RNA: Regulatory information in human disease and viral infection

RNA: Regulatorische Information bei humanen Krankheiten und viralen Infektionen

Inaugural-Dissertation

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

vorgelegt von

Lisa Müller
aus Idar-Oberstein

Düsseldorf, August 2021

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

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

Berichterstatter:

1. Prof. Dr. Heiner Schaal

2. Prof. Dr. Michael Feldbrügge

Tag der mündlichen Prüfung: 24.11.2021

Für meine Eltern

Content

Abstract

RNA mediates the flow of information between DNA and protein synthesis but also carries regulatory information required in a plethora of biological processes. Therefore, the precise regulation of its processing is necessary. A central step in pre-mRNA processing, especially in eukaryotes, is splicing. The spliceosome, a large, macromolecular machinery, removes introns from the pre-mRNA and ligates exons. This orchestrated process is tightly regulated and in particular, requires the precise recognition of the 3' and 5' splice sites at the exon/intron borders. The recognition of the splice sites can be supported or repressed by splicing regulatory elements that offer binding sites for splicing regulatory proteins. This regulation is carried out in a position-dependent manner, which is in particular important when two competing splice sites are located in proximity. Single point mutations in these splicing regulatory elements or the splice sites themselves can drastically change the splicing outcome and can contribute to human pathogenic diseases. Therefore, reliable prediction and robust modeling of the splicing outcome by bioinformatics could aid the evaluation of mutations affecting splicing in human genetic diagnostics.

RNA as an information store of genetic information is also exploited by viruses and at the same time serves to transmit this information. Despite their limited genome size, all viruses have acquired features during evolution to extract maximum information from the limited genome for a complex viral proteome necessary for successful replication. These features, for example, include extensive alternative splicing using the eukaryotic host cellular splicing machinery to enable the translation of downstream open reading frames, as in the case of HIV-1, or discontinuous transcription in the case of SARS-CoV-2. A balanced regulation of HIV-1 alternative splicing by splicing regulatory elements is highly crucial for successful viral replication and can thus be potentially targeted for novel antiviral approaches.

In the light of the SARS-CoV-2 pandemic, a global effort for the simultaneous search for novel antiviral approaches as well as prophylactic vaccinations was made and successfully put in place. To further the understanding of the viral pathogenesis and the immunological features of SARS-CoV-2, infection-based assays and the assessment of the humoral immune response to infection and vaccination are important pillars.

Zusammenfassung

RNA vermittelt den Informationsfluss zwischen DNA und Proteinsynthese, trägt jedoch auch regulatorische Informationen, die zu einer Vielzahl von biologischen Prozessen beitragen. Daher ist eine präzise Regulierung der RNA Prozessierung notwendig. Ein wichtiger Schritt dabei, insbesondere bei Eukaryoten, ist das Spleißen. Das Spleißeosom entfernt Introns aus der prä-mRNA und ligiert Exons. Dieser Prozess ist streng reguliert und erfordert insbesondere eine präzise Erkennung der 3'- und 5'-Spleißstellen an den Exon/Intron-Grenzen. Die Erkennung der Spleißstellen kann durch spleißregulatorische Elemente unterstützt oder unterdrückt werden, die Bindungsstellen für spleißregulatorische Proteine bereitstellen. Diese Regulation ist positionsabhängig, was besonders wichtig ist wenn zwei konkurrierende Spleißstellen in unmittelbarer Nähe sind. Einzelne Punktmutationen in diesen spleißregulatorischen Elementen oder in Spleißstellen selbst können das Spleißergebnis stark verändern und zu humanpathogenen Krankheiten beitragen. Daher ist eine verlässliche Vorhersage und robuste Modellierung des Spleißergebnisses durch Bioinformatik bei der Bewertung von Mutationen in der humangenetischen Diagnostik hilfreich.

RNA als Speichermedium der genetischen Information wird auch von Viren genutzt und dient gleichzeitig der Informationsübertragung. Trotz ihrer begrenzten Genomgröße haben Viren im Laufe der Evolution Merkmale erworben, um ein Maximum an Information für ein komplexes virales Proteom zu extrahieren, das für eine erfolgreiche Replikation notwendig ist. Zu diesen Merkmalen gehört, wie im Fall von HIV-1, extensives alternatives Spleißen unter Verwendung der eukaryotischen Wirtszell-Spleißmaschinerie, um die Translation nachgeschalteter offener Leserahmen zu ermöglichen, oder diskontinuierliche Transkription im Fall von SARS-CoV-2. Eine ausgewogene Regulation des alternativen Spleißens von HIV-1 durch spleißregulatorische Elemente ist wichtig für die virale Replikation und kann daher potenziell für neuartige antivirale Ansätze genutzt werden.

Vor dem Hintergrund der SARS-CoV-2-Pandemie wurde weltweit gleichzeitig an neue antivirale Ansätze sowie prophylaktischen Impfstoffe geforscht, die nun im Einsatz sind. Um die virale Pathogenese und die immunologischen Eigenschaften von SARS-CoV-2 besser zu verstehen, sind infektionsbasierte Assays und die Bewertung der humoralen Immunantwort auf Infektion und Impfung ein wichtiges Werkzeug.

1. Introduction

1.1 RNA

The central dogma of molecular biology is that genetic information stored as DNA must first be transcribed into an mRNA template before it is translated into proteins. Retroviruses reverse this flow by carrying a reverse transcriptase that generates DNA from RNA. For a long time, however, it has also been found that DNA not only encodes proteins but also generates various types of functional, untranslated RNAs, which are becoming increasingly important for gene regulation (1).

1.1.1 RNA processing

Upon the generation of precursor messenger RNA (pre-mRNA) via the DNA-dependent RNA polymerase II (Pol II), a series of steps are required to generate the mature mRNA molecule, transport it to the cytoplasm, and translate it into proteins. The first step of gene expression is the transcription of DNA into pre-mRNA by Pol II in assembly with transcription factors (2). Upon binding of Pol II to the promotor, the two DNA strands are separated so that the template DNA strand is accessible for transcription. During elongation, the RNA strand is generated in the 5' to 3' direction. After approximately 25nt a 7-methyl guanosine cap is fitted to the 5’ end.

The segmentation of the forming RNA by introns and exons is eliminated by pre-mRNA splicing, mostly co-transcriptional. This is then followed by polyadenylation, a process in which certain sequence elements, in particular the polyA site, close to the 3’ end is recognized by endonucleolytic cleaving enzymes which allow the addition of a poly adenosine tail (Figure 1). Hence, different regulatory machineries are required to engage with each other to function properly.

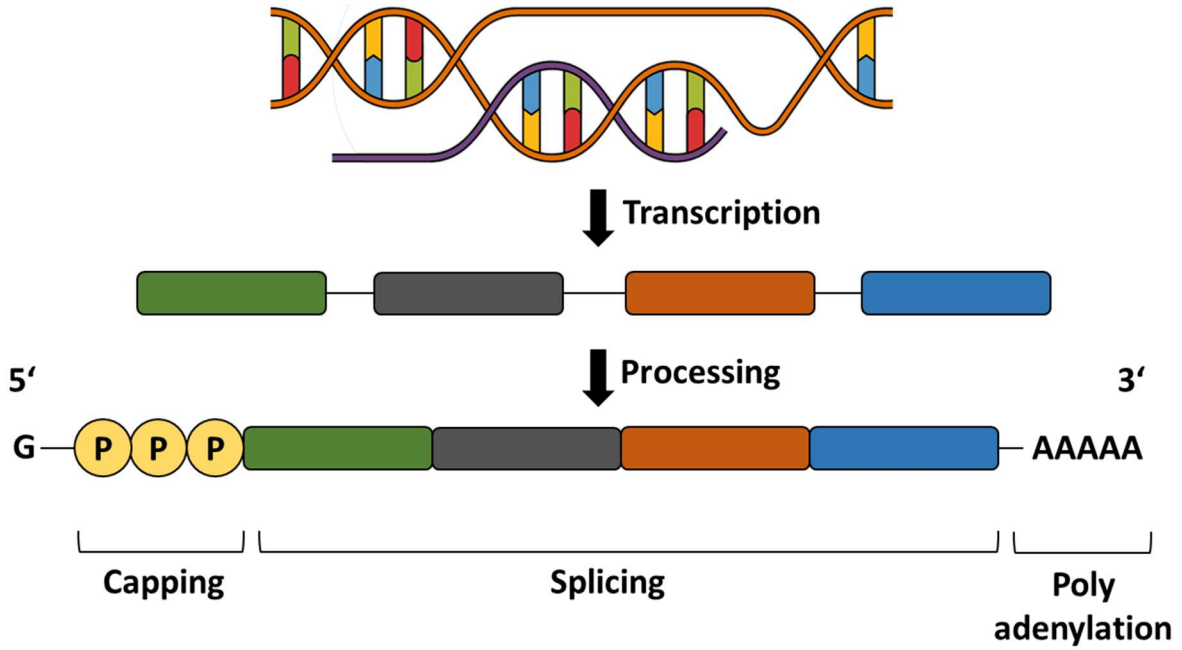


Figure 1: Pre-mRNA processing.

DNA is transcribed by the DNA-dependent RNA polymerase II. During transcription, the emerging pre-mRNA undergoes co-transcriptional processing in form of 5’ capping, splicing, and polyadenylation (Figure designed with Mind the Graph).

1.1.2 pre-mRNA splicing

For the vast majority of genes, before translation, intervening introns must be removed from the pre-mRNA and exons have to be ligated during the splicing process to produce a mature mRNA (3). The splicing process was first discovered in the 1970s in Adenovirus 2 late mRNA (4) and later in eukaryotes, more precisely in the IgM gene (5). Splicing, as an important form of gene regulation, is required to make full use of the modular gene structure. Alternative splicing (AS), first hypothesized in 1978 (6) in particular enables a single gene to increase its coding capacity during co- and post-transcriptional modifications by allowing the synthesis of transcript isoforms where exons, parts of exons, or intronic regions within a primary transcript are differentially joined, retained or skipped. It is hypothesized that around 95% of human genes are alternatively spliced (7) whereby there is mostly only one major variant (8). This can result in functional and structural distinct protein isoforms and tremendously extents the transcriptomic and subsequently proteomic repertoire of a single gene (9). Different protein isoforms arising from alternative splice variants can then ultimately influence protein-protein and protein-nucleic acid interactions as well as protein distribution and

enzymatic activity, demonstrating the ubiquitous role of splicing in various cellular processes, either physiological or in the case of misregulation, pathological (10,11). While this demonstrates the widespread and scope of splicing, some of the regulatory mechanisms and the full “splicing code” that orchestrates the process remains to be fully understood.

1.1.2.1 Key elements

The removal of intervening intronic sequences and the ligation of exons requires the recognition of intron-exon-boundaries with nucleotide precision. Therefore, various sequence elements, either directly or indirectly, orchestrate the early assembly of the spliceosome, the macromolecular machinery that facilitates the splicing process.

1.1.2.2 Spliceosome

The spliceosome, a highly dynamic multi-megadalton ribonucleoprotein (RNP) complex, carries out the splicing process together with a plethora of associated proteins. Mass spectroscopy analyses revealed that up to 200 core proteins are associated with the spliceosome (12,13). Uridine-rich small nuclear RNPs (snRNPs) are the major building blocks of the spliceosome. In general, two types of spliceosomes are available, the major U2-dependent spliceosome that recognizes GT-AG type introns and the minor U-12 dependent spliceosome that does not require the highly conserved canonical GT-AG dinucleotides but rather recognizes AT-AC type II introns (14).

The major U2-dependent spliceosome consists of the U1, U2, U5, and U4/U6 snRNPs, each with an associated snRNA (or two for U4/U6), while the U1 snRNP and U2 snRNP in the minor U-12 dependent spliceosome are replaced by the U11 and U12 snRNP, which are known to assemble as U11/U12 di-snRNP. Each snRNP (except for U6 snRNP) is associated with a set of seven Sm-proteins (B/B’, D3, D3, D1, E, F, and G) that assemble in a hetero-heptameric ring around the Sm site to constitute the snRNP particle core together with various particle specific proteins. The U6 snRNP, in contrast, is constituted of a hetero-heptameric ring of like-Sm proteins (LSm2-8) that assemble at the U6 snRNA 3’-terminus (15-17).

The biogenesis of RNPs requires an intermediate state including stable but inactive pre-RNPs which are spatially distinguished in sub-cellular compartments away from their final site of function. Their associated snRNAs, which are non-coding, non-polyadenylated transcripts, are differentiated into two main classes, the major Sm-class spliceosomal RNAs (U1, U2, U4, U4 atac, U5, U11, U12) and the Sm-like class (U6 and U6 atac) (18). Transcription and processing of Sm-class snRNAs have many parallels to the processing of classic mRNAs but also show some distinct differences. Sm-class RNAs are transcribed from specialized RNA polymerase II promotors which rely on sequence elements structurally comparable to TATA-box and enhancer sequences and apart from general transcription factors, binding of the pentameric snRNA activating protein complex (SNAPc) is required (19). Both, 5' capping and 3' cleavage occur co-transcriptionally, again showing similarities to mRNA processing. However, for final 3' maturation, the multiunit integrator complex that recognizes a sequence element called 3' box is required, which is located 9–19 nucleotides downstream of the snRNA-coding region (20). Integrator complex associated proteins share only little similarity to those facilitating mRNA cleavage and processing except for high sequence similarities to the cleavage and polyadenylation specificity factors CPSF73 and CPSF100, which are constitutively involved other pre-mRNA cleavage complexes (21). The newly synthesized snRNAs are then first transferred into the cytoplasm. However, since both, mRNAs and snRNAs are 5' capped, the two types of RNAs need to be distinguished by the export machinery. This is mainly hypothesized to be based on transcript length and binding of heterogenous nuclear ribonucleoproteins (hnRNPs) C1-C2. Transcripts of more than 250 nucleotides are marked by hnRNP C and associated with the nuclear RNA export factor 1–NTF2-related export 1 (NXF1–NXT1). Shorter transcripts, including snRNAs, are exported via the CRM1 pathway. Therefore, the heterodimeric cap-binding complex (CBC) binds to PHAX (phosphorylated adaptor RNA export) which, acting as adaptor proteins, link the 5´ cap and the export receptor chromosome region maintenance 1 protein (CRM1, also exportin 1) (14,22,23). Once the snRNAs reach the cytoplasm, the export machinery disassembles and the cytoplasmic snRNA processing is taken over by the survival of motor neuron (SMN) which associates the snRNAs with seven Sm-proteins which protect and stabilize the snRNAs. Various studies have shown that mutations in

the human SMN1 gene result in neuromuscular disease, in particular spinal muscular atrophy (24). SMN complex then recruits trimethylguanosine synthase 1 (TGS 1) which induces the formation of a 2,2,7 trimethylguanosine (TMG) structure at the RNAs' 5' end. This serves as nuclear localization signal while in parallel, the 3' end is cleaved by endonucleases to prepare mature RNP assembly (25). For that, an snRNP import factor, snurportin binds and interacts with the import receptor importin ß which then facilitates nuclear translocation (26). The newly imported RNPs then first accumulate in cajal bodies before being transported to nuclear speckles, which are nuclear domains in the interchromatin region that are rich in pre-mRNA splicing factors (27).
To finally conduct splicing activity, the snRNAs associated with the RNPs base pair with *cis*-acting RNA sequence elements and interact with *trans*-factors to initiate full spliceosomal assembly and splicing.

1.1.2.3 5' Splice donor sites

The 5' splice site (splice donor, SD, 5'ss) located at the 3' end of exons is the first sequence element the spliceosome interacts with. The conserved RNA sequence element base pairs with the free 5' RNA tail of the U1 snRNP that is composed of the 164nt U1 snRNA and ten proteins, in particular the seven Sm proteins and three U1-specific proteins (U1-70K, U1-A, and U1-C) (28,29). Here, U1-C is the protein proximal towards the free 5' RNA tail of the U1 snRNP and it was shown to stabilize the interaction between U1 snRNP and the splice donor sequence which promotes E complex formation and thus, the initiation of the splicing reaction (30). The highly degenerated consensus sequence CAG\GURAGUNN (R = purine, N = purine or pyrimidine, \ = exon-intron border) is composed of the last three exonic and the eight starting intronic nucleotides (Figure 2). Despite the degeneracy of the last two dinucleotides, it has been shown that positions +7 and +8 can support splicing efficiency (31-33). Full sequence complementarity to the U1 snRNA binding sequence would be given by the CAG/GTAAGTAT sequence, however, the most conserved motif in the splice donor consensus sequence is the almost invariant GT-dinucleotide at the exon/intron-border that can be found in about 99% of annotated mammalian 5'ss (34), although non-canonical 5'ss lacking the GT-motif can be recognized as well. In fact, in

a recent study it was possible to rank non-canonical splice site usage as GC > TT > AT > GA > GG > CT (35).

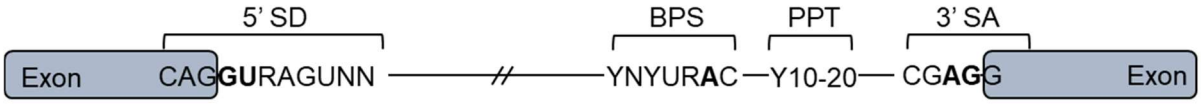


Figure 2: Key RNA sequence elements.

Two exons (grey boxes) are separated from each other by an intron. The 5' SD with the canonical GU dinucleotide, the branch point sequence (BPS), the polypyrimidine tract, and the 3' SA with the almost invariant AG dinucleotide are shown. N=any nucleotide; R=purine; Y=pyrimidine

1.1.2.4 3' Splice acceptor sites

As a counterpart of splice donors, splice acceptor sites (SA, 3'ss) are located at the 5' end of exons and interact with the early spliceosome, in particular the U2 snRNP. The splice acceptor is composed of the branch point sequence (BPS) within the intron upstream of the intron/exon border, a polypyrimidine-tract (PPT, ~15-50nt upstream of the intron/exon border), and the intronic terminal, almost invariant AG-dinucleotide at the intron/exon border which is characterized by the CAG/G (/ = intron-exon border) consensus motif (3,36,37). Non-canonical 3' splice sites are reported to have motifs containing AC- or TG-dinucleotides at the intron/exon border (38-40). A most crucial determinant of splice acceptor recognition is the polypyrimidine tract. While it is generally composed of pyrimidines in a highly degenerated sequence motif, uracils are preferred over cytosines for the constitution of strong PPTs (41). Upon initiation of the splicing process, the PTT is recognized by a 65kDa auxiliary factor of the U2 snRNP, namely U2AF65, whose binding is facilitated by two RNA recognition motifs (RRM1/RRM2) (42). In the case of the most common U2AF35-dependent 3' splice sites, a smaller subunit of U2AF, the 35kDa U2AF35 recognizes and binds the AG-dinucleotide. Here, both factors are crucial for further progression into an effective splicing reaction, in which U2AF35 and U2AF65 are present as heterodimers (43). An additional influence on splice acceptor recognition is exercised by the BPS that interacts with the U2 snRNA. The BPS has been shown to be constituted of a distinctive YNYUR**A**C motif (Y = pyrimidine, R = purine, N = any nucleotide) (14,44). While most BPS are located in relative proximity to the AG dinucleotide at the

intron/exon border, studies report about BPS located several hundred nucleotides within the upstream intron (Figure 2) (45). The BPS is recognized and bound by splicing factor 1/mammalian branch point binding protein (SF1/mBBP). Binding is often guided and supported by RS-RS domain interaction between SF1/mBBP and U2AF65 sitting on the PPT, however, independent binding of SF1/mBBP is also possible. Further progression of the early splicing reaction dissociates SF1/mBBP from the BPS and two ATPases, UAP56 and Prp5 together with SF3 promote U2snRNP binding, which is often stabilized by RS-RS domain interaction with U2AF65 (46,47).

1.1.2.5 Splicing regulatory elements and splicing regulatory proteins

The degeneracy of the splice site sequences underlines the requirement for additional regulatory support to successfully discriminate between “true” and “false” splice sites throughout the genome. *Cis*-acting RNA sequence elements that act as additional regulators are so-called splicing regulatory elements (SREs). These elements are short sequence stretches dispersed throughout the RNA that recruit splicing regulatory proteins (SRPs) (48,49). Two major protein families are involved in the recognition and binding of SREs: SR (Serine/Arginine rich) proteins and hnRNP (heterogeneous nuclear ribonucleoparticles), apart from several tissue-specific proteins.

The twelve human SR proteins are uniformly named SRSF 1 – 12 (50). They are composed of one or two amino-terminal RNA binding domains (RRM) as well as a carboxyl-terminal arginine-serine rich (RS) except for SRSF7, which contains a zinc finger motif. Here, the serine residues of the RS domain are targeted for phosphorylation, which highly influences SR protein activity. While the localization of most SR-protein family members is mainly confined to the nucleus, there is a subset including SRSF1, 3, and 7 that continuously shuttles between nucleus and cytoplasm, thereby being involved in various RNA metabolic processes, ranging from mRNA export and putative roles from nonsense-mediated decay (NMD) to translation (51-54). Additionally, a number of SR-like or SR-related proteins were identified (55).

The hnRNP protein family consists of 20 proteins, namely hnRNP A-U. In general, hnRNP proteins are diffusely localized throughout the nucleoplasm although there are exceptions. Some members of the hnRNP family exhibit rapid nucleo-cytoplasmic shuttling while other members, in particular, hnRNP C and U appear to be strictly

nuclear (56). Furthermore, their structural composition varies more widely compared to the SR protein family, but some similarities can be found. They include the RNA recognition motive (RRM) with a characteristic β1-α1-β2-β3-α2-β4 structure (57). Due to their structural diversity, hnRNPs are involved in various other cellular processes.
Apart from these ubiquitously expressed protein families, cell-type-specific splicing regulatory RBPs such as NOVA, ESRP, CELF, RBFOX, and the PTB family members are involved in the control of alternative splicing reactions and contribute to the development of cell-type-specific RNA isoforms in epithelial, muscle, and neuronal cells (58-61).

SRPs can interact directly or indirectly with the spliceosome and hence positively or negatively regulate snRNP binding and thus, the progression of the splicing reaction. Their mode of action follows a strict position dependency, hence the enhancing or repressing properties of an SRE are fully dependent on its position relative to a splice site (Figure 3) (62). Therefore, their precise prediction is a highly desirable tool in unraveling the splicing code, SREs, and SRP binding that can drastically influence splice site recognition and usage. Recently, using integrative approaches, a huge number of functional RNA elements have been added to the human catalog (63).
For SRE prediction, another layer of complexity is added by the high degeneracy of SRE motifs (64). In general, it is understood that SRPs prefer motifs of low compositional complexity (e.g., UUUUU/A for hnRNPCL1) (65). This significantly complicates the search for and prediction of functional SREs throughout the genome. Experimental approaches to unravel SRE motifs have been developed and progressed, they reach from the use of single splicing reporters to massive parallel splicing assays and lately, large-scale RNA sequencing approaches.

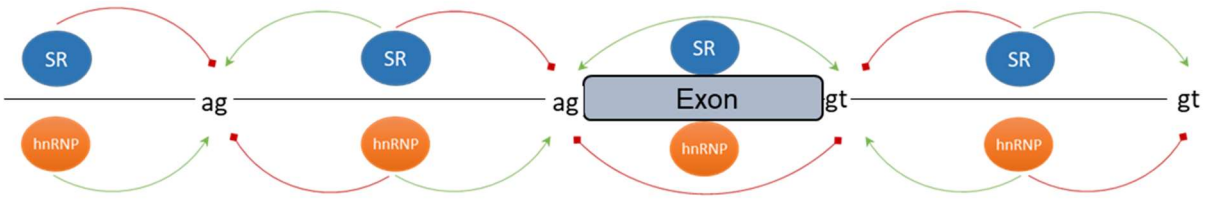


Figure 3: Position-dependent action of SREs.

Splicing regulatory elements offer binding sites for splicing regulatory proteins. Two protein families, SR proteins and hnRNPs are main binding partners. Their mode of action is dependent on their relative position towards the surrounding splice site (After Ptok and Müller 2019 (48), amended).

1.1.3 Splicing reaction

Overall, the sequential splicing process that is carried out by the spliceosome is characterized by two transesterification reactions and involves RNA–RNA, RNA–protein as well as protein-protein interactions. The first transesterification reaction involves the 2’ hydroxyl group of the branch point adenosine in the intron attacking the phosphodiester bond at the 5’ss which results in a 5’ exon intermediate structure and a lariat intermediate structure. During the second transesterification reaction, the phosphodiester bond at the 3’ss is attacked by the 3’ hydroxyl group of the 5’ exon, resulting in the replacement of the bond and the subsequent ligation of the exons (Figure 4) (13).

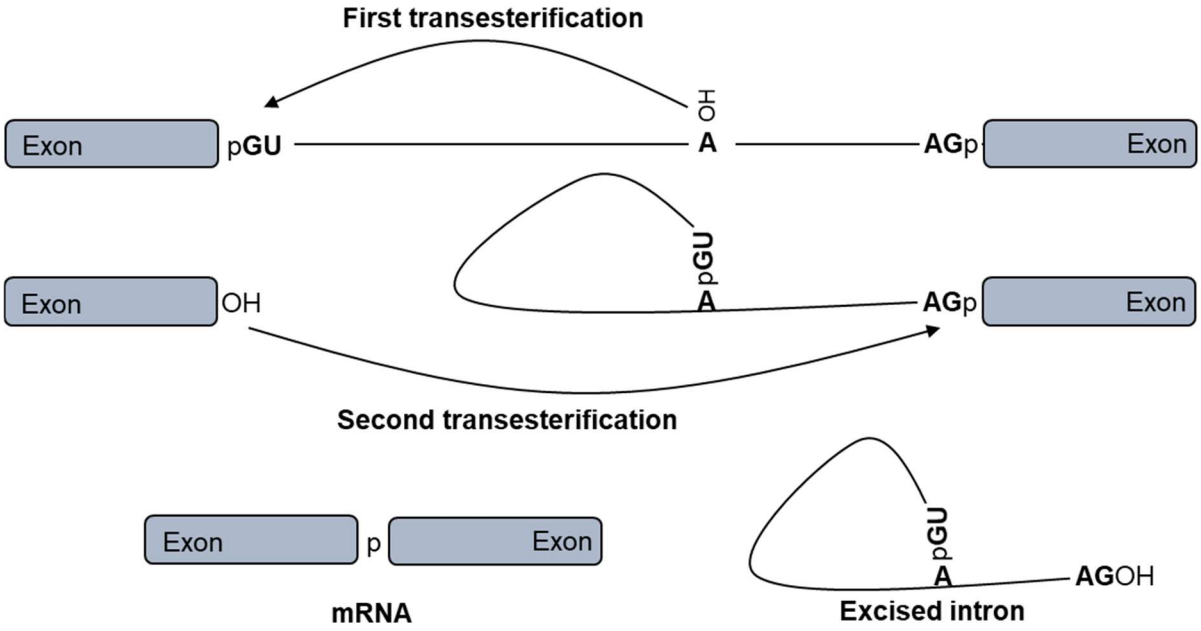


Figure 4: The splicing reaction involves two transesterification steps.

First, the 2’ hydroxyl group of the BPS-adenosine attacks the phosphodiester bond that links the 5’ exon and the intron, which generates a 5’ exon and a lariat intermediate structure. This is followed by the phosphodiester bond at the 3’ss being attacked by the hydroxyl group of the 5’ exon, which leads to a replacement of the bond and the generation of a lariat intron and the ligated exons.

Upon recognition of the 5’ss and the subsequent ATP-independent binding of the U1 snRNP, the Early (E) complex is formed. It is shortly followed by the recognition and the ATP-dependent binding of U2 snRNP to the BPS, catalyzed by the DExD/H helicases Prp5 and Sub2. As a result of U2 snRNP binding to the BPS and the interaction with the U1 snRNP binding the 5’ss, the branch point adenosine is bulged

out of the duplex and serves as a nucleophile for the first transesterification (66-68). This newly formed pre-spliceosomal A complex then recruits the pre-assembled tri-snRNPs U4/U6*U5 to form the still splicing inactive B complex. This is carried out by Prp28, another DExD/H helicase. Also involved at this stage is the NineTeen complex (Prp19, NTC) that stabilizes U6 snRNP and U5 snRNP binding to the pre-mRNA. Upon various ATP-dependent structural and compositional rearrangements of both RNA-RNA and RNA-protein interactions, U1 and U4 snRNPs are released which renders the complex into the catalytic active B* complex (69,70). Additionally, the 5' end of the U6 snRNA base pairs simultaneously with the 5'ss and interacts with the U2 snRNP, bringing the 5'ss in close proximity to the adenosine at the BPS which induces the first transesterification step. Again, several helicases, in particular Brr2, Snu114 and Prp2 are involved. The first catalytic step converts B* into the C complex, a process in which Prp8, Prp16, and Slu7 are involved (71). This generates a lariat intermediate structure where the first nucleotide of the 5'ss is attached to the branch point adenosine. Finally, in the post-spliceosomal P complex, the mRNA is released, the spliceosome dissociates and the components of the spliceosome are recycled for a further splicing reaction (Figure 5) (72). Additionally, upon the release of the mature mRNA, the Exon-Exon Junction Complex is recruited roughly 20nt upstream of the exon-exon junction. It is involved in mRNA export, translation, and quality control (73).

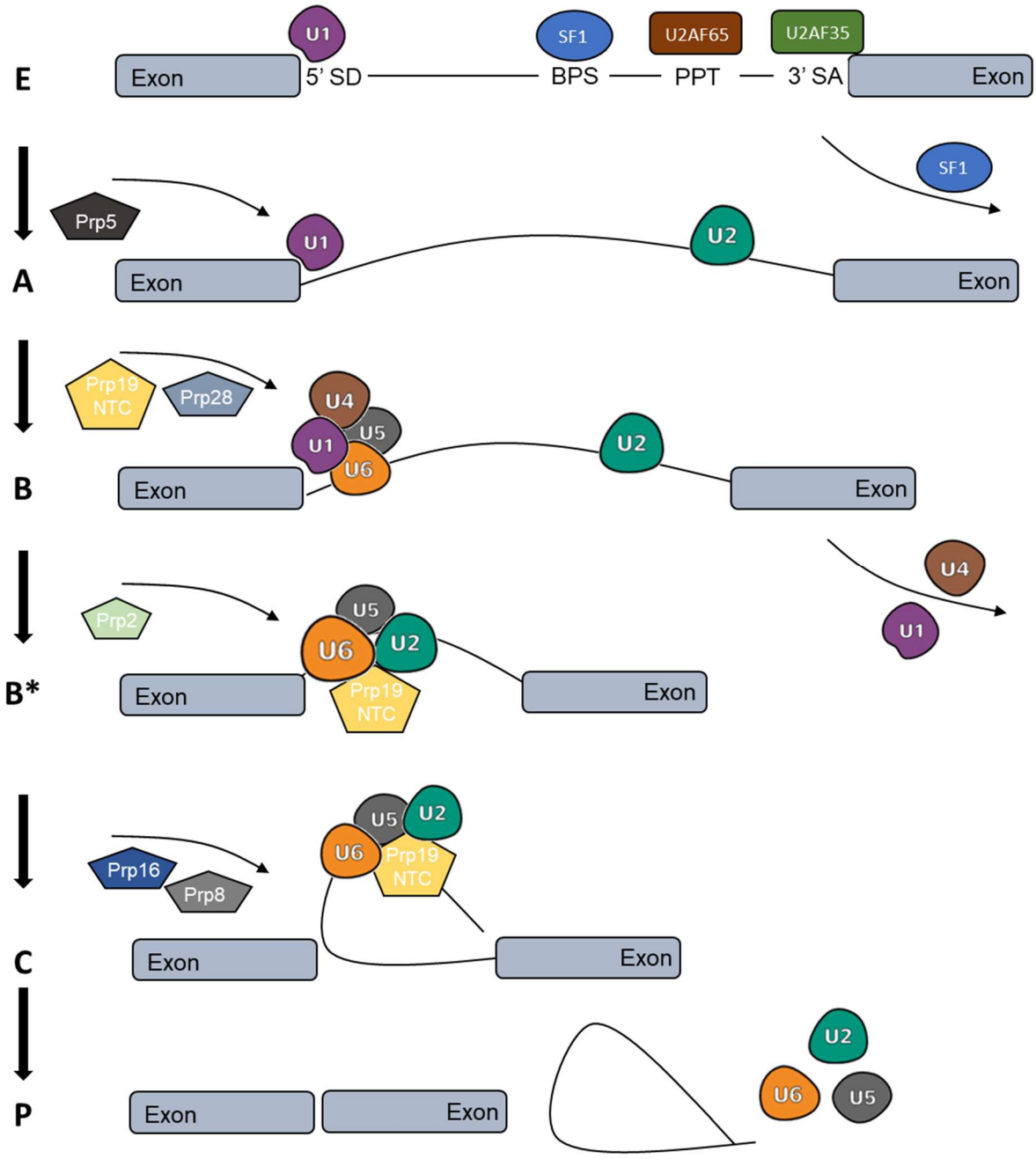


Figure 5: The splicing reaction.

*The stepwise spliceosomal assembly is initiated by the U1 snRNP interacting with the 5'ss and SF1, U2AF65, U2AF35 binding to the splice acceptor, which forms the E-complex. SF1 is displaced by ATP-dependent binding of the U2 snRNP which generates the A complex. Then, the pre-assembled tri-snRNP U4/U6*U5 binds, forming the B complex which is catalytically inactive. The active B* complex is formed through conformational changes and binding of the NineTeen complex (NTC) and other helicases. U1 snRNP and U4 snRNP are then released from the spliceosomal complex. The first transesterification reaction leads to an exon intermediate and lariat intron intermediate complex C. In the post-spliceosomal complex P that is formed after the second transesterification, the components dissociate while the exons are ligated and the intron is degraded (After Matera and Wang 2014 ((14).*

Despite the constitutive splicing process in which every intronic sequence is removed from the pre-mRNA, there is also an array of alternative splicing events (74). Multiple splice-isoforms derived from one primary transcript may lead to a broad variety of proteins that can be expressed by a single gene. More recent findings from large-scale mass spectrometry-based proteomics analyses, however, suggest that most human genes have only a single main protein isoform (8,40). Forms of alternative splicing include i) exon skipping, hence the exclusion of complete exons from the final transcript, ii) the choice of an alternative 3'ss or 5'ss resulting in a varied exon length, iii) the inclusion and exclusion of mutually exclusive exons or iv) the failing removal of (partial) intronic sequences (intron retention) (Figure 6) (75).

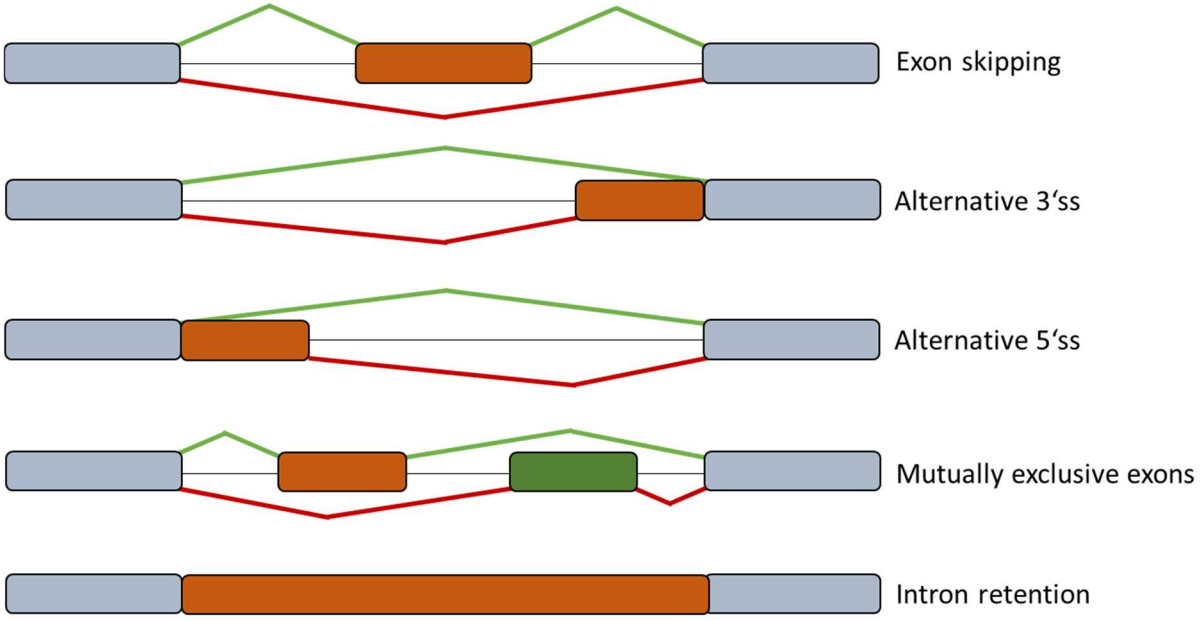


Figure 6: Modes of alternative splicing.

Apart from constitutive splicing where all sequential exons are included, alternative splicing leads to differential inclusion/exclusion of exons, the use of alternative splice sites, or the retention of introns.

1.1.4 Bioinformatic prediction of the splicing outcome

The evaluation of mutations affecting the splicing outcome is a desirable and essentially needed tool in diagnostics which can ultimately aid in designing more specific treatments for patients. Several publicly available algorithms are available and have been shown to quite robustly perform such predictions. One of them is the HEXplorer tool (https://www2.hhu.de/rna/html/hexplorer_score.php) (Erkelenz et al.,

2014). It is a bioinformatical approach that aims to analyze the occurrence of potential SRE motifs and that can also be used to indicate potential effects of mutations in the vicinity of splice sites or within SRE motifs. It is based on a RESCUE (Relative Enhancer and Silencer Classification by Unanimous Enrichment) concept (Fairbrother et al., 2002) and was established calculating the hexamer frequencies in datasets of 100nt long sequences up- or downstream of weak or strong 5'ss. From that, two Z-scores per hexamer were derived according to their location (Z_{EI} = exonic or intronic) towards respective splice sites (Z_{WS} = weak or strong). According to the position dependency of SREs, the calculation of Z_{EI} was considered more discriminative and was therefore used to calculate the HZ_{EI} score, which is the sum of Z_{EI} scores of overlapping hexamers of a respective index nucleotide (Erkelenz et al., 2014). Analysis output is a HEXplorer plot that reflects a potential binding landscape for SRPs. Furthermore, consequences of sequence changes by mutational alterations can be visualized.

While algorithms such as the HEXplorer or ESR-seq (76) allow the calculation of SRE profiles and thus, splicing properties in a sliding window approach, other algorithms such as ESEfinder, FAS-ESS, RESCUE-ESE, PESX are designed to identify previously described motifs as sequence blocks (77). Recent developments also include neural network or deep-learning-based approaches (MMSplice (78), SpliceAI (79)).

Apart from the evaluation of splicing regulatory properties of sequences, there are algorithms available to provide a calculation of the intrinsic strength of splice sites, hence their likelihood to pose as functional splice sites. For 5'ss, one experimentally based algorithm is the HBond score (HBS) algorithm. It evaluates the intrinsic strength by calculating the complementarity, hence the H-bonds that can be formed between the eleven nucleotide long splice donor and the free 5' end of the U1 snRNA (https://www2.hhu.de/rna// (32)).

Another commonly used is the MaxEnt score (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html, which is rather a statics-based approach. It considers the frequency of distribution of nucleotides at specific positions within 9 nucleotides of a splice donor. Accordingly, the strength of 3'

splice acceptor sites can be calculated similarly with the same algorithm but taking 20 intronic and 3 exonic nucleotides into account (80).

1.1.5 Splicing in disease

With up to 95% of multi-exonic genes being alternatively spliced, its role in cell homeostasis is evident (81). More than 200 human pathogenic diseases with underlying splicing aberrations are already identified, however, with the increasing use of genome sequencing in precision medicine, the number is likely to rise (82,83). Splicing alterations associated with diseases are primarily caused by nucleotide exchanges in *cis*-acting core splicing sequences such as the splice sites themselves or associated sequences, hence splicing regulatory elements that, via protein binding partners, interact directly with the spliceosome (84). Such nucleotide exchanges impairing splice site recognition causing activation of aberrant or cryptic splice sites or exon skipping can have different consequences (85). Changes that ultimately result in a frameshift in the coding sequence can lead to the introduction of premature stop codons which can cause the transcripts to be cleared via the nonsense-mediated decay pathway. Ultimately this can cause a defect in proteostasis which adds to diseases (86,87)

Apart from nucleotide exchanges in the RNA sequence to be spliced, aberrations in the spliceosome itself can have devastating effects and usually are not compatible with life. However, not all spliceosomal mutations are associated with a full loss of function but are still associated with human pathogenic diseases. A prominent example to this end is retinitis pigmentosa. While many tissue-specific genes are involved in disease progression, changes in core snRNP proteins, in particular, the pre-mRNA processing factors (Prpf) 3, 4, 6, 8, and 31 are linked to the disease (88). Additionally, several cancers such as myelodysplastic syndromes (MDS) or chronic myelomonocytic leukemia (CMML) can be attributed to such mutations (89).

1.2 RNA viruses

RNA as a central molecule of biological processes is also exploited by RNA viruses to store and process its genetic information. Here, viruses share the common hurdle of a limited genome size while they have to express their full transcriptome for successful replication. To this end, viruses share strategies that allow the expression of various mRNAs from a single RNA genome, which is reached by either pre-mRNA splicing or unconventional transcription. Additional strategies include but are not limited to transcriptional slippage in the form of polymerase frameshifting or translational slippage occurring as programmed ribosomal frameshifts to further increase the coding potential despite cap-dependent restriction of the translational mechanism.
Based on the general mechanism of viral RNA storage and mRNA production, the Baltimore classification places viruses into seven groups (90):

I: dsDNA viruses (e.g. Adenoviruses, Polyomaviruses)
II: (+)ssDNA viruses (e.g. Parvoviruses)
III: dsRNA viruses (e.g. Reoviruses)
IV: (+)ssRNA viruses (e.g. Coronaviruses, Picornaviruses, Flaviviruses)
V: (−)ssRNA viruses (e.g. Orthomyxoviruses, Rhabdoviruses, Filoviruses)
VI: (+)ssRNA-RT viruses with DNA intermediate in life-cycle (e.g. Retroviruses)
VII: dsDNA-RT with RNA intermediate in life-cycle (e.g. Hepadnaviruses)

The majority of the groups, in particular groups IV-VII, exploit RNA as at least an intermediate form for the storage of genetic information to generate mRNA for protein expression. Well-known representatives of RNA viruses are Hepatitis C Virus, Ebolavirus, SARS-CoV and SARS-CoV-2, Influenza virus as well as the retroviruses human T-cell lymphotropic virus type 1 (HTLV-1) and the human immunodeficiency virus (HIV) (91). Generally, RNA viruses require an RNA-dependent RNA polymerase (RdRp) to replicate their genome. Apart from the mRNA that is required for protein expression, RdRp generates copies of the viral genome in a replicase complex, an association of the RdRp with other factors (92). This allows most RNA viruses to replicate in the hosts' cytoplasm. Retroviruses, on the other hand, carry an RNA-dependent DNA polymerase (RdDp), the reverse transcriptase to reverse transcribe

the two copies of their single-strand RNA genomes into DNA, which integrates into the host' genome before it is transcribed into mRNA (93).

1.2.1 HIV-1

The human immunodeficiency virus (HIV) that belongs to the family of *Retroviridae* was first described in 1981. It was identified as the causative agent of the acquired immunodeficiency syndrome (AIDS) which is characterized by the loss of CD4+ lymphocytes which results in severe immunological failure (94,95). Two taxa are known, HIV-1 and HIV-2, with HIV-1 being the predominant variant in the western world while HIV-2 is primarily present in West Africa (96). HIV-1 can be grouped into four main classes, the major (M) group, the outliers (O) group, the non-M, non-O group (N), and P (alphabetical order, (97)). The major (M) group can then be subdivided into ten subgroups A, B, C, D, E, F, G, H, J and K, according to their global distribution. A model system for HIV research, however, is subgroup B despite not being the most dominant subgroup, but which is rather predominantly distributed in the western world.

