splice site selection. *Science* **262**:2035-2039.
155. **Kanopka, A., O. Muhlemann, and G. Akusjarvi.** 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**:535-538. doi:10.1038/381535a0 [doi].
156. **Kanopka, A., O. Muhlemann, S. Petersen-Mahrt, C. Estmer, C. Ohrmalm, and G. Akusjarvi.** 1998. Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* **393**:185-187. doi:10.1038/30277 [doi].
157. **Kashima, T., N. Rao, C. J. David, and J. L. Manley.** 2007. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum.Mol.Genet.* **16**:3149-3159. doi:ddm276 [pii];10.1093/hmg/ddm276 [doi].
158. **Kashima, T., N. Rao, C. J. David, and J. L. Manley.** 2007. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum.Mol.Genet.* **16**:3149-3159. doi:ddm276 [pii];10.1093/hmg/ddm276 [doi].
159. **Kashima, T., N. Rao, C. J. David, and J. L. Manley.** 2007. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum.Mol.Genet.* **16**:3149-3159. doi:ddm276 [pii];10.1093/hmg/ddm276 [doi].
160. **Ke, S. and L. A. Chasin.** 2010. Intronic motif pairs cooperate across exons to promote pre-mRNA splicing. *Genome Biol.* **11**:R84. doi:gb-2010-11-8-r84 [pii];10.1186/gb-2010-11-8-r84 [doi].

Referenzen

161. **Ke, S. and L. A. Chasin.** 2011. Context-dependent splicing regulation: exon definition, co-occurring motif pairs and tissue specificity. *RNA.Biol.* **8**:384-388. doi:14458 [pii].
162. **Kellenberger, E., G. Stier, and M. Sattler.** 2002. Induced folding of the U2AF35 RRM upon binding to U2AF65. *FEBS Lett.* **528**:171-176. doi:S0014579302032945 [pii].
163. **Kent, O. A. and A. M. MacMillan.** 2002. Early organization of pre-mRNA during spliceosome assembly. *Nat.Struct.Biol.* **9**:576-581. doi:10.1038/nsb822 [doi];nsb822 [pii].
164. **Kent, O. A., A. Reayi, L. Foong, K. A. Chilibeck, and A. M. MacMillan.** 2003. Structuring of the 3' splice site by U2AF65. *J.Biol.Chem.* **278**:50572-50577. doi:10.1074/jbc.M307976200 [doi];M307976200 [pii].
165. **Kinzler, K. W. and B. Vogelstein.** 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**:761, 763. doi:10.1038/386761a0 [doi].
166. **Kirschner, L. S., J. A. Carney, S. D. Pack, S. E. Taymans, C. Giatzakis, Y. S. Cho, Y. S. Cho-Chung, and C. A. Stratakis.** 2000. Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. *Nat.Genet.* **26**:89-92. doi:10.1038/79238 [doi].
167. **Knudson, A. G.** 1996. Hereditary cancer: two hits revisited. *J.Cancer Res.Clin.Oncol.* **122**:135-140.
168. **Kohn, D. B. and F. Candotti.** 2009. Gene therapy fulfilling its promise. *N.Engl.J.Med.* **360**:518-521. doi:360/5/518 [pii];10.1056/NEJMe0809614 [doi].
169. **Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley.** 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**:119-124. doi:10.1038/368119a0 [doi].
170. **Kol, G., G. Lev-Maor, and G. Ast.** 2005. Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum.Mol.Genet.* **14**:1559-1568. doi:ddi164 [pii];10.1093/hmg/ddi164 [doi].
171. **Kosowski, T. R., H. R. Keys, T. K. Quan, and S. W. Ruby.** 2009. DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *RNA.* **15**:1345-1362. doi:rna.1065209 [pii];10.1261/rna.1065209 [doi].
172. **Koufos, A., M. F. Hansen, N. G. Copeland, N. A. Jenkins, B. C. Lampkin, and W. K. Cavenee.** 1985. Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. *Nature* **316**:330-334.
173. **Krainer, A. R. and T. Maniatis.** 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42**:725-736. doi:0092-8674(85)90269-7 [pii].
174. **Kralovicova, J., M. B. Christensen, and I. Vorechovsky.** 2005. Biased exon/intron distribution of cryptic and de novo 3' splice sites. *Nucleic Acids Res.* **33**:4882-4898. doi:33/15/4882 [pii];10.1093/nar/gki811 [doi].

Referenzen

175. **Kramer, A.** 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu.Rev.Biochem.* **65**:367-409. doi:10.1146/annurev.bi.65.070196.002055 [doi].
176. **Krawczak, M., N. S. Thomas, B. Hundrieser, M. Mort, M. Wittig, J. Hampe, and D. N. Cooper.** 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum.Mutat.* **28**:150-158. doi:10.1002/humu.20400 [doi].
177. **Krummel, D. A., K. Nagai, and C. Oubridge.** 2010. Structure of spliceosomal ribonucleoproteins. *F1000.Biol.Rep.* **2**. doi:10.3410/B2-39 [doi].
178. **Kübart, S.** 2010. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
179. **Kutler, D. I., B. Singh, J. Satagopan, S. D. Batish, M. Berwick, P. F. Giampietro, H. Hanenberg, and A. D. Auerbach.** 2003. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood* **101**:1249-1256. doi:10.1182/blood-2002-07-2170 [doi];2002-07-2170 [pii].
180. **Kyriakopoulou, C., P. Larsson, L. Liu, J. Schuster, F. Soderbom, L. A. Kirsebom, and A. Virtanen.** 2006. U1-like snRNAs lacking complementarity to canonical 5' splice sites. *RNA.* **12**:1603-1611. doi:rna.26506 [pii];10.1261/rna.26506 [doi].
181. **Laggerbauer, B., T. Achsel, and R. Luhrmann.** 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc.Natl.Acad.Sci.U.S.A* **95**:4188-4192.
182. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe,**