1.2.1.1 HIV-1 structure and life cycle

The enveloped HIV-1 virions are composed of two copies of a 9.7kb plus stranded RNA genome within an internal nucleocapsid. Protein structures are derived from their respective polyprotein precursors which are encoded by the genes *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope glycoprotein). Additionally, HIV carries regulatory proteins that are encoded by *rev* (regulator of virion expression), *vpu* (viral protein u), *tat* (trans-activator of transcription), *nef* (negative regulatory factor), *vpr* (viral protein r), and *vif* (viral infectivity factor) (98,99). The viral envelope is composed of a double lipid layer and a matrix where glycoprotein gp120 is present on the surface and gp41 is anchored in the membrane, both encoded by *env* (100). The homo-dimeric genome that is associated with nucleocapsid (p7) is embedded in the capsid structure (p24) together with the enzymes Reverse Transcriptase (RT), Protease (PR), and Integrase (IN) which are encoded by *pol* (101).

The HIV-1 replication cycle is initiated by the binding of the Env glycoprotein to its receptor embedded in the surface of host cells, in particular mononuclear CD4+ T-

cells, macrophages, and dendritic cells. In particular, HIV-1 mediates cellular entry via membrane fusion that requires cleavage of the inactive precursor gp160 into gp120 and gp41 (102). Furthermore, depending on the virus' tropism, the co-receptors CXCR4 (X4 or T-tropic) or CCR5 (R5 or M-tropic) are needed, however, some subtypes such as D, are most frequently dualtropic (103). M-tropic strains are characteristic for early infection; with disease progression, a tropism shift occurs which results in the emerging of T-tropic viruses in the later stage of the disease. This is also associated with more rapid disease progression (104). This is followed by glycoprotein-mediated membrane fusion and allows the capsid, containing the nucleocapsid covered RNA genome, the Reverse Transcriptase (RT), the Integrase (IN), and the Viral Protein R (Vpr) to enter the cells' cytoplasm. Recently it has been found that intact HIV-1 capsids can also be imported into the nucleus without previous unpacking (105). Either way, the viruses' RNA is then reverse transcribed, and the double-stranded DNA is transported into the nucleus via the pre-integration complex (PIC) or directly integrated into the hosts' DNA (106). Integration is carried out by the 32kDA Integrase that is encoded by the *pol* gene. It consists of three domains, the N-terminal domain (NTD) that stretches from amino acid 1 to 50, the Catalytic Core Domain (CCD, AS 50-212) and the C-terminal Domain (CTD AS 212-288) (107). First, IN binds the two LTR ends of the cDNA via its catalytic domain and removes two nucleotides which creates free 3' OH ends. These interact in particular with short palindromic CA-sequences within the host genome that is run through the integrase. Upon integration, the overhanging 5' ends are repaired by cellular ligases (108). The cellular transcription machinery, in particular Pol II, recognizes the promotor within the viral 5' LTR. However, for efficient transcription, Tat binding to the TAR (transactivation-responsive region) element located at the 5' end of HIV-1 transcripts is required. Tat interacts with the positive transcription elongation factor b (P-TEFb) which promotes elongation (109). This is followed by extensive alternative splicing before the transcripts undergo translation to allow for the expression of viral proteins required for the formation of new virions (Figure 7).

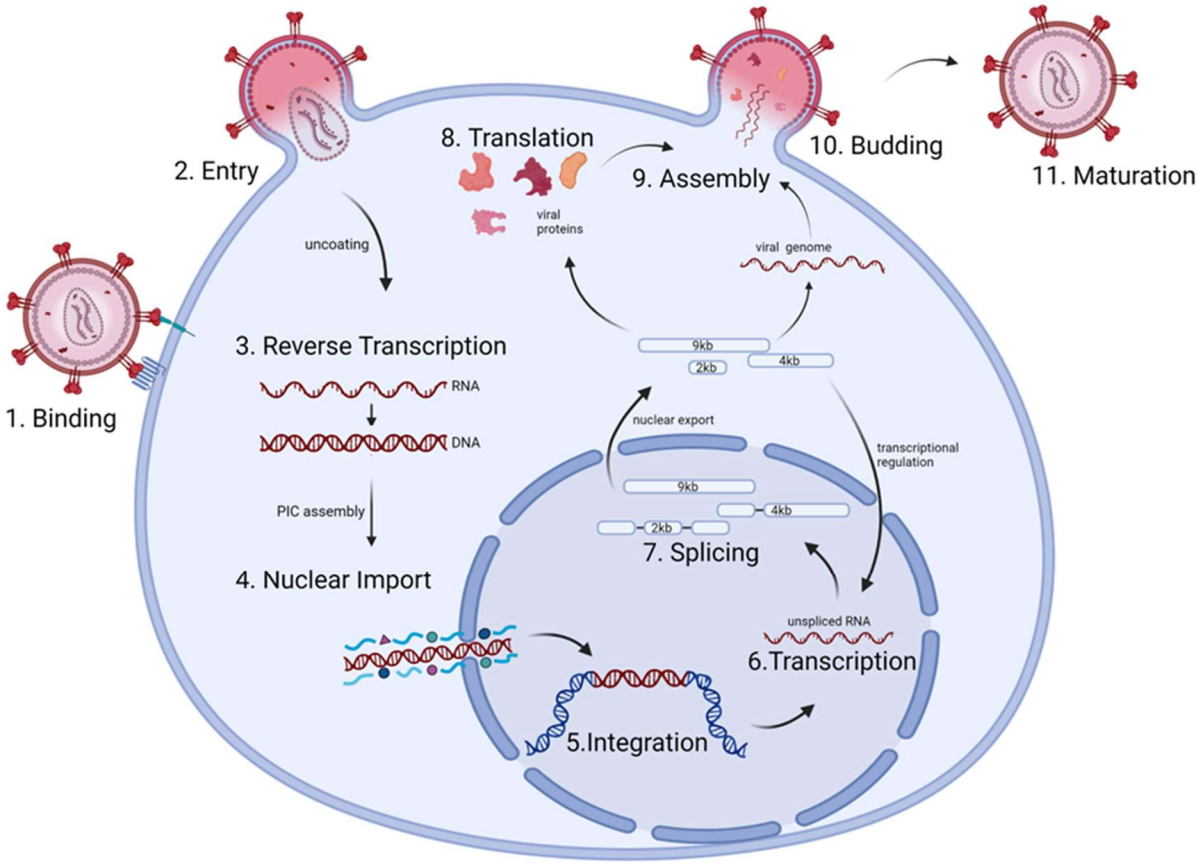


Figure 7: HIV-1 Life cycle.

Binding to the CD4-receptor and the respective co-receptor mediates entry into the host cell. There, the viral RNA is released and reverse transcribed either before (shown here) or after being transported into the nucleus (not shown here for clarity). The viral DNA is integrated into the hosts' genome where the DNA is transcribed into the viral pre-mRNA. Extensive alternative splicing leads to the generation of various mRNA species of the intronless 2kb, intron-containing 4kb, and unspliced 9kb class. Transcripts are translated into structural and enzymatic proteins while copies of the 9kb transcript also serve as genomic RNA, which is incorporated into budding virions. After the release of the viral particles, they mature into new infectious virus particles (After Pasternak,Lukashov and Berkhout 2013 (110), Figure prepared using Biorender.com).

1.2.1.2 HIV-1 mRNA processing

HIV-1 RNA is transcribed as a single 9 kb transcript from the U3 promotor after integration of the proviral DNA into the host genome. Since translation is generally started at the most CAP proximal start codon, the Gag/Pol open reading frame is in the position to be the one recognized. In order to fully express the viral genome, HIV-1 pre-mRNA therefore undergoes extensive alternative splicing to position downstream open reading frames closer to the 5' end of the mRNA, which leads to the generation of more than 50 distinct mRNAs (111-114). Exemptions are the programmed −1 ribosomal frameshift (−1 PRF) between the *gag* and *pol* open reading frames (ORFs)

and the bi-cistronic *vpu*/*env* transcript that is generated by discontinuous ribosome scanning (115,116). For alternative splicing, transcripts make use of the four major splice donors and eight splice acceptors (112). Their combinatorial usage allows for the generation of three different mRNA classes, the intronless 2kb, intron-containing 4kb and unspliced 9kb class (Figure 8). A minor class, the 1 kb class, has also recently been described after extensive next generation sequencing of viral transcripts (117). First, the small 2kb class transcripts including Tat are generated by extensive splicing. Tat and Rev then shuttle back to the nucleus to aid in transcription and, in case of Rev, to bind the RRE (Rev responsive element) that supports the nuclear export of the unspliced, hence intron containing 9kb class RNA via the CRM1 (exportin-1) pathway. This is then either used for translation or as genomic RNA. Furthermore, 4kb class transcripts are generated (112,118).

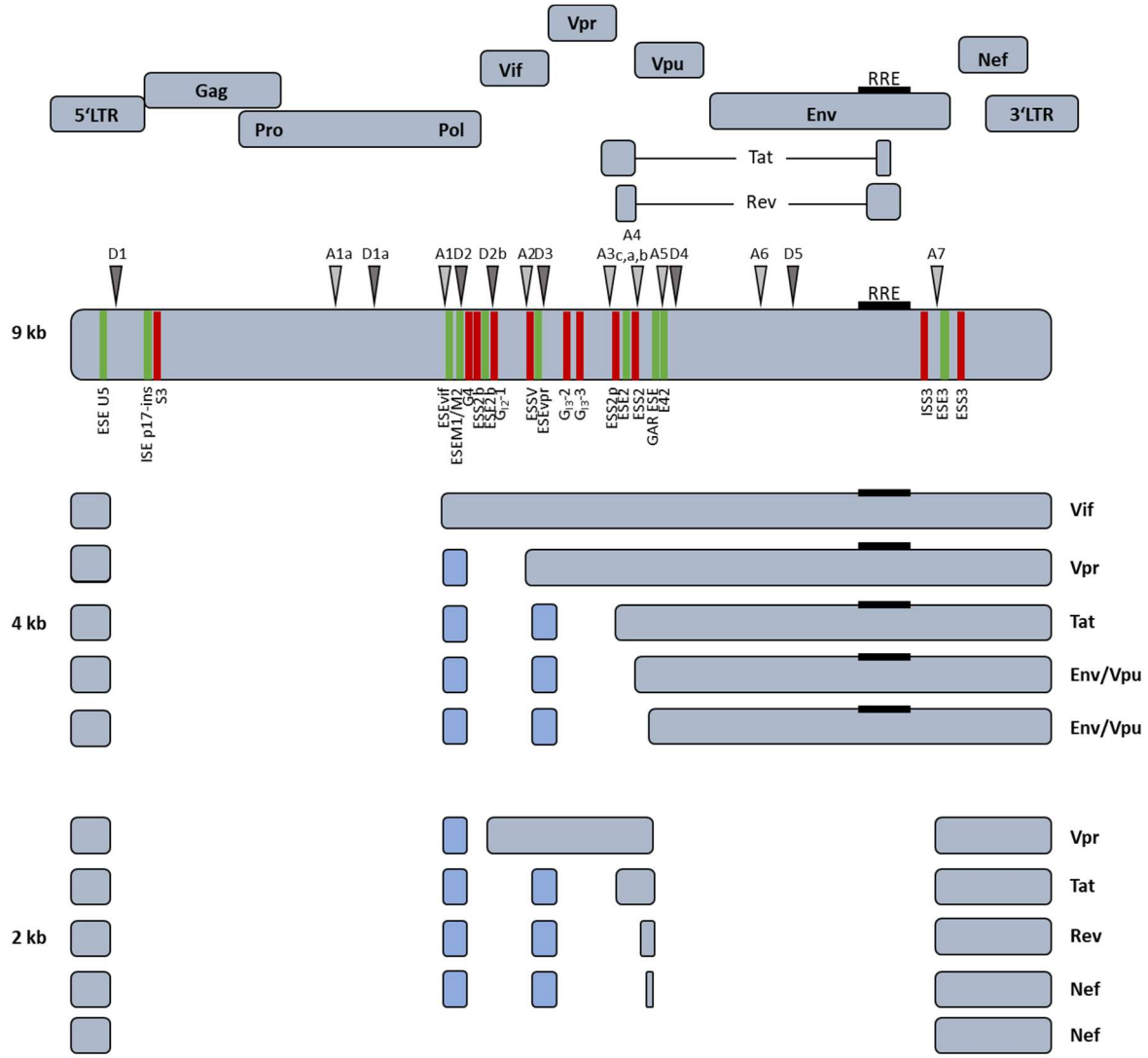


Figure 8: HIV-1 genome organization.

The HIV-1 genome contains eight open reading frames including Gag, Gag/Pol, Env, Vif, Vpr, Vpu, Tat, Rev, and Nef. The Rev Responsive Element (RRE) is localized within the Env ORF. The unspliced 9kb message is depicted with the position of all 5'ss and 3'ss (some are specific for certain HIV strains) as well as known SREs. All 4kb and 2kb mRNAs include the non-coding exon 1 while non-coding exons 2 and 3 are alternatively spliced. The 2kb mRNAs are additionally spliced at D4 and A7 (After Sertznig et. al 2018).

HIV-1 splice sites share their consensus sequence with human splice sites. The intrinsic strength of the viral splice sites, except for the mainly used D1 and D4 as well as A2 and A3, however, are generally low (119,120). The high degeneracy of these other splice sites allows for inefficient splicing reactions that are necessary for the generation of intron-containing transcripts. Furthermore, viral splice sites, similar to eukaryotic splice sites, highly rely on surrounding splicing regulatory elements for their regulation. The HIV-1 genome is interspersed with SREs and several studies have

shown that mutations in these SREs influence viral gene expression and thus, replication (121). Therefore, experimental approaches exploit viral splicing regulation as a potential new target for drug therapy (122).

1.2.1.3 HIV-1 antivirals and antiviral approaches

While there is still no sterile cure available, combinatory HAART therapy (highly active antiretroviral therapy) has drastically improved in the past years. The emergence of antiviral drugs for the treatment of HIV, in particular, the first drug Zidovudine/Retrovir (Azidothymidine, ATZ (123)) that was FDA approved in 1987 and all following developments highly aid in prolonging and improving patients' lives by reducing the viral burden which helps the immune system to recover and also prevents opportunistic infections. Ultimately, the development from mono- to dual therapy to the combinatorial use of HAART reduced HIV-associated morbidity and mortality in the past years (124-126). Unfortunately, secure access to medication and disease monitoring is still limited, especially in developing countries (127). Additionally, the increasing number of people receiving HAART and the prolonged duration of treatment, as well as non-adherence, contribute to the emergence of new drug resistances, especially since HIV shows a high genetic variability (128,129). The underlying cause for this variability is the viral reverse transcriptase which lacks proofreading activity and is estimated to have an error rate of approximately 2×10^{-5} per base per round of copying (130). Recent reports show drug resistances in therapy-naïve patients, underscoring not only the necessity for end-to-end clinical surveillance and resistance profiling of infected patients but also the importance of continued research into new drug targets (131,132).

The major drug classes currently target different essential viral enzymes. Available are in particular **reverse transcriptase inhibitors** (Nucleoside/nucleotide reverse transcriptase inhibitors NRTIs: abacavir, emtricitabine, lamivudine, tenofovir, zidovudine and Non-nucleoside reverse transcriptase inhibitors NNRTIs: doravirine, efavirenz, etravirine, nevirapine, rilpivirine), **protease inhibitors** (PIs: atazanavir, darunavir, fosamprenavir, lopinavir, ritonavir, tipranavir), **integrase inhibitors** (INSTIs: bictegravir, dolutegravir, elvitegravir, raltegravir) as well as cell **entry inhibitors** (including fusion inhibitors, post-attachment inhibitors, and CCR5 antagonists:

enfuvirtide, ibalizumab-uiyk, maraviroc). Furthermore, there are also minor drug targets that are recently under clinical evaluation (133,134). These include latent provirus reactivation drugs like the cancer treatment Vorinostat that was shown to reactivate latent viral reservoirs and in combination with other antivirals act in a “shock and kill” manner (135). Other drugs exploit the inhibition of integrated provirus transcription via Tat/TAR/P-TEFb complex (136) or, in the case of the designer recombinase (Brec1), aim to excises the HIV-1 provirus from the host cell genome (137).

1.2.2 SARS-CoV-2

In December 2019, a novel coronavirus emerged in the city of Wuhan, China (138). It was later named SARS-CoV-2 for its genetic and structural resemblance to another member of betacoronavirus family, SARS-CoV, which caused a pandemic of the 20th century. In 2002/2003 SARS-CoV spread to 26 countries, lead to roughly 8.000 infections, and had a fatality rate of 9.7%. This first pandemic of the new century was followed by an outbreak of MERS, the Middle Eastern respiratory syndrome coronavirus in 2012. With about 2500 reported cases it was not classified as a pandemic, however, the fatality rate of 34% posed a serious threat (139). Additionally, there are four other human pathogenic coronaviruses, HCoV-NL63 (alphacoronavirus), HCoV-229E (alphacoronavirus), HCoV-OC43 (betacoronavirus), and HCoV-HKU1 (betacoronavirus) that are considered to cause seasonal virus infections with mild clinical symptoms (140).

Although the novel corona virus SARS-CoV-2 and its associated COVID-19 disease are less lethal, the transmission rate is significantly higher compared to MERS-CoV or SARS-CoV. It is impossible to report current infection and death rates because the pandemic is still far from under control more than one year after the first report. With infections occurring worldwide, global research has shifted to quickly elucidate viral features and to rapidly develop therapies and prophylactic vaccinations.

1.2.2.1 SARS-CoV-2 structure and life cycle

The supramolecular structure of the positive-strand RNA virus that is roughly 125nm large in diameter is a pleomorphic shape with the extraordinarily large 30kb viral genome being embedded in a ribonucleoparticle complex (RNP). It encodes at least 29 proteins including four structural proteins (141). The nucleocapsid (N) is surrounded by a lipid envelope derived from the host and stubbed by a number of other proteins. One of the most remarkable structural proteins is the roughly 600 kDa large trimeric spike protein (S) that is critical for viral entry and is also exploited as a target for inhibitory drugs. It is split into two subunits, S1 and S2. The receptor-binding domain located in the S1 subunit is the direct interaction partner of the cellular angiotensin-converting enzyme 2 (ACE2) (142). Furthermore, the spike protein contains a furin-like cleavage site (PRRARS'V) between the S1 and S2 subunits which is found to contribute to rapid human-to-human transmission (141,143,144). The virus also carries a transmembrane matrix protein (M), whose tails are in contact with the RNP to link the genome to the envelope. Furthermore, the last structural protein is the minor envelope protein (E). Non-structural proteins include the papain-like protease nsp3, the main chymotrypsin-like protease nsp5 (or $3CL^{pro}$ or M^{pro}), the RdRp nsp12 and the helicase nsp13 (145) (Figure 9).

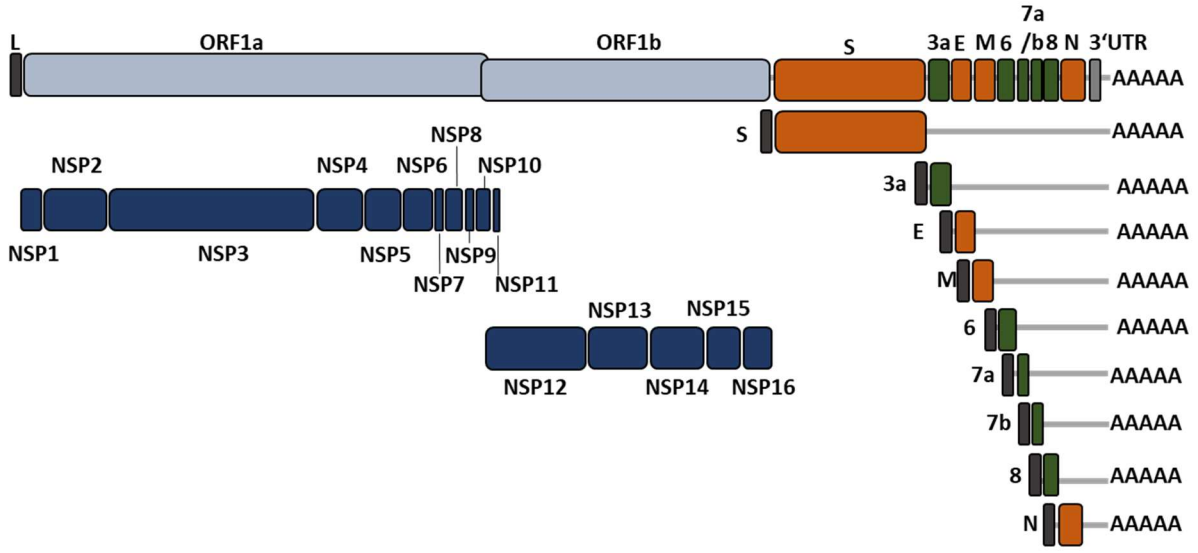


Figure 9: SARS-CoV-2 genome organization.

The major open reading frames ORF1a and ORF1b are translated into pp1a and pp1b that are cleaved into non-structural proteins (nsp 1-16, blue boxes). Nine subgenomic RNAs are produced that are translated into structural proteins (orange) and accessory proteins (green). The leader sequence is indicated by a black box. Modified after Kim et. al. 2020 (146) and Gordon et. al 2020 (147).

Viral entry into host cells is mediated via the trimeric spike glycoprotein on the virus' surface and the interaction with ACE2. The cellular serine protease TMPRSS2 facilitates proteolytic cleavage of the spike protein between the S1 and S2 subunit to initiate membrane fusion, however, cathepsin B or cathepsin L have also been shown to be involved in this process (148). Here, both, HIV-1 and SARS-CoV-2, rely on similar glycoprotein-mediated entry mechanisms (149). After the release into the hosts' cytoplasm, the positive-strand viral RNA is directed towards ribosomes for immediate translation of the two largest open reading frames, ORF1a and ORF1b to generate the polyproteins pp1a and pp1b (150). Translation of ORF1b requires a programmed −1 ribosomal frameshifting (-1 PRF) which is achieved by a slippery site and a pseudoknot secondary structure (151). Proteolytic cleavage is then carried out by the cysteine proteases nsp3 and nsp5. This generates the non-structural proteins, with nsp12–16 being involved in the main enzymatic functions such as RNA synthesis, RNA proofreading, and RNA modification. Nsp2–11 are furthermore required for the formation of the viral replication and transcription complex (RTC) and the interaction with the host cellular machinery (152). The structural and accessory proteins are generated via a discontinuous transcription process from nested subgenomic (g) RNAs which all terminate with the 3′ end of the full-length gRNA (146). The structural proteins are translated by ER-bound ribosomes and transit through the ER-to-Golgi intermediate compartment (ERGIC) before being transported towards the cell surface for assembly. An exception are the nucleocapsid proteins, which pre-assemble with the viral RNA in the cytoplasm to form the nucleocapsid. Release of the newly formed virions then takes place via exocytosis (152,153) (Figure 10).

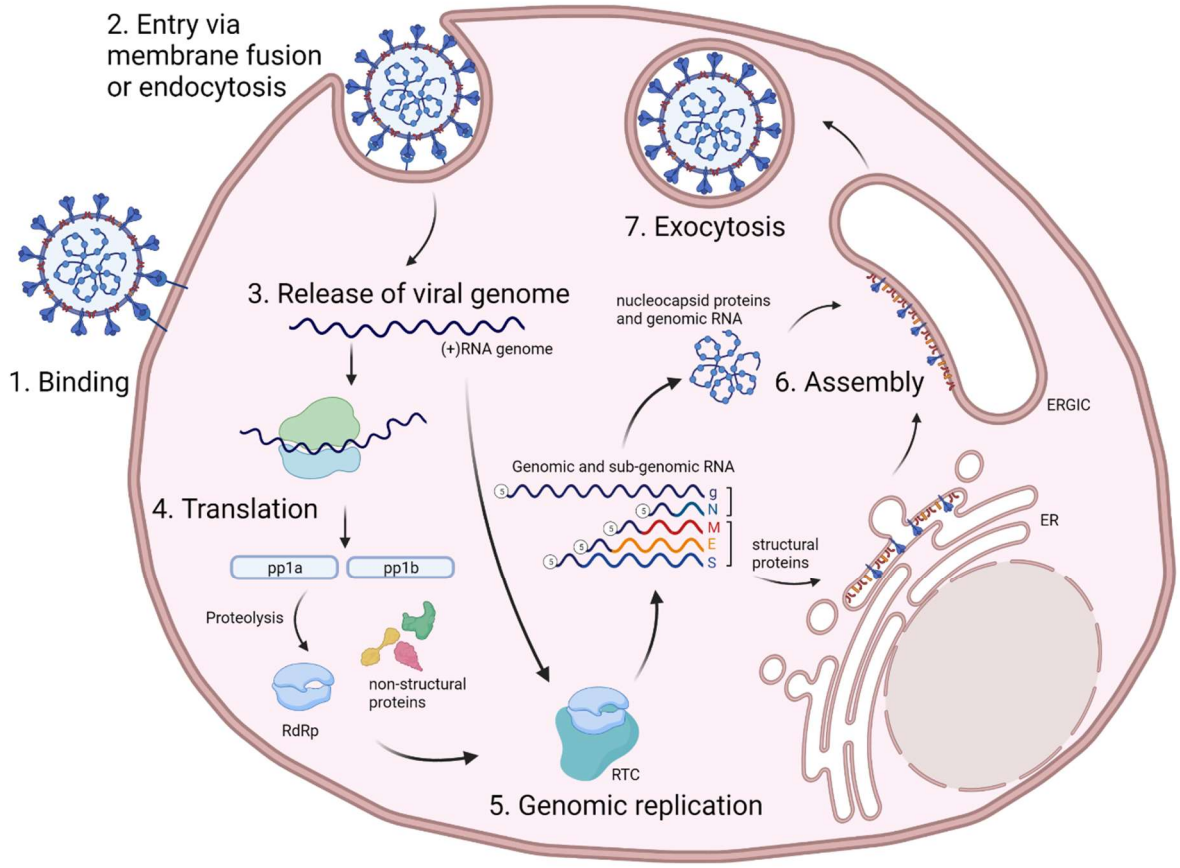


Figure 10: SARS-CoV-2 life cycle.

SARS-CoV-2 uses the angiotensin-converting enzyme 2 (ACE2) as host receptor. After binding, the spike glycoprotein is cleaved into S1 and S2 subunits by either TMPRSS2 or a furin protease. Then the viral particle is fused with the cell membrane or endocytosed. The viral genome is released into the cytoplasm and components of the replication and transcription complex (RTC) are translated from the genomic RNA and generate transcripts for the viral structural proteins. These are transported to the ER for translation and then moved to the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Nucleocapsid proteins assemble with the genomic RNA in the cytoplasm. Virions are assembled and then bud from the ERGIC before they are exocytosed (After Harriso, Lin and Wang 2020 (154), Figure prepared using Biorender.com).

1.2.2.2 COVID-19 treatment and prophylaxis

While there is still no singular specific antiviral cure for COVID-19, many promising treatment options were tested from the early beginning of the pandemic. Additionally, prophylactic vaccine developments were rapidly started which ultimately led to the emergency authority approval of several vaccine candidates in December 2020, one year after the first reports of the disease.

Early in the pandemic, several on-the-market compounds were tested for their activity against SARS-CoV-2 infections. These included the use of antivirals such as the combination HIV-1 protease inhibitors Lopinavir and Ritonavir (Kaletra). While the

combination of the drugs previously was shown to be effective against SARS-CoV-1 in tissue culture (155), it failed randomized clinical trials for SARS-CoV-2 infected patients (156). The antimalarial drug Hydroxychloroquine/Chloroquine was also under evaluation in clinical trials after it received approval for COVID-19 patients, however, studies saw no benefit of treatment (157). Also, medication unrelated to viral illnesses were exploited as a potential treatment for SARS-CoV-2 infections. This includes the use of camostat mesylate, a protease inhibitor drug approved for the treatment of pancreatitis that has shown to inhibit the protease TMPRSS2 in cell culture and that is under investigation in several clinical trials (158).

Remdesivir, a nucleoside analog that inhibits the viral RNA-dependent RNA polymerase and was initially designed to treat Ebola virus and Marburg virus infections showed lower respiratory tract infection and shorter time to recovery in a randomized trial (159). Also, Dexamethasone, a corticosteroid with immunosuppressive and anti-inflammatory features has been shown to be effective in critically ill patients, especially in those with a high likelihood to develop a systemic inflammatory response that might lead to multiple organ dysfunction syndrome (160). Current patient care guidelines also suggest benefits of the combinatory use of Dexamethasone and Remdesivir in hospitalized patients that require oxygen.

Another classic therapeutic approach used in the pandemic is the administration of convalescent plasma from recovered COVID-19 patients that contains neutralizing antibodies. While blood products certainly bear risks, the FDA argued that the benefits outweigh them. Due to the elaborate preparation, this treatment option is primarily recommended for severely ill patients as well as immunocompromised individuals (161).

Based on this concept, neutralizing monoclonal antibodies were rapidly developed and received authority approval. These antibodies such as Bamlanivimab or Etesevimab or combinatory formulations such as Casirivimab and Imdevimab (REGN-COV2) prevent RBD binding to the cellular ACE2 receptor and aid greatly in patient care, especially for patients with mild to moderate illness but with a high risk to develop severe COVID-19 (162,163).

Apart from therapeutic interventions, prophylactic vaccination was considered one of the main reliefs for the pandemic burden which led to the rapid development of several vaccines (164-167). Here, many different formulations are in use and novel development in vaccine formulation based on mRNA received FDA approval. Two mRNA technology vaccines Comirnaty (BNT162b2) by Biontech/Pfizer and mRNA-1273 by Moderna as well as the adenovirus-based AZD1222 (ChAdOx1) vaccine by AstraZeneca and the University of Oxford are most widely used in the EU. Sputnik V, a recombinant adenovirus vaccine is mainly used in eastern Europe but is currently under evaluation for use throughout the EU. In China, three inactivated vaccines, CoronaVac, BBIBP-CorV, WIBP-CorV as well as a recombinant adenovirus type 5 vector vaccine Convidicea and ZF2001, a recombinant vaccine are in use. With ongoing pandemic and continued clinical trials, more vaccines are likely to be approved.

2. Theses of this dissertation

Thesis I:

Splice site recognition is dependent on its sequence neighborhood

The generation of multiple mRNA isoforms from a single pre-mRNA transcript requires the precise recognition of exon-intron-borders. Key sequence elements in defining the ligation sites after intron removal are the 5' splice site and the 3' splice site. Additional layers of regulation such as transcript architecture and the sequence context, in particular the intronic and exonic proximal splicing regulatory element landscape, contribute to the selection of both, splice donor and splice acceptor sites. Due to the high level of influence on splice site selection by splicing regulatory elements, nucleotide changes in these sequence elements can have devastating effects. Therefore, bioinformatic prediction of the location and putative function of splicing regulatory elements aids in the evaluation of potential consequences of mutations with contributes to an improved evaluation of human pathogenic mutations near splice sites (**Chapters 1-4**).

Thesis II:

Impairment of cellular and viral SREs contributes to aberrant splicing

Human splice sites are highly dependent on splicing regulation elements in their vicinity. Any changes in the capacity of splice site surrounding SREs by missense or even silent mutations can drastically influence the splicing outcome, including full ablation of transcripts or shifts in splice isoforms. This in turn can contribute to human diseases and can have devastating effects on individuals. Similarly, HIV-1, which relies on the human splicing machinery for efficient processing of viral transcripts and generation of the entire viral proteome, also depends on functional splice site regulation by SREs that intersperse the viral genome. To express more than 50 mRNAs from a single transcript, HIV-1 RNA undergoes extensive alternative splicing. Accordingly, changes in splicing regulatory element capacity, due to mutations or masking of respective sequences can have severe effects on the balanced splicing outcome and might even contribute to a decreased susceptibility of the virus towards

antiviral agents. However, the sensitivity of viral gene expression to changes in splicing might also be exploited as a new intervention target (**Chapter 5-8**).

Thesis III:

Viral pathogenesis and immunological features of SARS-CoV-2

With the SARS-CoV-2 pandemic that started December 2019, global research shifted towards the rapid understanding of viral features including viral pathogenesis and the immunological features of the infection, or further on in the pandemic, vaccination respectively. SARS-CoV-2 infection and subsequent COVID-19 disease cause a wide range of symptoms, from mild flu-like symptoms to severe respiratory illness. However, rather atypical symptoms for respiratory illness were also reported and led to further research into the molecular background, revealing that the virus preferably infects neurons in human brain organoids. The extraordinary rapid spread of the virus and insufficient testing and containment strategies at the beginning of the pandemic quickly resulted in high numbers of seroprevalence that needed to be reliably monitored. Immunological monitoring was especially important in the light of convalescent plasma therapy, in particular early in the pandemic when no or only suboptimal drug interventions were available. Further along in the pandemic, in December 2020, approval of prophylactic vaccinations eased the pandemic burden, however, with a new class of vaccines based on mRNA, close monitoring of the strength and duration of the immune response is urgently needed (**Chapter 9-13**).

Thesen dieser Dissertation

These I:

Die Erkennung von Spleißstellen ist von ihrer Sequenzumgebung abhängig

Die Generierung multipler mRNA-Isoformen aus einem einzigen prä-mRNA-Transkript erfordert die präzise Erkennung der Exon-Intron-Grenzen. Schlüsselsequenzelemente bei der Definition der Ligationsstellen nach der Intron-Entfernung sind die 5' Spleißstelle und die 3' Spleißstelle. Zusätzliche Ebenen der Regulierung bilden die Transkriptarchitektur und der Sequenzkontext. Insbesondere die intronische und exonische Landschaft der benachbarten spleißregulatorischen Elemente tragen zur Auswahl der Spleißdonor- und Spleißakzeptorstellen bei. Aufgrund des hohen Einflusses von spleißregulatorischen Elementen auf die Auswahl der Spleißstellen können Nukleotidveränderungen in diesen Sequenzelementen verheerende Auswirkungen haben. Daher hilft die bioinformatische Analyse der Lage und der mutmaßlichen Funktion von putativen spleißregulatorischen Elementen bei der Bewertung der potentiellen Folgen von Mutationen und trägt zu einer verbesserten Bewertung von humanpathogenen Mutationen bei (**Kapitel 1-4**).

These II:

Die Beeinträchtigung von zellulären und viralen SREs trägt zu aberrantem Spleißen bei

Humane Spleißstellen sind in hohem Maße von spleißregulatorischen Elementen in ihrer Umgebung abhängig. Jegliche Veränderung in der Kapazität der SREs, sowohl durch mis-sense oder sogar durch stille Mutationen kann das Spleißergebnis drastisch beeinflussen und kann in der vollständigen Ablation von Transkripten oder der Verschiebung des Verhältnisses von Spleißisoformen führen. Dies wiederum kann zu humanen Krankheiten beitragen und verheerende Auswirkungen auf Individuen haben. In ähnlicher Weise ist HIV-1 in hohem Maße von der menschlichen Spleißmaschinerie abhängig, um die effiziente Verarbeitung viraler Transkripte und die Generierung des gesamten viralen Proteoms sicherzustellen. Hier beeinflussen ebenfalls SREs, die das virale Genom durchziehen, die funktionelle Spleißstellenregulation. Um mehr als 50 mRNAs aus einem einzigen Transkript zu

exprimieren, unterliegt die HIV-1-RNA einem umfangreichen alternativen Spleißen. Dementsprechend können Veränderungen in der Kapazität der spleißregulatorischen Elemente aufgrund von Mutationen oder Maskierung der entsprechenden Sequenzen schwerwiegende Auswirkungen auf das balancierte Spleißergebnis haben und sogar zu einer verminderten Empfindlichkeit des Virus gegenüber antiviralen Wirkstoffen beitragen. Die Empfindlichkeit der viralen Genexpression gegenüber Veränderungen im Spleißen könnte aber ebenso auch als Ziel für neuartige Therapien genutzt werden (**Kapitel 5-8**).

These III:

Virale Pathogenese und immunologische Eigenschaften von SARS-CoV-2

Mit der SARS-CoV-2-Pandemie, die im Dezember 2019 begann, verlagerte sich die weltweite Forschung, um schnell die Eigenschaften des Virus zu verstehen. Hier standen vor allem die virale Pathogenese und die Analyse der immunologischen Eigenschaften der Infektion beziehungsweise im weiteren Verlauf der Pandemie auch der Impfung im Vordergrund. Die SARS-CoV-2-Infektion und die daraus resultierende COVID-19-Erkrankung verursachen ein breites Spektrum an Symptomen, von leichten grippeähnlichen Symptomen bis hin zu schweren Atemwegserkrankungen. Aber auch für Infektionen mit respiratorischen Viren eher untypische Symptome, wie neurologische Störungen, wurden berichtet. Bei der Untersuchung der molekularen Hintergründe stellte sich heraus, dass das Virus in menschlichen Hirnorganoiden bevorzugt Neuronen infiziert kann. Die außergewöhnlich schnelle Ausbreitung des Virus und unzureichende Test- und Eindämmungsstrategien zu Beginn der Pandemie führten schnell zu hohen Seroprävalenzzahlen, die zuverlässig überwacht werden müssen. Die immunologische Überwachung war vor allem im Hinblick auf die rekonvaleszente Plasmatherapie wichtig, insbesondere zu Beginn der Pandemie, als keine oder nur suboptimale medikamentöse Interventionen verfügbar waren. Im weiteren Verlauf der Pandemie, im Dezember 2020, kam es dann durch die Zulassung prophylaktischer Impfungen zu einer Entlastung, jedoch ist bei einer neuen Klasse von Impfstoffen, die auf einer mRNA Technologie basieren, eine genaue Überwachung der Stärke und Dauer der Impfantwort dringend erforderlich (**Kapitel 9-13**).

3. Results

3.1. Splice site recognition is dependent on its sequence neighborhood

The central dogma of gene expression describes a directed flow of information from DNA to RNA to protein. Even though with the research on retroviruses it became clear that the flow of information can also be reversed, from RNA to DNA, the DNA to RNA flow contains various regulated processes including, capping, polyadenylation, and splicing with every process itself is highly regulated. Splice site recognition depends on several sequence elements that directly or indirectly interact with the spliceosome, which carries out the splicing process. Particularly crucial are the splice donor and splice acceptor sequences, which form the basis for the definition of exon-intron borders in the early splicing process. Here, the 5' splice site sequence, provides the landing platform for the U1 snRNP, which appears to act as a guide for the positioning of the spliceosome, since the U1 snRNP leaves the spliceosome even before the first catalytic reaction has occurred. As counterparts, but somewhat offset in time, 3' splice sites with their polypyrimidine tract and branch point sequence interact with auxiliary factors of the U2 snRNP or the U2 snRNP itself, respectively. In this chapter, the regulation of both sequence elements by their neighborhoods that offer binding sites for splicing regulatory proteins is reviewed.

3.1.1 Context matters: Regulation of splice donor usage (Chapter 1, Review)

In defining exon-intron borders during the early splicing process, the binding of U1 snRNP to the 5' splice site is particularly crucial. This reaction is dependent on the complementarity of the U1 snRNA and the splice donor sequence, which can be scored by the hydrogen bonds that can be formed between these two RNA sequences. However, as discussed in this review, sequence complementarity cannot be the sole basis for the strength of RNA duplex formation. It is rather determined by a vast interplay of mostly nearby sequences and the regulatory protein binding landscapes.
Especially concerning human pathogenic mutations, the important role of splice site regulation by splicing regulatory elements becomes evident because here, single nucleotide substitutions can influence splice site recognition and thereby strongly affect both the final transcripts as well as the ratio of alternatively spliced transcripts to each other. In the assessment of such potentially pathogenic nucleotide changes, computer-

assisted tools are often used. Although the predictions made with these tools are already quite robust but still limited, it is important to increase the reliability of the evaluations and predictions, especially concerning human diagnostics.

The following review is published in Biochim Biophys Acta Gene Regul Mech. 2019 Nov - Dec;1862(11-12):194391. doi: 10.1016/j.bbagrm.2019.06.002 by

Ptok J*, **Müller L***, Theiss S, Schaal H. (*Joint first authors)

Contributions:
J.P., L.M., S.T. and H.S. wrote the manuscript and revised it for important intellectual content.
Individual contribution: 25% L.M. wrote chapters 2.1., 2.2 and 3.

BBA - Gene Regulatory Mechanisms xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Context matters: Regulation of splice donor usage☆

Johannes Ptok[a,1], Lisa Müller[a,1], Stephan Theiss[b], Heiner Schaal[a,*]

[a] Institute of Virology, Medical Faculty, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany
[b] Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany

A B S T R A C T

Elaborate research on splicing, starting in the late seventies, evolved from the discovery that 5′ splice sites are recognized by their complementarity to U1 snRNA towards the realization that RNA duplex formation cannot be the sole basis for 5'ss selection. Rather, their recognition is highly influenced by a number of context factors including transcript architecture as well as splicing regulatory elements (SREs) in the splice site neighborhood. In particular, proximal binding of splicing regulatory proteins highly influences splicing outcome. The importance of SRE integrity especially becomes evident in the light of human pathogenic mutations where single nucleotide changes in SREs can severely affect the resulting transcripts. Bioinformatics tools nowadays greatly assist in the computational evaluation of 5'ss, their neighborhood and the impact of pathogenic mutations. Although predictions are already quite robust, computational evaluation of the splicing regulatory landscape still faces challenges to increase future reliability. This article is part of a Special Issue entitled: RNA structure and splicing regulation edited by Francisco Baralle, Ravindra Singh and Stefan Stamm.

1. Separating splice site from context

In order to address the impact of sequence context on splicing, we first need to specify what we understand by the “proper splice site” and its “context”. Historically, splice junction consensus sequences were first constrained to the highly conserved GT-AG dinucleotides at both intron ends [1,2], and U1 snRNA was assumed to bind to both intron ends across the excision-ligation point. However, in 1983 Mount and Steitz showed that the 11 nt long free 5′ end of U1 snRNA binds to the 5'ss but not to the 3'ss [3]. A tentative 5′ splice site motif derived from a collection of 139 5'ss showed clear nucleotide preferences in the nine “consensus” positions −3 to +6, but approximately equal probability for all four bases at the terminal intronic positions +7 and +8 of the possible RNA duplex region, leading to a 9 nt long 5'ss consensus motif [4]. A variety of mutational analyses confirmed that overall splicing efficiency is affected by many nucleotide exchanges within this 5′ splice site region [5–8], and that RNA duplex formation between the 5′ splice site and the terminal nucleotides of U1 snRNA is key to 5′ splice site recognition [9].

It was the first decade that paved the way for the concept that 5′ splice sites are recognized by their complementarity to U1 snRNA. The prevailing perception is that an RNA duplex is formed by a linear sequence of hydrogen bonds between U1 snRNA and the 5′ splice site nucleotides in the standard base-pairing register. Alternatively, Roca et al. described different ways of RNA duplex formation with either shifted [10] or bulged base-pairing registers [11], which are statistically hidden in consensus sequence motifs at the exon/intron border. By using massively parallel splicing assays, it was recently shown that 5'ss seem to be predominantly recognized *via* the normal register and that shifted registers might not always be productively recognized for splicing [12].

Since the large majority of 5'ss neither reflect alternative splicing registers nor U1 snRNA complementarity in positions +7 and +8, the 9 nt long 5'ss motif is the base for most scoring algorithms measuring the intrinsic strength of a 5′ splice site (for a review see [13]). Also, when applying tools measuring the information content of 5′ splice sites, information theory-based position weight matrices failed to identify additional nucleotides contributing to 5′ splice site strength [14].

There are several approaches to measuring splice site strength that go beyond independent position weight matrix scores. Derived by applying the statistical maximum entropy concept to separate sets of real and “decoy” splice sites, the 5′ and 3′ splice site MaxEnt scores take non-adjacent nucleotide dependencies into account when scoring 9 nt long sequences for 5'ss and 23 nt for 3'ss. Today, MaxEntScan is the most widely used splice site scoring tool, and it has been included in most online computational tools [15–17].

Using a complementary concept, the SD algorithm implements a dictionary approach measuring 5'ss strength by the logarithm of its sequence frequency in annotated 5'ss of the human genome. It has

☆ This article is part of a Special Issue entitled: RNA structure and splicing regulation edited by Francisco Baralle, Ravindra Singh and Stefan Stamm.

* Corresponding author.
E-mail address: schaal@uni-duesseldorf.de (H. Schaal).

[1] Joint first authors.

https://doi.org/10.1016/j.bbagrm.2019.06.002
Received 26 April 2019; Received in revised form 7 June 2019; Accepted 9 June 2019

reached 97% sensitivity at 95% specificity on a set of 179 previously reported splicing mutations, analyzed together with 32 minigenes [15–17].

The HBond score (HBS) measures the overall hydrogen bond pattern binding strength in the 11-nt long duplex between the 5'ss and the 5′ end of U1 snRNA. Assigning numerical weights to hydrogen bonds and "mismatches" in individual 5'ss positions, it takes into account inter-dependencies of up to seven neighboring nucleotides as well as G:U wobble base pairs. As a weighted number of hydrogen bonds, the 5'ss HBond score is dimensionless, and its values range from 1.8 (isolated GT) to 23.8 (full U1 snRNA complementarity CAG GTAAGTAT).

RNA duplex formation of a considerable number of 5′ splice sites [18] can in fact benefit from complementary nucleotides in the 7th and 8th positions of the intron [19–21]. Indeed, 25,289 (8.5%) out of 296,036 annotated human 5′ splices sites (GRCh38.91) contain a U1 snRNA complementary AT dinucleotide in positions +7 and +8, including 5′ splice sites of genes with significant diagnostic relevance, *e.g.* β-globin exon 1, SMN1/2 exon 4, BRCA1 exon 18 and 22, BRCA2 exon 6, 7 and 22. The significance of complementary nucleotides in positions +7 and +8 is supported by *in silico* substituting the AT dinucleotide in these 25,289 5'ss by non-complementary bases. Those 5'ss that exhibited large changes ΔHBS in their HBond scores, and thus were more vulnerable to mutations of positions +7 and +8, were stronger on average (higher MaxEnt and HBond scores) and had intron centered regions of U1 snRNA complementarity. In contrast, less vulnerable 5'ss were weaker on average (lower MaxEnt and HBond scores), and their 11 nucleotides wide motifs showed exon centered U1 snRNA complementarity, indicating a possible compensatory mechanism within the 9 nt long 5'ss consensus motif for mutations in positions +7 and +8 (Fig. 1).

RNA duplex formation alone, however, cannot explain 5′ splice site selection, as numerous sequences within an exon are not used as 5′ splice sites, even though their complementarity to U1 snRNA is even higher than the actually used nearby 5′ splice site.

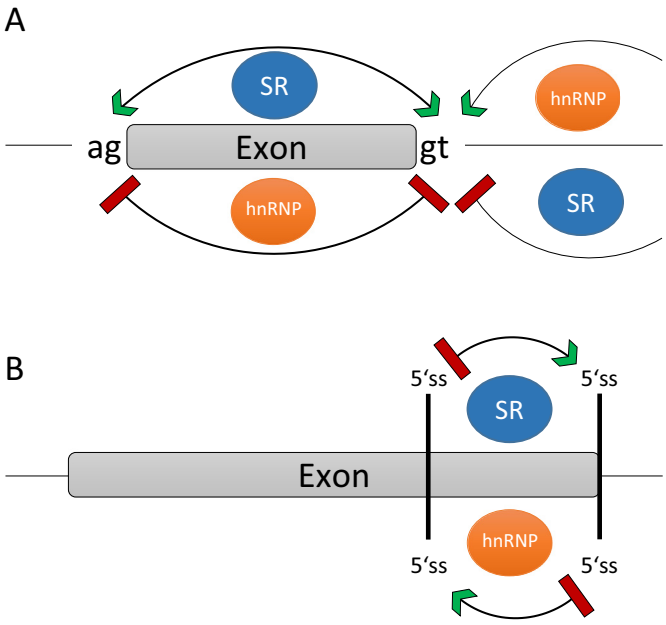


Fig. 2. Schematic drawing of the position dependent regulation of splice sites by SR and hnRNP proteins.
A While SR proteins (blue) enhance the usage of a 5'ss from an upstream position, they repress its recognition from a downstream position. Binding hnRNP proteins (orange) upstream of a 5'ss, on the other hand, repress its usage while they enhance from a downstream position.
B In a situation with two competing 5'ss, the same splicing regulatory element exhibits enhancing features on one splice site while it represses the other. Here, the widely used terms "exonic splicing enhancer (ESE)" and "exonic splicing silencer (ESS)" can be misleading.

2. Splicing regulatory elements: separating true from spurious splice sites

Sequences matching the 5′ and 3′ splice site consensus motifs can be

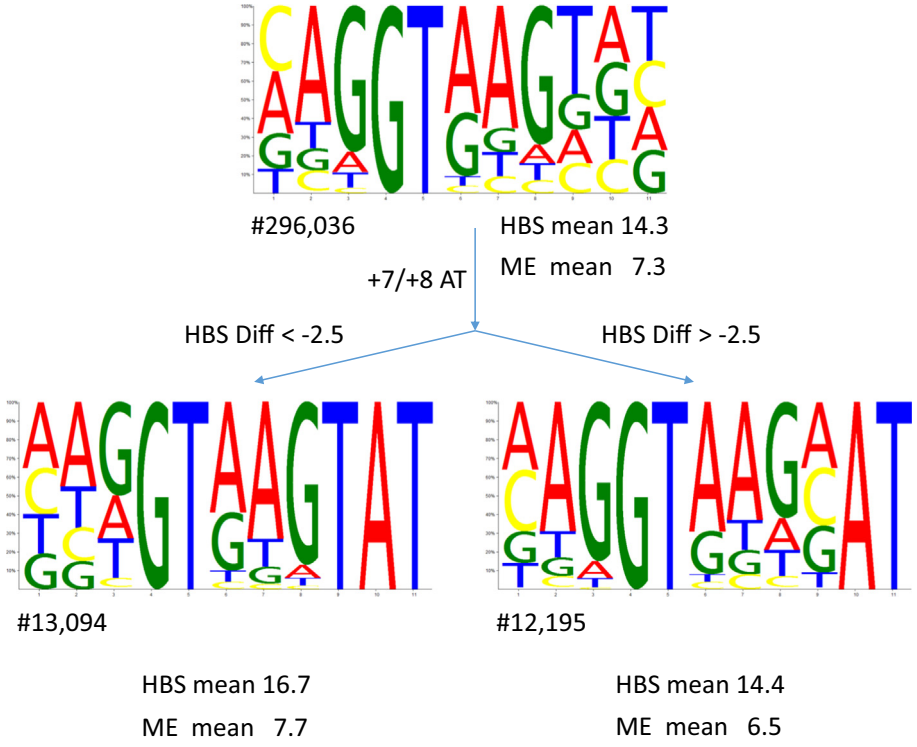


Fig. 1. Different vulnerability to loss of U1 snRNA complementarity in positions +7/+8. 25,289 out of 296,036 canonical human annotated GT 5'ss, with an AT dinucleotide in positions +7 and +8 were split into two groups according to their change in HBS when substituting AT by non-complementary CC. The sequence logos of the wt 5'ss with larger changes (ΔHBS < −2.5, N = 13,094) are depicted on the left side, those with smaller changes (ΔHBS > −2.5, N = 12,195) on the right side. The average MaxEnt scores (ME) and HBond scores (HBS) are shown beneath each sequence logo.

found almost anywhere in the human genome. If all were used as splice sites, transcriptome complexity would increase exponentially. However, a good match to a splice site consensus motif does not imply splice site use, and its recognition is influenced by various context factors: (i) location in first or last *vs.* internal exon [22–25], (ii) exon/intron architecture [26], (iii) regulatory sequences (reviewed in [27]) in close splice site proximity [28], or (iv) generally wider sequence contexts [12].

Frequently, splice site sequence context provides splicing regulatory *cis*-acting elements (SREs) as binding sites for splicing regulatory proteins (SRPs) supporting or inhibiting efficient spliceosomal assembly or (de-)stabilizing the RNA duplex, while interacting with each other and competing for binding sites [29,30] (Fig. 2A). Based on their activity and location in pre-mRNA, these *cis*-acting elements were historically classified as exonic splicing enhancers (ESEs) or silencers (ESSs), and intronic splicing enhancers (ISEs) or silencers (ISSs). However, this classification becomes ambiguous in the presence of several competing potential 5'ss, since the same sequence can be either exonic or intronic dependent on the actually used splice site (Fig. 2B).

2.1. RNA binding proteins and their role in 5'ss selection

RNA binding proteins (RBPs) play important roles in numerous post-transcriptional processes (reviewed in [31]), and they include splicing regulatory proteins (SRPs) that interact with *cis*-acting binding sites in the splicing reaction. Together with the proper splice sites, they contribute highly to the 'splicing code', but their mode of action is still not fully unraveled due to various layers of interaction (reviewed in [32]). It is generally understood that two distinct protein families, serine and arginine rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNP) are main interaction partners of SRE sequences [27,33]. The first group, SR proteins, represent a family of multifunctional RNA binding proteins that are involved in the regulation of constitutive and alternative pre-mRNA splicing [34]. Two distinct structural features characterize this family. First, they share a C-terminal arginine/serine (RS) domain, which is mainly but not exclusively responsible for protein-protein-interactions with the spliceosome, since it is also present in U1-70K, a protein associated with the U1 snRNP [35]. Furthermore, all family members display at least one RNA recognition motif (RRM) at the N-terminus that further provides RNA-binding specificity. The serine residues of the RS domain are targeted for phosphorylation, which in turn highly influences SR protein activity [36,37]. So far, twelve human SR proteins have been identified, and by now, they are uniformly termed serine arginine splicing factor (SRSF) 1–12, while their former names were often derived from their individual molecular weights [38]. Apart from spliceosome interaction, the RS domain is also capable of contacting the pre-mRNA directly either *via* the branch point sequence (BPS) or the 5'ss, which might constitute an alternative way to facilitate spliceosome assembly [39].

As counterparts of SR proteins, hnRNPs are the other key family of proteins involved in splicing regulation and nucleic acid metabolism with additional effects on translation and cellular transport (reviewed in [40,41]). Around 20 major types of hnRNP proteins, which are uniformly termed hnRNP A-U with molecular weights ranging from 34 kDa to 120 kDa, share structural features such as the RNA recognition motif (RRM) and auxiliary domains high in proline, glycine, tyrosine, arginine, glutamine or asparagine [40,42]. Due to their structural diversity, hnRNPs are involved in various cellular processes, though concerning their role in 5'ss selection, they were shown to have adverse behavior compared to SR proteins [28,43].

2.2. Finding a needle in a haystack – defining SRE binding motifs to map the regulatory landscape of 5'ss

Identification of precisely defined binding sites is one of the boxes that needs to be ticked on the way to a full understanding of the splicing code. During the past decades, elaborate approaches have been applied to elucidate the exact interaction sites between RNA and SRPs. In early stages, functional SREs were identified upon their disruption by pathogenic mutations in diseases. Over time, numerous experimental approaches, such as *in vivo* splicing and splicing reporter assays [44–46], and computational methods (*e.g.* [47–50]) were added. More recently, cross-linking and immunoprecipitation (CLIP) and various sub-methods as iClip [51], HITS-CLIP [52] and PAR-CLIP [53] have gained importance and are combined with other large-scale approaches as the identification of alternative splicing events by microarrays and RNA-seq [54]. Both *cis*-acting SRE motifs and *trans*-acting SRP binding domain motifs exhibit high sequence variability, which renders splice site usage prediction difficult. SREs are often degenerate in their sequences and are capable of binding multiple regulatory factors. SRPs, on the other hand, are as well capable of recognizing a wide variety of binding sites, which was shown in several studies [55,56]. While this feature contributes to the tightly regulated splicing process, it is a hurdle in the unraveling of the splicing code.

3. Misguided splicing regulation as the root of disease

Despite SRE motif degeneracy and complexity of splice site context, single nucleotide changes even outside the proper splice site sequence can have dramatic consequences for individuals due to their impact on splicing. For example, when located in SREs, mutations can disrupt binding of SRPs that are crucial for physiological exon/intron border recognition. Hence, mutations can have various effects on the splicing process such as generation of *de novo* splice sites, activation of cryptic splice sites, or decreased use of physiological splice sites [57,58]. In the following, we will give selected examples for these mechanisms:

Duchenne muscular dystrophy (DMD) is an X-linked disorder developing in early childhood and characteristically leaving the early teenage patients wheelchair bound due to muscle weakness caused by lack of dystrophin protein and resulting muscle degeneration. Patients die in their mid-twenties, often due to cardiac or respiratory muscle weakness [59]. One of the disease causing variants identified in DMD is a c.1684C > T mutation in exon 14 of the dystrophin gene that creates a *de novo* GT splice donor site with an HBS of 12.0, which results in a 22 base pair deletion in exon 14. Furthermore, it changes a CAA codon to a TAA stop codon leaving the resulting dystrophin protein dysfunctional [60].

Hutchinson-Gilford Progeria Syndrome (HGPS) is a severe laminopathy and a prominent example for spliceopathies. The main disease associated mutation that causes aberrant splicing of the LMNA gene is the translationally silent c.1824C > T that generates a *de novo* splice site located in exon 11 [61,62]. The mutation increases U1 snRNP complementarity (HBS WT = 12.90, HBS mt = 15.80) which renders this *de novo* splice site active. It generates an alternative lamin A transcript with a deletion of 150 nucleotides that results in a truncated protein (lamin AΔ150) [63].

In a child suffering from neonatal hypotonia, seizures, ataxia and a developmental delay [64], symptoms were related to the activation of a cryptic splice site in the E1α pyruvate dehydrogenase gene (PDH), which plays a key role in energy metabolism [65]. In this particular case, a G-to-A substitution in E1α PDH intron 7-8 was located 26 nucleotides downstream of the physiological splice site. ESEfinder analysis revealed that the mutation generates an SRSF2 protein binding site, that contributes to the use of the cryptic splice site at position 45 in the intron which is naturally inactive [64]. Additionally, the newly created SRE acts silencing on the upstream physiological splice donor which contributes to the aberrant splicing outcome [66].

Especially in a diagnostic setting, silent mutations that do not change the underlying coding potential of a sequence need to be taken into account. Such silent mutations are often ignored in routine diagnostics, since they do not change the resulting amino acids, but they still harbor the potential to severely interfere with splicing regulation.

One well studied example is the silent C > T transition (c.840C > T) in exon 7 of survival motor neuron gene 2 (SMN2) [67]. In patients suffering from spinal muscular atrophy (SMA), loss of survival motor neuron gene 1 (SMN1) could in principle be compensated by the nearly identical SMN2 gene. A critical difference between SMN1 and SMN2, however, lies in the single C > T change in SMN2 disrupting a splicing regulatory element and leading to skipping of SMN2 exon 7, and hence a non-functional SMN2 protein [68,69].

Publicly accessible databases that compile human pathogenic mutations, *e.g.* [70–73], are dominated by protein altering mutations and struggle far more with the inclusion of silent SNVs outside splice site motifs that still have disease-causing potential by mis-regulating splicing. This is due to the overall difficulty of current algorithms to reliably predict the influence of a SNV both on the binding affinity of a specific SRP and on the impact of the altered SRP binding on splice site recognition.

4. Computational evaluation of 5'ss

4.1. Open access tools

During the last two decades, various publicly available tools have been developed which can be used to analyze the interplay of splice site context and intrinsic splice site strength for a given sequence (Table 1). They can broadly be categorized into tools based on (1) a computational analysis of nucleotide motifs or k-mer distributions, (2) individual experimental data, (3) previously described motifs, (4) a combination of multiple tools or (5) neuronal networks (partially reviewed in [74]).

4.1.1. Tools based on computational analysis of nucleotide motifs or k-mer distributions

The HEXplorer [48,66], RBPmap [75], Splicing Factor Finder [76] and RESCUE-ESE [47,77] are based on computational analyses of nucleotide motifs or k-mer distributions. HEXplorer and RESCUE-ESE use the position-dependent effects of splicing regulatory proteins on splice site usage [28] and the corresponding difference in the abundance of hexamers upstream and downstream of splice donors and around intrinsically weak or strong splice sites. Based on a set of hexamer frequency weights from −73.27 (TTTTTT) to +34.35 (GAAGAA), the HEXplorer algorithm [48,66] provides a score profile of a genomic sequence, reflecting its ability to either enhance or silence nearby splice site usage. RESCUE-ESE scans a genomic sequence for the presence of 238 hexamers which were found more frequently within exons with weak 5′ splice sites than in exons with strong 5'ss or introns. These hexamers constitute potential binding sites for SR proteins enhancing downstream splice donor and upstream splice acceptor usage (“ESE”) [47,77]. RBPmap [75] as well as the Splicing Factor Finder [76] predict binding sites of RNA-binding proteins within a given genomic sequence using a collection of well-described binding motifs and their evolutionary conservation.

4.1.2. Tools based on individual experimental data

ESEfinder [49], ESRseq [50] and FAS-ESS [78] are built on grounds of individual experimental data. ESRseq is based on RNA-sequencing data of a three-exon minigene library, resulting in the scoring of 2272 of all 4096 hexamers for their potential position-dependent enhancing/repressing effects on splice site usage. ESRseq hexamer scores range from −1.06 (CTTTTA) to +1.03 (AGAAGA), while no scores are

Table 1
Tools to predict splicing regulatory elements.

Tool	Published	URL	Features
(1) Computational analysis of nucleotide motifs or k-mer distributions			
HEXplorer [48,66]	2014	https://www2.hhu.de/rna/html/hexplorer_score.php	Scores all hexamers for potential position-dependent enhancing/repressing effects on splice site usage
RBPmap [75]	2014	http://rbpmap.technion.ac.il	Calculates potential binding, using well described motifs of the literature and their conservation
Splicing Factor Finder [76]	2009	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691001	Mapping of splicing factor sites with the help of previously described binding motifs and their evolutionary conservation
RESCUE-ESE [47,77]	2002	http://genes.mit.edu/burgelab/rescue-ese/	Scans for 238 hexamers which are more frequent within exonic sequences of weak splice sites from the reference genome
(2) Based on individual experimental data			
ESRseq [50]	2011	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3149502/	Scores all hexamers for potential position-dependent enhancing/repressing effects on splice site usage, based on RNA-sequencing of a minigene library
FAS-ESS [78]	2004	http://genes.mit.edu/fas-ess/	Functional screening for 6-nt SELEX motifs of ESSs
ESEfinder [49]	2003	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home	Scans genomic sequence for functional SELEX motifs of SRSF1, SRSF2, SRSF5 and SRSF6
(3) Based on previously described motifs			
ATtRACT (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4823821/)	2016	http://attract.cnic.es	Scanning genomic sequence for hand-curated set of experimentally validated RBP binding motifs
SpliceAid [79]	2009	http://www.introni.it/splicing.html	Scans genomic sequences for validated binding motifs of human splice regulatory proteins
(4) Combining multiple tools			
EX-SKIP [80]	2011	http://ex-skip.img.cas.cz/	Tool predicting exon skipping based on result of multiple SRE prediction tools
Human Splicing Finder [81]	2009	http://www.umd.be/HSF3/	Tool predicting SREs, splice sites or branch sites
SROOGLE [82]	2009	http://sroogle.tau.ac.il	Tool predicting splice sites, SREs, branch sites and Polyadenylation sites
(5) Artificial neural network tool			
SpliceAI [83]	2019	https://github.com/Illumina/SpliceAI	Network predicting splice sites and taking SREs implicitly into account

assigned to 1824 hexamers predicted to have no influence on splicing [50]. The ESEfinder and FAS-ESS tools, on the other hand, are based on Systematic Evolution of Ligands by EXponential enrichment (SELEX) [84]. While ESEfinder scans genomic sequences for the presence of experimentally derived RNA binding motifs of recombinant SR proteins SRSF1, SRSF2, SRSF5 and SRSF6 [49], FAS-ESS correspondingly identifies RNA binding motifs of proteins repressing downstream splice donor and upstream splice acceptor usage ("ESS") [78].

4.1.3. Tools based on previously described motifs

ATtRACT [85] and SpliceAid [79] can be categorized as tools based on previously described motifs. SpliceAid, as well as the tissue-specific adaptation SpliceAid2 [86] are also based on experimentally validated binding motifs of splicing regulatory proteins which were selected from the literature. Similar to SpliceAid, ATtRACT is based on extensive literature search for general RBP binding motifs, thus not directly visually discriminating between SR and hnRNP proteins.

4.1.4. Tools combining multiple tools

The Human Splicing Finder [81], EX-SKIP [80] and SROOGLE [82] combine multiple tools for SRE prediction. Human Splicing Finder combines RESCUE-ESE and ESEfinder as well as an individual module for the detection of SRSF3 and SFRS10 binding motifs [81]. The EX-SKIP tool compares likelihood of exon skipping between exonic wild type and mutant sequences, based on the integration of RESCUE-ESE, PESE/PESS [87] or FAS-ESS [80]. Like Human Splicing Finder, SROOGLE combines tools like ESE-finder, RESCUE-ESE and FAS-ESS to predict potential binding sites of splicing regulatory proteins.

4.1.5. Artificial neural network tool

All above mentioned algorithms for computational identification of SREs follow a bottom-up approach, *i.e.* they collect specific SRE motifs ("atoms"), estimate their individual strength and assemble this information to predict their impact on splice site recognition.

A complementary, top-down approach that is completely agnostic to previously identified motifs has recently been presented by Jaganathan et al., who designed and trained a series of deep residual neuronal networks, SpliceAI, that directly classify each position in a pre-mRNA sequence as either splice donor, splice acceptor, or neither [83]. This type of analysis circumvents the *a-priori* identification of SREs, and rather lets the artificial neural network implicitly learn the splice site recognition rules.

For such a neural network, classification quality is measured by its top-*k* accuracy, which is the fraction of correctly predicted splice sites at the threshold where the number of predicted sites equals the number of true sites in the dataset. Interestingly, networks working on longer sequence segments of up to 10,000 nt exhibited much better splice site prediction than those working on 80 nt: top-*k* accuracy increased from 0.57 (80 nt) to 0.95 (10,000 nt). SpliceAI significantly outperforms GeneSplicer, MaxEntScan and NNSplice, and its source code is publicly hosted at https://github.com/Illumina/SpliceAI.

4.2. Exemplary evaluation of computational tools for assessment of context dependency

In this section, we will briefly demonstrate, on a feature-rich minigene model system, the context-related information that can be gathered from selected online tools of the above list. In a series of minigene reporter experiments, Lu et al. have systematically examined 5'ss context impact on splicing by inserting a set of eight splice sites with varying strengths into two different environments [88]. In detail, an exon of either gene TRIM62 or HMSD, framed by 500 upstream and downstream intronic nucleotides, was inserted between two exons of a minigene splicing reporter, and all combinations of 5'ss and context were assessed (Fig. 3). In the TRIM62 context, decreasing 5'ss strength from MaxEnt score 11 to 4.44 did not significantly disrupt splicing, while in the HMSD context, exon skipping occurred—and gradually increased—below MaxEnt score of 9.6. Thus, the TRIM62 5'ss context clearly supported splicing more effectively than its HMSD counterpart.

For our evaluation of online tools, we analyzed 50 nt up- and downstream of the 5'ss for potential impact on 5'ss usage (Table 2). To measure the "splicing regulatory effect" predicted for this 5'ss neighborhood, we calculated a "splice site enhancer weight" (SSEW) to capture both enhancing and silencing properties [66]: we assigned a weight of +1 or −1 for each exonic enhancer or silencer motif predicted by any of the algorithms, and subtracted the total sum of downstream weights from the sum of upstream weights. Accordingly, we calculated SSEW for ESRseq and HEXplorer as upstream ESRseq and HEXplorer totals minus downstream totals. Since different tools vary in algorithmic principle and value range, it is interesting to compare the different methods.

RESCUE-ESE and the ESEfinder consider subsets of SREs that promote downstream and repress upstream 5'ss usage. In the example above, RESCUE-ESE predicted no difference in 5'ss context, whereas, surprisingly, ESEfinder predicted a stronger 5'ss SRE support within the HMSD gene context, contrary to the experimental observation that the TRIM62 context supported weaker 5'ss more than HMSD.

Similar to the remaining tools, FAS-ESS, which scans sequences for a small set of SREs which repress downstream but enhance upstream 5'ss usage, correctly predicted the stronger TRIM62 5'ss context. RBPmap, HSF3 and EX-SKIP naturally have higher SSEWs than RESCUE-ESE or FAS-ESS, since they comprise predictions of multiple SRE tools. All three tools also graded TRIM62 context significantly more enhancing than HMSD context.

Rather than counting individual motifs, HEXplorer and ESRseq SSEWs total positive and negative hexamer-based scores in genomic regions impacting 5'ss usage. Both ESRseq and HEXplorer correctly assessed the TRIM62 context as more supportive for splicing than HMSD.

In their minigene model system, Lu et al. calculated intrinsic 5'ss strength using the MaxEnt score, based on nine 5'ss nucleotides. Calculating HBond scores of the 11 nucleotide long 5'ss obviously depended on the "context-positions" +7 and +8: within the HMSD context, 5'ss had equal or even slightly higher HBond scores than within the TRIM62 context (Table 3). Thus, in this model system, presence or absence of U1 snRNA complementarity in positions +7 and +8 cannot be responsible for the observed difference between both contexts.

However, for three 5'ss, HBond scores for hypothetically U1 snRNA complementary AT dinucleotides in positions +7/+8 are larger than in the TRIM62 context, so that nucleotide changes in these positions, that are usually included in the context, could in fact modify U1 snRNA complementarity and thus more directly impact splicing (Table 3, *).

Gene architecture could also play a role in differential splicing outcome between TRIM62 and HMSD contexts. The splice acceptor of the shorter 96 nucleotides long TRIM62 exon has a high MaxEnt score of 9.97, whereas the acceptor of the longer 174 nucleotides long HMSD exon is weaker (MaxEnt score 6.67) and thus potentially leads to less efficient exon definition. This is in line with recent findings of Wong et al., comparing the activity of 32,768 unique 5'ss sequences (NNN/GYNNNN) in three different gene contexts (BRCA2 exon 17, SMN1 exon 7, and ELP1 (IKBKAP) exon 20) by massively parallel splicing assays. In these experimental settings, context dependency of 5'ss recognition was strongly determined by the strength of the upstream 3'ss [12].

4.3. Selection of appropriate computational tools for different applications

Generally, specific scientific or diagnostic questions determine which tools are most suitable. RBPmap, RESCUE-ESE, FAS-ESS, SpliceAid, Human Splicing Finder, SROOGLE and especially the commonly used ESEfinder determine locations of highly validated SRP binding motifs within a given RNA sequence. These tools predict individual potential SRP binding sites and can be beneficial in single-case

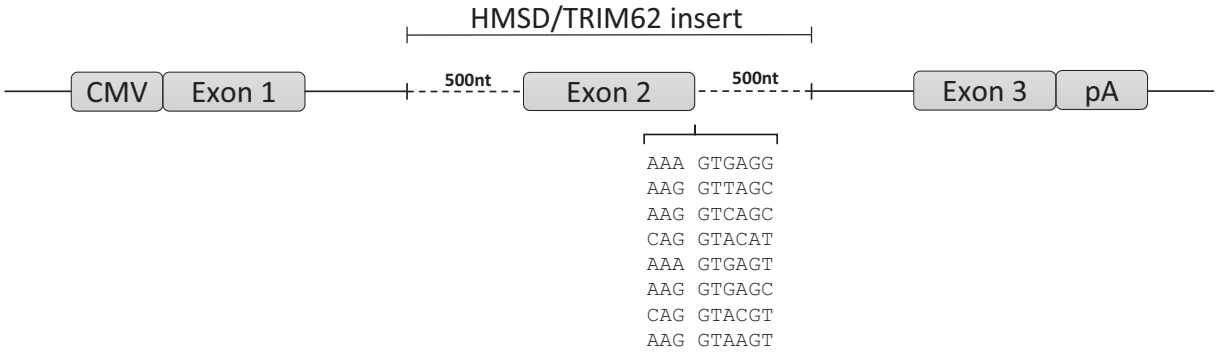


Fig. 3. TRIM62/HMSD minigene reporter. In order to compare two different SD contexts, exon 2 of HMSD and TRIM62 plus 500 nucleotides of the upstream and downstream introns were inserted into a three-exon minigene reporter. The respective wild type splice donor sequences were then replaced (pos. -3 to +6) with eight different splice donor sites, to measure splice donor strength dependent exon skipping in both contexts. CMV: CMV promotor; pA: polyadenylation site.

Table 2
Evaluation of TRIM62 or HMSD context using different tools.

For every tool, splice site enhancer weights were calculated as total upstream predicted SREs minus total downstream SREs, with exonic enhancers counted as +1 and silencers as −1. Higher splice site enhancer weights reflect higher expected enhancing effects on 5'ss usage. The difference between both contexts was calculated by subtracting the TRIM62 SSEW from the HMSD SSEW.

Tool	Splice site enhancer weight		Difference: HMSD-TRIM62
	TRIM62	HMSD	
RESCUE-ESE	−2	−2	0
RBPmap	8	−1	−9
FAS-ESS	3	−2	−5
ESE-finder	−1	7	8
EX-SKIP	51	−13	−64
HSF3	12	4	−8
HEXplorer	504	−92	−596
ESRseq	6,7	−0,5	−7,2

Table 3
Comparison of intrinsic 5'ss strength (HBond score, HBS) including positions +7/+8. Insertion of 9 nt long splice donor sequences into TRIM62 and HMSD contexts leads to different HBond scores depending on the context nucleotides in +7/+8 (CT for TRIM62, AA for HMSD, or AT in a hypothetical context providing U1 snRNA complementarity in positions +7/+8). Large HBond score differences are marked by *.

Inserted SD sequence	HBS: SD in TRIM62 +CT	HBS: SD in HMSD +AA	ΔHBS: HMSD – TRIM62	HBS: SD +AT	ΔHBS: +AT – TRIM62
AAAGTGAGG	10.5	10.7	0.2	12.0	1.5
AAGGTTAGC	13.2	13.4	0.2	14.7	1.5
AAGGTCAGC	13.2	13.4	0.2	14.7	1.5
CAGGTACAT	14.0	14.0	0.0	14.8	0.8
AAAGTGAGT	13.0	14.2	1.2	15.5	2.5*
AAGGTGAGC	15.7	15.9	0.2	17.2	1.5
CAGGTACGT	18.0	19.2	1.2	20.5	2.5*
AAGGTAAGT	19.6	20.8	1.2	22.1	2.5*

analyses of specific SNVs, but they do not explicitly quantify their impact on splicing, which limits their usefulness in a diagnostic setting.

On the other hand, EX-SKIP, the Human Splicing Finder or the state-of-the-art deep-learning based SpliceAI [83] estimate the expected splicing outcome from given sequences in a single, nontransparent "black box" procedure that does not make the incorporation of potential SREs explicit, and—except for HSF3—they do not provide information about the positions of potential SRP binding sites.

For a given sequence, ESRseq and HEXplorer aim at quantifying the potentially enhancing or silencing effects on 5'ss usage. Following a RESCUE-type approach, both algorithms are based on hexamer frequency differences between upstream and downstream sequences, and they reflect position-dependent effects of SREs. Both scores can be used to analyze changes in splice site context on a genome wide scale, but they do not predict individual potential SRP binding motifs like ESE-finder. Comparative studies validated that ESRseq and HEXplorer score changes significantly correlated with changes in splice site usage for known human gene variants as well as in splicing reporter experiments [89,90].

Human Splicing Finder constitutes a well-designed tool providing a clear overview about SREs predicted by various tools like RESCUE-ESE and ESEfinder, in addition to individual motifs for SRSF3 and SRSF10, derived from public data. It may be particularly suited to investigate whether overexpression of an SRP of interest might influence usage of a certain splice site.

5. Challenges in computational 5'ss evaluation

Although there has been a tremendous gain in knowledge on context dependency of splicing, there still remain important questions to be addressed, both locally on a small scale, and involving wider sequence neighborhoods of splice sites on a larger scale.

5.1. Small scale challenges

Generally, RNA secondary structure can impact splice site accessibility, and thus represents another context dependent factor [91–94]. Large scale RBP binding assays showed that although no RBP seems to strictly require a certain RNA secondary structure, some RBPs seem to prefer a binding site within or especially not within a hairpin loop [95,96]. Impact of secondary structure, however, has not been implemented yet in any of the currently available computational tools.

Tools based on SELEX methods [84] lead to the identification of comparatively few high-affinity motifs, leaving potential binding sites of lower affinity aside (*reviewed in* [97]). Additionally, tools mentioned above often predict overlapping or nearby SRP binding motifs which could indicate a protein-RNA binding competition situation, possibly involving steric hindrance. More generally, there is no explicit ("transparent") joint model of 5'ss and SREs addressing 5'ss usage. Quantitatively incorporating/modeling interactions and interdependencies between different SREs as well as between SREs and 5'ss still is a major challenge for computational prediction of splicing.

5.2. Large scale challenges

Gene architecture, *i.e.* genomic distribution of splice sites and SREs as well as exon and intron lengths, is currently not addressed by any of the above tools. Depending on the intrinsic strength and context of nearby splice sites, weakening an index splice site due to a pathogenic

mutation can result in different kinds of pathological disruptions of splicing, depending on gene structure [58].

Repression of 5'ss usage can either lead to skipping of the respective exon [98], alternative exon ends due to usage of cryptic splice sites [99], multiple exon skipping due to failed exon definition or changes in the order and dynamics of intron removal [100,101], intron inclusion, or a combination of the above [102]. In most exons below ~250 nt, splice site recognition is well explained by exon definition [103], describing the observation that binding of U1 snRNP at the 3′ end of an exon can enhance recognition of the 5′ exon end [104] and *vice versa* [105].

Mutually enhancing exon end recognition could explain the so-called proximity rule, that in case of competing equally strong 5'ss, the 5'ss closer to the 3'ss of the downstream exon is often predominantly chosen [106,107]. This proximity rule is still applicable after decreasing the proximal 5'ss strength, making it important to not only consider SRP binding sites during 5'ss context evaluation, but also presence and strength of surrounding potential splice site sequences [108]. In line with these findings, SpliceAI neural network classification was much better with up to 10,000 nt long input sequences than with short 80 nt neighborhoods, although SpliceAI does not explicitly identify SREs but implicitly learned their effects.

5.3. Cell-type specific computational prediction of 5'ss and SREs

A fundamental uncertainty of identifying SREs as potential SRP binding sites lies in the unknown availability of the specific SRP binding partners. It is known that expression of SRPs is cell-type specific. A possible approach to address cell-type specificity could be based on RNA-seq data from different cell types: For each cell type, exon junctions and 5'ss predominantly occurring in one cell type but not in others could be collected, and hexamer frequencies could be determined from their respective up- and downstream neighborhoods, along the lines of the RESCUE-ESE concept. From these frequency tables, normalized hexamer *Z*-weights could be calculated and enter into a cell-type specific HEXplorer score. Profiles generated with this HEXplorer score would then hypothetically represent cell-type specific (SRP-availability weighted) splicing enhancing or silencing effects.

6. Conclusion and outlook

Bioinformatics analyses of splice site environment can contribute significantly to understanding splice site usage. With the advent of RNA sequencing techniques, focus has moved to the impact of mutations on RNA processing regulation. This is particularly important in human mutation diagnostics in the vicinity of splice sites [109] as well as in development of therapeutic strategies.

FDA approved drugs have recently used this interventional pathway in SMA and DMD. In SMA patients, the use of antisense oligonucleotides (ASO) to mask an intronic splicing regulatory element has been shown to successfully restore exon inclusion (Nusinersen/Spinraza®) [110]. In DMD patients with loss of exons 49 and 50, the correct reading frame could be reinstated with Eteplirsen (Exondys 51®), leading to skipping of exon 51 through SRE masking [111].

While previously optimal ASO target sequences were typically identified by laborious and costly scanning of large target regions with overlapping ASOs, systematic *in silico* prediction of promising context dependent ASO targets might considerably speed up the development of RNA therapeutic agents.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (SCHA909/8-1; TI323/4-1), the Forschungskommission of the Heinrich Heine University Düsseldorf, the Heinz-Ansmann Stiftung für AIDS-Forschung, Düsseldorf (H.S.), and the Jürgen Manchot Stiftung (L.M., H.S.).

Transparency document

The Transparency document associated this article can be found, in online version.

Acknowledgements

We would like to thank all lab members for discussion and critical reading of the manuscript.

Author contributions

J.P., L.M., S.T. and H.S. wrote the manuscript and revised it for important intellectual content.