Referenzen

- F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, J. P. de, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, and Y. J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921. doi:10.1038/35057062 [doi].
183. **Langford, C. J. and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* **33**:519-527. doi:0092-8674(83)90433-6 [pii].
184. **Lear, A. L., L. P. Eperon, I. M. Wheatley, and I. C. Eperon.** 1990. Hierarchy for 5' splice site preference determined in vivo. *J.Mol.Biol.* **211**:103-115. doi:0022-2836(90)90014-D [pii];10.1016/0022-2836(90)90014-D [doi].
185. **Lee, C. G., P. D. Zamore, M. R. Green, and J. Hurwitz.** 1993. RNA annealing activity is intrinsically associated with U2AF. *J.Biol.Chem.* **268**:13472-13478.
186. **Lesser, C. F. and C. Guthrie.** 1993. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**:1982-1988.
187. **Lev-Maor, G., R. Sorek, N. Shomron, and G. Ast.** 2003. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* **300**:1288-1291. doi:10.1126/science.1082588 [doi];300/5623/1288 [pii].
188. **Levitt, N. C. and I. D. Hickson.** 2002. Caretaker tumour suppressor genes that defend genome integrity. *Trends Mol.Med.* **8**:179-186. doi:S1471491402022980 [pii].
189. **Lewis, H. A., H. Chen, C. Edo, R. J. Buckanovich, Y. Y. Yang, K. Musunuru, R. Zhong, R. B. Darnell, and S. K. Burley.** 1999. Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure.* **7**:191-203.
190. **Lim, S. R. and K. J. Hertel.** 2004. Commitment to splice site pairing coincides with A complex formation. *Mol.Cell* **15**:477-483. doi:10.1016/j.molcel.2004.06.025 [doi];S1097276504003739 [pii].
191. **Liu, H. X., S. L. Chew, L. Cartegni, M. Q. Zhang, and A. R. Krainer.** 2000. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol.Cell Biol.* **20**:1063-1071.
192. **Liu, H. X., M. Zhang, and A. R. Krainer.** 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**:1998-2012.
193. **Lund, M. and J. Kjems.** 2002. Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *RNA.* **8**:166-179.

Referenzen

194. **MacMillan, A. M., C. C. Query, C. R. Allerson, S. Chen, G. L. Verdine, and P. A. Sharp.** 1994. Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev.* **8**:3008-3020.
195. **Maddon, P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and R. Axel.** 1986. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**:333-348. doi:0092-8674(86)90590-8 [pii].
196. **Madhani, H. D. and C. Guthrie.** 1992. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**:803-817. doi:0092-8674(92)90556-R [pii].
197. **Madhani, H. D. and C. Guthrie.** 1994. Dynamic RNA-RNA interactions in the spliceosome. *Annu.Rev.Genet.* **28**:1-26. doi:10.1146/annurev.ge.28.120194.000245 [doi].
198. **Maeder, C., A. K. Kutach, and C. Guthrie.** 2009. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat.Struct.Mol.Biol.* **16**:42-48. doi:nsm.1535 [pii];10.1038/nsm.1535 [doi].
199. **Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm, and R. Luhrmann.** 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**:2205-2208. doi:10.1126/science.1077783 [doi];1077783 [pii].
200. **Manceau, V., C. L. Kielkopf, A. Sobel, and A. Maucuer.** 2008. Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *J.Mol.Biol.* **381**:748-762. doi:S0022-2836(08)00729-8 [pii];10.1016/j.jmb.2008.06.026 [doi].
201. **Manceau, V., M. Swenson, J. P. Le Caer, A. Sobel, C. L. Kielkopf, and A. Maucuer.** 2006. Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *FEBS J.* **273**:577-587. doi:EJB5091 [pii];10.1111/j.1742-4658.2005.05091.x [doi].
202. **Manley, J. L. and R. Tacke.** 1996. SR proteins and splicing control. *Genes Dev.* **10**:1569-1579.
203. **Marquis, J., K. Meyer, L. Angehrn, S. S. Kampfer, B. Rothen-Rutishauser, and D. Schumperli.** 2007. Spinal muscular atrophy: SMN2 pre-mRNA splicing corrected by a U7 snRNA derivative carrying a splicing enhancer sequence. *Mol.Ther.* **15**:1479-1486. doi:6300200 [pii];10.1038/sj.mt.6300200 [doi].
204. **Mathew, R., K. Hartmuth, S. Mohlmann, H. Urlaub, R. Ficner, and R. Luhrmann.** 2008. Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat.Struct.Mol.Biol.* **15**:435-443. doi:nsm.1415 [pii];10.1038/nsm.1415 [doi].
205. **Mathews, D. H. and D. H. Turner.** 2002. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J.Mol.Biol.* **317**:191-203. doi:10.1006/jmbi.2001.5351 [doi];S0022283601953513 [pii].
206. **Mathews, D. H. and D. H. Turner.** 2002. Experimentally derived nearest-neighbor parameters for the stability of RNA three- and four-way multibranch loops. *Biochemistry* **41**:869-880. doi:bi011441d [pii].

Referenzen

207. **Mattaj, I. W.** 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**:905-911. doi:0092-8674(86)90072-3 [pii].
208. **Mayeda, A., J. Badolato, R. Kobayashi, M. Q. Zhang, E. M. Gardiner, and A. R. Krainer.** 1999. Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J.* **18**:4560-4570. doi:10.1093/emboj/18.16.4560 [doi].
209. **McPheeters, D. S. and P. Muhlenkamp.** 2003. Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. *Mol.Cell Biol.* **23**:4174-4186.
210. **Mefford, M. A. and J. P. Staley.** 2009. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA.* **15**:1386-1397. doi:rna.1582609 [pii];10.1261/rna.1582609 [doi].
211. **Meindl, A., H. Hellebrand, C. Wiek, V. Erven, B. Wappenschmidt, D. Niederacher, M. Freund, P. Lichtner, L. Hartmann, H. Schaal, J. Ramser, E. Honisch, C. Kubisch, H. E. Wichmann, K. Kast, H. Deissler, C. Engel, B. Muller-Myhsok, K. Neveling, M. Kiechle, C. G. Mathew, D. Schindler, R. K. Schmutzler, and H. Hanenberg.** 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat.Genet.* **42**:410-414. doi:ng.569 [pii];10.1038/ng.569 [doi].
212. **Merendino, L., S. Guth, D. Bilbao, C. Martinez, and J. Valcarcel.** 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**:838-841. doi:10.1038/45602 [doi].
213. **Meyer, K., J. Marquis, J. Trub, N. R. Nlend, S. Verp, M. D. Ruepp, H. Imboden, I. Barde, D. Trono, and D. Schumperli.** 2009. Rescue of a severe mouse model for spinal muscular atrophy by U7 snRNA-mediated splicing modulation. *Hum.Mol.Genet.* **18**:546-555. doi:ddn382 [pii];10.1093/hmg/ddn382 [doi].
214. **Modafferi, E. F. and D. L. Black.** 1997. A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. *Mol.Cell Biol.* **17**:6537-6545.
215. **Moore, M. J.** 2000. Intron recognition comes of AGE. *Nat.Struct.Biol.* **7**:14-16. doi:10.1038/71207 [doi].
216. **Murphree, A. L. and W. F. Benedict.** 1984. Retinoblastoma: clues to human oncogenesis. *Science* **223**:1028-1033.
217. **Muto, Y., K. D. Pomeranz, C. Oubridge, H. Hernandez, C. V. Robinson, D. Neuhaus, and K. Nagai.** 2004. The structure and biochemical properties of the human spliceosomal protein U1C. *J.Mol.Biol.* **341**:185-198. doi:10.1016/j.jmb.2004.04.078 [doi];S0022283604006096 [pii].
218. **Nagai, K., C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans.** 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**:515-520. doi:10.1038/348515a0 [doi].
219. **Nakai, K. and H. Sakamoto.** 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**:171-177.