References

[1] J.F. Catterall, B.W. O'Malley, M.A. Robertson, R. Staden, Y. Tanaka, G.G. Brownlee, Nucleotide sequence homology at 12 intron–exon junctions in the chick ovalbumin gene, Nature 275 (1978) 510–513.
[2] R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, 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 (1978) 4853–4857.
[3] S.M. Mount, I. Pettersson, M. Hinterberger, A. Karmas, J.A. Steitz, The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site in vitro, Cell 33 (1983) 509–518.
[4] S.M. Mount, A catalogue of splice junction sequences, Nucleic Acids Res. 10 (1982) 459–472.
[5] M. Aebi, H. Hornig, C. Weissmann, 5′ cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5′ splice region, not by the conserved 5′ GU, Cell 50 (1987) 237–246.
[6] S. Weber, M. Aebi, In vitro splicing of mRNA precursors: 5′ cleavage site can be predicted from the interaction between the 5′ splice region and the 5′ terminus of U1 snRNA, Nucleic Acids Res. 16 (1988) 471–486.
[7] M. Aebi, H. Hornig, R.A. Padgett, J. Reiser, C. Weissmann, Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA, Cell 47 (1986) 555–565.
[8] C. Montell, A.J. Berk, Elimination of mRNA splicing by a point mutation outside the conserved GU at 5′ splice sites, Nucleic Acids Res. 12 (1984) 3821–3827.
[9] Y. Zhuang, A.M. Weiner, A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation, Cell 46 (1986) 827–835.
[10] X. Roca, A.R. Krainer, Recognition of atypical 5′ splice sites by shifted base-pairing to U1 snRNA, Nat. Struct. Mol. Biol. 16 (2009) 176–182, https://doi.org/10.1038/nsmb.1546.
[11] X. Roca, M. Akerman, H. Gaus, A. Berdeja, C.F. Bennett, A.R. Krainer, Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides, Genes Dev. 26 (2012) 1098–1109, https://doi.org/10.1101/gad.190173.112.
[12] M.S. Wong, J.B. Kinney, A.R. Krainer, Quantitative activity profile and context dependence of all human 5′ splice sites, Mol. Cell 71 (2018) 1012–1026 e1013 https://doi.org/10.1016/j.molcel.2018.07.033.
[13] X. Roca, A.R. Krainer, I.C. Eperon, Pick one, but be quick: 5′ splice sites and the problems of too many choices, Genes Dev. 27 (2013) 129–144, https://doi.org/10.1101/gad.209759.112.
[14] N. Caminsky, E.J. Mucaki, P.K. Rogan, Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis, F1000Res 3 (2014) 282, https://doi.org/10.12688/f1000research.5654.1.
[15] G. Yeo, C.B. Burge, Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals, J. Comput. Biol. 11 (2004) 377–394, https://doi.org/10.1089/1066527041410418.
[16] K. Sahashi, A. Masuda, T. Matsuura, J. Shinmi, Z. Zhang, Y. Takeshima, M. Matsuo, G. Sobue, K. Ohno, 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 (2007) 5995–6003, https://doi.org/10.1093/nar/gkm647.
[17] X. Roca, A.J. Olson, A.R. Rao, E. Enerly, V.N. Kristensen, A.L. Borresen-Dale, B.S. Andresen, A.R. Krainer, R. Sachidanandam, Features of 5′-splice-site efficiency derived from disease-causing mutations and comparative genomics, Genome Res. 18 (2008) 77–87, https://doi.org/10.1101/gr.6859308.
[18] L. Hartmann, S. Theiss, D. Niederacher, H. Schaal, Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Front. Biosci. 13 (2008) 3252–3272.
[19] S. Kammler, C. Leurs, M. Freund, J. Krummheuer, K. Seidel, T.O. Tange, M.K. Lund, J. Kjems, A. Scheid, H. Schaal, 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 (2001) 421–434.
[20] M. Freund, M.J. Hicks, C. Konermann, M. Otte, K.J. Hertel, H. Schaal, 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 (2005) 5112–5119, https://doi.org/10.1093/nar/gki824.
[21] M. Freund, C. Asang, S. Kammler, C. Konermann, J. Krummheuer, M. Hipp, I. Meyer, W. Gierling, S. Theiss, T. Preuss, D. Schindler, J. Kjems, H. Schaal, A novel approach to describe a U1 snRNA binding site, Nucleic Acids Res. 31 (2003) 6963–6975.
[22] H.V. Colot, F. Stutz, M. Rosbash, The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20 the small subunit of the nuclear cap-binding complex, Genes Dev. 10 (1996) 1699–1708, https://doi.org/10.1101/gad.10.13.1699.
[23] J.D. Lewis, E. Izaurralde, A. Jarmolowski, C. McGuigan, I.W. Mattaj, A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site, Genes Dev. 10 (1996) 1683–1698.
[24] M.M. Konarska, R.A. Padgett, P.A. Sharp, Recognition of cap structure in splicing in vitro of mRNA precursors, Cell 38 (1984) 731–736.
[25] H.G. Martinson, An active role for splicing in 3′-end formation, Wiley Interdiscip. Rev. RNA 2 (2011) 459–470, https://doi.org/10.1002/wrna.68.
[26] K.L. Fox-Walsh, Y. Dou, B.J. Lam, S.P. Hung, P.F. Baldi, K.J. Hertel, The architecture of pre-mRNAs affects mechanisms of splice-site pairing, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 16176–16181, https://doi.org/10.1073/pnas.0508489102.
[27] A.J. Matlin, F. Clark, C.W. Smith, Understanding alternative splicing: towards a cellular code, Nat. Rev. Mol. Cell Biol. 6 (2005) 386–398, https://doi.org/10.1038/nrm1645.
[28] S. Erkelenz, W.F. Mueller, M.S. Evans, A. Busch, K. Schoneweis, K.J. Hertel, H. Schaal, Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms, RNA 19 (2013) 96–102, https://doi.org/10.1261/rna.037044.112.
[29] J.F. Caceres, S. Stamm, D.M. Helfman, A.R. Krainer, Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors, Science 265 (1994) 1706–1709.
[30] I.C. Eperon, O.V. Makarova, A. Mayeda, S.H. Munroe, J.F. Caceres, D.G. Hayward, A.R. Krainer, 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 (2000) 8303–8318.
[31] E. Dassi, Handshakes and fights: the regulatory interplay of RNA-binding proteins, Front. Mol. Biosci. 4 (2017) 67, , https://doi.org/10.3389/fmolb.2017.00067.
[32] M. Baralle, F.E. Baralle, The splicing code, Biosystems 164 (2018) 39–48, https://doi.org/10.1016/j.biosystems.2017.11.002.
[33] Z. Wang, C.B. Burge, Splicing regulation: from a parts list of regulatory elements to an integrated splicing code, RNA 14 (2008) 802–813, https://doi.org/10.1261/rna.876308.
[34] M.L. Anko, Regulation of gene expression programmes by serine-arginine rich splicing factors, Semin. Cell Dev. Biol. 32 (2014) 11–21, https://doi.org/10.1016/j.semcdb.2014.03.011.
[35] P.J. Shepard, K.J. Hertel, The SR protein family, Genome Biol. 10 (2009) 242, https://doi.org/10.1186/gb-2009-10-10-242.
[36] W. van Der Houven Van Oordt, K. Newton, G.R. Screaton, J.F. Caceres, Role of SR protein modular domains in alternative splicing specificity in vivo, Nucleic Acids Res. 28 (2000) 4822–4831.
[37] V. Botti, F. McNicoll, M.C. Steiner, F.M. Richter, A. Solovyeva, M. Wegener, O.D. Schwich, I. Poser, K. Zarnack, I. Wittig, K.M. Neugebauer, M. Muller-McNicoll, Cellular differentiation state modulates the mRNA export activity of SR proteins, J. Cell Biol. 216 (2017) 1993–2009, https://doi.org/10.1083/jcb.201610051.
[38] J.L. Manley, A.R. Krainer, A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins), Genes Dev. 24 (2010) 1073–1074, https://doi.org/10.1101/gad.1934910.
[39] H. Shen, M.R. Green, A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly, Mol. Cell 16 (2004) 363–373, https://doi.org/10.1016/j.molcel.2004.10.021.
[40] T. Geuens, D. Bouhy, V. Timmerman, The hnRNP family: insights into their role in health and disease, Hum. Genet. 135 (2016) 851–867, https://doi.org/10.1007/s00439-016-1683-5.
[41] J. Jean-Philippe, S. Paz, M. Caputi, hnRNP A1: the Swiss army knife of gene expression, Int. J. Mol. Sci. 14 (2013) 18999–19024, https://doi.org/10.3390/ijms140918999.
[42] G. Dreyfuss, V.N. Kim, N. Kataoka, Messenger-RNA-binding proteins and the messages they carry, Nat. Rev. Mol. Cell Biol. 3 (2002) 195–205, https://doi.org/10.1038/nrm760.
[43] A. Busch, K.J. Hertel, Evolution of SR protein and hnRNP splicing regulatory factors, Wiley Interdiscip. Rev. RNA 3 (2012) 1–12, https://doi.org/10.1002/wrna.100.
[44] H.X. Liu, M. Zhang, A.R. Krainer, Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins, Genes Dev. 12 (1998) 1998–2012.
[45] H. Tian, R. Kole, Selection of novel exon recognition elements from a pool of random sequences, Mol. Cell. Biol. 15 (1995) 6291–6298.
[46] L.R. Coulter, M.A. Landree, T.A. Cooper, Identification of a new class of exonic splicing enhancers by in vivo selection, Mol. Cell. Biol. 17 (1997) 2143–2150.
[47] W.G. Fairbrother, R.F. Yeh, P.A. Sharp, C.B. Burge, Predictive identification of exonic splicing enhancers in human genes, Science 297 (2002) 1007–1013, https://doi.org/10.1126/science.1073774.
[48] S. Erkelenz, S. Theiss, M. Otte, M. Widera, J.O. Peter, H. Schaal, Genomic HEXploring allows landscaping of novel potential splicing regulatory elements, Nucleic Acids Res. 42 (2014) 10681–10697, https://doi.org/10.1093/nar/gku736.
[49] L. Cartegni, J. Wang, Z. Zhu, M.Q. Zhang, A.R. Krainer, ESEfinder: a web resource to identify exonic splicing enhancers, Nucleic Acids Res. 31 (2003) 3568–3571.
[50] S. Ke, S. Shang, S.M. Kalachikov, I. Morozova, L. Yu, J.J. Russo, J. Ju, L.A. Chasin, Quantitative evaluation of all hexamers as exonic splicing elements, Genome Res. 21 (2011) 1360–1374, https://doi.org/10.1101/gr.119628.110.
[51] J. Konig, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D.J. Turner, N.M. Luscombe, J. Ule, iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution, Nat. Struct. Mol. Biol. 17 (2010) 909–915, https://doi.org/10.1038/nsmb.1838.
[52] K. Charizanis, K.Y. Lee, R. Batra, M. Goodwin, C. Zhang, Y. Yuan, L. Shiue, M. Cline, M.M. Scotti, G. Xia, A. Kumar, T. Ashizawa, H.B. Clark, T. Kimura, M.P. Takahashi, H. Fujimura, K. Jinnai, H. Yoshikawa, M. Gomes-Pereira, G. Gourdon, N. Sakai, S. Nishino, T.C. Foster, M. Ares Jr., R.B. Darnell, M.S. Swanson, Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy, Neuron 75 (2012) 437–450, https://doi.org/10.1016/j.neuron.2012.05.029.
[53] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr., A.C. Jungkamp, M. Munschauer, A. Ulrich, G.S. Wardle, S. Dewell, M. Zavolan, T. Tuschl, Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP, Cell 141 (2010) 129–141, https://doi.org/10.1016/j.cell.2010.03.009.
[54] S. De, M. Gorospe, Bioinformatic tools for analysis of CLIP ribonucleoprotein data, Wiley Interdiscip. Rev. RNA 8 (2017), https://doi.org/10.1002/wrna.1404.
[55] Y. Wang, X. Xiao, J. Zhang, R. Choudhury, A. Robertson, K. Li, M. Ma, C.B. Burge, Z. Wang, A complex network of factors with overlapping affinities represses splicing through intronic elements, Nat. Struct. Mol. Biol. 20 (2013) 36–45, https://doi.org/10.1038/nsmb.2459.
[56] Y. Wang, M. Ma, X. Xiao, Z. Wang, Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules, Nat. Struct. Mol. Biol. 19 (2012) 1044–1052, https://doi.org/10.1038/nsmb.2377.
[57] M. Krawczak, N.S. Thomas, B. Hundrieser, M. Mort, M. Wittig, J. Hampe, D.N. Cooper, Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing, Hum. Mutat. 28 (2007) 150–158, https://doi.org/10.1002/humu.20400.
[58] A. Abramowicz, M. Gos, Splicing mutations in human genetic disorders: examples, detection, and confirmation, J. Appl. Genet. 59 (2018) 253–268, https://doi.org/10.1007/s13353-018-0444-7.
[59] M.S. Falzarano, C. Scotton, C. Passarelli, A. Ferlini, Duchenne Muscular Dystrophy: from diagnosis to therapy, Molecules 20 (2015) 18168–18184, https://doi.org/10.3390/molecules201018168.
[60] A. Nishiyama, Y. Takeshima, Z. Zhang, Y. Habara, T.H. Tran, M. Yagi, M. Matsuo, Dystrophin nonsense mutations can generate alternative rescue transcripts in lymphocytes, Ann. Hum. Genet. 72 (2008) 717–724, https://doi.org/10.1111/j.1469-1809.2008.00468.x.
[61] S. Gonzalo, R. Kreienkamp, P. Askjaer, Hutchinson-Gilford Progeria Syndrome: a premature aging disease caused by LMNA gene mutations, Ageing Res. Rev. 33 (2017) 18–29, https://doi.org/10.1016/j.arr.2016.06.007.
[62] P. Scaffidi, T. Misteli, Lamin A-dependent nuclear defects in human aging, Science 312 (2006) 1059–1063, https://doi.org/10.1126/science.1127168.
[63] S. Rodriguez, F. Coppede, H. Sagelius, M. Eriksson, Increased expression of the Hutchinson-Gilford progeria syndrome truncated lamin A transcript during cell aging, Eur. J. Hum. Genet. 17 (2009) 928–937, https://doi.org/10.1038/ejhg.2008.270.
[64] M. Mine, M. Brivet, G. Touati, P. Grabowski, M. Abitbol, C. Marsac, Splicing error in E1alpha pyruvate dehydrogenase mRNA caused by novel intronic mutation responsible for lactic acidosis and mental retardation, J. Biol. Chem. 278 (2003) 11768–11772, https://doi.org/10.1074/jbc.M211106200.
[65] G.K. Brown, L.J. Otero, M. LeGris, R.M. Brown, Pyruvate dehydrogenase deficiency, J. Med. Genet. 31 (1994) 875–879.
[66] A.L. Brillen, K. Schoneweis, L. Walotka, L. Hartmann, L. Muller, J. Ptok, W. Kaisers, G. Poschmann, K. Stuhler, E. Buratti, S. Theiss, H. Schaal, Succession of splicing regulatory elements determines cryptic 5ss functionality, Nucleic Acids Res. 45 (2017) 4202–4216, https://doi.org/10.1093/nar/gkw1317.
[67] C.L. Lorson, E. Hahnen, E.J. Androphy, B. Wirth, A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 6307–6311.
[68] J. Seo, M.D. Howell, N.N. Singh, R.N. Singh, Spinal muscular atrophy: an update on therapeutic progress, Biochim. Biophys. Acta 1832 (2013) 2180–2190, https://doi.org/10.1016/j.bbadis.2013.08.005.
[69] R.N. Singh, N.N. Singh, Mechanism of splicing regulation of Spinal Muscular Atrophy genes, Adv. Neurobiol. 20 (2018) 31–61, https://doi.org/10.1007/978-3-319-89689-2_2.
[70] K. Nakai, H. Sakamoto, Construction of a novel database containing aberrant splicing mutations of mammalian genes, Gene 141 (1994) 171–177.
[71] E. Buratti, M. Chivers, G. Hwang, I. Vorechovsky, DBASS3 and DBASS5: databases of aberrant 3′- and 5′-splice sites, Nucleic Acids Res. 39 (2011) D86–D91, https://doi.org/10.1093/nar/gkq887.
[72] P.D. Stenson, M. Mort, E.V. Ball, K. Shaw, A. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, Hum. Genet. 133 (2014) 1–9, https://doi.org/10.1007/s00439-013-1358-4.
[73] M.J. Landrum, J.M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, D.R. Maglott, ClinVar: public archive of interpretations of clinically relevant variants, Nucleic Acids Res. 44 (2016) D862–D868, https://doi.org/10.1093/nar/gkv1222.

[74] K. Ohno, J.I. Takeda, A. Masuda, Rules and tools to predict the splicing effects of exonic and intronic mutations, Wiley Interdiscip. Rev. RNA 9 (2018), https://doi.org/10.1002/wrna.1451.

[75] I. Paz, I. Kosti, M. Ares Jr., M. Cline, Y. Mandel-Gutfreund, RBPmap: a web server for mapping binding sites of RNA-binding proteins, Nucleic Acids Res. 42 (2014) W361–W367, https://doi.org/10.1093/nar/gku406.

[76] M. Akerman, H. David-Eden, R.Y. Pinter, Y. Mandel-Gutfreund, A computational approach for genome-wide mapping of splicing factor binding sites, Genome Biol. 10 (2009) R30, https://doi.org/10.1186/gb-2009-10-3-r30.

[77] G. Yeo, S. Hoon, B. Venkatesh, C.B. Burge, Variation in sequence and organization of splicing regulatory elements in vertebrate genes, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 15700–15705, https://doi.org/10.1073/pnas.0404901101.

[78] Z. Wang, M.E. Rolish, G. Yeo, V. Tung, M. Mawson, C.B. Burge, Systematic identification and analysis of exonic splicing silencers, Cell 119 (2004) 831–845, https://doi.org/10.1016/j.cell.2004.11.010.

[79] F. Piva, M. Giulietti, L. Nocchi, G. Principato, SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans, Bioinformatics 25 (2009) 1211–1213, https://doi.org/10.1093/bioinformatics/btp124.

[80] M. Raponi, J. Kralovicova, E. Copson, P. Divina, D. Eccles, P. Johnson, D. Baralle, I. Vorechovsky, Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6, Hum. Mutat. 32 (2011) 436–444, https://doi.org/10.1002/humu.21458.

[81] F.O. Desmet, D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres, C. Beroud, Human Splicing Finder: an online bioinformatics tool to predict splicing signals, Nucleic Acids Res. 37 (2009) e67, https://doi.org/10.1093/nar/gkp215.

[82] S. Schwartz, E. Hall, G. Ast, SROOGLE: webserver for integrative, user-friendly visualization of splicing signals, Nucleic Acids Res. 37 (2009) W189–W192, https://doi.org/10.1093/nar/gkp320.

[83] K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J.F. McRae, S.F. Darbandi, D. Knowles, Y.I. Li, J.A. Kosmicki, J. Arbelaez, W. Cui, G.B. Schwartz, E.D. Chow, E. Kanterakis, H. Gao, A. Kia, S. Batzoglou, S.J. Sanders, K.K. Farh, Predicting splicing from primary sequence with deep learning, Cell 176 (2019) 535–548 e524 https://doi.org/10.1016/j.cell.2018.12.015.

[84] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, Science 249 (1990) 505–510.

[85] G. Giudice, F. Sanchez-Cabo, C. Torroja, E. Lara-Pezzi, ATtRACT-a Database of RNA-binding Proteins and Associated Motifs, Database: The Journal of Biological Databases and Curation 2016, (2016), https://doi.org/10.1093/database/baw035.

[86] F. Piva, M. Giulietti, A.B. Burini, G. Principato, SpliceAid 2: a database of human splicing factors expression data and RNA target motifs, Hum. Mutat. 33 (2012) 81–85, https://doi.org/10.1002/humu.21609.

[87] X.H. Zhang, L.A. Chasin, Computational definition of sequence motifs governing constitutive exon splicing, Genes Dev. 18 (2004) 1241–1250, https://doi.org/10.1101/gad.1195304.

[88] Z.X. Lu, P. Jiang, J.J. Cai, Y. Xing, Context-dependent robustness to 5′ splice site polymorphisms in human populations, Hum. Mol. Genet. 20 (2011) 1084–1096, https://doi.org/10.1093/hmg/ddq553.

[89] L. Grodecka, E. Buratti, T. Freiberger, Mutations of pre-mRNA splicing regulatory elements: are predictions moving forward to clinical diagnostics? Int. J. Mol. Sci. 18 (2017), https://doi.org/10.3390/ijms18081668.

[90] O. Soukarieh, P. Gaildrat, M. Hamieh, A. Drouet, S. Baert-Desurmont, T. Frebourg, M. Tosi, A. Martins, Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools, PLoS Genet. 12 (2016) e1005756, , https://doi.org/10.1371/journal.pgen.1005756.

[91] T.E. Abbink, B. Berkhout, RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor, J. Virol. 82 (2008) 3090–3098, https://doi.org/10.1128/JVI.01479-07.

[92] D. Zychlinski, S. Erkelenz, V. Melhorn, C. Baum, H. Schaal, J. Bohne, Limited complementarity between U1 snRNA and a retroviral 5′ splice site permits its attenuation via RNA secondary structure, Nucleic Acids Res. 37 (2009) 7429–7440, https://doi.org/10.1093/nar/gkp694.

[93] J. Tan, L. Yang, A.A.L. Ong, J. Shi, Z. Zhong, M.L. Lye, S. Liu, J. Lisowiec-Wachnicka, R. Kierzek, X. Roca, G. Chen, A disease-causing intronic point mutation C19G alters tau exon 10 splicing via RNA secondary structure rearrangement, Biochemistry (Mosc) 58 (2019) 1565–1578, https://doi.org/10.1021/acs.biochem.9b00001.

[94] R. Soemedi, K.J. Cygan, C.L. Rhine, D.T. Glidden, A.J. Taggart, C.L. Lin, A.M. Fredericks, W.G. Fairbrother, The effects of structure on pre-mRNA processing and stability, Methods 125 (2017) 36–44, https://doi.org/10.1016/j.ymeth.2017.06.001.

[95] D. Ray, H. Kazan, K.B. Cook, M.T. Weirauch, H.S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L.H. Matzat, R.K. Dale, S.A. Smith, C.A. Yarosh, S.M. Kelly, B. Nabet, D. Mecenas, W. Li, R.S. Laishram, M. Qiao, H.D. Lipshitz, F. Piano, A.H. Corbett, R.P. Carstens, B.J. Frey, R.A. Anderson, K.W. Lynch, L.O. Penalva, E.P. Lei, A.G. Fraser, B.J. Blencowe, Q.D. Morris, T.R. Hughes, A compendium of RNA-binding motifs for decoding gene regulation, Nature 499 (2013) 172–177, https://doi.org/10.1038/nature12311.

[96] D. Dominguez, P. Freese, M.S. Alexis, A. Su, M. Hochman, T. Palden, C. Bazile, N.J. Lambert, E.L. Van Nostrand, G.A. Pratt, G.W. Yeo, B.R. Graveley, C.B. Burge, Sequence, structure, and context preferences of human RNA binding proteins, Mol. Cell 70 (2018) 854–867 e859 https://doi.org/10.1016/j.molcel.2018.05.001.

[97] K.B. Cook, T.R. Hughes, Q.D. Morris, High-throughput characterization of protein-RNA interactions, Brief. Funct. Genomics 14 (2015) 74–89, https://doi.org/10.1093/bfgp/elu047.

[98] L. Zeng, W. Liu, W. Feng, X. Wang, H. Dang, L. Gao, J. Yao, X. Zhang, A novel donor splice-site mutation of major intrinsic protein gene associated with congenital cataract in a Chinese family, Mol. Vis. 19 (2013) 2244–2249.

[99] Y. Habara, Y. Takeshima, H. Awano, Y. Okizuka, Z. Zhang, K. Saiki, M. Yagi, M. Matsuo, In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G– > A mutations in introns of the dystrophin gene, J. Med. Genet. 46 (2009) 542–547, https://doi.org/10.1136/jmg.2008.061259.

[100] L.J. Fang, M.J. Simard, D. Vidaud, B. Assouline, B. Lemieux, M. Vidaud, B. Chabot, J.P. Thirion, A novel mutation in the neurofibromatosis type 1 (NF1) gene promotes skipping of two exons by preventing exon definition, J. Mol. Biol. 307 (2001) 1261–1270, https://doi.org/10.1006/jmbi.2001.4561.

[101] T. Hori, T. Fukao, K. Murase, N. Sakaguchi, C.O. Harding, N. Kondo, Molecular basis of two-exon skipping (exons 12 and 13) by c.1248+5g > a in OXCT1 gene: study on intermediates of OXCT1 transcripts in fibroblasts, Hum. Mutat. 34 (2013) 473–480, https://doi.org/10.1002/humu.22258.

[102] K. Takahara, U. Schwarze, Y. Imamura, G.G. Hoffman, H. Toriello, L.T. Smith, P.H. Byers, D.S. Greenspan, Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptor-site mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I, Am. J. Hum. Genet. 71 (2002) 451–465, https://doi.org/10.1086/342099.

[103] B.L. Robberson, G.J. Cote, S.M. Berget, Exon definition may facilitate splice site selection in RNAs with multiple exons, Mol. Cell. Biol. 10 (1990) 84–94, https://doi.org/10.1128/Mcb.10.1.84.

[104] B.E. Hoffman, P.J. Grabowski, U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon, Genes Dev. 6 (1992) 2554–2568.

[105] S. Ke, L.A. Chasin, Context-dependent splicing regulation: exon definition, co-occurring motif pairs and tissue specificity, RNA Biol. 8 (2011) 384–388.

[106] R. Reed, T. Maniatis, A role for exon sequences and splice-site proximity in splice-site selection, Cell 46 (1986) 681–690.

[107] A.B. Rosenberg, R.P. Patwardhan, J. Shendure, G. Seelig, Learning the sequence determinants of alternative splicing from millions of random sequences, Cell 163 (2015) 698–711, https://doi.org/10.1016/j.cell.2015.09.054.

[108] M.J. Hicks, W.F. Mueller, P.J. Shepard, K.J. Hertel, Competing upstream 5′ splice sites enhance the rate of proximal splicing, Mol. Cell. Biol. 30 (2010) 1878–1886, https://doi.org/10.1128/MCB.01071-09.

[109] D. Baralle, E. Buratti, RNA splicing in human disease and in the clinic, Clin Sci (Lond) 131 (2017) 355–368, https://doi.org/10.1042/CS20160211.

[110] N.N. Singh, M.D. Howell, E.J. Androphy, R.N. Singh, How the discovery of ISS-N1 led to the first medical therapy for spinal muscular atrophy, Gene Ther. 24 (2017) 520–526, https://doi.org/10.1038/gt.2017.34.

[111] E.H. Niks, A. Aartsma-Rus, Exon skipping: a first in class strategy for Duchenne Muscular Dystrophy, Expert. Opin. Biol. Ther. 17 (2017) 225–236, https://doi.org/10.1080/14712598.2017.1271872.

3.1.2. Modelling splicing outcome by combining 5’ss strength and splicing regulatory elements (Chapter 2)

As discussed above, precise recognition and usage of splice sites is critical for the correct processing of pre-mRNA. Unfortunately, single point mutations in splicing signals including splicing regulatory elements or splice sites themselves can have devastating effects on the splicing outcome and can contribute to severe diseases. The prediction and robust modeling of the splicing outcome could aid the evaluation of potentially pathogenic splicing mutations in human genetic diagnostics. This chapter aims to provide experimental foundations for the development of an algorithm to describe a functional splice site strength by combining its intrinsic 5’ss strength measured by U1 snRNA complementarity and scoring of the SREs neighborhood to create a 5’ss usage landscape.

To achieve this, *in silico* designed sequences with prescribed splicing regulatory properties were tested for the accuracy of the bioinformatics predictions using splicing reporter assays. Furthermore, the respective splicing regulatory proteins binding the *in silico* designed sequences were identified using a pull-down assay.

With these predictions being accurate, a novel RNA-seq-based 5'ss utilization landscape was developed based on the HEXplorer algorithm used to design the sequences. It maps the competition of over 320,000 pairs of 5'ss and adjacent exonic GT dinucleotides along the axis of splice site strength and SRE strength. The resulting 5'ss utilization landscape provides a unified view of the impact of 5' splice site strength and the role of neighboring SREs on splice site recognition and the distinction between comparably strong but unused exonic GT sites and utilized 5'ss. This can contribute to the improved evaluation of human pathogenic mutations.

The following article is under extended revision in Nucleic Acids Research.

Müller L, Ptok J, Antemann J, Hillebrand F, Brillen AL, Theiss S*, Schaal H*.

Contributions:

L.M. S.T. and H.S. conceived the study and designed the experiments. L.M., J.A. A.L.B. and F.H. performed cloning, transfection experiments, RNA-pull-down analyses

and (q)RT-PCR analyses. S.T. and J.P. carried out bioinformatical analyses. L.M., J.P., S.T. and H.S. wrote the manuscript.

Individual contribution: 40% L.M. designed and cloned the constructs and performed the experiments that are shown in Figure 1 and 2. These provide the basis for constructs shown in Figure 3 that were generated during a master's project under supervision of L.M. L.M. performed the experiment in Figure 4. L.M. wrote the first draft of the manuscript.

Modeling splicing outcome by combining 5'ss strength and splicing regulatory elements

Lisa Müller[1], Johannes Ptok[1], Jennifer Antemann[1], Frank Hillebrand[1], Anna-Lena Brillen[1], Stephan Theiss[1]* and Heiner Schaal[1]*.

[1] Institute of Virology, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, 40225 Germany

* To whom correspondence should be addressed. Tel: +49 211 81 12393; Fax: +49 211 81 10856; Email: schaal@uni-duesseldorf.de. Correspondence may also be addressed to theiss@uni-duesseldorf.de.

Present Address: Anna-Lena Brillen, Institute of Virology, Faculty of Medicine, University of Bonn, Bonn, 53127, Germany

ABSTRACT

Correct pre-mRNA processing in higher eukaryotes vastly depends on splice site recognition. Beyond conserved 5'ss and 3'ss motifs, splicing regulatory elements (SREs) play a pivotal role in this recognition process. Here, we present *in silico* designed sequences with arbitrary *a priori* prescribed splicing regulatory HEXplorer properties that can be concatenated to arbitrary length without changing their regulatory properties. We confirmed *in silico* predictions in splicing reporter assays and identified computationally expected SRE binding proteins. Aiming at a unified "functional splice site strength" encompassing both U1 snRNA complementarity and impact from neighboring SREs, we developed a novel RNA-seq based 5'ss usage landscape, mapping the competition of 320,601 pairs of high confidence 5'ss and neighboring exonic GT sites along HBond and HEXplorer score coordinate axes. For appropriate aggregation of RNA-seq reads, we applied a two-tier gene-and-sample reads normalization. This served as basis for a logistic 5'ss usage prediction model, which greatly improved discrimination between strong but unused exonic GT sites and highly used 5'ss. Our 5'ss usage landscape offers a unified view on 5'ss and SRE neighborhood impact on splice site recognition, and may contribute to improved mutation assessment in human genetics.

INTRODUCTION

For almost all human primary protein coding transcripts recognition of splice sites, the borders between exons and introns, is key in deciphering their open reading frames. In order to accurately ligate exons after intron removal, splice sites at exon-intron-borders need to be recognized with single nucleotide precision during early assembly of the spliceosome. Splice site recognition depends upon conserved sequence motifs at both intron ends, and the first step in the splicing process is splice donor recognition by the U1 snRNP at a highly conserved GT dinucleotide (1).

Formation of an RNA duplex between up to 11 nucleotides (nt) of the splice donor (5'ss) with the 5' end of U1 snRNA is a main determinant in 5'ss selection (2-4). The statistical likelihood of a 9 nt long potential 5'ss sequence being used as 5'ss is frequently quantified by its maximum entropy based MaxEnt score (5), while the HBond score (HBS) algorithm based on all 11 nt quantifies the U1 snRNA complementarity of a potential 5'ss (https://www2.hhu.de/rna// (6,7)). However, exons and introns contain numerous GT sites with high MaxEnt and HBond scores indicating potential 5'ss, which are not used as exon-intron-borders.

Thus, 5'ss complementarity to U1 snRNA cannot alone be responsible for splice site use (8). The efficiency with which splice sites are recognized additionally depends on proximal *cis*-acting splicing regulatory elements (SREs) and their protein binding partners including SR (serine-arginine-rich) (9,10) and hnRNP (heterogeneous nuclear ribonucleoparticle) proteins (11,12). Generally, proteins bound by SREs act in a position dependent way: SR proteins have enhancing properties on downstream located 5'ss and repress upstream located 5'ss, while hnRNP proteins act reversely (13,14). Mechanistically, splicing regulatory proteins (SRPs) may impact U1 snRNA duplex stability due to allosteric regulation of U1 snRNP structure (15). Through these combined SRP binding effects, the sequence neighborhood of a splice site can have a significant impact on splice site recognition and hence splicing efficiency (16-19). Especially with regard to an estimated at least 25% of human inherited diseases caused by mutations either directly altering splice sites or disrupting SREs in their vicinity (20,21), computational evaluation of a possibly pathogenic impact of individual SNVs is important for human genetics (22-27).

Various algorithms and corresponding computational tools have been developed and made publicly available to analyze splicing regulatory elements: some algorithms identify previously described hexamer or octamer motifs (e.g. ESEfinder, FAS-ESS, RESCUE-ESE, PESX, cf. e.g. (28,29)), others provide e.g. hexamer weights quantifying their splice enhancing or silencing properties, and enabling the calculation of SRE profiles in moving windows along genomic sequences (ESR-seq (30), HEXplorer (31)). Most recently, neural network or deep-learning based algorithms for splicing prediction have been developed that take splice sites and their neighborhoods or very wide sequence contexts into account (MMSplice (32), SpliceAI (33)) (4,34).

Minigene splicing reporters are widely used model systems to experimentally examine splicing. In particular, massively parallel splicing assays (MPSA) permit screening the impact on splicing for a large number of randomly generated sequences in a single experiment. These random sequences can e.g. cover a 5'ss position, various specific exonic k-*mer* positions, or be spread out across an entire exon. For each individual "input" sequence, an RNA-seq based enrichment index quantifies the sequence impact on splicing, the "output" (30,35).

Here, we followed the inverse route of an *in-silico* sequence design process for a controlled SRE containing segment between two competing 5'ss in a splicing reporter. In particular, we designed segments with well-defined HEXplorer profiles representing average splice enhancing, silencing or neutral properties that are adaptable to different lengths, and we observed switching between the competing 5'ss. In addition, we systematically co-varied 5'ss complementarity and SRE properties of 5'ss neighborhoods. Complementarily, we examined splice site competition in our large whole transcriptome RNA-seq dataset and derived a two-dimensional splice site usage landscape for all combinations of 5'ss complementarity and SRE neighborhood.

MATERIAL AND METHODS

Expression plasmids

pXGH5 (hGH) (36) was cotransfected to monitor transfection efficiency.

Oligonucleotides

All oligonucleotides used were obtained from Metabion GmbH (Planegg, Germany) (see Supplementary File S1).

Cloning

A reporter construct based on the HIV-1 glycoprotein/eGFP expression plasmid (6,13) as well as a 3-exon minigene based on the fibrinogen Bß subunit under the control of a cytomegalovirus immediate early (CMVie) promoter (37) were used in this study. All sequences were cloned using either PCR-products of the respective forward and reverse primer pairs or DNA fragments. Detailed cloning strategies and primer sequences can be found in Supplementary File S1.

Cell culture and RT-PCR analysis

HeLa cells (ATCC® CCL-2™, mycoplasma free) were cultivated in Dulbecco's high-glucose modified Eagle's medium (Gibco #41966) supplemented with 10% fetal calf serum (PAN Biotech #P30-3031) and 50 µg/ml penicillin and streptomycin each (Gibco #15140-122). Transient-transfection experiments were performed with six-well plates at 2.5×10^5 cells per well by using TransIT®-LT1 transfection reagent (Mirus Bio LLC US #MIR2305) according to the manufacturer's instructions. Total RNA was isolated 24 h post-transfection by using acid guanidinium thiocyanate-phenol-chloroform as described previously (38). For (q)RT-PCR analyses, RNA was reversely transcribed by using Superscript III Reverse Transcriptase (Invitrogen #18080-085) and Oligo(dT) primer (Roche #10814270001). For the analyses of the splicing constructs either primer pair #3210/#3211(#640) or #2648/2649 was used and PCRs were separated on non-denaturing 10% polyacrylamide gels. Quantitative RT-PCR analysis was performed by using the qPCR MasterMix (PrimerDesign Ltd #PPLUS-CL-SY-10ML) and Roche LightCycler 1.5. For normalization, primers #1224/#1225 were used to monitor the level of the transfection control hGH present in each sample.

Protein isolation by RNA affinity chromatography

Substrate RNAs were in vitro transcribed using theT7 RiboMaxTM Express Large Scale RNA Production System (Promega #P1320) according to the manufacturer's recommendations. Three thousand picomoles of the substrate RNA oligonucleotides for each octamer (+10.32 #5648, –0.15 #5647, –10.35 #5846) were covalently coupled to adipic acid dihydrazideagarose beads (Sigma #40802-10ML). 60% of HeLa nuclear extract (SKU: CC-01-20-50, Cilbiotech/now Ipracell #CC-01-20-50) was added to the immobilized RNAs. After stringent washing with buffer D containing different concentrations of KCl (20mM HEPES-KOH [pH 7.9], 5%[vol/vol] glycerol, 0.1-0.5 M KCl, 0.2 M ethylenediaminetetraacetic acid, 0.5 mM dithiothreitol, 0.4M MgCl2), precipitated proteins were eluted in protein sample buffer. Samples were heated up to 95°C for 10 min and either submitted to LC-MS/MS-analysis or loaded onto sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE) for western blot analysis. Samples were transferred to a nitrocellulose membrane probed with primary and secondary antibodies (SRSF3 (Abcam ab198291, 1:000), PTB (kind gift from Douglas Black, 1:1000), hnRNPD (Merk Millipore AUF-1 07-260, 1:1000), MS2 (Tetracore TC-7004-002, 1:1000), Goat anti-Rabbit IgG Superclonal™ Secondary Antibody (Invitrogen A27036, 1:2500) and developed with ECL chemiluminescence reagent (GE Healthcare #RPN2106).

HEXplorer score algorithm and Splice Site HEXplorer Weight (SSHW)

Based on a RESCUE-type approach, the HEXplorer score HZ_{EI} is calculated from different hexamer occurrences in exonic and intronic sequences in the neighborhood of splice donors, and it has been successfully used for the identification of exonic splicing regulatory elements (31,37,39). Briefly, from 43,464 constitutively spliced human exons with canonical 5'ss collected from ENSEMBL (29), Z-scores for all 4,096 hexamers were calculated from normalized hexamer frequency differences up- and downstream of weak and strong splice donors, ranging from -73 for TTTTT to +34 for GAAGAA.

The HEXplorer score HZ_{EI} of any index nucleotide in a genomic sequence is then calculated as average hexamer Z-score of all six hexamers overlapping with this index nucleotide. This algorithm permits plotting HEXplorer score profiles along genomic sequences, and these profiles reflect splice enhancing or silencing properties in the

neighborhood of a splice donor: HEXplorer score positive regions support downstream splice donors and repress upstream ones, and HZ_{EI} negative regions *vice versa*. HEXplorer score profiles of genomic sequences were calculated using the web interface (https://www2.hhu.de/rna/html/hexplorer_score.php).

As measure of SRE impact on 5'ss recognition, we calculated the 5' splice site HEXplorer weight SSHW as the total HZ_{EI} sum in a 50 nt upstream minus the symmetrical 50 nt downstream neighborhood (37,40), excluding all 11 nt of the 5'ss from the HZ_{EI} calculation: the 50 nt wide neighborhoods ended at exonic position -4 and started at intronic position +9, respectively. This definition has been made analogous to the "exonic splicing motif difference" ESMD introduced by Ke et al. and to the "splice site enhancer weight" by Brillen et al. (37,40), and it captures both enhancing and silencing properties of 50 nt wide up- and downstream regions that have been used before and are plausibly considered to contain relevant SREs.

When comparing SSHW of pairs of exonic GT sites and 5'ss, we carefully adapted the selection of appropriate neighborhoods depending on the GT-site-to-5'ss distance: If GT site and 5'ss were more than 60 nt apart, we used 50 nt wide neighborhoods A, B1, B2 and C as depicted in Suppl. Fig. 4B. For pairs of GT site and 5'ss that were between 61 nt and 111 nt apart, the neighborhoods B1 and B2 consequently overlapped. If GT site and 5'ss were closer than 61 nt, we chose B1 = B2 as the entire region between but excluding the two sites. We then calculated the SSHW difference between GT site and 5'ss as $\Delta SSHW = (\sum_A - \sum_{B1} - \sum_{B2} + \sum_C) HZ_{EI}$ (Suppl. Fig. 4B).

The receiver operating characteristic curves (ROC) obtained for the three regression models was generated using the R-package ROCit (version 1.1.1).

Mass spectrometric analysis

Protein samples were shortly separated over about 4 mm running distance in a 4–12% polyacrylamide gel. After silver staining, protein containing bands were excised and prepared for liquid chromatography-tandem mass spectrometry (LC-MS/MS) as described previously (37).

RNA Sequencing data generation and processing

We re-analyzed a RNA sequencing data set originating from 46 samples of primary fibroblasts, that was previously described (41). Briefly, the cDNA libraries were created using TruSeq RNA SamplePrep kit (Illumina) after poly(A) enrichment according to the manufacturer's protocol. Afterwards, the samples were amplified on 9 Illumina flow cells and sequenced on a Illumina HiSeq 2000 sequencer. Subsequently the resulting 101-nt sequence segments were converted to Fastq by CASAVA (1.8.2). The samples were checked for base calling quality during sequencing, sub-sequences of a single read with low average base calling quality as well as left over adapters from library preparation were removed using Trimmomatic version 0.36 (42). Trimmed reads shorter than 75 bases were discarded since this length is an established threshold in the analysis concerning exon junctions (43). The tool sortMeRNA was used to validate complete rRNA removal during poly(A) RNA enrichment (44). Throughout the different steps of FASTQ file processing, the quality of the reads was assessed using the tools FASTQC and MultiQC. After processing the FASTQ files, the reads were mapped to the ENSEMBL human reference genome (version 91) using the STAR software package (2.5.4b). The reads were aligned to the reference following the two-pass mapping protocol recommended for splice site usage analysis (45,46). After Alignment with STAR, the BAM files were summarized to a single gap file using CRAN package rbamtools (47) and Bioconductor package spliceSites (48). Additional packages were used during the analysis. Fastq file preparation and alignment, as well as the first part of BAM file processing in R was accomplished using custom BASH shell scripts in the environment of the High Performing Cluster of the Heinrich-Heine University Düsseldorf. Computational support and infrastructure was provided by the "Centre for Information and Media Technology" (ZIM) at the University of Düsseldorf (Germany).

RESULTS

Periodic sequences: hexamers and octamers

Computational tools for the identification of splicing regulatory elements, such as ESEfinder, RESCUE-ESE, PESX, FAS-ESS, ESRseq (QUEPASA), and—more indirectly—SPANR (49) or HAL (50) can be used to quantify the splicing regulatory properties of a given 5'ss neighborhood. Assigning a weight of +1 or –1 to each predicted splicing enhancer or silencer motif, a 5' splice site enhancer weight (SSEW)

can be calculated as the total of all upstream SRE weights minus all downstream SRE weights (37,40). By definition, this SSEW is a measure of SRE support for 5'ss recognition, and it takes position- and direction dependent SRE action into account. Here, we took an alternative approach, using the HEXplorer algorithm that provides a numerical weight HZEI for each nucleotide of a genomic sequence, reflecting the average exonic splicing regulatory properties of all six hexamers overlapping with this nucleotide. Analogous to SSEW, we calculated the 5' splice site HEXplorer weight SSHW as the total HZEI of the upstream minus the downstream neighborhood.

Since single RNA-recognition motifs (RRMs) of splicing regulatory proteins are thought to bind up to eight nucleotides (51), and in line with motif lengths applied by computational tools, we aimed at in silico designing 5'ss neighborhoods with specific a priori prescribed splicing regulatory properties (SSHW) by concatenating octamers rather than inserting SRP binding motifs into presumed neutral 5'ss neighborhoods. In particular, we sought to find periodic octamer sequences with approximately constant HEXplorer score profile (HZEI ≈ const.). By definition, HEXplorer score profiles of periodic sequences (with period ≥ 6 nt) have the same periodicity as these sequences. Thus, for octamer repeats, up to eight different HZEI values occur in the HEXplorer profile, and they repeat every eight nucleotides. Suppl. Fig. 1A exemplarily shows HEXplorer profiles of two 48nt long sequences consisting of six octamer repeats with approximately the same average HZEI (5.22—5.26), but different amplitude variations around the average. Concatenation of different octamers, however, can induce large amplitude variations in the transition region. In Suppl. Fig. 1B, HEXplorer profiles of two sequences, each composed of six octamer repeats with nearly constant (CV = 0.10) average HZEI = 5.22 are displayed (Suppl. Fig. 1B, i and ii). Concatenating three octamer repeats of each of the two octamers leads to a pronounced "dip" in the HEXplorer profile at the transition (Suppl. Fig. 1B, iii).

We therefore systematically searched for octamer sequences that upon concatenation show little HEXplorer score variation around their average HZEI value. To this end, we calculated average and standard deviation of HZEI values for all 65,536 possible octamers from 5-fold concatenations (i.e. 40nt long sequences). In order to avoid accidentally creating 5'ss or 3'ss in the designed sequences, we excluded octamers containing a GT or AG dinucleotide, or creating one by concatenation, with 23,120

octamers remaining. Limiting HZEI variation to standard deviation below 2 still leaves 18,925 octamers in the HZEI/nt range from –20 to +14. The octamer histogram in Suppl. Fig. 1C displays the number of different octamers for all HZEI/nt intervals. Each bin contains sequences with standard deviation below 0.5.

From this set of extremely low HZEI variability octamers, we selected three test octamers with average HZEI of +10.32 (ACCACCGG), –0.15 (CCTATTGG) and –10.35 (AATTCTCT). Note that there are only four nucleotides different between octamers +10.32 and –0.15 (Fig. 1A).

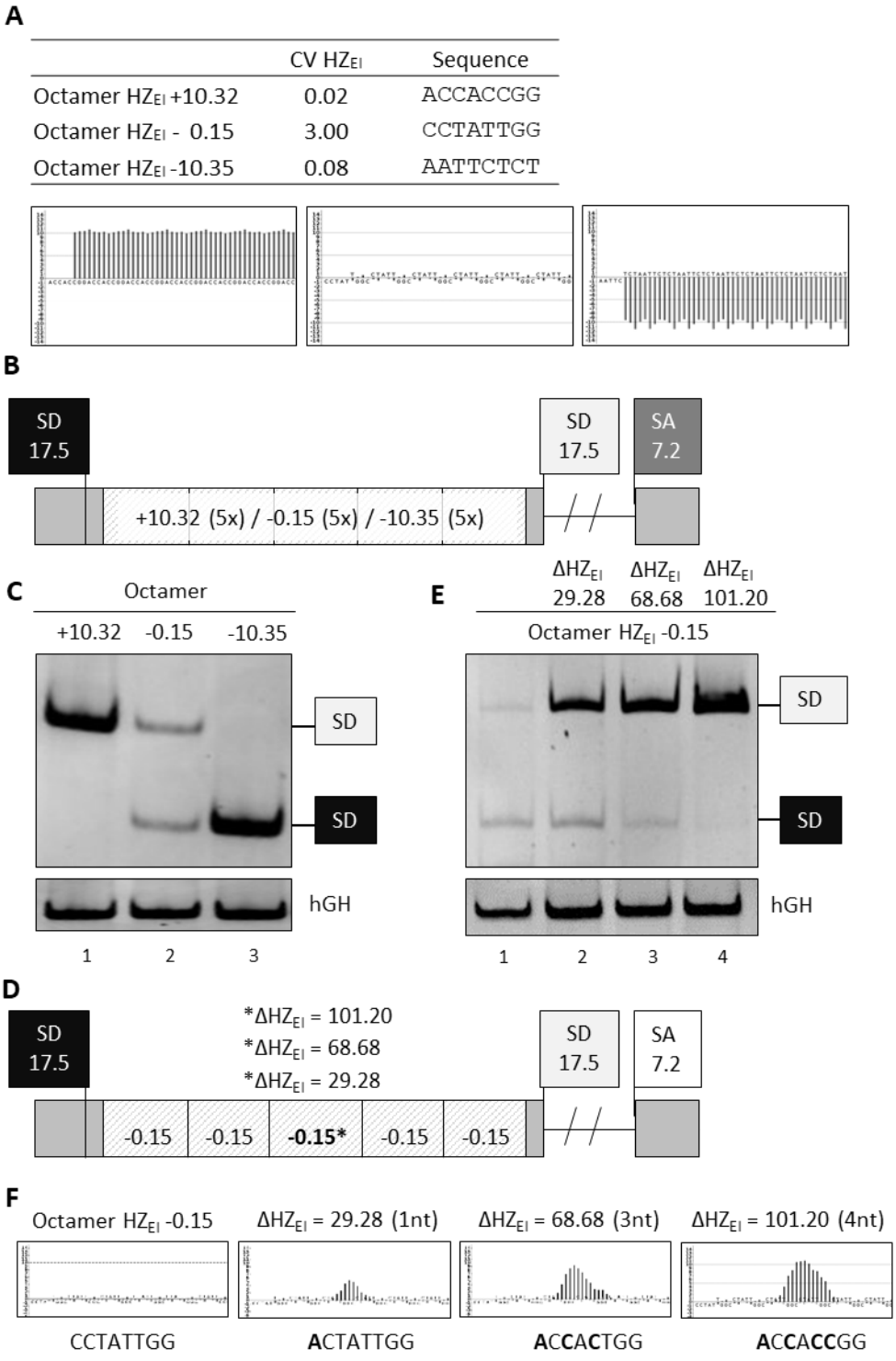


Figure 1: HEXplorer guided sequences mutations shift splice donor use. A *HEXplorer predicted positive, neutral and negative periodic octamer sequences.* ***B*** *Schematic drawing of the reporter construct that contains two equally strong splice donors with an HBond score of 17.5 (MaxEnt 10.10) and is used to detect up- or downstream enhancing or silencing properties of the HEXplorer predicted sequences.* ***C*** *2.5 × 10^5 HeLa cells were transiently transfected with 1µg of each construct together with 1µg of pXGH5 (hGH) to monitor transfection efficiency. Twenty-four hours after transfection, RNA was*

isolated and subjected to RT-PCR analysis using primer pairs #3210/#3211 and #1224/#1225 (hGH). PCR products were separated by a 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide. ***D*** *Schematic drawing of the reporter construct that contains two equally strong splice donors SD with an HBond score of 17.5 (MaxEnt 10.10) and is used to detect up- or downstream enhancing or silencing properties of the HEXplorer predicted sequences. Point mutations in the central splicing neutral octamer –0.15 sequence increase the positive HEXplorer plot area indicated by the positive* ΔHZ_{EI} *and lead to a stepwise increase in downstream (gray) SD use (right).* ***E*** *2.5 × 10^5 HeLa cells were transiently transfected with 1µg of each construct together with 1µg of pXGH5 (hGH) to monitor transfection efficiency. Twenty-four hours after transfection, RNA was isolated and subjected to RT-PCR analysis using primer pairs #3210/#3211 and #1224/#1225 (hGH). PCR products were separated by a 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide (left).*

Splicing reporter test of in silico designed octamers

In order to experimentally test the behavior of the *in silico* predicted 40nt sequences (Fig. 1A), they were inserted between two identical copies of a strong 5'ss sequence with an HBond score HBS = 17.5 (http://www2.hhu.de/rna/html/hbond_score.php (7)). These competing 5'ss define the 3' end of the first exon of the HIV-based two-exon splicing reporter whose RNA level depends on U1 snRNP binding to either the upstream or downstream 5'ss (Fig. 1B, (6)). The impact of the inserted 40nt sequences on splice site selection was analyzed by RT-PCR following transient transfection assays. Insertion of octamer +10.32 repeats resulted solely in recognition of the downstream 5'ss, while the use of the upstream donor was completely repressed, confirming its predicted splicing regulatory activity (Fig. 1C, lane 1). While the splicing neutral octamer -0.15 mediated between the two splice donors on a basal level (Fig. 1C, lane 2), insertion of octamer -10.35 repeats resulted in exclusive selection of the upstream located 5'ss (Fig. 1C, lane 3), confirming the predicted splicing regulatory activities of both sequences. In general, insertion of octamer +10.32 and octamer -10.35 drastically elevated the overall splice donor recognition while strictly following the position dependent action of splicing regulatory elements, whereas the splicing neutral octamer -0.15 led to less efficient overall splice site recognition and did not show any splice donor preference in this reporter.

In the next series of experiments, we implemented a stepwise transition in the central octamer from the splicing neutral octamer -0.15 to octamer +10.32 by nucleotide substitutions (Fig. 1D). The first single nt substitution increased the HEXplorer score by ΔHZ_{EI} = 29.28, a three-nt substitution by ΔHZ_{EI} = 68.68, and a four-nt substitution by ΔHZ_{EI} = 101.2.

Increasing the HEXplorer score HZ_{EI} of the neutral octamer -0.15 by ~30 led to an increase of overall splicing efficiency and shifted 5'ss usage to the downstream 5'ss (Fig. 1E, lane 2). Further increasing the HEXplorer score HZ_{EI} by ~40 (total change ΔHZ_{EI} ~70 from baseline), reduced upstream donor usage and increased downstream 5'ss usage even more (Fig. 1E, lane 3). Finally, substituting the central octamer -0.15 with a single +10.32 octamer further increased the enhancer capacity by ~30 (total change ΔHZ_{EI} ~100 from baseline), leading to the exclusive usage of the downstream splice donor site, while upstream donor usage could not be detected (Fig. 1E, lane 4). Thus, in this setting even a single octamer +10.32 within the otherwise HEXplorer neutral sequence -0.15 led to a complete switch to the downstream 5'ss, similar to the previously tested five octamer copies (cf. Fig. 1C, lane 1).

Interaction between splicing regulatory elements and splice site

In the previous experiments, we examined the competition between two identical 5'ss, and SRE impact on 5'ss selection for a single fixed 5'ss. In the next step, we determined 5'ss usage for different upstream splicing regulatory sequences and 5' splice sites in a modified HIV-based splicing reporter without the upstream competing 5'ss (Fig. 2A).

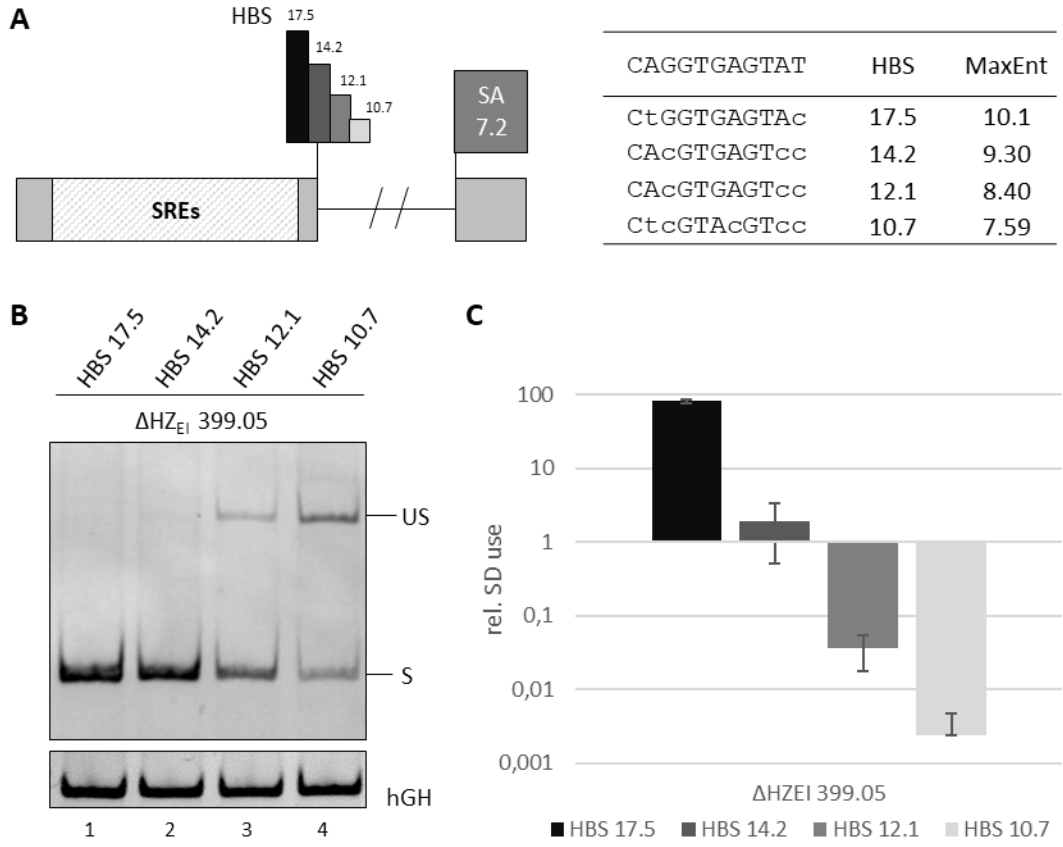


Figure 2: Interaction between SRE, intrinsic splice donor strength and splice site usage. A Sketch of the parental plasmid with differently strong SREs upstream of four different splice donors with decreasing intrinsic strength. Stepwise decrease in splice donor strength by nucleotide substitutions in comparison to wild-type U1 snRNA is measured by HBond score (HBS) and MaxEnt score. **B** 2.5×10^5 HeLa cells were transfected with 1µg of the reporter plasmids and 1µg of pXGH5 (hGH) that was used as a control for transfection efficiency. RNA was harvested and reverse transcribed into cDNA 24 hours post transfection with primer pair #3210/#3211/#640. PCR products were run on a 10% non-denaturing polyacrylamide gel. Insertion of octamer 10.32 with an ΔHZ_{EI} score of 399.05 in comparison to the splicing neutral octamer –0.15 upstream of the splice donor with decreasing HBS led to a stepwise decrease in SD usage, while the detection of unspliced message was increased respectively. **C** Relative SD use measured by qPCR confirmed the stepwise decrease in SD usage with decreasing HBS for octamer +10.32.

Specifically, we analyzed all 16 combinations of the above octamers with the following four 5'ss strengths (in decreasing order: HBS 17.5 > 14.2 > 12.1 > 10.7; Fig. 4B), comparing spliced and unspliced transcripts. In particular, we tested the central octamers with ΔHZ_{EI} ~30, ~70, ~100 flanked by two copies of octamer –0.15 on both sides, as well as five repeats of octamer +10.32 with ΔHZ_{EI} ~400 (Fig. 2A). HeLa cells were transfected with the respective plasmids, and after a 24h incubation period,

splice site usage was analyzed by RT-PCR. By using a primer trio for RT-PCR analysis, it was possible to detect both, the amount of 5'ss use and the amount of the unspliced message.

In Fig. 2B, we exemplarily present splice site usage decrease for all four 5'ss in combination with the strongest enhancing sequence of five octamer +10.32 repeats (ΔHZ_{EI} ~400). RT-PCR revealed a concomitant stepwise decrease in 5'ss usage along reduction of 5'ss strength, while simultaneously, an increase in the amount of unspliced message was detected (Fig. 2B). As expected, five octamer +10.32 repeats upstream of a 5'ss with an HBS of 17.5 showed the highest amount of 5'ss usage suggesting that this sequence strongly facilitated U1 snRNP binding to the splice site (Fig. 2B, lane 1). Decreasing 5'ss U1 snRNA complementarity to an HBond score of 14.2 (Fig. 2B, lane 2) only slightly decreased 5'ss usage in the RT-PCR, while the more sensitive qPCR detected quite a marked drop in 5'ss usage (Fig. 2C). Decreasing U1 snRNA complementarity down to HBS 10.7 led to a smaller amount of spliced message and more unspliced transcripts (Fig. 2B, lane 4). Note that in a dataset of annotated human 5'ss, HBond scores of 10.7 or lower are very rare (31), which underscores the strong activation by the five octamer +10.32 repeats.

Insertion of the neutral octamer -0.15 sequence containing four nucleotide substitutions (ΔHZ_{EI} ~100, see Fig. 1C, E) resulted in a similar, stepwise decrease of 5'ss usage with decreasing 5'ss strength (Fig. 2A, C). However, in this setting, 5'ss recognition was already weakened, indicating a general requirement for an SRE. By further decreasing the HZ_{EI} score of the upstream located SRE with the insertion of the ΔHZ_{EI} ~70 or ΔHZ_{EI} ~30 octamers (Fig. 1C, E), general splice donor recognition was further decreased even for the splice donor with the highest 5'ss strength 17.5 (Fig. 2C). While for the two strongest 5'ss (HBS 17.5 and 14.2), stepwise decrease of upstream SREs led to a monotonous decrease in 5'ss usage, qPCR reached its quantitative detection limit for the two weaker 5'ss (HBS 12.1 and 10.7) at ΔHZ_{EI} ~70.

These results show that 5'ss usage can be quantitatively titrated by varying ΔHZ_{EI} as measure of SRE activity and HBond score for intrinsic splice donor strength.

Octamer sequences mediate exon inclusion

In order to analyze the effect of the three octamer sequences on general exon recognition instead of 5'ss selection, the five octamer repeats were inserted into a previously described 3-exon minigene splicing reporter (37) (Fig. 3A). As expected from its predicted splicing regulatory properties, insertion of five octamer +10.32 repeats resulted in full exon inclusion (Fig. 3B, lane 1), consistent with the view that exonic SR protein binding supports both splice donor and splice acceptor recognition and hence, general exon inclusion (12). Similarly, insertion of five negative octamer –10.35 repeats resulted in full exon skipping (Fig. 3B, lane 3) as predicted. Since in our splice site competition experiments in Fig. 1C, the overall amount of splicing decreased from octamer +10.32 repeats to octamer –0.15 repeats, we would have expected less exon inclusion for octamer –0.15 also in this three-exon reporter. Surprisingly however, insertion of the splicing neutral octamer –0.15 repeats resulted in a seemingly unchanged amount of exon inclusion, while at the same time some exon skipping was detectable (Fig. 3B, lane 2). Thus, in the context of this reporter, the high intrinsic strength of both 3'ss (MaxEnt 11.7) and 5'ss (HBS 17.5, MaxEnt 10.10) seemed to be sufficient for exon definition.

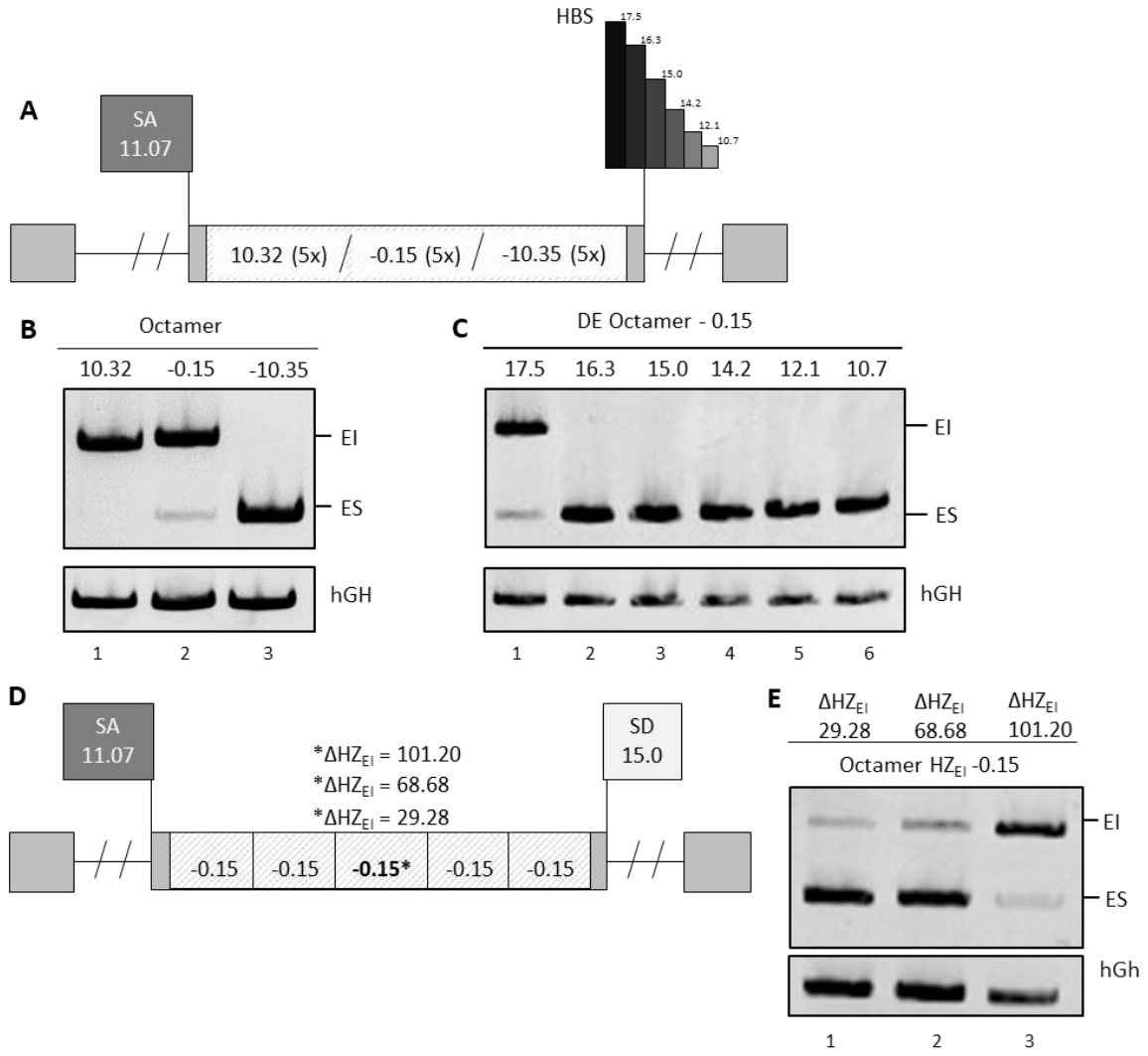


Figure 3: Artificial octamer sequences contribute to exon recognition.

A *Sketch of the 3-exon minigene reporter plasmid. The middle exon contains an insertion site for SREs which are flanked by an intrinsically strong splice acceptor (SA MaxEnt 11.07) and splice donor (SD HBS 17.5, MaxEnt 10.10).* ***B*** *2.5 × 10^5 HeLa cells were transfected with 1µg of the reporter plasmids and 1µg of pXGH5 (hGH) that was used as a control for transfection efficiency. RNA was harvested and reverse transcribed into cDNA 24 hours post transfection with primer pair #2648/#2649. PCR products were run on a 10% non-denaturing polyacrylamide gel. While insertion of octamer +10.32 resulted in the expected exon inclusion and insertion of octamer –10.35 led to complete exon skipping, insertion of the splicing neutral octamer –0.15 led to a higher amount of exon inclusion than exon skipping.* ***C*** *Analysis of exon inclusion in the presence of the neutral octamer –0.15 upstream of six different splice donors with HBond scores ranging from 17.5 down to 10.7. Without SRE support, lowering the HBond score from 17.5 to 16.3 results in full exon skipping.* ***D*** *Sketch of the 3-exon minigene reporter plasmid with the central octamer containing HEXplorer guided mutations (Fig. 1D). In this construct, the intrinsic splice donor strength was lowered from HBS 17.5 to 15.0 so that the*

stepwise induction of exon inclusion by the insertion of the point mutations could be measured. ***E*** *The reduction of intrinsic splice donor strength resulted in full exon skipping upon insertion of the splicing neutral octamer –0.15. Insertion of nucleotide substitutions that elevate* ΔHZ_{EI} *led to a stepwise increase in exon inclusion while exon skipping was stepwise decreased.*

In order to further examine the dependency of exon inclusion on 5'ss strength even without the support of exonic SREs, we analyzed exon inclusion in the presence of the neutral octamer –0.15 in combination with six splice donors with HBond scores from 17.5 down to 10.7 (Fig. 3C). Even for a slightly weaker 5'ss with HBS 16.3, we observed full exon skipping, confirming that yet in the presence of a strong 3'ss, a strong 5'ss is required for exon recognition in absence of any exonic splicing enhancer indicated in the HEXplorer profile.

Selecting a 5'ss that exhibited full exon skipping, we further titrated dependency of exon inclusion using the central octamers with ΔHZ_{EI} ~30, ~70 and ~100, upstream of the 5'ss with HBS 15.0 (Fig. 3D). In this series of minigene reporter experiments, we observed simultaneous decrease in exon skipping and increase in exon inclusion when going from octamer –0.15 to central octamers with ΔHZ_{EI} ~30, ~70 and ~100 (Fig. 3E).

Our results confirm in the three-exon splicing assay that exon inclusion can be quantitatively titrated by varying ΔHZ_{EI} and splice site strength.

SR- and hnRNP proteins bind to HEXplorer-designed octamer sequences

To further analyze the mechanism of splicing regulation conducted by the non-evolutionary *in silico* designed artificial octamer sequences, we performed an RNA affinity purification assay to identify splicing regulatory proteins binding to the sequences. To this end, we incubated 40nt long RNA oligonucleotides (five octamer repeats) with HeLa nuclear extract (52). After several washing steps, the remaining specifically bound proteins were eluted and subjected to MS-analysis. Results were analyzed using Perseus software (53). When filtering for highest MS/MS counts and searching for splicing related proteins, a binding preference of SRSF3 to the downstream enhancing splicing regulatory octamer +10.32 was revealed (Fig. 4A). The negative octamer –10.35 was preferably bound by the PTB isoforms PTBP1 and

PTBP2, as well as hnRNPDL and TIA-1, all known repressors of downstream splice donors (13). The neutral octamer –0.15 showed no preferred binding for any splicing related proteins (Fig. 4A, Supplementary File S2). Validation of these results was performed via western blot using antibodies specifically detecting the splicing related binding proteins SRSF3 for octamer +10.32, PTB and hnRNPD for octamer –10.35 and the control MS2 coat (Fig. 4B).

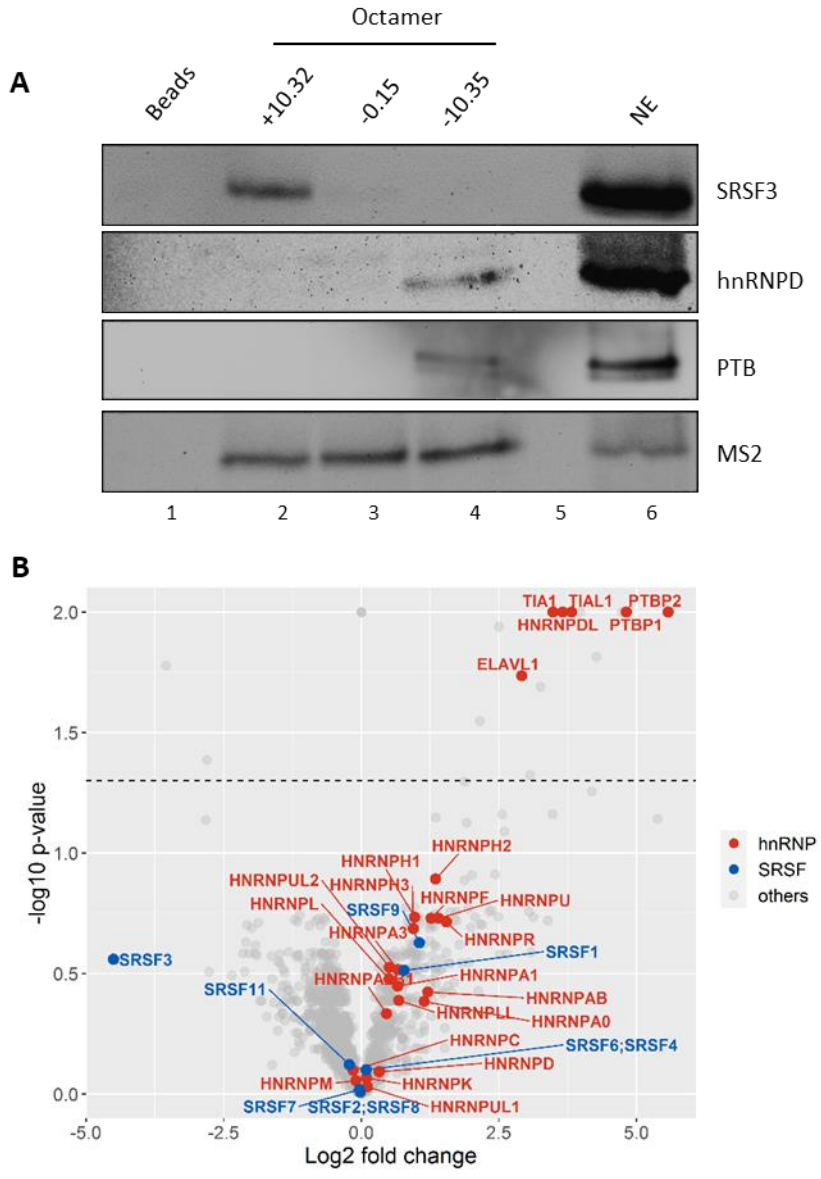


Figure 4: MS-analysis reveals protein binding partners of the artificial octamer sequences. A *In an RNA affinity chromatography assay, substrate RNAs containing a bacteriophage MS2 sequence and either octamer +10.32, octamer –0.15 or octamer –10.35, sequences were covalently linked to adipic acid dihydrazide-agarose beads and incubated with HeLa cell nuclear protein extract. Recombinant bacteriophage MS2 coat protein was added to monitor RNA input. Samples were subjected to MS-*

analysis. Volcano plot with splicing regulatory proteins marked in red (hnRNP or significantly different hnRNP-like proteins) and blue (SR or significantly different SR-like proteins). ***B*** *For MS-analysis validation, precipitated proteins were resolved by SDS-PAGE (12%) and detected by immunoblot analysis using anti-SRSF3, anti-hnRNPD, anti-PTB or anti-MS2 coat protein antibodies to validate candidates that were revealed by MS-analysis.*

When inspecting the octamer sequences and their protein binding partners, some partial recognition motif overlaps became obvious. Octamer +10.32 exhibits an SRSF3 recognition sequence, a four-nucleotide CANC motif, as it was published by Hargous and colleagues (54). Furthermore, the negative octamer –10.35 shares short sequence stretches with binding sites of proteins related to the hnRNP family, PTB (CTCT repeats, (55)) and TIA-1 (partial TCT (56)). Several studies have tried to elucidate distinct binding sites for both SR and hnRNP proteins, and these factors seem to recognize highly degenerate, variable RNA sequences (57). In the next section, we systematically tested the SRSF3 motif CANC, both as repeats and inserted in the octamer –0.15 sequence.

CANC—Different splicing regulatory properties for different "N"

In order to assess potential HEXplorer score profile differences between the four CANC sequences, we first built four 40 nt long sequences by concatenating 10 repeats of every CANC. As expected for potential SRSF3 binding sites, HEXplorer score profiles were exclusively positive for CAAC (HZ_{EI} ~10), CAGC (HZ_{EI} ~7) and CATC (HZ_{EI} ~5) repeats. Surprisingly however, ten repeats of CACC, which is actually part of the octamer +10.32, showed an entirely negative HEXplorer score profile with HZ_{EI} ~ –4, suggesting suppressing properties (Suppl. Fig. 2A). We therefore experimentally tested these four 40 nt long sequences in our two-exon competition splicing reporter containing two equally strong splice donor sites, in order to assess the validity of HEXplorer score profile prediction (Suppl. Fig. 2A).

The impact on splice site selection was analyzed by RT-PCR following transient transfection assays. Consistent with the unexpected HEXplorer score prediction, insertion of CAAC, CATC and CAGC repeats led to the exclusive use of the downstream located donor (Suppl. Fig. 2B, lanes 1, 2 and 4), while insertion of CACC repeats led to a complete switch to the upstream 5'ss (Suppl. Fig. 2B, lane 3). For the

CACC motif, in fact, concatenation creates a cytosine-rich CACCC motif which may be bound by the exonic splicing silencer hnRNP K (58), consistent with the negative HEXplorer profile.

In order to prevent concatenation artifacts, we next examined the impact of single CANC motifs on exon inclusion in our three-exon splicing reporter (Suppl. Fig. 2C). To this end, we inserted CANC in the middle of the central octamer between pairs of –0.15 octamers CCTATTGG. We additionally modified the two flanking nucleotides of the central octamer on either side of CANC to either maximize or minimize total HEXplorer score (Suppl. Fig. 2D). HZ_{EI} was maximized on average for ACCANCAA ("strong neighborhood") and minimized for CACANCTA ("weak neighborhood") as central octamers (Suppl. Fig. 2D).

While the neutral central octamer –0.15 led to complete exon skipping, in the weak neighborhood CA—TA, the four CANC SRSF3 binding sites in the central octamer primarily resulted in exon skipping and only a low level of exon inclusion (Suppl. Fig. 2D, lanes 1, 2—5). Strengthening the neighborhood by substituting AC—AA as flanking dinucleotides around the same CANC binding sites increased HZ_{EI} by 70.95 up to 91.71 (Suppl. Fig. 2D) and resulted in a high level of exon inclusion (Suppl. Fig. 2D, lanes 6—9). These experiments confirmed that all four CANC binding sites act as exonic splicing enhancers, when CANC concatenation artifacts are avoided. Furthermore, the neighboring dinucleotides additionally impact exon inclusion level beyond the central CANC motif, in accordance with HEXplorer prediction for the *in silico* designed weak and strong neighborhoods.

2.4 billion high-confidence RNA-seq reads from 46 human fibroblast samples

Complementary to our experimental analysis, we also examined 5'ss context impact on splice site competition using data from our large human fibroblast RNA-seq transcriptome dataset (1,41). In order to mimic the 5'ss competition situation experimentally examined above (cf. Fig. 1), we analyzed pairs of 5'ss and nearby exonic GTs, using the ratio of RNA-seq reads detected on each as relative usage measure. Comparing RNA-seq reads across many genes from different samples, however, requires careful normalization of reads and removal of potentially noisy read counts, which we address below.

In each of 46 human fibroblast samples, we separately collected (gapped) exon junction reads detected at any given 5'ss coordinate (*5'ss c*). In the entire RNA-seq dataset, a total of 2,408,714,500 (2.41 billion) reads were found in a total of 27,177 genes, using only exon junctions with gap quality score gqs ≥ 400 and gap length < 26,914 (95% of human introns are shorter) as described in (41,48). The majority of genes were very reliably expressed in many—on average in 30—samples: 12,850 genes (47.3%) containing 99.7% of all reads were detected in all 46 samples, and 7,076 genes totaling only 0.002% of all reads were detected in less than 10 samples. The number of samples that a gene was detected in followed a U-shaped distribution (Suppl. Fig. 3A, red line), and those genes detected in few samples each had very few reads. Genes detected in more samples also had more reads *per sample*, not just in total (Suppl. Fig. 3A, blue bars).

The number of genes detected in any one of the 46 samples varied slightly around an average of 17,828 (± SD 619, median 17,848, range 15,915—18,828), and each gene had an average of 2937 reads in each single sample. Suppl. Tab. 1 gives a schematic overview of the arrangement of reads on exon junctions and 5'ss across genes and samples.

Normalization then proceeded in three steps. In order to account for differential RNA-seq detection between samples, we normalized all 5'ss reads by the total number (in millions) of exon junction reads in this sample, obtaining sample normalized RPMG (*reads per million gapped reads*) values for the 5'ss usage in each sample.

$$RPMG(5'ss\ c, gene\ g, sample\ s) = 10^6 \cdot \#reads(c,g,s) / Total\ \#reads(s)$$

In the second normalization step, we factored in differential gene expression in each sample. For a specific gene in a given sample, we determined the most used 5'ss in this gene and defined its RPMG value as gene-expression measure MRIGS (*maximum RPMG in gene and sample*).

$$MRIGS(gene\ g, sample\ s) = \max_{5'ss\ c} RPMG(c,g,s)$$

Supplementary Fig. 3A indicates that there were many genes with very few reads (e.g. below a total of 10 reads per sample on all exon junctions of the entire gene). If such genes were detected in samples with an overall high level of technical RNA-seq read

coverage (large sequencing library size), they may have been false-positive detections due to RNA-seq technique limitations, and could be identified by low MRIGS values. Since by definition MRIGS values were separately normalized in each sample, we considered a lower MRIGS cutoff suitable to separate biologically reliable from noisy reads.

To this end, we determined the distribution of exon junction reads on genes with different expression levels (MRIGS), and aggregated all 2.41 billion reads into logarithmically equidistant MRIGS bins. For low gene expression at the left-hand flank of the read-weighted MRIGS distribution, we found a pronounced drop in the total sum of exon junction reads for genes below MRIGS = 1 (Suppl. Fig. 3B, red bar). By definition, in a gene with MRIGS < 1, the most used 5'ss had less than one read for every million exon junction reads in the entire sample. We subsequently kept genes in those samples, where they were detected with MRIGS ≥ 1, as *high-confidence genes* in our analysis, and discarded genes as *noise candidates* in those samples where they were low expressed with MRIGS < 1. This lower cutoff region of MRIGS < 1 contained the 47% lowest expressed gene detections in a single sample, but held only 22.4 million reads (0.93%) in total.

Including only high-confidence gene detections in each sample, we extracted 2.39 billion exon junction reads (99.1%) in 11,948 genes, with an average of 5489 (median 1453) reads per gene in each single sample, indicating that noise removal has been effective on the gene level. These genes were the basis for extracting pairs of 5'ss and exonic GT sites. To permit an appropriate 5'ss selection, we eventually extended the "high-confidence" criterion from genes to splice sites.

Gene-and-sample read normalization permits 5'ss usage assessment

In order to allow 5'ss usage comparison across genes with different expression levels in a single sample, we normalized all 5'ss reads by the individual gene expression MRIGS in the specific sample. We thus obtained gene-and-sample normalized reads (GSNR) for each 5'ss in each sample:

$$GSNR(5'ss\ c, gene\ g, sample\ s) = RPMG(c, g, s)/MRIGS(g, s)$$

By definition, GSNR values were between 0 and 100%, and in each sample each gene contained one 5'ss with GSNR = 100%: the 5'ss with this gene's maximum (MRIGS) number of reads in this sample. Finally, to obtain a measure of the overall expression of a given 5'ss in our RNA-seq fibroblast dataset, we averaged the different GSNRs of a 5'ss across all samples, where the gene was sufficiently expressed:

$$GNR(5'ss\ c, gene\ g) = \langle GSNR(5'ss\ c, gene\ g, sample\ s)\rangle_{sample\ s}$$

Since the "most-used" 5'ss of a given gene could differ from sample to sample, there was not necessarily a single 5'ss with GNR = 100% in every gene.

In order to determine the GNR distribution in the above high-confidence genes (MRIGS ≥ 1), we extracted 92,493 internal exons with canonical 5'ss that were *Ensembl* annotated in at least one TSL1 transcript and contained at least one exonic GT site. These exons had a median exon length of 166 nt (average 417 nt), and the 5'ss GNR distribution was composed of three parts (Suppl. Fig. 4): (1) a narrow peak at low GNR indicating noisy reads, (2) a Gaussian part between 20% and 97% with mean 72% and standard deviation 18% ($r^2 = 0.995$), and (3) a peak at 98—100% reflecting the maximally used 5'ss in each gene. Similar to our approach in (1), we considered 3240 5'ss (3.5%) detected below 2% of gene expression level (GNR < 2%) as potential noise candidates. For further analysis, we retained 89,253 *high-confidence* 5'ss (96.5%) with GNR ≥ 2% from genes with MRIG ≥ 1.

320,601 pairs of high-confidence 5'ss and exonic GTs from exons of TSL1 transcripts

For these high-confidence 5'ss, we then extracted all GT dinucleotides between 12 nt downstream of the 3'ss and 17 nt upstream of the 5'ss. This GT search region was chosen to ensure that there was at least a one-hexamer wide potential SRP binding site not overlapping the 11 nt long 5'ss or GT-site, as well as the 23 nt long 3'ss.

We further excluded potential U12 splice donors, defined by the list of confirmed U12-dependent 5'ss reported in (59), and those 5'ss with a GTT trinucleotide at positions +1/+2/+3 which may bind U1 snRNP by bulging the T nucleotide in position +2. In order to better mimic our splice site competition experiments in splicing reporters with short exons, we only included GT sites less than 150 nt from the 5'ss. Collecting all

GT dinucleotides in this search region (SA+12 nt to SD-17 nt) while applying these strict filter conditions, we obtained a total of 320,601 GT-and-5'ss pairs in 89,008 exons. Note that only 8833 GT sites (2.8%) actually had RNA-seq reads. In each pair, we then compared U1 snRNA complementarity (HBS) and splice site HEXplorer weight (SSHW) between GT sites and annotated 5'ss.

Exonic GT sites have lower U1 snRNA complementarity than annotated 5'ss used in fibroblasts

As expected, exonic GT sites had much lower U1 snRNA complementarity than 5'ss (GT HBS 6.2 ± 3.0, mean ± SD, vs. 5'ss HBS 15.1 ± 2.5; N = 320,601 pairs; cf. Fig. 5A i for individual GT- and 5'ss-HBS distributions). In Figure 5A ii, light gray bars show the HBond score difference distribution $\Delta HBS = HBS_{GT} — HBS_{5'ss}$, and indeed, in 98.9% of pairs, the exonic GT site was weaker than the 5'ss. For the subset of GT sites with RNA-seq reads, the ΔHBS distribution was significantly shifted to higher values (Fig. 5A ii, dark vs. light gray bars). In overlays of HBS distributions (a) for all pairs together with (b) only GT sites with RNA-seq reads, it is obvious that only the HBS distribution of GT sites with reads is shifted (dark vs. light gray bars in Fig. 5A iii), while for 5'ss the distributions nearly coincide (dark vs. light gray bars in Fig. 5A iv). Thus, GT sites with RNA-seq reads were not as much weaker compared to their 5'ss as those that were not used in fibroblasts.

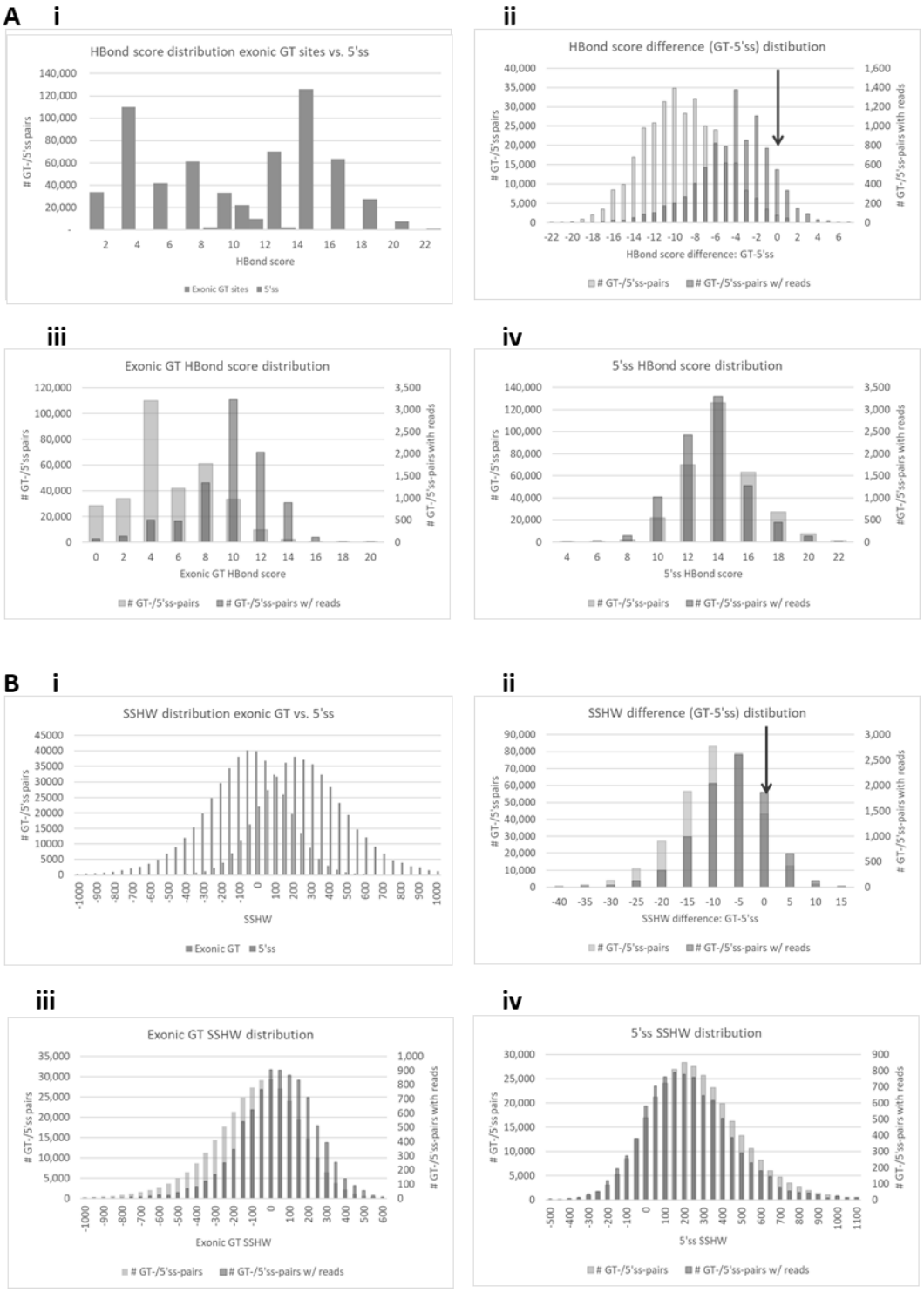


Figure 5: Exonic GT sites have lower U1 snRNA complementarity and weaker SRE support than nearby 5'ss. A i *HBond score distributions for 320,601 pairs of high-confidence 5'ss and exonic GT sites closer than 150 nt.* ***A ii*** *HBond score difference* HBS_{GT}–$HBS_{5'ss}$ *distribution. For 99% of all pairs, the 5'ss HBS was higher than the exonic GT HBS, indicating a stronger 5'ss compared to competing exonic GTs. Arrow indicates* $\Delta HBS=0$. ***A iii, iv*** *HBond score distributions for 8833 pairs with RNA-seq reads on GT compared to all 320,601 pairs.* ***B*** *SSHW distributions for the same datasets.*

Exonic GT sites have weaker SRE neighborhood than annotated 5'ss used in fibroblasts

In order to compare splice site HEXplorer weight SSHWs for GT sites and 5'ss as measures of their respective SRE support for recognition as splice sites, we carefully selected appropriate neighborhoods of these sites. Excluding the 11 nt long proper 5'ss or exonic GT site sequence, we calculated the SSHW as the total HEXplorer score ($\sum_{up} HZ_{EI}$) in a 50 nt wide upstream region, starting at exonic position -4 minus the corresponding downstream total ($\sum_{dn} HZ_{EI}$), starting at intronic position +9. For GT sites closer than 100 nt to the 5'ss, the downstream region of the GT site overlapped with the upstream region of the 5'ss, and for GT sites closer than 50 nt to the 5'ss, we used the entire sequence between position +9 of the GT site and position -4 of the 5'ss (cf. Methods and schematic in Suppl. Fig. 4B).

In the 320,601 GT-and-5'ss pairs, exonic GT sites also had lower splice site HEXplorer weights than 5'ss (GT SSHW -1.1 ± 4.8, mean ± SD, vs. 5'ss SSHW 5.8 ± 5.0; cf. Fig. 5B i for individual SSHW distributions). However, the two SSHW distributions overlapped to a much higher degree than the respective HBS distributions, indicating higher importance of HBS for splice site recognition than SSHW (cf. Fig. 5B i for individual GT- and 5'ss-SSHW distributions).

In Figure 5B ii, light gray bars show the SSHW difference distribution $\Delta SSHW = SSHW_{GT} — SSHW_{5'ss}$, and in 82.0% of pairs, the exonic GT site had lower SSHW than the 5'ss. For the subset of GT sites with RNA-seq reads, the ΔSSHW distribution was only slightly shifted to higher values (Fig. 5B, ii, dark vs. light gray bars). This is also reflected in only minor shifts between individual SSHW distributions for GT sites with RNA-seq reads compared to all GT sites (Fig. 5B iii and iv, cf. light and dark bars).

From these analyses we conclude that in our RNA-seq fibroblast dataset, exonic GT sites have significantly lower HBond scores than their associated 5'ss, and HBond scores of GT sites with RNA-seq reads are higher than those of GT sites without reads.

Splicing regulatory properties of 50 nt wide neighborhoods, quantified by SSHW, exhibit the same tendencies, albeit to a much lower degree.

5' splice site usage dependence on 5'ss strength and SRE support

After separately identifying HBond score and SSHW differences between GT-sites and 5'ss with and without RNA-seq reads in 320,601 pairs, we set out to determine relative GT usage dependency both on U1 snRNA complementarity and SRE support simultaneously. This is a tentative approach to a comprehensive "functional splice site strength" concept encompassing splice site U1 snRNA complementarity and SRE neighborhood.