Referenzen

220. **Nelissen, R. L., C. L. Will, W. J. van Venrooij, and R. Luhrmann.** 1994. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13**:4113-4125.
221. **Nelson, K. K. and M. R. Green.** 1989. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev.* **3**:1562-1571.
222. **Neveling, K.** 2004. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
223. **Neveling, K., D. Endt, H. Hoehn, and D. Schindler.** 2009. Genotype-phenotype correlations in Fanconi anemia. *Mutat.Res.* **668**:73-91. doi:S0027-5107(09)00168-7 [pii];10.1016/j.mrfmmm.2009.05.006 [doi].
224. **Newby, M. I. and N. L. Greenbaum.** 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat.Struct.Biol.* **9**:958-965. doi:10.1038/nsb873 [doi];nsb873 [pii].
225. **Newman, A. and C. Norman.** 1991. Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell* **65**:115-123. doi:0092-8674(91)90413-S [pii].
226. **Newman, A. J. and C. Norman.** 1992. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**:743-754. doi:0092-8674(92)90149-7 [pii].
227. **Northemann, W., M. Scheurlen, V. Gross, and P. C. Heinrich.** 1977. Circular dichroism of ribonucleoprotein complexes from rat liver nuclei. *Biochem.Biophys.Res.Commun.* **76**:1130-1137. doi:0006-291X(77)90973-1 [pii].
228. **Norton, P. A.** 1994. Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res.* **22**:3854-3860.
229. **O'Keefe, R. T. and A. J. Newman.** 1998. Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. *EMBO J.* **17**:565-574. doi:10.1093/emboj/17.2.565 [doi].
230. **O'Keefe, R. T., C. Norman, and A. J. Newman.** 1996. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86**:679-689. doi:S0092-8674(00)80140-3 [pii].
231. **Ochman, T.** 2011. Institut für Virologie, Heine-Heine-Universität, Düsseldorf.
232. **Ohe, K., T. Watanabe, S. Harada, S. Munesue, Y. Yamamoto, H. Yonekura, and H. Yamamoto.** 2010. Regulation of alternative splicing of the receptor for advanced glycation endproducts (RAGE) through G-rich cis-elements and heterogenous nuclear ribonucleoprotein H. *J.Biochem.* **147**:651-659. doi:mvp207 [pii];10.1093/jb/mvp207 [doi].
233. **Oubridge, C., N. Ito, P. R. Evans, C. H. Teo, and K. Nagai.** 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**:432-438. doi:10.1038/372432a0 [doi].
234. **Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**:898-903.

Referenzen

235. **Padgett, R. A., S. M. Mount, J. A. Steitz, and P. A. Sharp.** 1983. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell* **35**:101-107. doi:0092-8674(83)90212-X [pii].
236. **Parker, R. and P. G. Siliciano.** 1993. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**:660-662. doi:10.1038/361660a0 [doi].
237. **Patel, A. A. and J. A. Steitz.** 2003. Splicing double: insights from the second spliceosome. *Nat.Rev.Mol.Cell Biol.* **4**:960-970. doi:10.1038/nrm1259 [doi];nrm1259 [pii].
238. **Patton, J. R. and T. Pederson.** 1988. The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc.Natl.Acad.Sci.U.S.A* **85**:747-751.
239. **Paul, S., W. Dansithong, D. Kim, J. Rossi, N. J. Webster, L. Comai, and S. Reddy.** 2006. Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO J.* **25**:4271-4283. doi:7601296 [pii];10.1038/sj.emboj.7601296 [doi].
240. **Peled-Zehavi, H., J. A. Berglund, M. Rosbash, and A. D. Frankel.** 2001. Recognition of RNA branch point sequences by the KH domain of splicing factor 1 (mammalian branch point binding protein) in a splicing factor complex. *Mol.Cell Biol.* **21**:5232-5241. doi:10.1128/MCB.21.15.5232-5241.2001 [doi].
241. **Pena, V., A. Rozov, P. Fabrizio, R. Luhrmann, and M. C. Wahl.** 2008. Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* **27**:2929-2940. doi:emboj2008209 [pii];10.1038/emboj.2008.209 [doi].
242. **Perriman, R. J. and M. Ares, Jr.** 2007. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev.* **21**:811-820. doi:21/7/811 [pii];10.1101/gad.1524307 [doi].
243. **Pinotti, M., D. Balestra, L. Rizzotto, I. Maestri, F. Pagani, and F. Bernardi.** 2009. Rescue of coagulation factor VII function by the U1+5A snRNA. *Blood* **113**:6461-6464. doi:blood-2009-03-207613 [pii];10.1182/blood-2009-03-207613 [doi].
244. **Pinotti, M., L. Rizzotto, D. Balestra, M. A. Lewandowska, N. Cavallari, G. Marchetti, F. Bernardi, and F. Pagani.** 2008. U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. *Blood* **111**:2681-2684. doi:blood-2007-10-117440 [pii];10.1182/blood-2007-10-117440 [doi].
245. **Ponder, B.** 1988. Cancer. Gene losses in human tumours. *Nature* **335**:400-402. doi:10.1038/335400a0 [doi].
246. **Ponthier, J. L., C. Schluepen, W. Chen, R. A. Lersch, S. L. Gee, V. C. Hou, A. J. Lo, S. A. Short, J. A. Chasis, J. C. Winkelmann, and J. G. Conboy.** 2006. Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. *J.Biol.Chem.* **281**:12468-12474. doi:M511556200 [pii];10.1074/jbc.M511556200 [doi].
247. **Popp, H., R. Kalb, A. Fischer, S. Lobitz, I. Kokemohr, H. Hanenberg, and D. Schindler.** 2003. Screening Fanconi anemia lymphoid cell lines of non-A, C, D2, E,