In our RNA-seq dataset, gene-normalized reads (GNR) reflect GT site or 5'ss usage likelihood, and we therefore quantified GT usage relative to 5'ss by their GNR log-odds ratio $\mathrm{LGNRr} = \log_{10}\left(\mathrm{GNR}_{\mathrm{GT}}/\mathrm{GNR}_{5'\mathrm{ss}}\right)$. In order to tabulate LGNRr as a function of both ΔHBS and ΔSSHW, we first binned these variables to obtain GT-/5'ss-pair groups of approximately equal sizes. Rather than choosing equidistant ΔHBS- and ΔSSHW-bin intervals, we focused on adequate resolution in the important regime of GT sites with RNA-seq reads. From the two ΔHBS and ΔSSHW distributions shown in Fig. 5A ii and Fig. 5B ii (dark bars), we obtained ten 10%-wide bins each for ΔHBS and ΔSSHW, splitting the sample of 8833 pairs with RNA-seq reads into 10 × 10 two-dimensional bins containing about 8833/(10 × 10) GT-/5'ss-pairs each. On average, each two-dimensional bin contained 3206 pairs overall and 88 pairs with RNA-seq reads. For every ΔHBS- and ΔSSHW-bin, we then calculated the average LGNRr of all pairs, and color-coded cells with low (high) relative GT site usage in red (green). In this table, GT-site usage relative to 5'ss covered three orders of magnitude from 10^{-3} to 10^{-6} in statistically reliable values: the median coefficient of variation (CV_{LGNRr} = standard deviation / mean LGNRr) of the LGNRr values in each two-dimensional bin was 0.21 (average CV_{LGNRr} = 0.25, standard deviation CV_{LGNRr} = 0.18). We further averaged the two-dimensional LGNRr table with an exponential smoothing algorithm using 0.7 × average of all eight neighboring bins. Eventually, to obtain a LGNRr representation on an equidistant square grid, we applied cubic spline interpolation in ΔHBS steps of 0.2 and ΔSSHW steps of 0.5 (Fig. 6A).

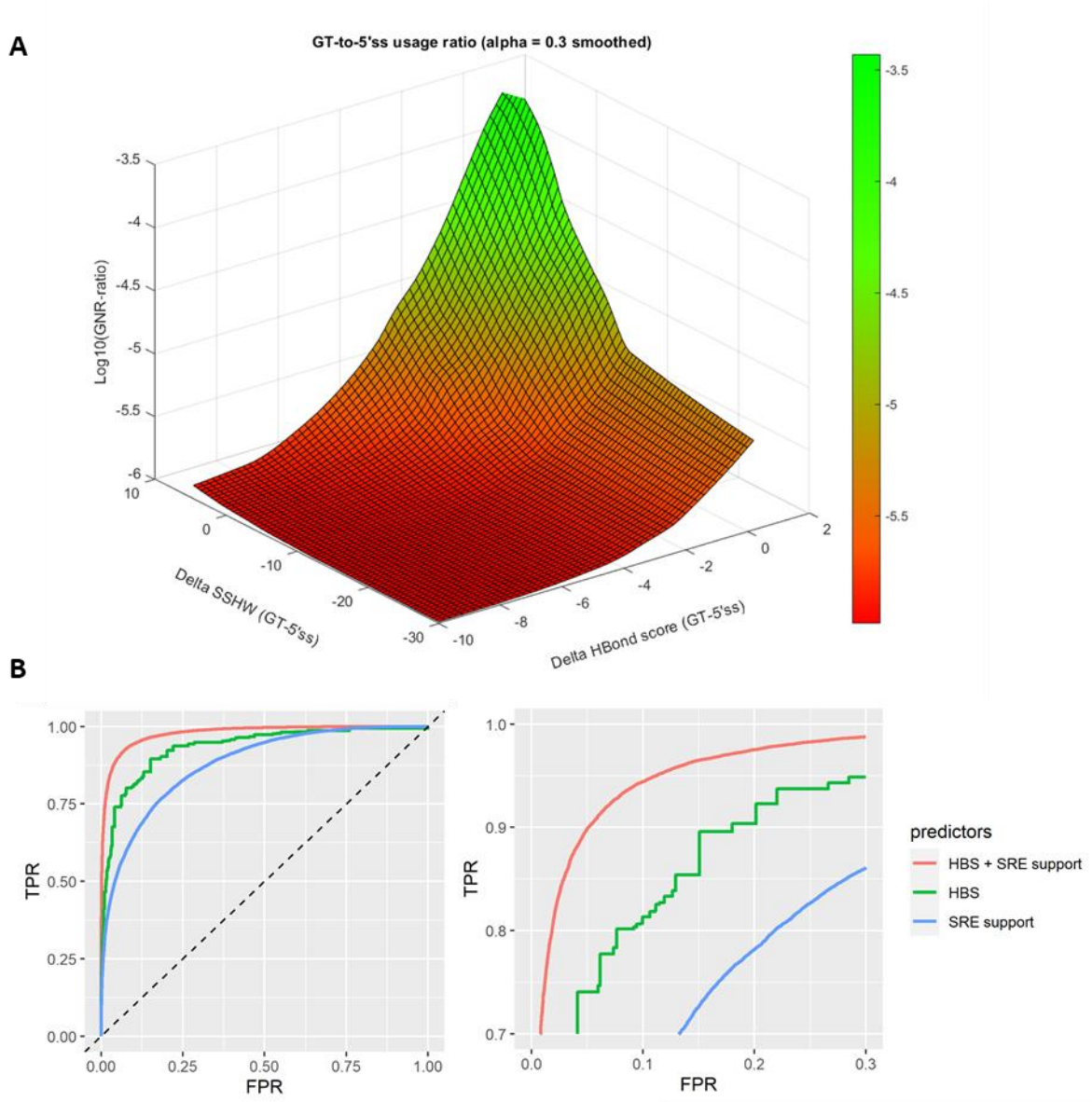


Figure 6: Combination of HBS and SSHW improves classification of GT sites and 5'ss. A: *Average LGNRr=log$_{10}$(GNR$_{GT}$/GNR$_{5'ss}$) as measure of GT-site usage relative to 5'ss (vertical z-axis), plotted as function of HBond score difference ΔHBS=HBS$_{GT}$–HBS$_{5'ss}$ and splice site HEXplorer weight difference ΔSSHW=SSHW$_{GT}$–SSHW$_{5'ss}$. Color-coding shows a monotonous transition from exclusive 5'ss usage (front corner, red) to higher GT-site usage (back corner, green).* ***B:*** *Receiver operating characteristic curves of three logistic regression models for the classification of 15,163 annotated 5'ss and 15,029 exonic GT sites closer than 150 nt and with HBS≥10, but less than 1% RNA-seq reads of the associated nearby 5'ss. ROC curves for logistic model based only on SSHW (blue, AUC 0.88), based only on HBS (green, AUC 0.93) and based on both HBS and SSHW (red, AUC 0.98) show stepwise improvement of classification accuracy.*

Both the two-dimensional surface plot (Fig. 6A) and the color-coded table (Suppl. Table 2) showed a clear picture of relative GT-to-5'ss-usage dependence on both U1 snRNA complementarity and on SRE support. There is a region of low GT site usage for both large negative ΔHBS and ΔSSHW (red), mirrored by a region of higher GT usage in the opposite corner with higher, positive ΔHBS and ΔSSHW (green), and a smooth, diagonal transition region (yellow). This behavior is also reflected in the 3D-plot in Fig. 6B. A sufficiently large negative ΔHBS cannot be compensated by even the strongest SRE-containing neighborhood (high SSHW), while for positive or only slightly negative ΔHBS, GT sites can be used despite lack of SRE support (negative ΔSSHW). This result underscores that 5'ss complementarity to U1 snRNA is the dominant feature in splice site recognition, and SRE support plays a secondary, auxiliary part.

Combination of HBS and SSHW improves classification of GT sites and 5'ss

In order to further examine the power of HBond score and SSHW to distinguish annotated 5'ss from exonic GT-sites in a classification task, we selected 60,384 pairs with low usage GT-sites (GNRr = GNR_{GT} / $GNR_{5'ss}$ < 1%) that had medium-to-high U1 snRNA complementarity (HBS ≥ 10). In competition with their respective 5'ss, these GT sites were barely used, although they had reasonable complementarity with an HBond score of at least 10. In this dataset, we expected SRE neighborhoods of both 5'ss and GT site to possibly play a stronger part in splice site selection.

We then split the pairs and pooled both GT sites and 5'ss into a single set of 120,768 potential splice sites. Randomly splitting this entire dataset into a training set (75%) and a validation set (25%), we fit three different logistic models for the binary prediction of true 5'ss in a balanced sample of 45,165 GT sites and 45,411 5'ss. In the first model, we used only the SSHW as single predictor variable, in the second model we used only the HBond score, and finally we entered both SSHW and HBS simultaneously into the regression model. In all three regressions, the coefficients of SSHW and HBS were highly statistically significant ($p < 10^{-6}$), indicating that these variables significantly contributed to distinguishing true 5'ss from GT sites in the training dataset.

We then tested the three regression models on the remaining 25% of the entire dataset, containing 15,163 annotated 5'ss and 15,029 exonic GT sites. Fig. 6B shows the

receiver operating characteristic curves (ROC) obtained for the three regression models, plotting *sensitivity* (true positive rate, TPR) versus *1—specificity* (false positive rate, FPR) upon variation of the cutoff of the prediction scores obtained from the regressions. All three models achieved good classification results for separating true 5'ss from GT sites in the validation dataset, indicated by all ROC curves extending far into the upper left corner of the diagram. Using the *area-under-the-curve* ($0 < AUC < 1$; $AUC = 0.5$ for random assignment) as overall measure to compare the regression models, we found a clear hierarchy for goodness of classification: the model using only the HBond score increased AUC to 0.93 from $AUC = 0.86$ for SSHW only, and entering both variables into the model again improved the classification to $AUC = 0.98$. Thus, in terms of the ROC curves, there is a nearly even AUC spacing of 0.05 each from SSHW < HBS < SSHW + HBS. To complete the model, we also added an interaction term HBS × SSHW to the logistic regression, but this term did not acquire a significant coefficient and thus could not improve the classification. This classification shows that for 5'ss and GT sites, the HBond score is more informative than the "SRE neighborhood parameter" SSHW alone, but SSHW adds as much classification value to HBS as HBS adds to SSHW.

DISCUSSION

In this manuscript, we present *in silico* designed sequences with arbitrary *a priori* prescribed splicing regulatory properties, quantitatively represented by a constant HEXplorer score profile. We comprehensively confirmed *in silico* predictions on splice site recognition in single intron splicing competition as well as exon inclusion assays. From an MS analysis of proteins binding to these *in silico* designed SRE sequences, we exemplarily confirmed splicing regulatory proteins binding specifically to enhancing, neutral or silencing sequences. We complementarily selected 320,601 pairs of high confidence 5'ss and neighboring exonic GT sites from our large human fibroblast RNA-seq dataset, and derived a two-dimensional splice site usage landscape from gene-and-sample normalized RNA-seq reads. This GNR landscape served as a basis for a logistic 5'ss usage prediction model, depending on both U1 snRNA complementarity and HEXplorer score differences. This model greatly improved 5'ss discrimination between strong but unused exonic GT sites and highly used 5'ss by adding the splice site HEXplorer weight calculated from a 50 nt neighborhood to the classification algorithm based solely on the HBond score.

In principle, sequences with prescribed splicing regulatory properties could be obtained by inserting single known SRE motifs into assumed splicing neutral sequences, like the octamer "CCAAACAA" that has been proposed and tested as a building block for splicing neutral sequences (19,60). However, even in this seemingly simple case, concatenation of the octamer "CCAAACAA" accidentally creates a "CANC" motif as potential SRSF3 binding site (54), altering the splicing regulatory properties of the single octamer (61). In this study, we used the HEXplorer algorithm (31) to design splice enhancing, silencing and neutral octamers, *ab initio* avoiding accidental HEXplorer profile fluctuations possibly introduced by concatenation. Reversing the above sketched process, we generated putative binding sites by using the HEXplorer algorithm without restricting the sequences to single SR- or hnRNP binding sites, and we experimentally confirmed the splicing regulatory properties of *in silico* designed octamer sequences in single- or two-intron splicing reporters.

Analyzing the splicing outcome of the *in silico* designed sequences in the HIV-1 based single intron splicing reporter (Fig. 1, 2), it was obvious that the total RNA amount was

not conserved, which was clearly visible e.g. upon inserting the splicing neutral sequence (Fig. 1C). U1 snRNA can impact steady-state RNA levels via two pathways: on the one hand, U1 snRNP binding supports RNA stability and on the other hand it can also stimulate transcription initiation, leading to more or less primary transcript from the outset. With viral RNAs, whether Polyoma (62) or HIV-1 (6,63), reduced steady-state RNA quantities could be detected depending on weakening or deleting the 5'-cap-proximal 5'ss. Furthermore, Damgaard et al. found evidence that interaction with the 5'-cap-proximal bound U1 snRNP can >5-fold stimulate transcription initiation from both HIV-1 and ß-globin reporters by enhancing the assembly of the RNA polymerase preinitiation complex, regardless of whether the U1 snRNA binding site was splicing active (64). Thus, insertion of splicing neutral octamers (Fig. 1C) or weakening RNA duplex formation (Fig. 2, HBS 17.5 down to 10.7) leads to lower U1 snRNA binding and consequently to lower steady-state RNA levels as observed here. In the three-exon reporter context, however, the presence of the promoter-proximal 5'ss seemed to be sufficient for comparable RNA amounts as seen in Fig. 3B (lane 2).

Assuming a proportional interplay between 5'ss strength (HBS) and SRE impact (SSHW), a rough guesstimate of an equivalence between HBS and SSHW can be gleaned from the experiments shown in Fig. 3D, E: We observed that in the presence of just the splicing neutral octamer, an HBS of 17.5 was required for exon inclusion. For a weaker 5'ss with HBS = 15.0, SRE neighborhoods with ΔSSHW $\leq$ 70 did not suffice to support exon inclusion while ΔSSHW = 100 did, so that 2.5 HBS units seem to correspond to ΔSSHW ~ 100. This conclusion is only valid in the context of the concrete 3'ss in our splicing reporter, though, and it is not clear, if this tentative equivalence relation also holds independent of 3'ss.

RNA affinity chromatography (pulldown) and MS analysis (Fig. 4, Supplementary File S2) unveiled trans-acting binding partners of the inserted octamer sequences. In particular, the downstream enhancing octamer +10.32 showed a high prevalence for binding SRSF3, the smallest SR protein with a molecular weight of about 19 kDa. In fact, octamer +10.32 contains a CANC motif, identified as *in vivo* binding motif for SRSF3 (54). In our MS analysis, octamer +10.32 also bound the splicing related protein SREK1. In a recent study, SREK1 was shown to augment the splicing regulatory activity of SFRS3 and also repress other SR proteins (65,66). The splicing

neutral octamer -0.15 showed no binding of splicing regulatory proteins, and was rather bound by ubiquitous RNA binding proteins (Fig. 4, Supplementary File S2). This behavior is also reflected when applying ESEfinder (67) to identify exonic splicing enhancer motifs, which showed many hits for octamer +10.32, none for octamer –10.35, and one SRSF5 for octamer -0.15. For octamer -10.35, we confirmed binding of two splicing related proteins found in the MS analysis by Western blot: the polypyrimidine tract binding protein PTBP1 (hnRNP I) and hnRNPDL, which act as downstream silencers of splice donor sites according to the position dependency of splicing regulatory elements. SELEX analysis of the binding preferences of PTBP1 revealed the recognition of UCUU motifs going together with the recognition of the AAUUCUCU motif of octamer -10.35 (55).

For our complementary analysis of RNA-seq data from 46 human fibroblast samples, we applied rather strict 5'ss filter criteria to obtain a dataset of 89,253 *high confidence* 5'ss forming a total of 320,601 pairs with neighboring exonic GT sites. We obtained this dataset by a two-tier normalization process, taking both differential sequencing efficiency across samples (library size) and differential gene expression within a sample into account. Aiming to ascertain reliably detected 5'ss RNA-seq reads above biological and sequencing noise, we first discarded genes in those samples, where they were very weakly expressed, and then additionally discarded 5'ss below 2% of the gene expression level. In this way, we systematically improved the removal of noisy reads introduced in (61).

Our novel RNA-seq based 5'ss usage landscape concept quantifies the usage of exonic GT sites relative to their nearby 5'ss by their log-gene-normalized read ratio LGNRr, as function of both HBS and SSHW differences "GT-site – 5'ss". In this definition, we used the HBond score as measure for 5'ss complementarity to U1 snRNA. However, since HBond and MaxEnt scores achieved similar accuracies in discriminating between annotated and "mock" splice sites (29), we would expect a similar structure of the 5'ss usage landscape plotted vs. MaxEnt score instead. As measure of splicing regulatory properties, we used the total HEXplorer weights from 50 nt regions up- minus downstream of 5'ss. Necessarily, any choice of neighborhood size is to some extent arbitrary. However, several studies indicate only weak dependence on neighborhood size: Putative exonic splicing enhancer and silencer

octamer (PESX) frequencies have been shown to remain rather constant in 100 nt long composite exons (50 nt center and 25 nt ends) and introns (68). Similarly, the distributions of the top 400 ESEseqs and ESSseqs showed little variation in 100 nt long composite exons and introns (30). Eventually, individual hexamer weights used in the HEXplorer definition were highly correlated when derived from 100 nt or 30 nt wide 5'ss neighborhoods. Therefore, we expect to capture relevant SRP binding sites within the chosen 50 nt neighborhoods.

In recent years, state-of-the-art machine learning algorithms for splice site prediction have been developed and evaluated. Using a modular architecture, MMSplice encompasses six neural network modules covering donor and acceptor sites, as well as their respective exonic and intronic neighborhoods, and processes the neural network output in linear and logistic regression models. MMSplice outperformed previous splicing prediction models in the "Critical Assessment of Genome Interpretation" (CAGI) challenge (32,69-71). Designed as a 32-layer deep neural network built from residual blocks, the deep learning tool SpliceAI achieved an impressive 95% top-*k* accuracy in identifying splice sites from DNA sequence alone, however using features from a very wide region of 10,000 nt around the index site (33). As all machine learning algorithms, these models appear as black boxes to the user, and their splice site usage predictions are not transparent in terms of biological mechanisms: they may well successfully apply features with no biological meaning. In contrast, our RNA-seq based GT-site-to-5'ss usage ratio landscape model clearly shows both effects of 5'ss strength and neighboring splicing regulatory elements.

AVAILABILITY

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) data are included in Supplementary File S2.

ACCESSION NUMBERS

The RNA-seq dataset (41) analyzed in this study is available through ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-4652.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

We thank Björn Wefers for technical assistance and Philipp Peter for implementing the HEXplorer algorithm on our RNA website. We also thank Gereon Poschmann for helping with MS-analysis evaluation. We thank all lab members for discussion and critical reading of the manuscript. We would like to thank Douglas Black for providing the PTB antibody.

FUNDING

This work was supported by Deutsche Forschungsgemeinschaft (DFG) to HS [SCHA 909/4-1]; Jürgen Manchot Stiftung, Düsseldorf to LM, ALB and HS; Stiftung für AIDS-Forschung, Düsseldorf to HS, and Forschungskommission of the Medical Faculty, Heinrich Heine Universität Düsseldorf [2020-12] to HS.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Erkelenz, S., Theiss, S., Kaisers, W., Ptok, J., Walotka, L., Muller, L., Hillebrand, F., Brillen, A.L., Sladek, M. and Schaal, H. (2018) Ranking noncanonical 5' splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. *Genome Res.*, **28**, 1826-1840.

2. Zhuang, Y. and Weiner, A.M. (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*, **46**, 827-835.

3. Wong, M.S., Kinney, J.B. and Krainer, A.R. (2018) Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. *Mol. Cell*, **71**, 1012-1026 e1013.

4. Ptok, J., Muller, L., Theiss, S. and Schaal, H. (2019) Context matters: Regulation of splice donor usage. *Biochim Biophys Acta Gene Regul Mech*, **1862**, 194391.

5. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, **11**, 377-394.

6. Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K., Tange, T.O., Lund, M.K., Kjems, J., Scheid, A. and Schaal, H. (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.

7. Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., Meyer, I., Gierling, W., Theiss, S., Preuss, T. *et al.* (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.*, **31**, 6963-6975.

8. Sun, H. and Chasin, L.A. (2000) Multiple splicing defects in an intronic false exon. *Molecular and cellular biology*, **20**, 6414-6425.

9. Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical journal*, **417**, 15-27.

10. Anko, M.L. (2014) Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin. Cell Dev. Biol.*, **32**, 11-21.

11. Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T. and Chabot, B. (2007) hnRNP proteins and splicing control. *Advances in experimental medicine and biology*, **623**, 123-147.

12. Busch, A. and Hertel, K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley interdisciplinary reviews. RNA*, **3**, 1-12.

13. Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schoneweis, K., Hertel, K.J. and Schaal, H. (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA*, **19**, 96-102.

14. Reber, S., Stettler, J., Filosa, G., Colombo, M., Jutzi, D., Lenzken, S.C., Schweingruber, C., Bruggmann, R., Bachi, A., Barabino, S.M. *et al.* (2016) Minor intron splicing is regulated by FUS and affected by ALS-associated FUS mutants. *The EMBO journal*, **35**, 1504-1521.

15. Shenasa, H., Movassat, M., Forouzmand, E.J. and Hertel, K.J. (2020) Allosteric Regulation of U1 snRNP by Splicing Regulatory Proteins Controls Spliceosomal Assembly. *RNA*.

16. Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M. and Baralle, F.E. (2004) hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic acids research*, **32**, 4224-4236.

17. Matera, A.G. and Wang, Z. (2014) A day in the life of the spliceosome. *Nature reviews. Molecular cell biology*, **15**, 108-121.

18. Ptok, J., Muller, L., Theiss, S. and Schaal, H. (2019) Context matters: Regulation of splice donor usage. *Biochim Biophys Acta Gene Regul Mech*.

19. Zhang, X.H., Arias, M.A., Ke, S. and Chasin, L.A. (2009) Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *Rna*, **15**, 367-376.

20. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J. and Fairbrother, W.G. (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11093-11098.

21. Sterne-Weiler, T., Howard, J., Mort, M., Cooper, D.N. and Sanford, J.R. (2011) Loss of exon identity is a common mechanism of human inherited disease. *Genome research*, **21**, 1563-1571.

22. Caminsky, N., Mucaki, E.J. and Rogan, P.K. (2014) Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Research*, **3**, 282.

23. Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., Tosi, M. and Martins, A. (2016) Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS genetics*, **12**, e1005756.

24. Hartmann, L.T., Stephan; Niederacher, Dieter; Schaal, Heiner. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Frontiers in bioscience : a journal and virtual library*, **13**, 3252-3272.

25. Wai, H.A., Lord, J., Lyon, M., Gunning, A., Kelly, H., Cibin, P., Seaby, E.G., Spiers-Fitzgerald, K., Lye, J., Ellard, S. *et al.* (2020) Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genetics in Medicine*, **22**, 1005-1014.

26. Grodecka, L., Buratti, E. and Freiberger, T. (2017) Mutations of Pre-mRNA Splicing Regulatory Elements: Are Predictions Moving Forward to Clinical Diagnostics? *Int. J. Mol. Sci.*, **18**.

27. Canson, D., Glubb, D. and Spurdle, A.B. (2020) Variant effect on splicing regulatory elements, branchpoint usage, and pseudoexonization: strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Hum. Mutat.*

28. Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802-813.

29. Hartmann, L., Theiss, S., Niederacher, D. and Schaal, H. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front. Biosci.*, **13**, 3252-3272.

30. Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J. and Chasin, L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.*, **21**, 1360-1374.

31. Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O. and Schaal, H. (2014) Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic acids research*, **42**, 10681-10697.

32. Cheng, J., Nguyen, T.Y.D., Cygan, K.J., Celik, M.H., Fairbrother, W.G., Avsec, Z. and Gagneur, J. (2019) MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.*, **20**, 48.

33. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B. *et al.* (2019) Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, **176**, 535-548 e524.

34. Rowlands, C.F., Baralle, D. and Ellingford, J.M. (2019) Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells*, **8**.

35. Braun, S., Enculescu, M., Setty, S.T., Cortes-Lopez, M., de Almeida, B.P., Sutandy, F.X.R., Schulz, L., Busch, A., Seiler, M., Ebersberger, S. *et al.* (2018) Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. *Nature communications*, **9**, 3315.

36. Selden, R.F., Howie, K.B., Rowe, M.E., Goodman, H.M. and Moore, D.D. (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Molecular and cellular biology*, **6**, 3173-3179.

37. Brillen, A.L., Schoneweis, K., Walotka, L., Hartmann, L., Muller, L., Ptok, J., Kaisers, W., Poschmann, G., Stuhler, K., Buratti, E. *et al.* (2017) Succession of splicing regulatory elements determines cryptic 5ss functionality. *Nucleic Acids Res.*, **45**, 4202-4216.

38. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry*, **162**, 156-159.

39. Brillen, A.L., Walotka, L., Hillebrand, F., Muller, L., Widera, M., Theiss, S. and Schaal, H. (2017) Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b. *J. Virol.*, **91**, e00389-00317.

40. Ke, S., Zhang, X.H. and Chasin, L.A. (2008) Positive selection acting on splicing motifs reflects compensatory evolution. *Genome research*, **18**, 533-543.

41. Kaisers, W., Boukamp, P., Stark, H.J., Schwender, H., Tigges, J., Krutmann, J. and Schaal, H. (2017) Age, gender and UV-exposition related effects on gene expression in in vivo aged short term cultivated human dermal fibroblasts. *PLoS One*, **12**, e0175657.

42. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

43. Chhangawala, S., Rudy, G., Mason, C.E. and Rosenfeld, J.A. (2015) The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome biology*, **16**, 131.

44. Kopylova, E., Noe, L. and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211-3217.

45. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.

46. Dobin, A. and Gingeras, T.R. (2015) Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics*, **51**, 11 14 11-19.

47. Kaisers, W., Schaal, H. and Schwender, H. (2015) rbamtools: an R interface to samtools enabling fast accumulative tabulation of splicing events over multiple RNA-seq samples. *Bioinformatics*, **31**, 1663-1664.

48. Kaisers, W., Ptok, J., Schwender, H. and Schaal, H. (2017) Validation of Splicing Events in Transcriptome Sequencing Data. *International journal of molecular sciences*, **18**.

49. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R. *et al.* (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, **347**, 1254806.

50. Rosenberg, A.B., Patwardhan, R.P., Shendure, J. and Seelig, G. (2015) Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell*, **163**, 698-711.

51. Afroz, T., Cienikova, Z., Clery, A. and Allain, F.H.T. (2015) One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol.*, **558**, 235-278.

52. Erkelenz, S., Hillebrand, F., Widera, M., Theiss, S., Fayyaz, A., Degrandi, D., Pfeffer, K. and Schaal, H. (2015) Balanced splicing at the Tat-specific HIV-1 3'ss A3 is critical for HIV-1 replication. *Retrovirology*, **12**, 29.

53. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M. and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*, **13**, 731-740.

54. Hargous, Y., Hautbergue, G.M., Tintaru, A.M., Skrisovska, L., Golovanov, A.P., Stevenin, J., Lian, L.Y., Wilson, S.A. and Allain, F.H. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *The EMBO journal*, **25**, 5126-5137.

55. Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. *et al.* (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054-2057.

56. Dember, L.M., Kim, N.D., Liu, K.Q. and Anderson, P. (1996) Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *The Journal of biological chemistry*, **271**, 2783-2788.

57. Baralle, D. and Baralle, M. (2005) Splicing in action: assessing disease causing sequence changes. *Journal of medical genetics*, **42**, 737-748.

58. Cyphert, T.J., Suchanek, A.L., Griffith, B.N. and Salati, L.M. (2013) Starvation actively inhibits splicing of glucose-6-phosphate dehydrogenase mRNA via a bifunctional ESE/ESS element bound by hnRNP K. *Biochimica et biophysica acta*, **1829**, 905-915.

59. Alioto, T.S. (2007) U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.*, **35**, D110-115.

60. Arias, M.A., Lubkin, A. and Chasin, L.A. (2015) Splicing of designer exons informs a biophysical model for exon definition. *RNA*, **21**, 213-229.

61. Erkelenz, S., Theiss, S., Kaisers, W., Ptok, J., Walotka, L., Muller, L., Hillebrand, F., Brillen, A.L., Sladek, M. and Schaal, H. (2018) Ranking noncanonical 5' splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. *Genome research*.

62. Barrett, N.L., Li, X. and Carmichael, G.G. (1995) The sequence and context of the 5' splice site govern the nuclear stability of polyoma virus late RNAs. *Nucleic Acids Res.*, **23**, 4812-4817.

63. Hammarskjold, M.L., Li, H., Rekosh, D. and Prasad, S. (1994) Human immunodeficiency virus env expression becomes Rev- independent if the env region is not defined as an intron. *The Journal of Virology*, **68**, 951-958.

64. Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H. and Kjems, J. (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol. Cell*, **29**, 271-278.

65. Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., Jr. and Fu, X.D. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell*, **50**, 223-235.

66. Barnard, D.C., Li, J., Peng, R. and Patton, J.G. (2002) Regulation of alternative splicing by SRrp86 through coactivation and repression of specific SR proteins. *Rna*, **8**, 526-533.

67. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.*, **31**, 3568-3571.

68. Zhang, X.H. and Chasin, L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.*, **18**, 1241-1250.

69. Rhine, C.L., Neil, C., Glidden, D.T., Cygan, K.J., Fredericks, A.M., Wang, J., Walton, N.A. and Fairbrother, W.G. (2019) Future directions for high-throughput splicing assays in precision medicine. *Hum. Mutat.*, 40, 1225-1234.

70. Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J. and Fairbrother, W.G. (2017) Pathogenic variants that alter protein code often disrupt splicing. Nat. Genet., 49, 848-855.

71. Cheng, J., Celik, M.H., Nguyen, T.Y.D., Avsec, Z. and Gagneur, J. (2019) CAGI 5 splicing challenge: Improved exon skipping and intron retention predictions with MMSplice. Hum. Mutat.

3.1.3. Position dependent regulation of splice acceptor sites (Chapter 3)

Splice acceptors sites pose as counterparts for 5' splice sites, as both sequences are sequentially recognized in the early splicing process before they are subsequently ligated in the course of the splicing reaction. While the 11nt long 5'ss recognition motif is rather short, the regulatory sequences required for efficient splice acceptor recognition can span hundreds of nucleotides into the upstream intron to the respective branch point sequence. Additional regulatory sequences around the CAG/G consensus motif of the intron/exon border is the polypyrimidine tract, which is located in the upstream intron and is located about 5–40 base pairs upstream of the canonical AG dinucleotide of the 3'ss. The MaxEnt score, a bioinformatics tool that provides information about a sequences' likelihood to pose a splicing signal, evaluates putative 3'ss intrinsic strength by calculating their (23nt: 20 intronic / 3 exonic) similarity to used 3'ss/consensus 3'ss. However, this is only one side of the medal: Additional information about splice acceptor strength and the likelihood of usage can be drawn when the MaxEnt score algorithm is combined with HEXplorer predictions that allow the analyses of protein binding partners, as well as the concept of position dependency since 3' splice sites, follow similar rules of regulation by RNA binding proteins as 5' splice sites.

The following manuscript draft is unpublished.

The final list of authors will be determined after completion of the project.
Individual contribution: L.M. designed and cloned the constructs that are shown in Figure 1, 2, 3. These provide the basis for constructs shown in Figure 4 and 5 that were generated during a masters project under supervision of L.M. Figure 6 was generated during a bachelors project under supervision of L.M. The manuscript draft was written by L.M.

Equidirectional support to both exon-defining splice sites reverses position-dependent splicing activation and repression by SR and hnRNP proteins.

Introduction:

Correct recognition of exon-intron boundaries, namely the splice donor (5'ss, SD) and splice acceptor sites (3′ss, SA) is key for nucleotide-accurate removal of intron sequences.

The 3' splice site is primarily characterized by its intronically located polypyrimidine tract (PPT) consisting of about 10-20 pyrimidine nucleotides and the AG dinucleotide that constitutes the intron/exon border by the CAG/G consensus motif (Cartegni, Chew et al. 2002). Additional regulation of the splice acceptor is achieved by the branch point sequence constituted by the YNYURAC (Y = pyrimidine, R = purine, N = any nucleotide) consensus motif and in particular, the branch point adenosine mostly located within 15-50 nucleotides in the upstream intron (Wahl, Will et al. 2009).

The main interaction partners of splice acceptors within the early spliceosome are the U2 snRNP auxiliary factors (U2AF). The larger 65-kDa subunit of the U2AF heterodimer, namely U2AF65, binds to the PPT (Valcarcel, Gaur et al. 1996) while the smaller 35-kDa subunit (U2AF35) interacts directly with the conserved AG dinucleotide (Wu, Romfo et al. 1999) in almost 90% of cases (Wu and Fu 2015). Furthermore, U2AF35 plays an important role in interactions with splicing regulatory proteins (SRPs), including SR proteins, that in turn can also positively act on U1 snRNP binding to the SD across the exon by stabilizing spliceosomal components at splice sites through protein-protein interactions (Zuo and Maniatis 1996, Graveley 2000). Apart from U2AF heterodimer subunits, homologs such as U2AF26, Urp and PUF60 have been reported to participate in heterodimer formation and interaction with SR proteins (Tronchere, Wang et al. 1997, Page-McCaw, Amonlirdviman et al. 1999, Shepard, Reick et al. 2002). In recent years, more and more factors interacting with the polypyrimidine tract in a positively or negatively regulatory manner have been described (Kovacova, Soucek et al. 2020).

Generally, it is known that both constitutive and alternative splicing, is regulated by cis-acting splicing regulatory elements (SREs) offering binding sites for trans-acting splicing regulatory proteins that can act from both, intronic and exonic positions (Busch

and Hertel 2012). Major protein families involved in this regulation are serine/arginine (SR)-rich proteins and heterogeneous nuclear ribonucleoproteins hnRNP proteins (Wang and Burge 2008). For 5' splice sites, a strict position-dependent regulation has been described where SR proteins act enhancing from an upstream position and repressing from a downstream position and hnRNP proteins vice versa (Erkelenz, Mueller et al. 2013). In a competition situation with more than one putative splice donor site in close vicinity, SREs highly regulate the usage and repression of one or the other respectively (Ptok, Müller et al. 2019). Here we show that this concept of position dependency and the regulation of competing splice sites can also be transferred to splice acceptor sites.

Furthermore, to qualitatively and quantitatively assess the interaction between SREs and splice sites, various bioinformatics approaches aiming to score the likelihood of particular splice donor and acceptor usage have been developed (Rosenberg, Patwardhan et al. 2015, Moles-Fernandez, Duran-Lozano et al. 2018). One approach to scoring the intrinsic strength of 3' splice sites (SA) is the 3'ss MaxEnt scoring. This algorithm is based on maximum entropy, and the maximum value for the strength of a SA is 13.59 (Eng, Coutinho et al. 2004). The MaxEnt score of a SA is calculated by taking 23 nucleotides into account (20 intronic nucleotides at an AG intron/exon border and the first three exonic nucleotides) (Yeo and Burge 2004). A tool that assesses whether sequences are generally likely to bind either SR or hnRNP proteins is the HEXplorer tool. It analyses overlapping hexamers and calculates a score according to their exonic and intronic frequency in vicinity of SD. A sequence with an HEXplorer score (HZEI) > 0 is considered to putatively act as a binding site for SR or related proteins while sequences with negative HZEI scores are considered to primarily bind proteins related to the hnRNP family (Erkelenz, Hillebrand et al. 2015). Both tools have been used throughout several studies and have been found to robustly predict splicing outcomes (Soukarieh, Gaildrat et al. 2016, Grodecka, Buratti et al. 2017).

Current models describe the positive regulation of splice acceptors and subsequent exon inclusion by SR proteins from an exonic position while hnRNP proteins exhibit an adverse behavior. This goes in line with the regulation of splice donors, however, here it has been shown that splice donor recognition can be regulated by hnRNP proteins from an intronic position as well. The role of SREs and SRPs becomes more complex

in the regulation of competing splice sites, where one SRE can have two functions depending on the actual position towards a splice site. This regulation has been thoroughly shown (Erkelenz, Mueller et al. 2013, Brillen, Schoneweis et al. 2017).
In the present work, we showed that the concept of position-dependent SRE regulation, which is already established for splice donor regulation, is also transferable to splice acceptor sites. To demonstrate that the concept of position-dependent SRE regulation is transferable to splice acceptor regulation, a series of 3-exon minigene splicing reporters with two competing splice acceptor sites was employed. Furthermore, additional regulation through exon-length dependent exon definition in combination with varying intrinsic splice donor strength was analyzed. Furthermore, we tested whether a combination of the intrinsic strength of a SA as determined by the MaxEnt score algorithm and the surrounding SRE landscape assessed via the HEXplorer tool with regard to a position-dependent regulation of SA can reliably describe its use. By combining information from both bioinformatic tools, MaxEnt score, and HEXplorer algorithm, a robust approximation of the recognition and usage potential of a splice acceptor could be drawn.

Results

For the middle exon HEXplorer in silico designed positive (10.32) / neutral (-0.15) / negative (-10.35) SREs (octamers) were paired with intrinsically strong splice sites (Figure 1A and 1B). The mode of action of the *in silico* designed SRE sequences is confirmed in Müller et. al., (in revision). Following the concept of position-dependent regulation, insertion of octamer 10.32, which has been shown to bind SR proteins, led to full exon inclusion of the middle exon, as expected. Similarly, insertion of octamer -10.35, which is assumed to bind hnRNP and hnRNP-like proteins, resulted in full skipping of the middle exon. Surprisingly, insertion of the neutral octamer -0.15 which has been shown to bind neither SR nor hnRNP proteins, resulted in a higher level of exon inclusion than exon skipping. This suggests that the intrinsically strong splice sites (SA MaxEnt 11.7, SD Hbond score 17.5) used in this splicing reporter are generally well recognized and exon inclusion occurs almost without additional support from exonic SREs (Figure 1C).

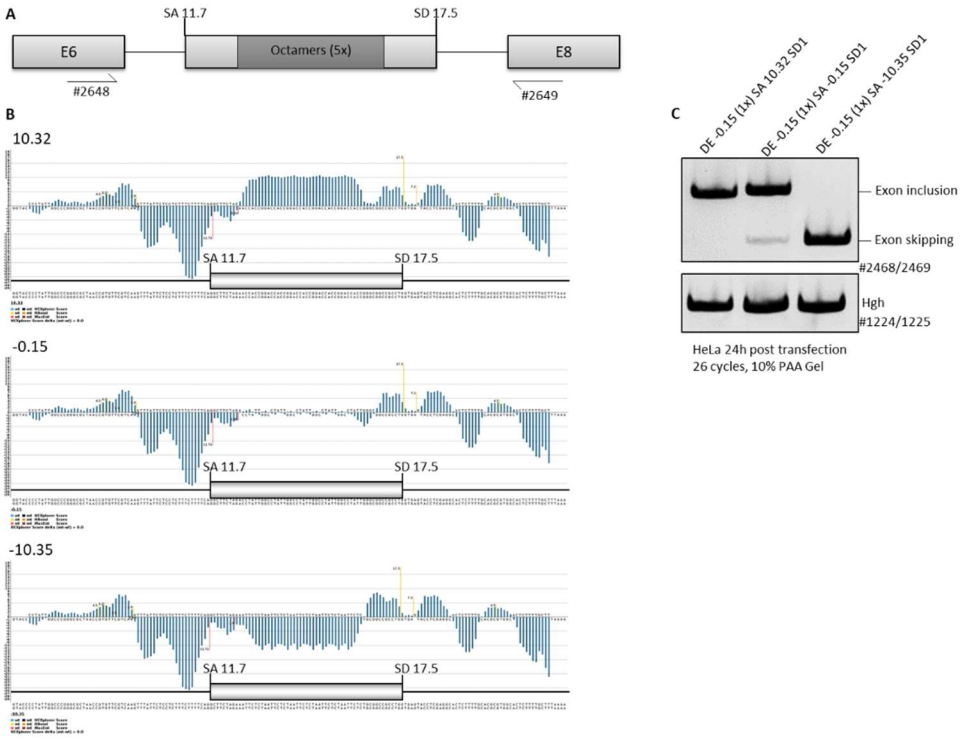


Figure 1: Regulation of splice sites from an exonic position.

A *Schematic drawing of the 3-exon minigene splicing reporter. The middle exon carries five copies of the in silico designed octamers, framed by an SA with a MaxEnt score of 11.7 and an SD with an HBond score of 17.5.* ***B*** *HEXplorer plots of the used octamer sequences in the construct.* ***C*** *HeLa cells were transfected with the constructs and hgh was used to monitor transfection efficiency. RNA was harvested 24h post-transfection, subjected to cDNA synthesis and subsequent RT-PCR to analyze the splicing pattern using primer pair #2648/#2649 via a 10% non-denaturing PAA gel.*

Thus, based on their position in this test exon, it can be concluded that SREs and the proteins bound thereto equally support and repress the upstream and downstream splice sites, respectively. Hence, in the case of splice acceptors, it is hypothesized that they are supported by SR protein binding from an exonic position while they repress its usage from an intronic position and that hnRNP proteins act vice versa. The underlying principle is a simultaneous regulation of both splice sites when a respective SREs is located in between, which ultimately leads to either supported or inhibited exon recognition. This ultimately means that in the case of two proximal, competing splice acceptor sites, SREs binding splicing regulatory proteins would regulate the use of one or another splice acceptor in a relative position-dependent way.

To systematically test this hypothesis, another variation of the previously used 3-exon minigene reporter was generated.

The middle exon was extended at its 5' end by a copy of the splice acceptor (MaxEnt 11.7) with the same PPT and putative BPS. In the case of using this newly added splice acceptor, the original splice acceptor is located in the exon. Whether in this reporter the position of an SRE between these competing sites is exonic or intronic depends on the respective use of the competing splice acceptor. But what is unaffected in any case is the position of the SRE, namely upstream or downstream from the competing splice acceptor.

The SD of this exon with competing SA is an intrinsically strong SD with an HBS of 17.5 (Figure 2A). To simulate the regulation of the competing SA from up and downstream positions, previously described synthetic octamer SREs were inserted in all possible combinations (Figure 2B).

Depending on which splice acceptor, the distal (SAd) or the proximal (SAp) is recognized by the splicing machinery, either the long middle exon (325 bp), the short exon (227 bp), or none, i.e. only the two flanking exons (168bp) are included in the final mRNA transcript. Testing the synthetic octamer sequences in this splicing reporter, RT-PCR revealed that depending on the SRE combination, splicing outcome varied: i) the middle exon was skipped (-10/-10), ii) only the longer (using SAd, -10/0) or shorter exon (using SAp, 0/+10) was included, iii) both, exon inclusion and skipping (-10/0) could be observed or iv) both exons (SAd and Sap, +10/-10) were included (Figure 2C). Thus, regulation of the two splice acceptors strictly follows the hypothesized rules of position dependency of the SREs.