Referenzen

- F, G subtypes for defects in BRCA2/FANCD1. *Cytogenet.Genome Res.* **103**:54-57. doi:10.1159/000076289 [doi];76289 [pii].
248. **Query, C. C., M. J. Moore, and P. A. Sharp.** 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**:587-597.
249. **Raghunathan, P. L. and C. Guthrie.** 1998. A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* **279**:857-860.
250. **Raghunathan, P. L. and C. Guthrie.** 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr.Biol.* **8**:847-855. doi:S0960-9822(07)00345-4 [pii].
251. **Rappsilber, J., U. Ryder, A. I. Lamond, and M. Mann.** 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**:1231-1245. doi:10.1101/gr.473902 [doi].
252. **Rasheed, S., W. A. Nelson-Rees, E. M. Toth, P. Arnstein, and M. B. Gardner.** 1974. Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* **33**:1027-1033.
253. **Rebeck, T. R. and S. M. Domchek.** 2008. Variation in breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res.* **10**:108. doi:bcr2115 [pii];10.1186/bcr2115 [doi].
254. **Reddy, R., D. Henning, and H. Busch.** 1981. Pseudouridine residues in the 5'-terminus of uridine-rich nuclear RNA I (U1 RNA). *Biochem.Biophys.Res.Commun.* **98**:1076-1083. doi:0006-291X(81)91221-3 [pii].
255. **Reed, R.** 1989. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**:2113-2123.
256. **Reed, R.** 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr.Opin.Cell Biol.* **12**:340-345. doi:S0955-0674(00)00097-1 [pii].
257. **Reese, M. G., F. H. Eeckman, D. Kulp, and D. Haussler.** 1997. Improved splice site detection in Genie. *J.Comput.Biol.* **4**:311-323.
258. **Reyes, J. L., E. H. Gustafson, H. R. Luo, M. J. Moore, and M. M. Konarska.** 1999. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA.* **5**:167-179.
259. **Reyes, J. L., P. Kois, B. B. Konforti, and M. M. Konarska.** 1996. The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA.* **2**:213-225.
260. **Rinke, J., B. Appel, H. Blocker, R. Frank, and R. Luhrmann.** 1984. The 5'-terminal sequence of U1 RNA complementary to the consensus 5' splice site of hnRNA is single-stranded in intact U1 snRNP particles. *Nucleic Acids Res.* **12**:4111-4126.
261. **Ritchie, D. B., M. J. Schellenberg, and A. M. MacMillan.** 2009. Spliceosome structure: piece by piece. *Biochim.Biophys.Acta* **1789**:624-633. doi:S1874-9399(09)00104-7 [pii];10.1016/j.bbagr.2009.08.010 [doi].

Referenzen

262. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol.Cell Biol.* **10**:84-94.
263. **Roca, X. and A. R. Krainer.** 2009. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. *Nat.Struct.Mol.Biol.* **16**:176-182. doi:nsmb.1546 [pii];10.1038/nsmb.1546 [doi].
264. **Roca, X., R. Sachidanandam, and A. R. Krainer.** 2003. Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res.* **31**:6321-6333.
265. **Roca, X., R. Sachidanandam, and A. R. Krainer.** 2005. Determinants of the inherent strength of human 5' splice sites. *RNA.* **11**:683-698. doi:11/5/683 [pii];10.1261/rna.2040605 [doi].
266. **Rodriguez, J. R., C. W. Pikielny, and M. Rosbash.** 1984. In vivo characterization of yeast mRNA processing intermediates. *Cell* **39**:603-610. doi:0092-8674(84)90467-7 [pii].
267. **Rogan, P. K., B. M. Faux, and T. D. Schneider.** 1998. Information analysis of human splice site mutations. *Hum.Mutat.* **12**:153-171. doi:10.1002/(SICI)1098-1004(1998)12:3<153::AID-HUMU3>3.0.CO;2-I [pii];10.1002/(SICI)1098-1004(1998)12:3<153::AID-HUMU3>3.0.CO;2-I [doi].
268. **Rogan, P. K. and T. D. Schneider.** 1995. Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum.Mutat.* **6**:74-76. doi:10.1002/humu.1380060114 [doi].
269. **Ruskin, B. and M. R. Green.** 1985. An RNA processing activity that debranches RNA lariats. *Science* **229**:135-140.
270. **Ruskin, B., P. D. Zamore, and M. R. Green.** 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**:207-219. doi:0092-8674(88)90509-0 [pii].
271. **Rutz, B. and B. Seraphin.** 1999. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA.* **5**:819-831.
272. **Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler.** 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**:928-933. doi:10.1038/35057149 [doi].
273. **Saeyns, Y., S. Degroeve, D. Aeyels, P. Rouze, and Y. Van de Peer.** 2004. Feature selection for splice site prediction: a new method using EDA-based feature ranking. *BMC.Bioinformatics.* **5**:64. doi:10.1186/1471-2105-5-64 [doi];1471-2105-5-64 [pii].
274. **Saeyns, Y., S. Degroeve, D. Aeyels, Y. Van de Peer, and P. Rouze.** 2003. Fast feature selection using a simple estimation of distribution algorithm: a case study on splice site prediction. *Bioinformatics.* **19 Suppl 2**:ii179-ii188.