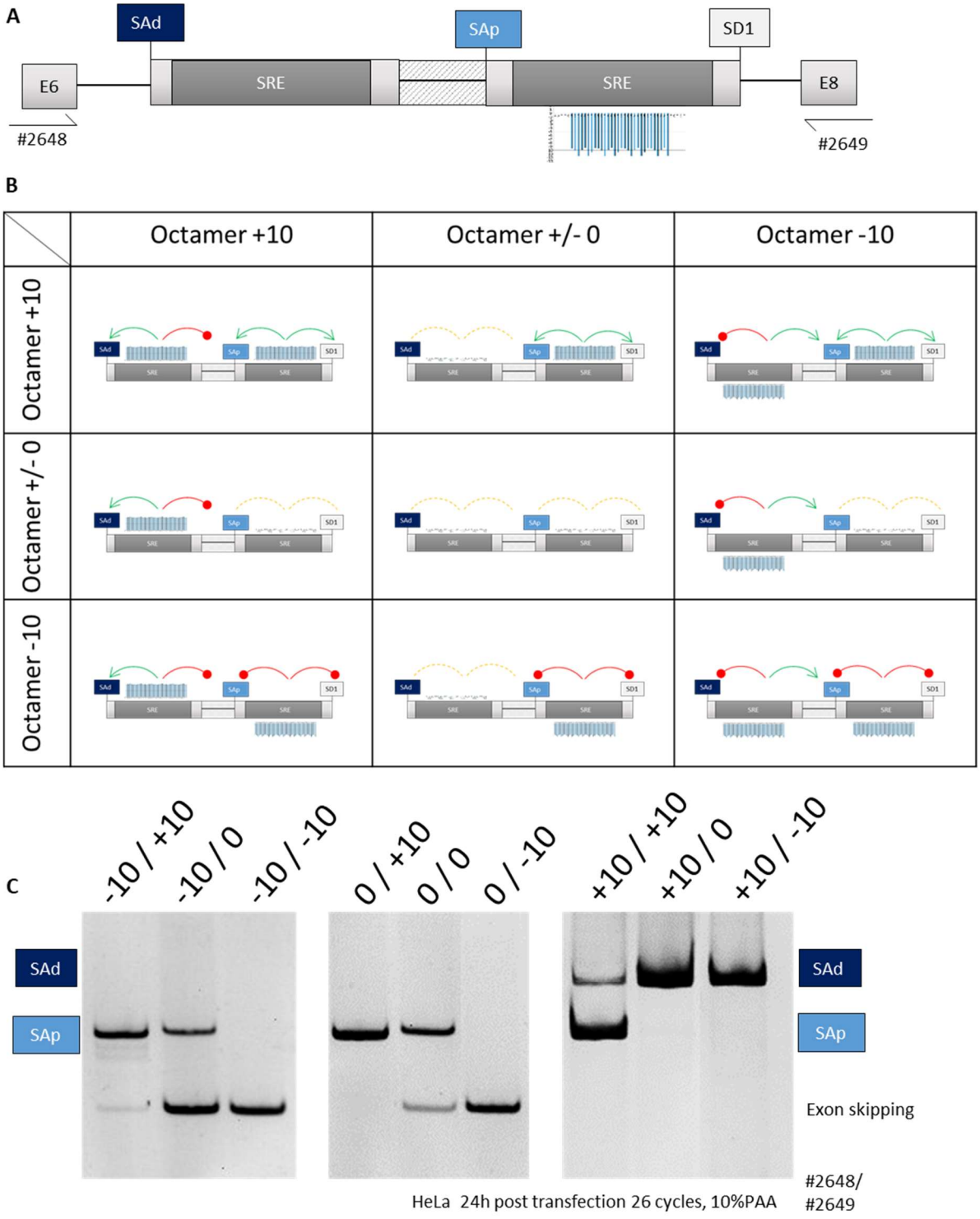


Figure 2: Combinations of SREs in a splicing reporter system.

A *Systematic combination of in silico SREs that either pose as binding sites for SR proteins (Octamer +10.32), hnRNP proteins (Octamer -10.35), or none splicing regulatory protein, i.e. neutral (Octamer -0.15) with the predicted effect on splice site recognition and exon inclusion.* ***B*** *HeLa cells were transfected with the respective reporter plasmids and RNA was harvested 24h post-transfection.*

Splicing patterns were analyzed by RT-PCR with primer pair #2648/#2649 and gel-electrophoresis using a 10% non-denaturing polyacrylamide gel.

Next, we addressed the question to what extent the intrinsically strong splice donor in this reporter construct contributes to the recognition of either the distal or the proximal splice acceptor and the subsequent recognition and inclusion of the exon. Therefore, we paired constructs that showed full recognition of SAd (+10/-10) or SAp (-10/+10) with splice donors with decreasing intrinsic strength measured by HBS (HBS 17.5 – HBS 10.7) (Figure 3A). The long exon (SAd) was more susceptible to changes in the intrinsic SD strength than the short SAp exon fitting the idea of a cross-talk between splice donor and splice acceptor which could be disturbed by further protein binding on the RNA between them. The middle exon containing SREs that allow the recognition of SAd lost full recognition when paired with an SD with an HBS of 16.3 and was fully skipped when paired with an SD with an HBS of only 12.1 (Figure 3B left panel), whereas for the inclusion of the short exon with SREs that enhanced the use of the SAp, a splice donor with an HBS of only 12.1 was completely sufficient (Figure 3B right panel). However, the decreased susceptibility of the short exon could be explained by the dual function of the proximal SRE. The potent octamer +10.32 which was shown to bind SRSF3 is in this set of constructs not only in close proximity of the SA but could also have enhancing effects on the closely downstream located SD, supporting the recognition and subsequent use of both, the 3’ and 5’ splice sites. Adversely, the construct in which SAd is used contains a strong hnRNP binding site downstream of the second SA repressing it, however, that means that it is located directly upstream of the SD, potentially influencing its recognition by the U1 snRNP as well.

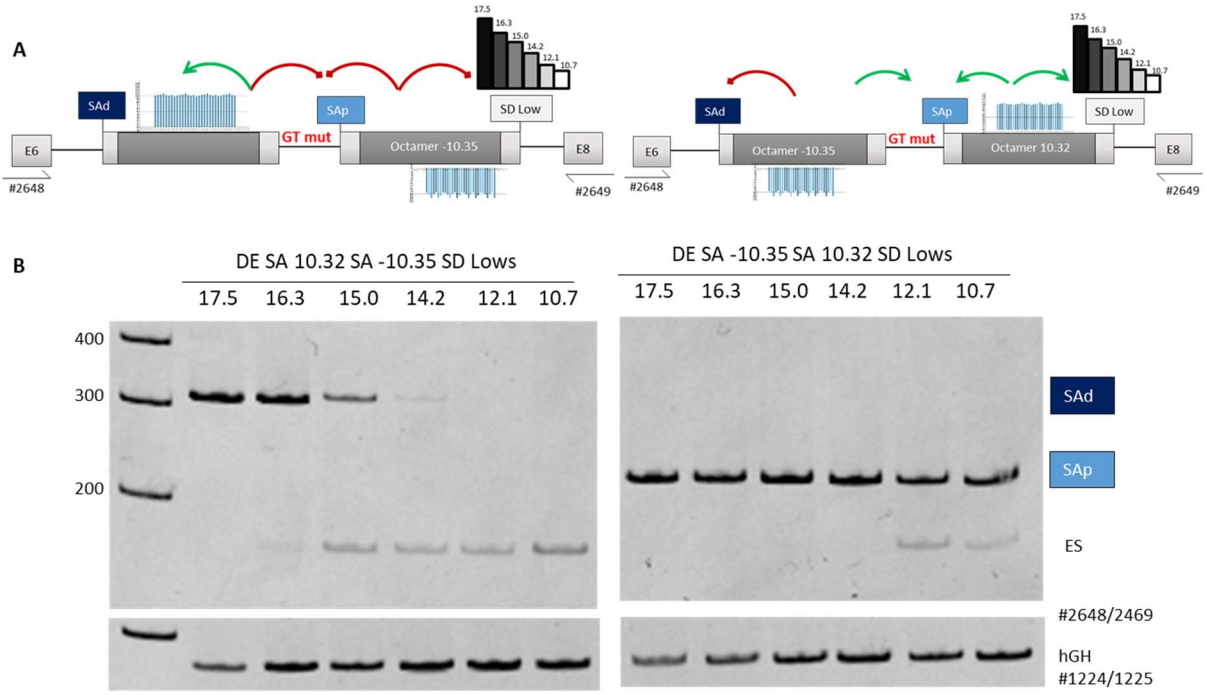
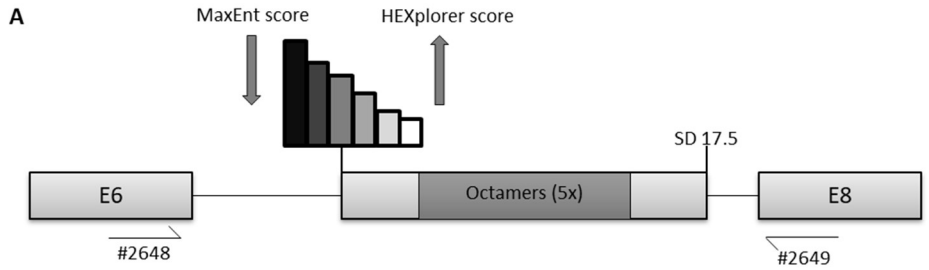


Figure 3: Influence of decreasing splice donor strengths on proximal and distal splice acceptor use.

A *Schematic drawing of the 3-exon minigene splicing reporter constructs. SRE were either combined to enhance the distal or proximal splice acceptor.* ***B*** *HeLa cells were transfected with the respective constructs and a plasmid encoding hgh was used to monitor transfection efficiency. 24-h post-transfection, RNA was harvested and splicing patterns were analyzed via RT-PCR using primer pairs #2648/#2649 and #1224/#1225 for hgh. When enhancing the distal splice acceptor (left panel), the loss of full exon inclusion can already be seen for the combination with SD 16.3. In the construct where the proximal splice acceptor is supported, full exon recognition is achieved even for the weak splice donor with an HBond score of 10.7 (right panel).*

Equivalent to the experiments with decreasing splice donor strengths we systematically tested twelve decreasing splice acceptor strengths, which were calculated using the MaxEnt score algorithm, and their susceptibility towards changes in SRE strength. When designing these SA, one main focus was that the HEXplorer ΔHZEI score was increasing simultaneously with decreasing MaxEnt score. An increasingly positive HZEI score indicates an increase in the likelihood of SR and -like protein binding while the likelihood of hnRNP and -like protein binding decreases. Accordingly, SA with an HEXplorer positive upstream sequence are less likely to be recognized and used due to the increase of potential SR binding that represses them

from an upstream position. In line, a decreasing MaxEnt score indicates a decreasing likelihood of the sequence to be used as a splice acceptor (Figure 4A and B).



B

MaxEnt	SA ΔHZ_{EI}
11,60	0
10,65	76,3
10,00	102,93
9,19	115,29
8,60	117,17
7,47	154,53
6,14	177,45
5,49	183,07
4,27	210,21
2,99	258,24
2,26	288,81
1,10	298,97
0,50	323,44

Figure 4: Splice acceptor sites with decreasing MaxEnt score and increasing HEXplorer score. ***A*** *Schematic drawing of the 3-exon minigene splicing reporter with amended splice acceptor sites.* ***B*** *MaxEnt scores of the used splice acceptor sites and respective changes in the HEXplorer score (∆HZEI) compared to the native splice site.*

First, the splicing neutral octamer -0.15 was used as an exonic sequence, hence neither the splice acceptor with its decreasing MaxEnt scores (10.65 to 0.50) nor the splice donor with an Hbond score of 17.5 received SRE support from an exonic position. Surprisingly, with respect to the previous experiment, in this construct, no combination led to exon recognition (Figure 5A). In its native form as described in Figure 1, insertion of the splicing neutral octamer -0.15 into this construct led to a high level of exon recognition in combination with a splice acceptor with a MaxEnt score of 11.7 and a splice donor of 17.5. In the above-used construct (Figure 1), however, the used splice acceptor was an optimized version of SA7 derived from HIV 1, carrying an optimal branch point sequence. For calculation of the intrinsic strength of splice acceptors used in this systematic evaluation, only the 23 MaxEnt scored nucleotides were used without an optimized branchpoint sequence, which might contribute to the reduced exon recognition.

By increasing the SRE support from an exonic position with only three point mutations in the splicing neutral sequence (∆HZEI of 107.3 compared to the splicing neutral octamer -0.15), this alternation induced exon recognition for constructs with a splice acceptor with a MaxEnt score > 6.14, with exon recognition being fully lost below this

score. Furthermore, exon recognition seemed to be highly dependent on the SA strength as a stepwise decrease of recognition and a stepwise increase in exon skipping was correlated with the MaxEnt score and the respective ΔHZEI difference (Figure 5B).

Insertion of octamer 10.32 in these constructs could underpin this observation. Again, an SA strength dependent induction of exon inclusion was seen, however, the cut-off MaxEnt score where exon recognition is lost highly shifted to a MaxEnt score of 0.50 (Figure 5C).

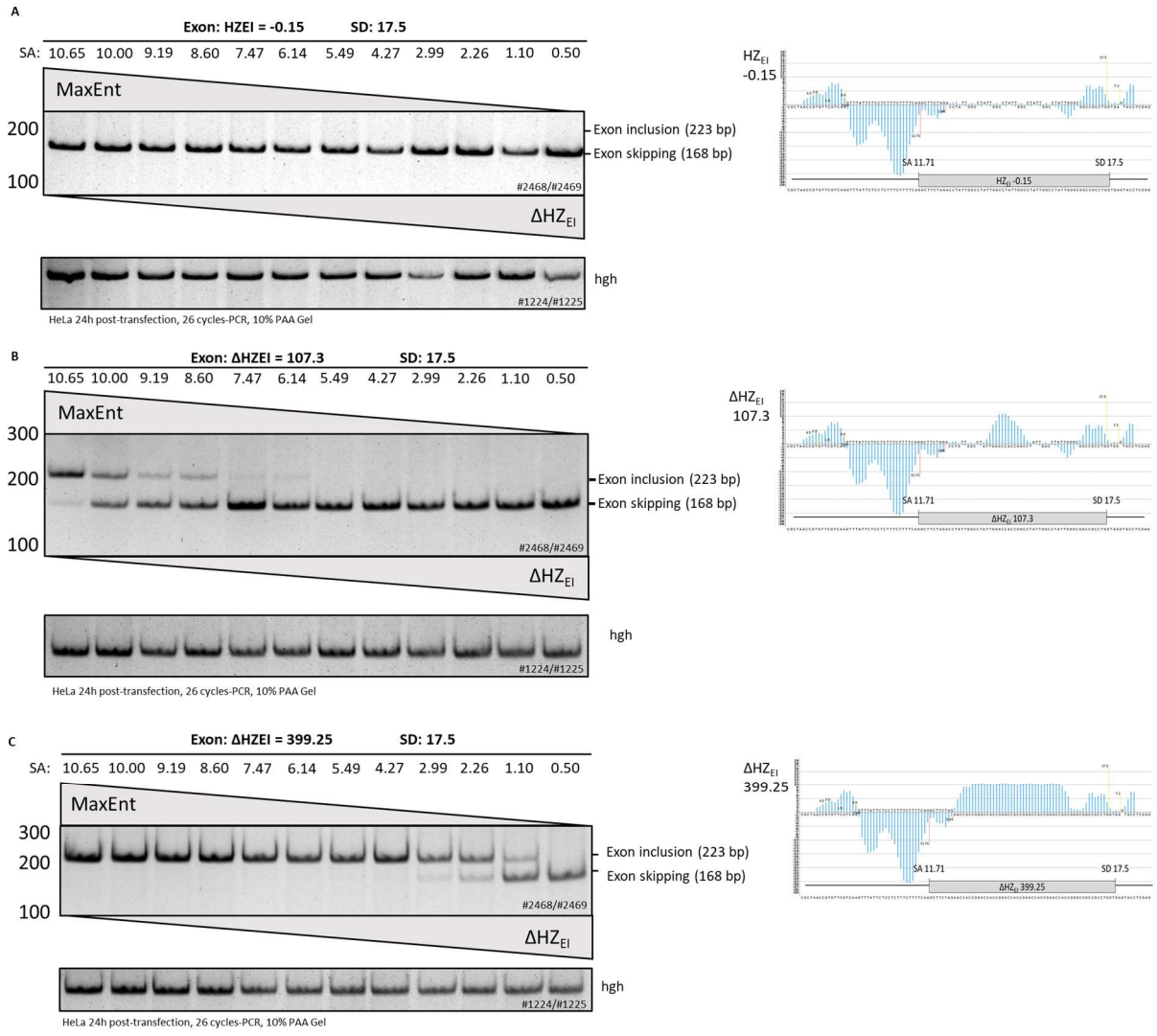


Figure 5: SRE and SA strength dependent titration of exon recognition.

A *Constructs with splice acceptor sites with decreasing MaxEnt score/increasing ΔHZEI were combined with a splicing neutral SRE.* ***B*** *Constructs with splice acceptor sites with decreasing MaxEnt score were combined with a moderate enhancing SRE with a ΔHZEI of 107.3 compared to the splicing neutral SRE*

(A). ***C*** *Constructs with splice acceptors with decreasing MaxEnt score were combined with a highly enhancing SRE with a ∆HZEI of 399.25 compared to the splicing neutral SRE (A). For all constructs, HeLa cells were transfected with the respective constructs and a plasmid encoding hgh as transfection control. 24-h post-transfection, RNA was harvested and splicing patterns were analyzed via RT-PCR using primer pairs #2648/#2649 and #1224/#1225 for hgh.*

To further dissect the concept of SA regulation by its PPT and to test the functional prediction of the MaxEnt score and the HEXplorer score, a second reporter system based on HIV 1 pNL4-3 was employed. The 4-exon3-intron construct derived from HIV-1 is under the control of an LTR promoter and at the 3' end is an SV40 polyadenylation signal. It contains the HIV 1 native 5' splice sites SD1, SD2, SD3, and the native 3' splice sites SA1, SA2, SA3. Immediately downstream of exon 4 is the coding region for chloramphenicol acetyltransferase (CAT), which was used as an internal control (Figure 6A).

The main focus was on the regulation of SA3 by its upstream located polypyrimidine tract, which was not included in the MaxEnt score calculation in this case because of its position. In its wild type sequence, the pyrimidine content of the 29 nucleotides upstream of SA3 is 69.0%, the 18 scored intronic nucleotides (upstream of the consensus AG) of the acceptor have a MaxEnt score of 9.8. First, the pyrimidine content of the upstream sequence was stepwise decreased (constructs 2-5) without changing the MaxEnt score since the nucleotide exchanges were located outside of the scoring region of the algorithm. A change in the splicing outcome upon these changes could already be seen for construct #3 compared to construct #1 WT (Figure 6B, lane 3). While in the wild-type situation, SA3 was fully used, recognition and use was gradually decreased with decreasing pyrimidine content starting with construct #3, as indicated by the retention of the respective intron 3. Construct #5 with a pyrimidine content of 48.3% and a still unchanged MaxEnt score already showed a switch from high SA use to a high extent of intron retention (Figure 6B, lane 5). Upon further decreasing the pyrimidine content but then also simultaneously decreasing the MaxEnt score a full switch towards SA3 being not recognized and intron 3 being fully retained could be seen (Figure 6B lanes 6-9).

Interestingly, the change in pyrimidine content was fully depicted by a correlating ∆HZEI. Hence, in this case, the MaxEnt score did not fully pick up the prediction of

changes in the splicing outcome due to the restriction of 18 intronic nucleotides excluding the AG dinucleotide. Therefore, it could be argued that a combination of MaxEnt score and HEXplorer algorithm could contribute to a more reliable assessment of splice acceptor strength.

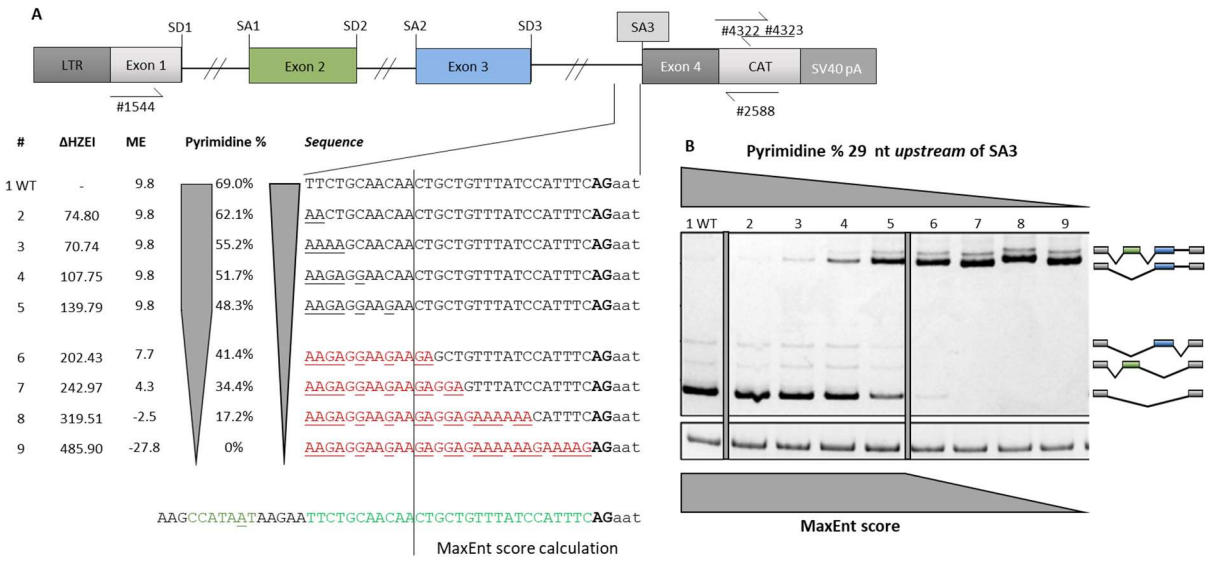


Figure 6: Additional intronic regulation of SA through PPT is not reflected by MaxEnt scoring.
A *Schematic overview of the splicing reporter construct. The 4-exon3-intron contains HIV-1 native 5' and 3' splice sites, downstream of exon 4 is the coding region for chloramphenicol acetyltransferase (CAT). It was focused on the region 29 nucleotides upstream of SA3. In a native setting, this region has a pyrimidine content of 69.0%, the splice acceptor strength is 9.8 (1 WT). For the following constructs, the MaxEnt score was kept constant while the pyrimidine content was decreased (#2-#5). The decrease in pyrimidine content was continued, in the following constructs with a simultaneous decrease in MaxEnt score (#6-#9).* ***B*** *RT-PCR analysis was run on a non-denaturing 10% PAA gel. 2.5x105 HeLa cells were co-transfected with 1 µg of each LTR 4-exon reporter and 0.2 µg each of SVctat to transactivate the LTR promoter. After 24 h incubation, RNA was isolated and reverse transcribed into cDNA using oligo d(T) primers. Semi-quantitative PCR was performed using primers #1544 and #2588. For comparison of spliced and total reporter mRNA, CAT sequences were amplified using primers #4322 and #4323.*

To confirm the hypothesis that a combination of MaxEnt score and HEXplorer algorithm might render the most accurate prediction of SA usage, we picked a random AG dinucleotide from the HIV-1 genome with a high MaxEnt score of 10.07, but poor splice site HEXplorer weight (SSHW). The 23 MaxEnt scored nucleotides were inserted into the previously employed three-exon minigene splicing reporter with an HIV-1 derived artificial exon and an SD with an HBS of 17.5. Additionally, two

HEXplorer designed SA were tested in the same construct. They were designed to have similar MaxEnt scores, but according to the HEXplorer plot analysis, more potent PPTs with a negative ΔHZEI compared to the randomly selected AG site. For the HEXplorer designed SA with a MaxEnt score of 9.9, the ΔHZEI compared to the random AG is -128.69, the HEXplorer designed SA with a MaxEnt score of 10.3 has an ΔHZEI of -174.1 (Figure 7A). While insertion of the randomly selected AG site with unaltered PPT has been found to lead to full exon skipping as expected due to the suboptimal PPT, insertion of the two HEXplorer designed SA led to an increase in exon inclusion (Figure 7B). This supports the suggestion that a combination of HEXplorer prediction and MaxEnt score could be advantageous for describing the functional strength of a SA.

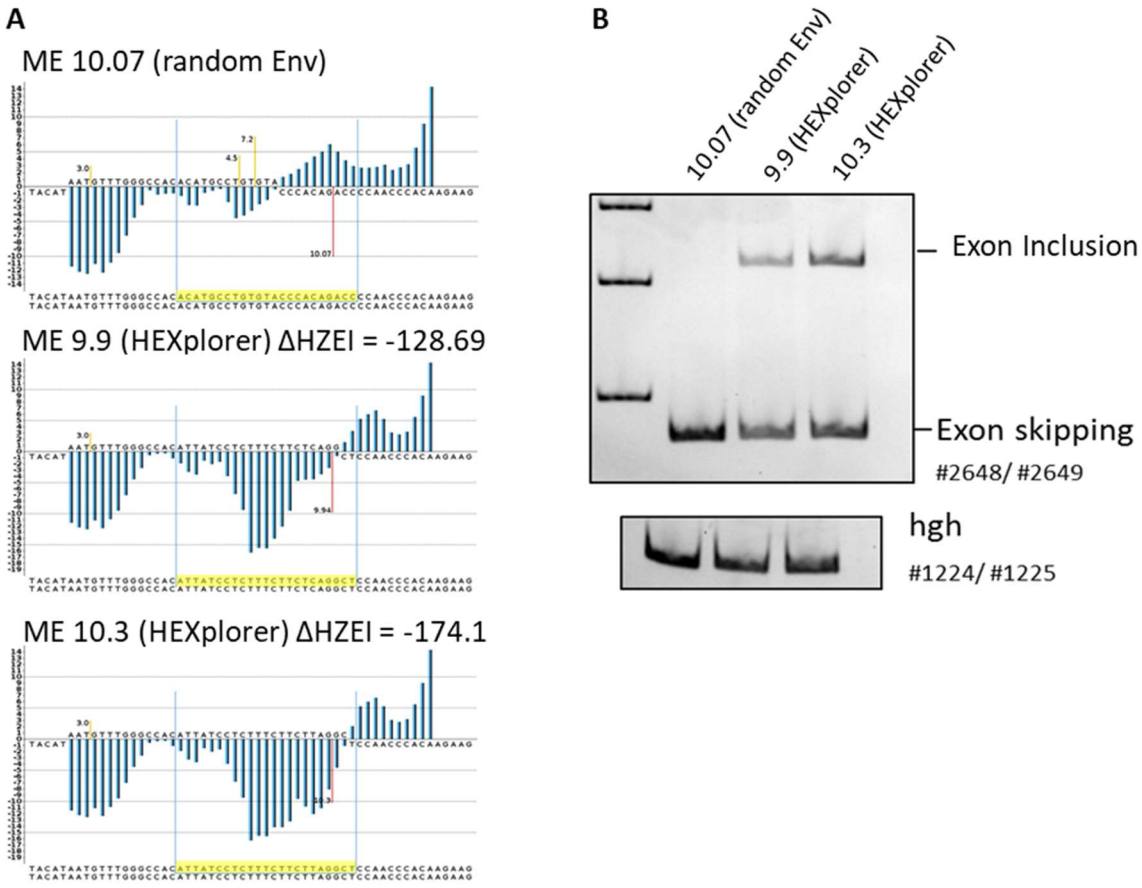


Figure 7: HEXplorer optimized PPTs contribute to SA use.

A *HEXplorer plots of a randomly chosen AG dinucleotide with an HEXplorer predicted suboptimal PPT and a MaxEnt (ME) score of 10.07. Below are two splice acceptors with similar MaxEnt scores (9.9 and 10.3 respectively) but HEXplorer optimized PPTs.* ***B*** *2.5x105 HeLa cells were co-transfected with 1 µg*

of the reporter construct and 1 µg of hgh as a control. After 24 h incubation, RNA was isolated and reverse transcribed into cDNA before being used for RT-PCR analysis with primer pair #2648/2649 or #1224/1225 for hgh respectively.

Discussion

Splice site regulation is highly dependent on neighboring splicing regulatory elements and as shown in this manuscript, regulation of not only splice donor sites but also splice acceptor sites relies on the position-dependent influence of SREs.

To further elucidate this concept, in silico designed SRE sequences with prescribed regulatory features were employed in different settings. First, they were tested in a simple middle exon of a three-exon-minigene construct. There, as expected, octamer 10.32 with a previously shown SR protein binding affinity acted positively on both splice sites and thus, exon recognition. Surprisingly, insertion of a splicing neutral sequence (-0.15) still resulted in a high level of exon recognition. This is possibly due to the intrinsically strong splice sites in the construct, which are sufficiently recognized even without the additional support of SREs, according to the concept that weaker splice sites require more SRE support than stronger splice sites (Erkelenz, Theiss et al. 2014). Insertion of octamer -10.35 that has been shown to preferably bind hnRNP and related proteins resulted in the expected loss of exon recognition due to the negative regulation of the upstream located splice acceptor and the downstream located splice donor (Figure 1).

In such a construct, however, it is not possible to distinguish singular splice site regulation from general exon recognition. Therefore, another splicing reporter model system was employed to further elucidate the position-dependent regulation of two equally strong competing splice acceptors. In particular, two identical copies of the HIV-1 derived SA7 were inserted into the three-exon-minigene reporter as well as insertion sites for the in silico designed octamer sequences (Figure 2A). This allowed testing of combinatorial regulation of the distal or proximal splice acceptor. In Figure 2B, the hypothesized regulation pattern of the two splice acceptors is depicted as dependent on the combination of SREs. Interestingly, these predictions were met as shown in Figure 2C, which displays that depending on the inserted SRE combination, either the distal or proximal acceptor was recognized, resulting in either the inclusion

of the short or the long exon variant. The understanding of the regulation of competing splice sites, both SA and SD, is in particular crucial in the evaluation of human pathogenic mutations. Various human pathogenic splicing defects are based on mutations located in SREs that can change their regulatory capacity, which then results in the loss of recognition of constitutive splice sites or increased recognition of alternative/cryptic sites. For competing splice acceptors, a case of a patient has been described where a single mutation within intron 2 of the dystrophin gene, whose misregulation is the main cause of dystrophinopathies, resulted in the inclusion of an alternative exon due to activation of a cryptic splice acceptor (Yagi, Takeshima et al. 2003, Abramowicz and Gos 2019). Interestingly, it has been proposed that in a setting with two closely located AG-dinucleotides, the one closer to the BPS is more frequently used (Smith, Chu et al. 1993), however, this is rather a trend than a rule (Gooding, Clark et al. 2006). For human genetic disorders, the majority of cryptic splice acceptors are located at -1 or -2, however, splice site switches to SA located in the further distance are also reported which are also likely regulated by SREs (Vorechovsky 2006).

As mentioned earlier, SREs located in close proximity between a SA and an SD may not only affect one or the other splice site but may affect exon recognition in general by equidirectionally supporting both exon-defining splice sites but in reverse position dependency. Therefore, we tested the interplay of the distal and proximal splice acceptors in a setting with differentially strong splice donor sites (Figure 3A). Analysis of the splicing pattern showed that the distal splice acceptor in this setting was more susceptible to a weakened splice donor compared to the proximal splice acceptor. The underlying cause might be the restraint of exon length for exon recognition versus intron recognition, hence a decreased interaction of the U1 snRNP and the U2AF in the distal acceptor setting versus the proximal acceptor setting. Generally, it is understood that both, exon and intron size highly influence the splicing outcome (Sterner, Carlo et al. 1996). It has been described that a minimal length of exons is required to prevent sterical hindrance of the spliceosomal components (Dominski and Kole 1991). Accordingly, a maximal exon size for efficient exon definition is given due to the hindrance of the interaction of both splice sites and their spliceosomal binding partners by in-between binding splicing regulatory proteins. Therefore, the majority of

exons displays a length of less than 200bp (Sakharkar, Chow et al. 2004, De Conti, Baralle et al. 2013)

The intrinsic strength of splice acceptors is primarily characterized by the polypyrimidine tract as well as its distance to the AG dinucleotide at the exon/intron border and towards the branchpoint sequence (Akerman and Mandel-Gutfreund 2007). In particular, longer uninterrupted PPTs lead to higher affinity binding sites for spliceosomal components and thus to more efficient recognition of splice acceptors (Dominski and Kole 1991). The SA MaxEnt score algorithm aims to score the intrinsic strength of SA by calculating the probability of the sequence to be used as a splice site according to its resemblance to used SA. Here, three exonic and 20 intronic nucleotides are taken into account, accordingly, the PPT upstream of the intron/exon border highly influences the evaluation of the sequence (Yeo and Burge 2004). In a systematic approach, SA with decreasing MaxEnt scores were tested in combination with differentially strong SREs located within the exon. While decreasing the MaxEnt score, simultaneous changes in the HEXplorer score (ΔHZEI) compared to the parental construct were as well-considered, since the HEXplorer score indicates putative binding of SRP including those regulating SA (Figure 4). This systematic approach demonstrated the combinatorial regulation of SA by intrinsic splice site strength and exonic SREs (Figure 5). Furthermore, it showed a combined discriminatory ability of the HEXplorer score and the MaxEnt score to predict functional SA.

In a reporter construct that allows for the evaluation of the regulation of a SA located at the last exon, the intronic regulation by the respective PPT was further dissected. Here, 29nt upstream of the AG dinucleotide were taken into account which exceeds the region that is included in the MaxEnt score calculation. The splicing pattern was monitored while the pyrimidine content of the intronic region was stepwise decreased from 69% to 0% by point mutations from the 5’ direction (Figure 6). The pyrimidine content upstream of SA is described as an important regulator of efficient SA use (Fu, Ge et al. 1988). In the reporter construct used here, the first indication of decreased SA use became visible when the pyrimidine content was reduced to 55.2%, however, this change was not picked up by the MaxEnt calculation as it remained at 9.8 as for the wildtype construct. A further reduction of the pyrimidine content to 41.4% provoked a decrease in the MaxEnt score from 9.8 to 7.7. Here, an almost complete switch from

SA use to the loss of recognition was already seen, hence further stepwise decrease of the MaxEnt score and the pyrimidine content had no additional effect. Interestingly, changes in the PPT correlated with increasing HEXplorer scores. Since the HEXplorer tool has the advantage that it is not restricted to a certain length for the analysis of a sequence, the combination of the two bioinformatics tools might provide additional beneficial information for the prediction of SA use.

To test this, a random AG dinucleotide from the HIV-1 genome with a high MaxEnt score (10.07) but suboptimal PPT according to HEXplorer predictions was tested as a SA in a three-exon-minigene reporter. Since the insertion of this sequence did not lead to exon inclusion, two splice acceptors with similar MaxEnt scores but optimized HEXplorer predicted PTTs were designed and tested. Other than the original sequence, the HEXplorer guided amendment of the PTT rendered the SA functional which resulted in exon inclusion. This strengthens the suggestion that a combination of the tools provides extended information for the prediction of SA use. More precise in silico splicing predictions including the consequences of mutations and the likelihood of use of cryptic splice sites are a highly desired tool in human genetics, especially in the light of evolving precision medicine (Jian, Boerwinkle et al. 2014, Rhine, Neil et al. 2019).

In this study, model systems were used to analyze the intricate interplay between splice acceptor strength, splice donor strength, and subsequent exon recognition and inclusion. To pave the way for an approach that allows the prediction of functional splice acceptor strength, two known bioinformatic tools evaluating the intrinsic splice site strength and the location and strength of splicing regulatory elements have been combined and were shown to provide more precise predictions. Discriminatory predictions of the splicing outcome are a highly desired tool in human genetics and the evaluation of pathogenic mutations.

Material and Methods

Bioinformatics.

The MaxEnt SA score algorithm is available under http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html and the HEXplorer tool is available under https://www2.hhu.de/rna/html/hexplorer_score.php.

Three-exon minigenes.

The three-exon minigenes which are derived from the fibrinogen Bß-minigene pT-Bß-IVS7ß 1G>T were previously described PMID: 28039323. The middle exon is amendable due to various restriction sites. Dual splice acceptor plasmids were cloned using two overlapping gene fragments ordered from Eurofins Genomics to generate Designer exon SA7 plain, which only contained two equal splice acceptors with their respective PPT and spacers with restriction sites in between. Splicing regulatory octamer sequences were inserted by either restriction digestion with XbaI and NotI using previously described plasmids (Müller et. al, in revision) or PCR products using NheI/XmaI (For Primers see Sup. Table 1). All oligonucleotides used were obtained from Metabion GmbH. SD Lows were inserted via restriction digestion with XbaI/XhoI using previously described plasmids (Müller et. al, in revision).

The HIV-based LTR 4-Exon-3-Intron-Reporter plasmids were derived from previously described plasmids (Widera, Erkelenz et al. 2013). All SA sequences were PCR generated (for primers see Sup. Table 1) and inserted into an intermediate pUC19 vector via NdeI/SalI. Positive clones were then again NdeI/SalI digested and inserted into the LTR vector.

SA with decreasing MaxEnt scores were generated by PCR and restriction digestion using KpnI/XbaI into previously published plasmids carrying the respective SREs (Müller et. al, in revision).

Expression plasmids.

pXGH5 expresses human growth hormone (hgh) under control of the mouse metallothionein-1 promoter (mMT1+) (Selden, Howie et al. 1986). Because the hormone is naturally synthesized only in cells of the anterior lobe of the pituitary gland, the plasmid is suitable for normalizing transfection efficiency in cell culture experiments.

SVctat encodes the protein of NL4-3 Tat (AS 1-86) from pUHctat (Schaal, Pfeiffer et al. 1993, Krummheuer, Lenz et al. 2001). The gene is under control of the SV40 early promoter and is followed at the 3' end by an SV40 polyadenylation signal. The plasmid

was co-transfected with the HIV-1 NL4-3 LTR 4-exon 3 intron reporter for transactivation of the HIV-1 LTR promoter in cell culture experiments.

Cell culture and transfection.

HeLa cells were cultured in Dulbecco's high-glucose modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and 50 g/ml penicillin-streptomycin. For transient transfection, 2x105 cells per well were plated in six-well plates. Transient-transfection experiments were performed using Mirus TransIT-LT1 transfection reagent according to the manufacturer's instructions.

RNA isolation and RT-PCR.

Either 24h post-transfection or 72h post-infection, total cellular RNA was isolated by using acid guanidinium thiocyanate-phenol-chloroform. RNA was reverse transcribed by using Superscript III reverse transcriptase (Invitrogen) and oligo(dT) primers (Invitrogen) for semiquantitative RT-PCR with the denoted primer pairs.

Sup. Table 1:
Gene Strand Part A:
TTTGGTACCC GCTAACCGTG TTCGTCAAGT TTATTCTCCT CTTTCTTTTC AGGCTGCTAG CCTAAGCGCT ATACCCGGGC GC
Gene Strand Part B:
ATACCCGGGC GCTAACCGTG TTCGTCAAGT TTATTCTCCT CTTTCTTTTC AGGCTTCTAG AATAGTCGAC GCTGCGGCCG CCTGGTGAGT ACCTCGAGGC ACTCTTTTGC ACGCGTGGC

Octamer 10.32
#5820 FW: 5 ' GCTGCTAGCACCACCGGACCACCGGACCACCGGACCA
#5821 REV: 5' GCGCCCGGGCCGGTGGTCCGGTGGTCCGGTGGTCC

Octamer -0.15
#5822 FW: 5' GCTGCTAGCCCTATTGGCCTATTGGCCTATTGGCCTATT
#5823 REV: 5' GCGCCCGGGCCAATAGGCCAATAGGCCAATAGGCCAATAGG,

Octamer -10.35
#5824 FW: 5 ' GCTGCTAGCAATTCTCTAATTCTCTAATTCTCTAATTCTCTAATT
#5825 REV: 5' GCGCCCGGGAGAGAATTAGAGAATTAGAGAATTAGAGAATTAGAGA

#3751: 5' CACCATATGTATATTTCAAGGAAAGCTAAGGA
#4167: 5'
TGCTATGTCGACACCCAATTCTGAAATGGATAAAGAGGAGTTGTTGGAGAATTCTTATTATGGC TTCCACTCCTGCCCA

#3753: 5‘
TGCTATGTCGACACCCAATTCTGAAATGTATAAACAGCAGTTGTTGCAGTATTCTTATTAGGCTT
CCACTCCTGCCCA
#4170: 5’
TGCTATGTCGACACCCAATTCTCTTTTCTTTTTTTCTCCTCTTCTTCCTCTTTTCTTATTATGGCTTC
CACTCCTGCCCA
#4285: 5‘
TGCTATGTCGACACCCAATTCTGAAATGTTTTTTCTCCTCTTCTTCCTCTTTTCTTATTATGGCTTC
CACTCCTGCCCA
#4286: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACTCCTCTTCTTCCTCTTTTCTTATTATGGCT
TCCACTCCTGCCCA
#4369: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACAGCTCTTCTTCCTCTTTTCTTATTATGGCT
TCCACTCCTGCCCA
#4287: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACAGCAGTTCTTCCTCTTTTCTTATTATGGCT
TCCACTCCTGCCCA
#4314: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACAGCAGTTGTTCCTCTTTTCTTATTATGGCT
TCCACTCCTGCCCA
#4288: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACAGCAGTTGTTGCTCTTTTCTTATTATGGCT
TCCACTCCTGCCCA
#4371: 5‘
TGCTATGTCGACACCCAATTCTGAAATGGATAAACAGCAGTTGTTGCAGTTTTCTTATTATGGC
TTCCACTCCTGCCCA

LTR SD1 SA2Ex2SD2 kompl. SA3 Py+ Ex3SD3 SA4 (pNL4-3) #3751, #4167
LTR SD1 SA2Ex2SD2 kompl. SA3 Py- Ex3SD3 SA4 (pNL4-3) #3751, #3753
LTR SD1 SA2Ex2SD2 kompl. SA3 Py-- Ex3SD3 SA4 (pNL4-3) #3751, #4170
LTR SD1 SA2Ex2SD2 kompl. +6 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4285
LTR SD1 SA2Ex2SD2 kompl. +13 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4286
LTR SD1 SA2Ex2SD2 kompl.+16 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4369
LTR SD1 SA2Ex2SD2 kompl. +20 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4287
LTR SD1 SA2Ex2SD2 kompl. +23 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4314
LTR SD1 SA2Ex2SD2 kompl. + 25 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4288
LTR SD1 SA2Ex2SD2 kompl. + 27 bp wt Ex3SD3 SA4 (pNL4-3) #3751, #4371

#6357
5‘TTTGGTACCATTCTCCTCTTACTTTTCAGGCT TCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 10.65

#6358
5‘TTTGGTACCATTCTCCTCTTTCTATTCAGGCT TCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 10.00

#6359
5’TTTGGTACCCGGCTAATCTTTCTTTTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 9.19

#6360
5’TTTGGTACCCGGATAATCTTTCTTTTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 8.60

#6361
5’TTTGGTACCATTCTCCTCAAACTTTTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 7.47

#6362
5‘TTTGGTACCCGGAGAATCTTTCTTTTCAGGCT TCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 6.14

#6363
5’TTTGGTACCAAACTCCTCAAACTTTTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 5.49

#6364
5‘TTTGGTACCAAACTCCGCAAACTTTTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 4.27

#6365
5‘TTTGGTACCATTCTCCACAAACAATTCAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 2.99

#6366
5‘TTTGGTACCATTATCCGCAAACAATACAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 2.26

#6367
5‘TTTGGTACCAAACGAAGCTTTCAAAACAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 1.10

#6368
5‘TTTGGTACCAAACTCCGCAAACAAAACAGGCTTCTAGACCTATTGGCCTATTGG‘3
FW DE SA7 SALOW 0.50

References

Abramowicz, A. and M. Gos (2019). "Correction to: Splicing mutations in human genetic disorders: examples, detection, and confirmation." J Appl Genet 60(2): 231.

Akerman, M. and Y. Mandel-Gutfreund (2007). "Does distance matter? Variations in alternative 3' splicing regulation." Nucleic Acids Res 35(16): 5487-5498.

Brillen, A. L., K. Schoneweis, L. Walotka, L. Hartmann, L. Muller, J. Ptok, W. Kaisers, G. Poschmann, K. Stuhler, E. Buratti, S. Theiss and H. Schaal (2017). "Succession of splicing regulatory elements determines cryptic 5ss functionality." Nucleic Acids Res 45(7): 4202-4216.

Busch, A. and K. J. Hertel (2012). "Evolution of SR protein and hnRNP splicing regulatory factors." Wiley Interdiscip Rev RNA 3(1): 1-12.

Cartegni, L., S. L. Chew and A. R. Krainer (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." Nat Rev Genet 3(4): 285-298.

De Conti, L., M. Baralle and E. Buratti (2013). "Exon and intron definition in pre-mRNA splicing." Wiley Interdiscip Rev RNA 4(1): 49-60.

Dominski, Z. and R. Kole (1991). "Selection of splice sites in pre-mRNAs with short internal exons." Mol Cell Biol 11(12): 6075-6083.

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(1): 67-76.