Referenzen

275. **Sahashi, K., A. Masuda, T. Matsuura, J. Shinmi, Z. Zhang, Y. Takeshima, M. Matsuo, G. Sobue, and K. Ohno.** 2007. In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Res.* **35**:5995-6003. doi:gkm647 [pii];10.1093/nar/gkm647 [doi].
276. **Sandoval, N., M. Platzer, A. Rosenthal, T. Dork, R. Bendix, B. Skawran, M. Stuhmann, R. D. Wegner, K. Sperling, S. Banin, Y. Shiloh, A. Baumer, U. Bernthaler, H. Sennefelder, M. Brohm, B. H. Weber, and D. Schindler.** 1999. Characterization of ATM gene mutations in 66 ataxia telangiectasia families. *Hum.Mol.Genet.* **8**:69-79. doi:ddc009 [pii].
277. **Sashital, D. G., G. Cornilescu, C. J. McManus, D. A. Brow, and S. E. Butcher.** 2004. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat.Struct.Mol.Biol.* **11**:1237-1242. doi:nsmb863 [pii];10.1038/nsmb863 [doi].
278. **Savitsky, K., M. Platzer, T. Uziel, S. Gilad, A. Sartiel, A. Rosenthal, O. Elroy-Stein, Y. Shiloh, and G. Rotman.** 1997. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of ATM gene expression. *Nucleic Acids Res.* **25**:1678-1684. doi:gka314 [pii].
279. **Schaffert, N., M. Hossbach, R. Heintzmann, T. Achsel, and R. Luhrmann.** 2004. RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J.* **23**:3000-3009. doi:10.1038/sj.emboj.7600296 [doi];7600296 [pii].
280. **Schaub, M. C., S. R. Lopez, and M. Caputi.** 2007. Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. *J.Biol.Chem.* **282**:13617-13626. doi:M700774200 [pii];10.1074/jbc.M700774200 [doi].
281. **Scherer, S.** 2008. A short guide to the human genome. Cold Spring Harbor Press, New York.
282. **Scherly, D., W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm, and I. W. Mattaj.** 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**:4163-4170.
283. **Schneider, M., C. L. Will, M. Anokhina, J. Tazi, H. Urlaub, and R. Luhrmann.** 2010. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like pre-catalytic splicing complexes. *Mol.Cell* **38**:223-235. doi:S1097-2765(10)00212-1 [pii];10.1016/j.molcel.2010.02.027 [doi].
284. **Schöneweis, K.** 2010. Institut für Virologie, Heinrich-Heine-Universität, Düsseldorf.
285. **Schwer, B.** 2008. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol.Cell* **30**:743-754. doi:S1097-2765(08)00331-6 [pii];10.1016/j.molcel.2008.05.003 [doi].
286. **Schwer, B. and C. Guthrie.** 1991. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**:494-499. doi:10.1038/349494a0 [doi].
287. **Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, and M. Wigler.** 2004.

Referenzen

- Large-scale copy number polymorphism in the human genome. *Science* **305**:525-528. doi:10.1126/science.1098918 [doi];305/5683/525 [pii].
288. **Segault, V., C. L. Will, M. Polycarpou-Schwarz, I. W. Mattaj, C. Branlant, and R. Luhrmann.** 1999. Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. *Mol.Cell Biol.* **19**:2782-2790.
289. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. *Nucleic Acids Res.* **6**:3387-3398.
290. **Selenko, P., G. Gregorovic, R. Sprangers, G. Stier, Z. Rhani, A. Kramer, and M. Sattler.** 2003. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol.Cell* **11**:965-976. doi:S1097276503001151 [pii].
291. **Selvakumar, M. and D. M. Helfman.** 1999. Exonic splicing enhancers contribute to the use of both 3' and 5' splice site usage of rat beta-tropomyosin pre-mRNA. *RNA.* **5**:378-394.
292. **Senapathy, P., M. B. Shapiro, and N. L. Harris.** 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**:252-278.
293. **Serra, M. J. and D. H. Turner.** 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol.* **259**:242-261.
294. **Shapiro, M. B. and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**:7155-7174.
295. **Sharma, S., L. A. Kohlstaedt, A. Damianov, D. C. Rio, and D. L. Black.** 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat.Struct.Mol.Biol.* **15**:183-191. doi:nsmb.1375 [pii];10.1038/nsmb.1375 [doi].
296. **Sharp, P. A.** 1994. Split genes and RNA splicing. *Cell* **77**:805-815. doi:0092-8674(94)90130-9 [pii].
297. **Shen, H. and M. R. Green.** 2006. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev.* **20**:1755-1765. doi:gad.1422106 [pii];10.1101/gad.1422106 [doi].
298. **Shen, H., J. L. Kan, and M. R. Green.** 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol.Cell* **13**:367-376. doi:S1097276504000255 [pii].
299. **Shepard, P. J., E. A. Choi, A. Busch, and K. J. Hertel.** 2011. Efficient internal exon recognition depends on near equal contributions from the 3' and 5' splice sites. *Nucleic Acids Res.* **39**:8928-8937. doi:gkr481 [pii];10.1093/nar/gkr481 [doi].
300. **Siatecka, M., J. L. Reyes, and M. M. Konarska.** 1999. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**:1983-1993.

Referenzen

301. **Sickmier, E. A., K. E. Frato, H. Shen, S. R. Paranawithana, M. R. Green, and C. L. Kielkopf.** 2006. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol.Cell* **23**:49-59. doi:S1097-2765(06)00340-6 [pii];10.1016/j.molcel.2006.05.025 [doi].
302. **Sigel, R. K., A. Vaidya, and A. M. Pyle.** 2000. Metal ion binding sites in a group II intron core. *Nat.Struct.Biol.* **7**:1111-1116. doi:10.1038/81958 [doi].
303. **Singh, R. and J. Valcarcel.** 2005. Building specificity with nonspecific RNA-binding proteins. *Nat.Struct.Mol.Biol.* **12**:645-653. doi:nsmb961 [pii];10.1038/nsmb961 [doi].
304. **Singh, R., J. Valcarcel, and M. R. Green.** 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**:1173-1176.
305. **Sironi, M., G. Menozzi, L. Riva, R. Cagliani, G. P. Comi, N. Bresolin, R. Giorda, and U. Pozzoli.** 2004. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.* **32**:1783-1791. doi:10.1093/nar/gkh341 [doi];32/5/1783 [pii].
306. **Small, E. C., S. R. Leggett, A. A. Winans, and J. P. Staley.** 2006. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol.Cell* **23**:389-399. doi:S1097-2765(06)00413-8 [pii];10.1016/j.molcel.2006.05.043 [doi].
307. **Smith, C. W., T. T. Chu, and B. Nadal-Ginard.** 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol.Cell Biol.* **13**:4939-4952.
308. **Soares, L. M., K. Zanier, C. Mackereth, M. Sattler, and J. Valcarcel.** 2006. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**:1961-1965. doi:312/5782/1961 [pii];10.1126/science.1128659 [doi].
309. **Sontheimer, E. J. and J. A. Steitz.** 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**:1989-1996.
310. **Sorek, R. and G. Ast.** 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**:1631-1637. doi:10.1101/gr.1208803 [doi];13/7/1631 [pii].
311. **Sorek, R., G. Lev-Maor, M. Reznik, T. Dagan, F. Belinky, D. Graur, and G. Ast.** 2004. Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol.Cell* **14**:221-231. doi:S1097276504001819 [pii].
312. **Sorek, R., R. Shemesh, Y. Cohen, O. Basechess, G. Ast, and R. Shamir.** 2004. A non-EST-based method for exon-skipping prediction. *Genome Res.* **14**:1617-1623. doi:10.1101/gr.2572604 [doi];14/8/1617 [pii].
313. **Spadaccini, R., U. Reidt, O. Dybkov, C. Will, R. Frank, G. Stier, L. Corsini, M. C. Wahl, R. Luhrmann, and M. Sattler.** 2006. Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA.* **12**:410-425. doi:12/3/410 [pii];10.1261/rna.2271406 [doi].

Referenzen

314. **Spena, S., S. Duga, R. Asselta, M. Malcovati, F. Peyvandi, and M. L. Tenchini.** 2002. Congenital afibrinogenemia: first identification of splicing mutations in the fibrinogen Bbeta-chain gene causing activation of cryptic splice sites. *Blood* **100**:4478-4484. doi:10.1182/blood-2002-06-1647 [doi];2002-06-1647 [pii].
315. **Spena, S., M. L. Tenchini, and E. Buratti.** 2006. Cryptic splice site usage in exon 7 of the human fibrinogen Bbeta-chain gene is regulated by a naturally silent SF2/ASF binding site within this exon. *RNA*. **12**:948-958. doi:rna.2269306 [pii];10.1261/rna.2269306 [doi].
316. **Sporn, M. B. and A. B. Roberts.** 1985. Autocrine, paracrine and endocrine mechanisms of growth control. *Cancer Surv.* **4**:627-632.
317. **Staley, J. P. and C. Guthrie.** 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**:315-326. doi:S0092-8674(00)80925-3 [pii].
318. **Staley, J. P. and C. Guthrie.** 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol.Cell* **3**:55-64. doi:S1097-2765(00)80174-4 [pii].
319. **Stanbridge, E. J.** 1976. Suppression of malignancy in human cells. *Nature* **260**:17-20.
320. **Stanek, D., J. Pridalova-Hnilicova, I. Novotny, M. Huranova, M. Blazikova, X. Wen, A. K. Sapra, and K. M. Neugebauer.** 2008. Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol.Biol.Cell* **19**:2534-2543. doi:E07-12-1259 [pii];10.1091/mbc.E07-12-1259 [doi].
321. **Stehelin, D., H. E. Varmus, J. M. Bishop, and P. K. Vogt.** 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**:170-173.
322. **Stenson, P. D., E. V. Ball, K. Howells, A. D. Phillips, M. Mort, and D. N. Cooper.** 2009. The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum.Genomics* **4**:69-72. doi:U8K3X868GR637691 [pii].
323. **Sterner, D. A., T. Carlo, and S. M. Berget.** 1996. Architectural limits on split genes. *Proc.Natl.Acad.Sci.U.S.A* **93**:15081-15085.
324. **Stevens, S. W., D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson.** 2002. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol.Cell* **9**:31-44. doi:S1097276502004367 [pii].
325. **Strauss, E. J. and C. Guthrie.** 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev.* **5**:629-641.
326. **Sturchler, C., P. Carbon, and A. Krol.** 1992. An additional long-range interaction in human U1 snRNA. *Nucleic Acids Res.* **20**:1215-1221.
327. **Sun, H. and L. A. Chasin.** 2000. Multiple splicing defects in an intronic false exon. *Mol.Cell Biol.* **20**:6414-6425.
328. **Sun, J. S. and J. L. Manley.** 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**:843-854.

Referenzen

329. **Surowy, C. S., V. L. van Santen, S. M. Scheib-Wixted, and R. A. Spritz.** 1989. Direct, sequence-specific binding of the human U1-70K ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. *Mol.Cell Biol.* **9**:4179-4186.
330. **Tacke, R., Y. Chen, and J. L. Manley.** 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc.Natl.Acad.Sci.U.S.A* **94**:1148-1153.
331. **Tange, T. O., C. K. Damgaard, S. Guth, J. Valcarcel, and J. Kjems.** 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* **20**:5748-5758. doi:10.1093/emboj/20.20.5748 [doi].
332. **Tanner, G., E. Glaus, D. Barthelmes, M. Ader, J. Fleischhauer, F. Pagani, W. Berger, and J. Neidhardt.** 2009. Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Hum.Mutat.* **30**:255-263. doi:10.1002/humu.20861 [doi].
333. **Tarn, W. Y. and J. A. Steitz.** 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**:801-811. doi:S0092-8674(00)81057-0 [pii].
334. **Tarn, W. Y. and J. A. Steitz.** 1996. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**:1824-1832.
335. **Teigelkamp, S., A. J. Newman, and J. D. Beggs.** 1995. Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14**:2602-2612.
336. **Teigelkamp, S., E. Whittaker, and J. D. Beggs.** 1995. Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res.* **23**:320-326. doi:4c0229 [pii].
337. **Teraoka, S. N., M. Telatar, S. Becker-Catania, T. Liang, S. Onengut, A. Tolun, L. Chessa, O. Sanal, E. Bernatowska, R. A. Gatti, and P. Concannon.** 1999. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am.J.Hum.Genet.* **64**:1617-1631. doi:S0002-9297(07)63663-5 [pii];10.1086/302418 [doi].
338. **Tilghman, S. M., D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, and P. Leder.** 1978. Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc.Natl.Acad.Sci.U.S.A* **75**:725-729.
339. **Timmers, C., T. Taniguchi, J. Hejna, C. Reifsteck, L. Lucas, D. Bruun, M. Thayer, B. Cox, S. Olson, A. D. D'Andrea, R. Moses, and M. Grompe.** 2001. Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol.Cell* **7**:241-248. doi:S1097-2765(01)00172-1 [pii].
340. **Tonegawa, S., C. Brack, N. Hozumi, and V. Pirrotta.** 1978. Organization of immunoglobulin genes. *Cold Spring Harb.Symp.Quant.Biol.* **42 Pt 2**:921-931.
341. **Ule, J., G. Stefani, A. Mele, M. Ruggiu, X. Wang, B. Taneri, T. Gaasterland, B. J. Blencowe, and R. B. Darnell.** 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**:580-586. doi:nature05304 [pii];10.1038/nature05304 [doi].

Referenzen

342. **Valcarcel, J., R. K. Gaur, R. Singh, and M. R. Green.** 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273**:1706-1709.
343. **Valentine, C. R.** 1998. The association of nonsense codons with exon skipping. *Mutat.Res.* **411**:87-117.
344. **Valenzuela, P., A. Venegas, F. Weinberg, R. Bishop, and W. J. Rutter.** 1978. Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA. *Proc.Natl.Acad.Sci.U.S.A* **75**:190-194.
345. **Varani, G. and K. Nagai.** 1998. RNA recognition by RNP proteins during RNA processing. *Annu.Rev.Biophys.Biomol.Struct.* **27**:407-445. doi:10.1146/annurev.biophys.27.1.407 [doi].
346. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, F. Di, V, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, and M. Nodell.** 2001. The sequence of the human genome. *Science* **291**:1304-1351. doi:10.1126/science.1058040 [doi];291/5507/1304 [pii].

Referenzen

347. **Verlander, P. C., A. Kaporis, Q. Liu, Q. Zhang, U. Seligsohn, and A. D. Auerbach.** 1995. Carrier frequency of the IVS4 + 4 A-->T mutation of the Fanconi anemia gene FAC in the Ashkenazi Jewish population. *Blood* **86**:4034-4038.
348. **Vorechovsky, I.** 2006. Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* **34**:4630-4641. doi:gkl535 [pii];10.1093/nar/gkl535 [doi].
349. **Wahl, M. C., C. L. Will, and R. Luhrmann.** 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**:701-718. doi:S0092-8674(09)00146-9 [pii];10.1016/j.cell.2009.02.009 [doi].
350. **Walsh, T. and M. C. King.** 2007. Ten genes for inherited breast cancer. *Cancer Cell* **11**:103-105. doi:S1535-6108(07)00025-6 [pii];10.1016/j.ccr.2007.01.010 [doi].
351. **Wang, G. S. and T. A. Cooper.** 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat.Rev.Genet.* **8**:749-761. doi:nrg2164 [pii];10.1038/nrg2164 [doi].
352. **Wang, J., Y. F. Chang, J. I. Hamilton, and M. F. Wilkinson.** 2002. Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay. *Mol.Cell* **10**:951-957. doi:S1097276502006354 [pii].
353. **Wang, Z., M. E. Rolish, G. Yeo, V. Tung, M. Mawson, and C. B. Burge.** 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**:831-845. doi:S0092867404010566 [pii];10.1016/j.cell.2004.11.010 [doi].
354. **Wang, Z., X. Xiao, N. E. Van, and C. B. Burge.** 2006. General and specific functions of exonic splicing silencers in splicing control. *Mol.Cell* **23**:61-70. doi:S1097-2765(06)00333-9 [pii];10.1016/j.molcel.2006.05.018 [doi].
355. **Wassarman, D. A. and J. A. Steitz.** 1992. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science* **257**:1918-1925.
356. **Weinberg, R. A.** 1981. Use of transfection to analyze genetic information and malignant transformation. *Biochim.Biophys.Acta* **651**:25-35.
357. **Weinberg, R. A.** 2007. *The biology of cancer.* Garland Science, Taylor & Francis Group, New York.
358. **Whitney, M. A., P. Jakobs, M. Kaback, R. E. Moses, and M. Grompe.** 1994. The Ashkenazi Jewish Fanconi anemia mutation: incidence among patients and carrier frequency in the at-risk population. *Hum.Mutat.* **3**:339-341. doi:10.1002/humu.1380030402 [doi].
359. **Wijk, R., A. C. van Wesel, A. A. Thomas, G. Rijksen, and W. W. van Solinge.** 2004. Ex vivo analysis of aberrant splicing induced by two donor site mutations in PKLR of a patient with severe pyruvate kinase deficiency. *Br.J.Haematol.* **125**:253-263. doi:10.1111/j.1365-2141.2004.04895.x [doi];BJH4895 [pii].
360. **Will, C. L. and R. Luhrmann.** 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol.Chem.* **386**:713-724. doi:10.1515/BC.2005.084 [doi].
361. **Will, C. L., S. Rumpler, G. J. Klein, W. J. van Venrooij, and R. Luhrmann.** 1996. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein

Referenzen

- enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**:4614-4623. doi:6w0172 [pii].
362. **Will, C. L., C. Schneider, A. M. MacMillan, N. F. Katopodis, G. Neubauer, M. Wilm, R. Luhrmann, and C. C. Query.** 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.* **20**:4536-4546. doi:10.1093/emboj/20.16.4536 [doi].
363. **Will, C. L., C. Schneider, R. Reed, and R. Luhrmann.** 1999. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* **284**:2003-2005. doi:7567 [pii].
364. **Will, C. L., H. Urlaub, T. Achsel, M. Gentzel, M. Wilm, and R. Luhrmann.** 2002. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.* **21**:4978-4988.
365. **Wimmer, K., X. Roca, H. Beiglbock, T. Callens, J. Etzler, A. R. Rao, A. R. Krainer, C. Fonatsch, and L. Messiaen.** 2007. Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. *Hum.Mutat.* **28**:599-612. doi:10.1002/humu.20493 [doi].
366. **Wolf, E., B. Kastner, J. Deckert, C. Merz, H. Stark, and R. Luhrmann.** 2009. Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28**:2283-2292. doi:emboj2009171 [pii];10.1038/emboj.2009.171 [doi].
367. **Wu, J. Y. and T. Maniatis.** 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**:1061-1070. doi:0092-8674(93)90316-I [pii].
368. **Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green.** 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**:832-835. doi:10.1038/45590 [doi].
369. **Wyatt, J. R., E. J. Sontheimer, and J. A. Steitz.** 1992. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**:2542-2553.
370. **Xiao, X., Z. Wang, M. Jang, and C. B. Burge.** 2007. Coevolutionary networks of splicing cis-regulatory elements. *Proc.Natl.Acad.Sci.U.S.A* **104**:18583-18588. doi:0707349104 [pii];10.1073/pnas.0707349104 [doi].
371. **Xu, Y. Z., C. M. Newnham, S. Kameoka, T. Huang, M. M. Konarska, and C. C. Query.** 2004. Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J.* **23**:376-385. doi:10.1038/sj.emboj.7600050 [doi];7600050 [pii].
372. **Yean, S. L., G. Wuenschell, J. Termini, and R. J. Lin.** 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**:881-884. doi:10.1038/35048617 [doi].
373. **Yeo, G. and C. B. Burge.** 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J.Comput.Biol.* **11**:377-394. doi:10.1089/1066527041410418 [doi].

Referenzen

374. **Yeo, G., S. Hoon, B. Venkatesh, and C. B. Burge.** 2004. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc.Natl.Acad.Sci.U.S.A* **101**:15700-15705. doi:0404901101 [pii];10.1073/pnas.0404901101 [doi].
375. **Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth.** 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**:219-222.
376. **Zamore, P. D. and M. R. Green.** 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc.Natl.Acad.Sci.U.S.A* **86**:9243-9247.
377. **Zamore, P. D. and M. R. Green.** 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc.Natl.Acad.Sci.U.S.A* **86**:9243-9247.
378. **Zamore, P. D., J. G. Patton, and M. R. Green.** 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**:609-614. doi:10.1038/355609a0 [doi].
379. **Zeitlin, S. and A. Efstratiadis.** 1984. In vivo splicing products of the rabbit beta-globin pre-mRNA. *Cell* **39**:589-602. doi:0092-8674(84)90466-5 [pii].
380. **Zhang, X. H., M. A. Arias, S. Ke, and L. A. Chasin.** 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA*. **15**:367-376. doi:rna.1498509 [pii];10.1261/rna.1498509 [doi].
381. **Zhang, X. H., K. A. Heller, I. Hefter, C. S. Leslie, and L. A. Chasin.** 2003. Sequence information for the splicing of human pre-mRNA identified by support vector machine classification. *Genome Res.* **13**:2637-2650. doi:10.1101/gr.1679003 [doi];13/12/2637 [pii].
382. **Zhou, Z., L. J. Licklider, S. P. Gygi, and R. Reed.** 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**:182-185. doi:10.1038/nature01031 [doi];nature01031 [pii].
383. **Zhou, Z., J. Sim, J. Griffith, and R. Reed.** 2002. Purification and electron microscopic visualization of functional human spliceosomes. *Proc.Natl.Acad.Sci.U.S.A* **99**:12203-12207. doi:10.1073/pnas.182427099 [doi];182427099 [pii].
384. **Zhuang, Y. and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**:827-835. doi:0092-8674(86)90064-4 [pii].
385. **Zhuang, Y. A., A. M. Goldstein, and A. M. Weiner.** 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc.Natl.Acad.Sci.U.S.A* **86**:2752-2756.

ANHANG

PUBLIKATIONEN (MIT EIGENER BETEILIGUNG)

Hartmann L, Neveling K, Borkens S, Schneider H, Freund M, Grassman E, Theiss S, Wawer A, Burdach S, Auerbach AD, Schindler D, Hanenberg H, Schaal H

Correct mRNA processing at a mutant TT splice donor in FANCC ameliorates the clinical phenotype in patients and is enhanced by delivery of suppressor U1 snRNAs

American Journal of Human Genetics 2010 October 8

Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, **Hartmann L**, Schaal H, Ramser J, Honisch E, Kubisch C, Wichmann HE, Kast K, Deissler H, Engel C, Muller-Myhsok B; Neveling K, Kiechle M, Mathew CG, Schindler D, Schmutzler RK, Hanenberg H

Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene

Nature Genetics 2010 April 18

Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, **Hartmann L**, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG

Mutation of the RAD51C gene in a Fanconi anemia-like disorder

Nature Genetics 2010 April 18

Hartmann L, Theiss S, Niederacher D, Schaal H

Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases?

Frontiers in Bioscience 2008 May 1

DANKSAGUNG

Danken möchte ich an dieser Stelle

Herrn Prof. Dr. Rolf Wagner für die freundliche Übernahme der Betreuung und sein Interesse an der Arbeit.

Herrn Prof. Dr. Heiner Schaal für die Übernahme des Ko-Referats sowie für die individuelle Betreuung und seine ansteckende Begeisterungsfähigkeit.

Herrn Stephan Theiss für die exzellente Unterstützung bei allen bioinformatischen Aspekten der Arbeit.

Herrn Dr. Werner Bouschen und Frau Dr. Sabine Metzger für die Hilfe bei der massenspektrometischen Proteinanalyse.

Herrn Prof. Dr. Helmut Hanenberg für das interessante Thema und die rund um die Uhr Unterstützung.

allen Mitarbeiterinnen und Mitarbeitern der Virologie und Kinderklinik Düsseldorf für die gute Zusammenarbeit.

Meiner Familie
und meinen Freunden

ERKLÄRUNG

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe. Die Arbeit wurde bisher noch nicht anderweitig als Dissertation eingereicht oder veröffentlicht.

Düsseldorf, den 27.04.2012

Linda Hartmann