Erkelenz, S., F. Hillebrand, M. Widera, S. Theiss, A. Fayyaz, D. Degrandi, K. Pfeffer and H. Schaal (2015). "Balanced splicing at the Tat-specific HIV-1 3'ss A3 is critical for HIV-1 replication." Retrovirology 12: 29.

Erkelenz, S., W. F. Mueller, M. S. Evans, A. Busch, K. Schoneweis, K. J. Hertel and H. Schaal (2013). "Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms." RNA 19(1): 96-102.

Erkelenz, S., S. Theiss, M. Otte, M. Widera, J. O. Peter and H. Schaal (2014). "Genomic HEXploring allows landscaping of novel potential splicing regulatory elements." Nucleic Acids Res 42(16): 10681-10697.

Fu, X. Y., H. Ge and J. L. Manley (1988). "The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3' splice site in alternative splicing." EMBO J 7(3): 809-817.

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(1): R1.

Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." RNA 6(9): 1197-1211.
Grodecka, L., E. Buratti and T. Freiberger (2017). "Mutations of Pre-mRNA Splicing Regulatory Elements: Are Predictions Moving Forward to Clinical Diagnostics?" Int J Mol Sci 18(8).

Jian, X., E. Boerwinkle and X. Liu (2014). "In silico prediction of splice-altering single nucleotide variants in the human genome." Nucleic Acids Res 42(22): 13534-13544.

Kovacova, T., P. Soucek, P. Hujova, T. Freiberger and L. Grodecka (2020). "Splicing Enhancers at Intron-Exon Borders Participate in Acceptor Splice Sites Recognition." Int J Mol Sci 21(18).

Krummheuer, J., C. Lenz, S. Kammler, A. Scheid and H. Schaal (2001). "Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression." Virology 286(2): 276-289.

Moles-Fernandez, A., L. Duran-Lozano, G. Montalban, S. Bonache, I. Lopez-Perolio, M. Menendez, M.

Santamarina, R. Behar, A. Blanco, E. Carrasco, A. Lopez-Fernandez, N. Stjepanovic, J. Balmana, G. Capella, M. Pineda, A. Vega, C. Lazaro, M. de la Hoya, O. Diez and S. Gutierrez-Enriquez (2018). "Computational Tools for Splicing Defect Prediction in Breast/Ovarian Cancer Genes: How Efficient Are They at Predicting RNA Alterations?" Front Genet 9: 366.

Page-McCaw, P. S., K. Amonlirdviman and P. A. Sharp (1999). "PUF60: a novel U2AF65-related splicing activity." RNA 5(12): 1548-1560.

Ptok, J., L. Müller, S. Theiss and H. Schaal (2019). "Context matters: Regulation of splice donor usage." Biochim Biophys Acta Gene Regul Mech 1862(11-12): 194391.

Rhine, C. L., C. Neil, D. T. Glidden, K. J. Cygan, A. M. Fredericks, J. Wang, N. A. Walton and W. G. Fairbrother (2019). "Future directions for high-throughput splicing assays in precision medicine." Hum Mutat 40(9): 1225-1234.

Rosenberg, A. B., R. P. Patwardhan, J. Shendure and G. Seelig (2015). "Learning the sequence determinants of alternative splicing from millions of random sequences." Cell 163(3): 698-711.
Sakharkar, M. K., V. T. Chow and P. Kangueane (2004). "Distributions of exons and introns in the human genome." In Silico Biol 4(4): 387-393.

Schaal, H., P. Pfeiffer, M. Klein, P. Gehrmann and A. Scheid (1993). "Use of DNA end joining activity of a Xenopus laevis egg extract for construction of deletions and expression vectors for HIV-1 Tat and Rev proteins." Gene 124(2): 275-280.

Selden, R. F., K. B. Howie, M. E. Rowe, H. M. Goodman and D. D. Moore (1986). "Human growth hormone as a reporter gene in regulation studies employing transient gene expression." Mol Cell Biol 6(9): 3173-3179.

Shepard, J., M. Reick, S. Olson and B. R. Graveley (2002). "Characterization of U2AF(6), a splicing factor related to U2AF(35)." Mol Cell Biol 22(1): 221-230.

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(8): 4939-4952.

Soukarieh, O., P. Gaildrat, M. Hamieh, A. Drouet, S. Baert-Desurmont, T. Frebourg, M. Tosi and A. Martins (2016). "Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools." PLoS Genet 12(1): e1005756.

Sterner, D. A., T. Carlo and S. M. Berget (1996). "Architectural limits on split genes." Proc Natl Acad Sci U S A 93(26): 15081-15085.

Tronchere, H., J. Wang and X. D. Fu (1997). "A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA." Nature 388(6640): 397-400.

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(5282): 1706-1709.

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(16): 4630-4641.

Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell 136(4): 701-718.

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA 14(5): 802-813.

Widera, M., S. Erkelenz, F. Hillebrand, A. Krikoni, D. Widera, W. Kaisers, R. Deenen, M. Gombert, R. Dellen, T. Pfeiffer, B. Kaltschmidt, C. Munk, V. Bosch, K. Kohrer and H. Schaal (2013). "An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA." J Virol 87(5): 2707-2720.

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(6763): 832-835.

Wu, T. and X. D. Fu (2015). "Genomic functions of U2AF in constitutive and regulated splicing." RNA Biol 12(5): 479-485.

Yagi, M., Y. Takeshima, H. Wada, H. Nakamura and M. Matsuo (2003). "Two alternative exons can result from activation of the cryptic splice acceptor site deep within intron 2 of the dystrophin gene in a patient with as yet asymptomatic dystrophinopathy." Hum Genet 112(2): 164-170.

Yeo, G. and C. B. Burge (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." J Comput Biol 11(2-3): 377-394.

Zuo, P. and T. Maniatis (1996). "The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing." Genes Dev 10(11): 1356-1368.

3.1.4 Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing (Chapter 4)

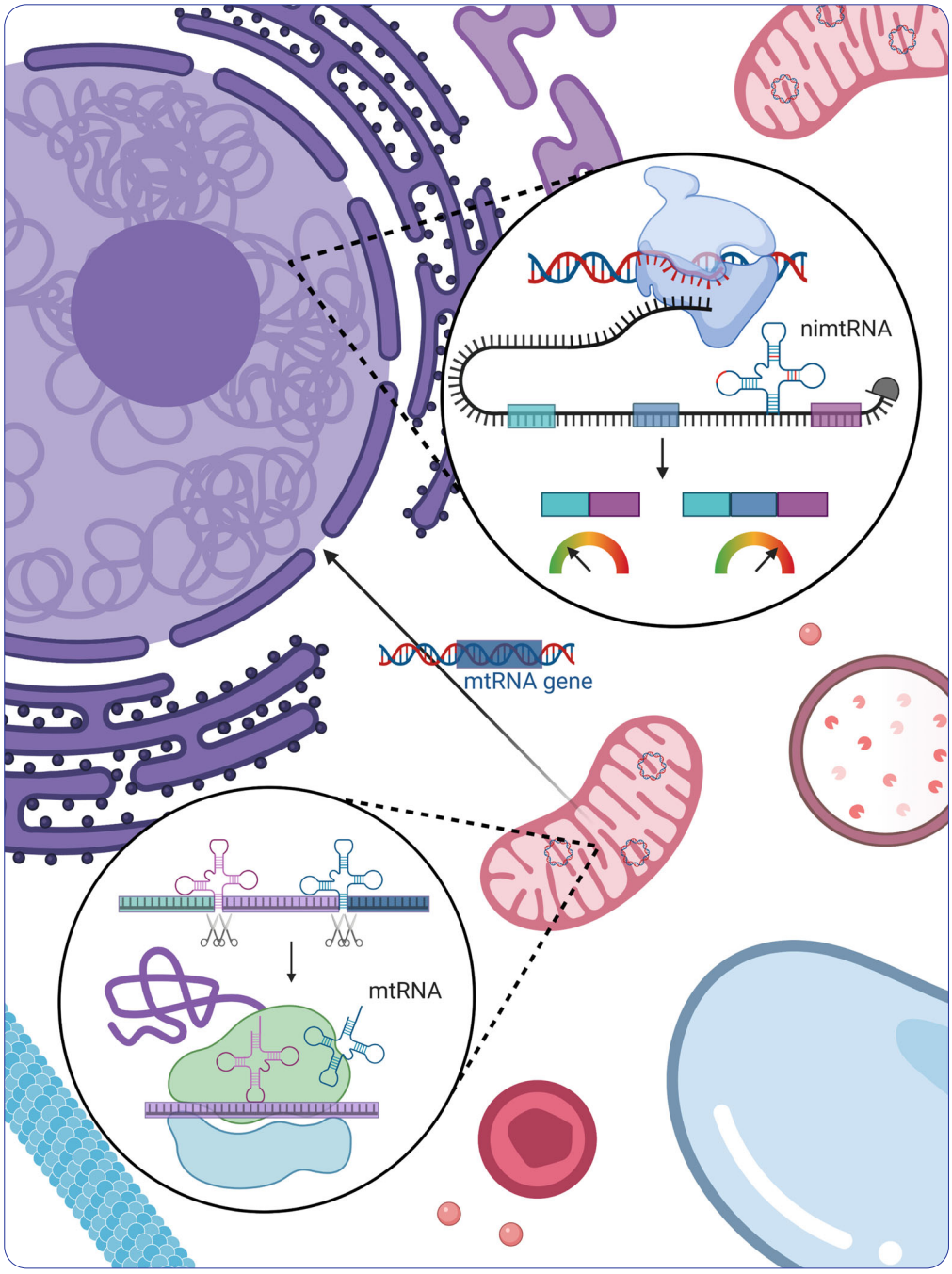
Apart from classic splicing regulatory elements usually located near splice sites of 4 to 6nt length, different classes of intronic non-coding RNAs can also regulate splicing if they possess binding sites for splicing regulatory proteins. By *in silico* and *in vitro* analysis it has been possible to identify nuclear intronic mitochondrial-derived tRNAs homologs (termed nimtRNAs) within introns of the human genome, that are found to promote splice site recognition from an intronic position. This adds another layer of regulation for splice site recognition.

The following article is published Genome Biol. 2020 Dec 8;21(1):299. doi: 10.1186/s13059-020-02199-6. by

Hoser SM, Hoffmann A, Meindl A, Gamper M, Fallmann J, Bernhart SH, **Müller L**, Ploner M, Misslinger M, Kremser L, Lindner H, Geley S, Schaal H, Stadler PF, Huettenhofer A.

Contributions:
A.H. and P.F.S. contributed to the conceptualization; A.H., P.F.S., and A.H. contributed to the methodology; S.M.H., A.M., M.G., L.K., A.H., J.F., S.B., M.P., and L.M. contributed to the investigation; A.H., S.M.H., P.F.S., and A.H. contributed to the writing of the original draft; A.H., S.M.H., L.M., M.M., S.G., and H.S. contributed to the writing and review and editing; A.H. and P.F.S. contributed to the funding acquisition; A.H., P.F.S., H.S., S.G., and H.L. contributed to the resources; A.H. and P.F.S contributed to the supervision. The author(s) read and approved the final manuscript.
Individual contribution: 5% L.M. provided the basis for the constructs that are shown in Figure 3, 4, S2 and S3. L.M. performed bioinformatic predictions using the HEXplorer tool.

Genome Biology



Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing

Hoser *et al.*

BMC

Hoser *et al. Genome Biology* (2020) 21:299
https://doi.org/10.1186/s13059-020-02199-6

Hoser et al. Genome Biology (2020) 21:299
https://doi.org/10.1186/s13059-020-02199-6

Genome Biology

RESEARCH Open Access

Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing

Simon M. Hoser[1*], Anne Hoffmann[2,3], Andreas Meindl[1], Maximilian Gamper[1], Jörg Fallmann[3], Stephan H. Bernhart[3], Lisa Müller[4], Melanie Ploner[1], Matthias Misslinger[5], Leopold Kremser[6], Herbert Lindner[6], Stephan Geley[7], Heiner Schaal[4], Peter F. Stadler[3,8] and Alexander Huettenhofer[1*]

* Correspondence: simon.hoser@i-med.ac.at; alexander.huettenhofer@i-med.ac.at
Alexander Huettenhofer is Lead Contact.
[1]Division of Genomics and RNomics, Biocenter, Medical University of Innsbruck, 6020 Innsbruck, Austria
Full list of author information is available at the end of the article

Abstract

Background: The presence of nuclear mitochondrial DNA (numtDNA) has been reported within several nuclear genomes. Next to mitochondrial protein-coding genes, numtDNA sequences also encode for mitochondrial tRNA genes. However, the biological roles of numtDNA remain elusive.

Results: Employing in silico analysis, we identify 281 mitochondrial tRNA homologs in the human genome, which we term nimtRNAs (nuclear intronic mitochondrial-derived tRNAs), being contained within introns of 76 nuclear host genes. Despite base changes in nimtRNAs when compared to their mtRNA homologs, a canonical tRNA cloverleaf structure is maintained. To address potential functions of intronic nimtRNAs, we insert them into introns of constitutive and alternative splicing reporters and demonstrate that nimtRNAs promote pre-mRNA splicing, dependent on the number and positioning of nimtRNA genes and splice site recognition efficiency. A mutational analysis reveals that the nimtRNA cloverleaf structure is required for the observed splicing increase. Utilizing a CRISPR/Cas9 approach, we show that a partial deletion of a single endogenous nimtRNA^{Lys} within intron 28 of the *PPFIBP1* gene decreases inclusion of the downstream-located exon 29 of the *PPFIBP1* mRNA. By employing a pull-down approach followed by mass spectrometry, a 3′-splice site-associated protein network is identified, including KHDRBS1, which we show directly interacts with nimtRNA^{Tyr} by an electrophoretic mobility shift assay.

Conclusions: We propose that nimtRNAs, along with associated protein factors, can act as a novel class of intronic splicing regulatory elements in the human genome by participating in the regulation of splicing.

Keywords: nimtRNA, tRNA-lookalikes, Intronic splicing enhancer (ISE), Splicing regulatory element, numtDNA, tRNA, Mitochondrial tRNA, Splicing, Alternative splicing, Constitutive splicing

Background

In the course of evolution, splicing has been demonstrated to be an increasingly important step in eukaryotic gene expression [1]. Orchestrated by spliceosomal complexes, intronic regions are excised from a pre-mRNA transcript, while exonic regions are joined [2]. Splice acceptor and donor sites define exon/intron borders, which is achieved by base complementarity to small nuclear RNAs (snRNAs), which are part of spliceosomal small nuclear ribonucleo-proteins (snRNPs) [2]. By means of splice site selection, a single pre-mRNA transcript can be employed to generate several distinct splice products, a process designated as alternative splicing. This is achieved by modulating splice site strength through *cis*-regulatory elements within the pre-mRNA transcript. These splicing regulatory elements (designated as SREs) are recognized by *trans*-acting proteins in a sequence- and structure-dependent manner which directly or indirectly interact with the spliceosome in a position-dependent manner [3, 4].

In addition to SREs, intronic ncRNAs such as microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs) have been demonstrated to affect host gene splicing. Due to an interplay between spliceosomal components and miRNA processing enzymes, intronic miRNA processing has been shown to be able to counteract host pre-mRNA splicing *in cis* [5]. In contrast, processing of miRNA-211 was demonstrated to promote splicing of its hosting intron [6]. Additionally, intronic snoRNAs have been shown to be co-transcriptionally/pre-splicing bound by snoRNA processing enzymes, thus indicating a potential mechanism of interaction between intronic *cis*-acting snoRNAs and the spliceosome [7–10]. Accordingly, pre-mRNA splicing of NOP56, a component of canonical snoRNP complexes, is autoregulated *in cis* by the intron-hosted snoRNA SNORD86 [11].

While the majority of snoRNA genes and a large number of miRNAs are located within introns of nuclear protein-coding genes [12], nuclear tRNAs are generally transcribed by RNA polymerase III as independent transcription units, employing internal promoter sequences, i.e., boxes A and B, respectively. Nuclear-encoded tRNAs are transcribed as precursor sequences and are subsequently processed by two endonucleases, i.e., RNase P and RNase Z, at their 5′- and 3′-terminus [13], respectively, resulting in mature RNA species of approximately 70–90 nt in length [14].

In contrast to the nuclear genome, the human mitochondrial genome contains 22 mitochondrial tRNA genes (mtRNAs), interspersed between 13 protein-coding genes which predominantly encode for proteins of the respiratory chain. Three different polycistronic transcripts are generated by a single mitochondrial RNA polymerase [15]. Subsequent cleavage of these polycistronic transcripts mediated by the two tRNA processing enzymes (i.e., mitochondrial RNase P and RNase Z) generates mature mtRNAs in a process which concomitantly releases intervening mitochondrial mRNAs [13]. Both nuclear and mitochondrial tRNAs exhibit a characteristic, cloverleaf-shaped secondary structure, which among other functions is also important for their processing. However, while most mtRNAs still show a canonical cloverleaf structure, they lack some of the features that are highly conserved in nuclear tRNAs, in particular sequences characteristic of highly conserved D-loops and/or T-loops [16]. In some cases, they may also lack entire tRNA structural domains [17, 18]. Compared to nuclear tRNAs, mtRNAs show a different sequence bias and exhibit tertiary interactions distinct from nuclear-encoded tRNAs [19].

According to the endosymbiotic theory, eukaryotic mitochondria originated from the progressive transfer of ancient α-proteobacteria DNA into the eukaryotic

genome [20]. Thus, the mitochondrial genomes of higher organisms are 100- to 300-fold smaller than bacterial genomes but still carry the hallmarks of a bacterial ancestor [21]. Interestingly, mammalian genomes harbor a large number of genomic regions designated as "nuclear mitochondrial DNA" (numtDNA) [22]. It can therefore be seen that the integration of numtDNA into the nuclear genome is a rapid and ongoing process [23] that is fast enough to render human haplotypes polymorphic for numtDNA. Insertions appear approximately uniformly across the genome [24] and are favored in locations exhibiting DNA curvature and adjacent to A/T oligomers [25]. They are enriched near retrotransposable elements [25], whose genomic distribution can be explained by random insertion and duplications [26]. In particular, numtDNAs do not appear in clusters and are not enriched on particular chromosomes [25]. Thus, insertions of numtDNA are independent, random events that serve no known purpose [27–29]. Nevertheless, a small number of numtDNA sequences have been implicated in human genetic diseases [30].

NumtDNAs display variations in size, the position of the fragment from which the numtDNA is derived in the mitogenome and in evolutionary age. At the time of insertion, the numtDNA sequence is identical to its counterpart in the mitogenome. Subsequent to its insertion, numtDNA and mitogenomic sequences evolve independently. The mitogenomic sequence (shown in red in Fig. 1) remains subject to the selection pressures in the mitochondrion. On the other hand, different fates are possible for numtDNA sequences: (i) The insertion disrupts cellular functions, the genome variant carrying the numtDNA is quickly removed by selection and no genomic record of the insertion event survives. (ii) In the most likely scenario, the newly inserted numtDNA does not affect the cell's functions and is hence, from an evolutionary standpoint, neutral. In this case, the numtDNA accumulates substitutions at the same rate as other neutrally evolving DNA sequences. This process is slow enough for numtDNA sequences to remain recognizable by sequence similarity on timescales comparable to the radiation of the placental mammals. Eventually, however, all traces of an ancient numtDNA insertion are eradicated by the accumulation of random mutations. (iii) In some cases, numtDNA sequences and in particular the mtRNAs contained within them may acquire novel functions in the nuclear genome. In this case, the functional sequence is subject to the influence of stabilizing selection for its new function and persists in the nuclear genome. Occasional duplications of numtDNA in the nuclear genome can further complicate the picture [31].

MtRNA genes are inserted into the nuclear genome as components of numtDNAs. The nuclear genome had previously been scanned for full-length mitochondrial tRNAs, which were named mitochondrial tRNA-lookalikes (MTLs) [32, 33]. The analysis of selection pressures acting on MTLs must therefore take their peculiar evolutionary history into account. In this context, we face two important issues: (i) A comparison of different MTLs with the corresponding mtRNAs must take into account that the insertion events potentially occurred at different points in time (consider, e.g., the three numtDNAs 8, 9, and b versus the mtRNA of species A in Fig. 1). The sequence divergence of MTL and mtRNA confounds the selection pressure on the mtRNA within the mitochondrion since the insertion event with the selective effects the MTL may have experienced in the nuclear genome. This can be accounted for by using the surrounding sequence of the numtDNA as a "molecular clock" that implicitly has recorded the insertion time. (ii) When comparing two MTLs that derive from different insertion events of the same mtRNA, the degree of

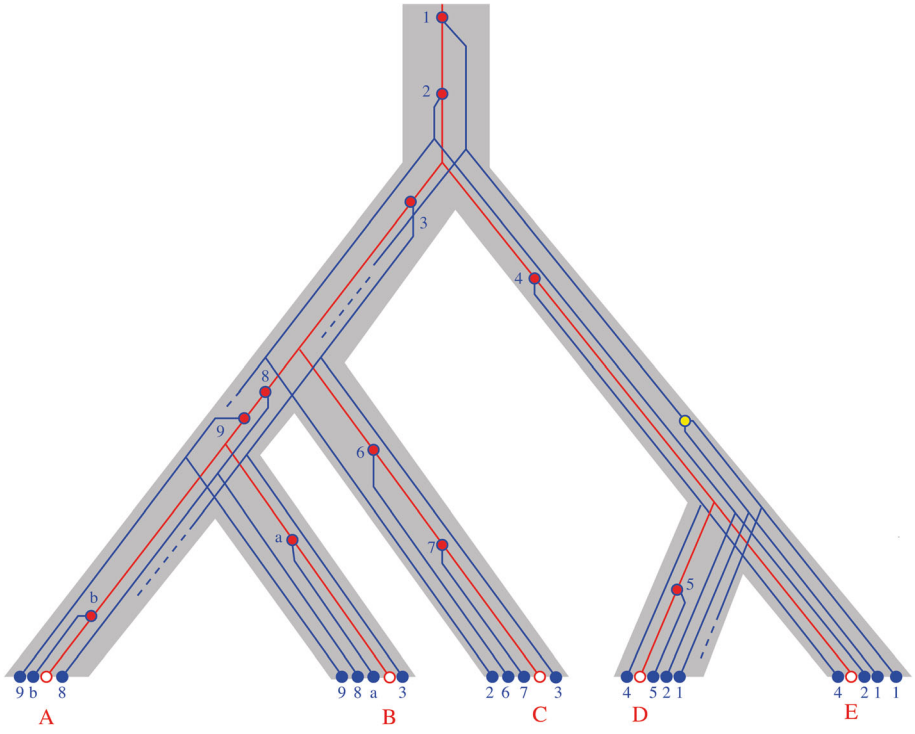


Fig. 1 Evolution of mtRNAs and mtRNA insertions in the nuclear genome. The gray outline shows the phylogenetic relationship of five species A, B, C, D, and E as a dated tree, i.e., the "*y*-axis" corresponds to time before the present. Mitochondrial DNA, usually transmitted through the maternal lineages, faithfully follows the species tree (red tree); its leaves are the extant mtRNAs. Insertions of mitochondrial material occur at random time points independently in the different lineages (red circles). Following an insertion, the inserted material (numtDNA) evolves independently, as shown by blue trees within the species trees. Blue bullets denote the MTLs, and dashed lines indicate MTLs that have been removed by genomic events or that have mutated beyond the detection threshold. Each species contains MTLs that have been inserted at different times. Orthologous MTLs derive from the same insertion event (here denoted by the same blue numbers). Sequence comparison between MTLs or MTLs and mtRNAs shows the substitutions accumulated along the path between them. Orthologous MTLs are connected by blue path only, along which only selective pressures in the nuclear genome have left their traces. Comparisons between MTLs and mtRNAs in the same species almost always trace back to the insertion event. Their sequence differences thus record both the selective pressures acting on mtRNAs (red part of the path) and those acting on the inserted MTL (blue part of the path). Rare exceptions are duplications of MTLs after their insertion (yellow ball on the right). A comparison of arbitrary pairs of MTLs from different species in general corresponds to MTLs deriving from distinct insertion events. Their common history is a mixture of red paths (between the two insertion events) and blue paths (following the insertion events) and thus a mixture of selective pressures on mtRNA and MTL. These cannot be disentangled without exact dating of the insertion events

sequence divergence is a composite function of the selection pressures faced by the two numtDNAs after insertion and selection pressures operating in the mitochondrion on the mtRNA between the first insertion event and the second (consider, e.g., the path from numtDNA 6 to numtDNA 7 in species B, which involves the red (mitochondrial) segment between insertion events 6 and 7 as well as the blue (numtDNA) segments connecting the numtDNAs to the insertion events). As a consequence, MTL/MTL comparisons within the same genome cannot separate selection pressures acting in the nuclear genomes from selection pressures in the mitochondrion. To overcome this problem, one has to identify orthologous MTLs in different species, i.e., MTLs that derive from the same numtDNA insertion event. These can be identified reliably by considering homology of DNA outside of the inserted numtDNA to determine syntenic MTLs. It should be noted, however, that

they cannot be identified by simple sequence comparison, since all MTLs with the same codon are homologs and sequences that were inserted more recently by distinct events will be more similar to present-day mtRNAs than old orthologous MTLs. This is because mtRNAs (red paths) evolve more slowly than inserted MTLs (blue paths), as long as an inserted MTL has not acquired a new function in the nuclear genome, which would subject it to strong negative selection and thus results in its conservation. Only sufficiently old numtDNA insertions, namely those that pre-date speciation events that separate species with sequenced genomes, therefore are informative of selective pressures on MTLs as revealed by direct comparison of MTLs.

As a consequence, we therefore have to focus on MTLs that are embedded in recognizable larger numtDNA sequences. This also allows us to distinguish bona fide MTLs from degraded copies of nuclear tRNAs or tRNA-associated short interspersed nuclear elements (SINEs) [34], which, at least for old insertions, cannot be separated cleanly on sequence similarity scores alone.

Notably, MTL sequences often differ substantially from their mitochondrial counterparts. Also, MTLs of the same tRNA isotype can vary extensively in their sequence. In the human genome, there are only eight MTLs that are still identical in sequence to their primordial mtRNA counterparts, while the remaining 489 MTLs show up to 25 mismatches [32]. At present, the biological function and relevance of MTLs is still unknown. About 20% of known human MTLs have been reported by the group of Telonis and coworkers to be located in introns of protein-coding or noncoding RNA transcripts [32].

Currently, a single MTL annotation strategy was published [32, 33] based on a BLAST search of the known nuclear and mitochondrial tRNA sequences against the nuclear genome with the intention of identifying full-length tRNA-like sequences in the nuclear genome. Since structural conservation is not included in this previous approach, MTLs that have diverged at their sequence level but may have retained tRNA-like structures are not annotated. Applying the computational annotation workflow presented here, we were able to identify numerous novel MTLs and nuclear-encoded intronic mitochondrial-derived tRNA genes (designated as nimtRNAs) in humans and mice. Notably, nimtRNAs were always flanked by sequences of mitochondrial origin.

Strikingly, the canonical tRNA secondary structure was conserved as observed mutations relative to their mitochondrial counterparts were found either in loop regions or as compensatory base changes in stem domains. In this study, we thus aimed to investigate the potential function(s) of nimtRNAs located within the introns of nuclear-encoded pre-mRNAs. We demonstrate that nimtRNAs interact with specific RNA-binding proteins (RBPs) and participate in regulation of splice site usage by a mechanism comparable to that of bona fide SREs.

Results

Numerous, so far unidentified nimtRNAs are present in nuclear genomes

To scan, in particular, the human and mouse genomes for MTL sequences, we applied different combinations of annotation tools (tRNAscan-SE and Infernal) and strategies (*NUMT*-based and *genome*-based), see Additional File 1: Fig. S1A. Within the *NUMT*-based approach (using published numtDNA sequences as reference only) for the human genome, we obtained 775 hits from Infernal [35] and 726 hits from tRNAscan-SE. In

contrast, the *genome*-based approach (using the whole genome as reference), we received 367 hits, only, from Infernal, whereas tRNAscan-SE [36] scored about 2.65 times more hits (977 hits). The analysis of the mouse genome yielded very similar results. We got 105 hits from Infernal and 79 hits from tRNAscan-SE within the *NUMT*-based approach. The hits from the *genome*-based approach vary from 75 (Infernal) to 246 (tRNAscan-SE).

Since for each numtDNA the original mtDNA sequence is known, we used this synteny information to validate our results. For each method, we classified the detected hits as true positives (TPs) if they were found in the corresponding numtDNA as described by their synteny of the originating mitochondrial DNA. The remaining hits were designated as false positives (FPs). As shown in Additional file 1: Fig. S1A, Infernal found 2% more TPs in human than tRNAscan-SE (true positive rate (TPR) of 0.91) within the *NUMT*-based approach. Despite the lower sensitivity of tRNAscan-SE, the tool counts only 29 false positives (FPs) compared to the 68 FP hits of Infernal. The difference is even more pronounced in the *NUMT*-based approach for mouse, where Infernal identified 13% more TPs, but also 11% more FPs compared to tRNAscan-SE. tRNAscan-SE shows the highest sensitivity in the *genome*-based approach with a TPR of 0.88 and 0.72 in human and mouse, respectively. Infernal delivers much less TP in both species for the *genome*-based method. In both the *NUMT*- and the *genome*-based approach, tRNAscan-SE shows the best balance between TPs and FPs. For downstream analysis, the final MTL set is composed of all detected TPs (MTLs within recognizable numtDNA) regardless of the method and tool used.

Finally, we identified 731 MTLs within recognizable numtDNA (42 MTLs (*NUMT*-based method) + 684 (*NUMT*- and *genome*-based method) + 5 (*genome*-based method)) and 92 MTLs within recognizable numtDNA (16 MTLs (*NUMT*-based method) + 73 (*NUMT*- and *genome*-based method) + 3 (*genome*-based method)) in human and mouse genomes, respectively (Fig. 2a). Thereof are 355 MTLs in human and 44 MTLs in mouse novel discoveries. Our MTL annotation strategy is more sensitive (TPR of 0.93 in human and 0.85 in mouse) compared to previous MTL annotations (TPR of 0.48 in human and 0.47 in mouse) [32, 33] (Additional file 1: Fig. S1A). Previous computational studies have demonstrated that about 20% of the MTLs are located within introns which we designate as nimtRNAs of nuclear protein-coding genes in humans [33, 37]. We observed comparable results with our analysis. In humans, we identified a total of 281 nimtRNAs of all types in the introns of 76 different host genes, of which 30 were protein-coding, 28 were specifying long intergenic noncoding RNAs (lincRNAs), 13 were coding for short ncRNAs, and 5 were pseudogenes. In total, 121 of the identified nimtRNAs in human are novel. Compared to previous surveys [33, 37], we identified 12 novel nimtRNAs (of total 34) in 11 different host genes (9 different protein-coding genes and 2 different lncRNAs) in the mouse. A complete list of all annotated MTLs and nimtRNAs found in mice and humans can be obtained from Additional file 2: Table S1 and Additional file 3: Table S2, respectively.

Conservation of MTLs and nimtRNAs

In several cases, large clusters of nimtRNA genes with extensive sequence similarities to the mitochondrial genome were present within introns of nuclear genes but were completely absent from exonic regions. Four examples of nimtRNA host gene introns in humans and one in mouse are shown (Fig. 2b). The observed mitochondrial clusters are

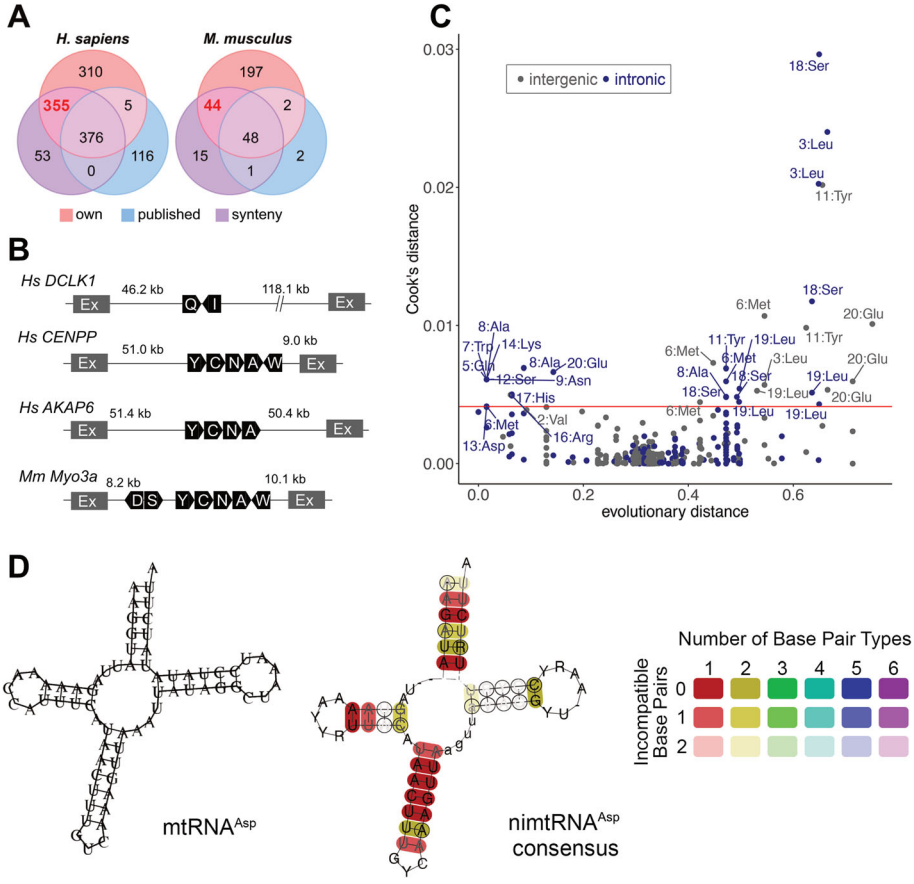


Fig. 2 Computational analyses of MTLs within the human and mouse genome. **a** Overview of the annotated MTLs in humans and mice. Venn diagram illustrating the overlap between our annotated (own) MTLs and those that are already known (published), as well as which of them are located within previously annotated numtDNAs (synteny). Although the synteny information indicates which mtRNAs have been integrated into the nuclear genome, not all of them can be annotated due to their strong degradation (53 in human and 15 in mouse). In total, we identified 731 and 92 MTLs within recognizable numtDNA in humans and mice, respectively. Of these, 355 are newly annotated in humans and 44 in the mouse. **b** The genes *DCLK1*, *CENPP*, and *AKAP6* in *H. sapiens* and the gene *Myo3a* in *M. musculus* were analyzed by tRNAscan. Additionally, the 5′- and 3′-distances from nimtRNA clusters to the intron termini were determined. White letters in black boxes represent the single letter amino acid code of the respective nimtRNA (Q = nimtRNA^{Gln}, I = nimtRNA^{Ile}, Y = nimtRNA^{Tyr}, C = nimtRNA^{Cys}, N = nimtRNA^{Asn}, A = nimtRN^{Ala}, W = nimtRNA^{Trp}, D = nimtRNA^{Asp}, S = nimtRNA^{Ser}). **c** Evolutionary conserved MTLs. Outliers of MTLs are subject to a stronger stabilizing selection after their insertion into the nuclear genome relative to numtDNAs and are shown above the red line. Outliers were measured by Cook's distance. The majority (25 of 36) of the more extreme outliers are nimtRNAs. **d** Preservation of the secondary structure of nimtRNAs. As an example, multiple sequence alignments along with the consensus sequence-structure RNA motif are shown for all nimtRNAs of type nimtRNA^{Asp}. The sequences of nimtRNA^{Asp} exhibit base changes compared to their primordial mtRNA^{Asp}, but the secondary structure is maintained. The different colors provide information concerning the number of distinct base pairs occurring whereas the shading indicates how many sequences or structures in the alignment do not form a particular base pair

located in different genes in the mouse when compared to the human genomes. We found a similar degree of evolutionary conservation between nimtRNAs and the corresponding mitochondrial sequences among different mammals. Since their PhyloP scores are very low, the majority of MTLs within recognizable numtDNA show no evidence of negative selection in the host genomes. While we found that PhyloP scores are slightly enhanced in MTLs within recognizable numtDNA and nimtRNAs compared to the

surrounding numtDNA sequences (Additional file 1: Fig. S1B and C), the selection pressures are insufficient to identify individual MTLs within recognizable numtDNA or nimtRNAs that are under strong negative selection. Instead, only a few shorter elements are conserved. We interpret these as possible binding sites that have emerged from the inserted mtRNA sequence. Using a different method, we identified about a dozen MTLs within recognizable numtDNA and nimtRNAs that appear to have evolved significantly more slowly than the adjacent numtDNA sequences. It is also interesting to note nimtRNAs represent the majority of the more extreme outliers (Fig. 2c) as measured by Cook's distance. All these outliers are listed in Additional file 4: Table S3.

Based on the consensus structure of each type of nimtRNA, it is apparent that base changes in most nimtRNA types were located either in loop regions of tRNAs or, in several cases, were present in the form of compensatory base changes in stem structures (Fig. 2d, Additional file 5: Table S4). Accordingly, in most consensus structures, the mitochondrial secondary structure is largely retained and thus probably also their function (see below). Evolutionary conserved compensatory base changes are consistent with a functional role of nuclear-encoded nimtRNA genes. In a few cases, the consensus structures deviate strongly from their primordial mtRNAs. This is probably one reason why we cannot find all expected MTLs within a numtDNA, as is the case for mtRNA^{Pro} in particular.

Taken together, the insertion of nimtRNA genes in the respective introns of nuclear genes might be a very recent evolutionary event, which might have occurred independently in different species in addition to potential retainment of pre-existing nimtRNAs. Furthermore, this computational analysis points to MTLs within recognizable numtDNA and nimtRNAs as a source of functional binding sites. As expected in such a scenario, most MTLs within recognizable numtDNA and nimtRNAs have not attained functional significance because they are simply not present in a useful genomic context or there is no selective advantage to be gained from an MTL within recognizable numtDNA- or nimtRNA-derived binding site at the position of the insertion.

NimtRNAs located in introns of nuclear-encoded pre-mRNAs are not processed as bona fide tRNAs in 293T cells

Mitochondrial- as well as nuclear-encoded tRNAs are post-transcriptionally processed by RNase P and RNase Z at their 5′- and 3′-terminus, respectively (see above). In order to more closely investigate a potential cleavage, processing and function of nimtRNAs, we employed an eGFP splicing reporter, designated as Low0-eGFP, consisting of a non-coding exon, a 2.2-kb-long intron and a second exon, containing the coding sequence for the enhanced green fluorescent protein (eGFP; Fig. 3a) [38].

A cluster of five nimtRNA genes (out of the seven nimtRNAs of the *Myo3a* gene, Fig. 3a), previously reported to be present within intron 30 of the mouse *Myo3a* gene [33], was inserted into the intronic region of this splicing reporter. The rationale for not including all seven nimtRNAs was based on the fact that two nimtRNAs, i.e., nimtRNA^{Asp} and nimtRNA^{Ser}, respectively, are located 1.5 kb upstream from the cluster of five nimtRNAs, and the insertion of a region spanning additional 1.5 kb might have impaired canonical reporter splicing. Thereby, the mouse nimtRNA cluster exhibited distinct sequence differences compared to bona fide human mtRNAs or human nimtRNAs thus permitting their specific detection by northern blot analysis. HEK 293T

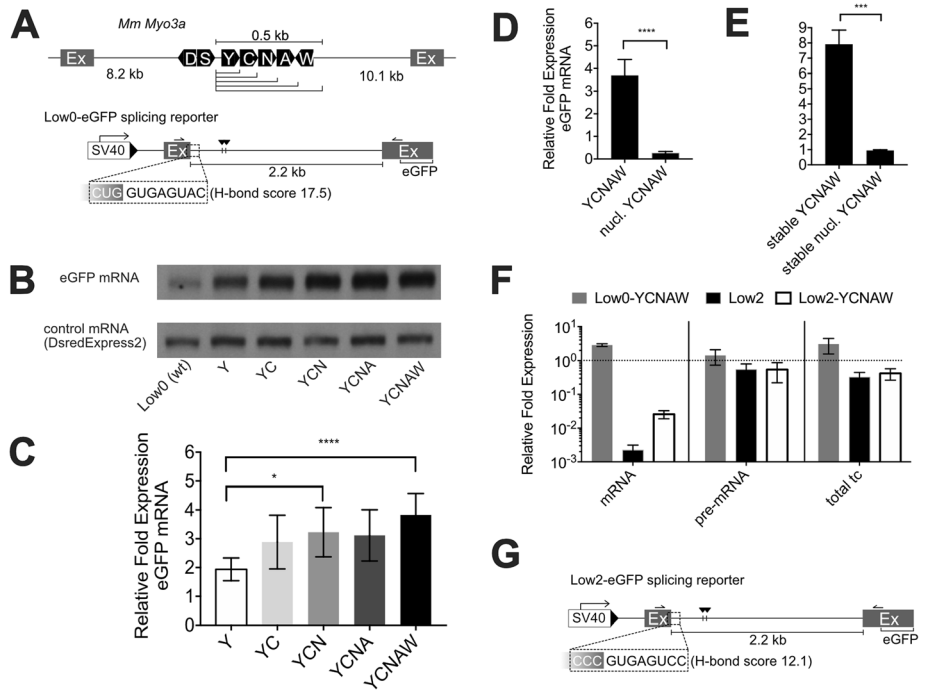


Fig. 3 NimtRNAs increase host gene mRNA abundance by increasing splicing efficiency. **a** Black boxes represent the single letter amino acid code of the respective nimtRNA: one nimtRNA (Y), two (YC), three (YCN), four (YCNA), or five (YCNAW) nimtRNAs from the mouse *Myo3a* intron 30 were cloned into the intronic region of the Low0-eGFP splicing reporter, exhibiting an efficient 5′-splice site. NimtRNA genes were inserted as indicated by triangles. **b** Northern blot analysis of eGFP mRNA constructs as depicted in **a** performed with digoxigenin (DIG)-labeled probes. Normalization of RT-qPCR and northern blot analyses was performed to Dsred2Express transfection control mRNA. Experiments were performed in triplicates. **c** RT-qPCR analysis of constructs as depicted in **a** was performed with the forward primer binding to exon 1 and the reverse primer binding to exon 2, depicted by black arrows shown in **a**. **d** Five nuclear tRNA genes (designated as nucl. YCNAW) were inserted into the Low0-eGFP splicing reporter and their influence on splicing was compared to the reporter containing five nimtRNAs (YCNAW). The Low0-eGFP splicing reporter was taken as a reference. **e** Stable cell lines were generated containing a single copy of the Low0-eGFP splicing reporter and either five nimtRNAs or five nuclear tRNAs. Normalization was performed to β-Actin. The stably integrated Low0-eGFP splicing reporter was taken as reference. **f** mRNA, pre-mRNA, and total transcript (mRNA+pre-mRNA) levels from cells transfected with the respective constructs were assessed by RT-qPCR. Error bars represent the standard deviation from the mean of three independent experiments. **g** The Low2-eGFP splicing reporter differs from the Low0-eGFP splicing reporter by possessing a less efficient 5′-splice site as indicated. Averages and standard deviations were determined from three independent sets of experiments. Error bars represent the SD and $*P < 0.05$; $***P < 0.001$; $****P < 0.0001$ (ANOVA)

cells were transiently transfected with the splicing reporters either lacking or containing nimtRNAs. The abundance and/or processing of nimtRNAs, i.e., nimtRNA^{Tyr}, nimtRNA^{Cys}, and nimtRNA^{Asn}, from the *Myo3a* nimtRNA cluster was investigated by northern blotting. However, no hybridization signal was detected corresponding to fully processed nimtRNAs of about 70 nt in length (Additional file 1: Fig. S2A).

We thus next investigated whether processing of another reported intron-encoded ncRNA, i.e., a snoRNA within the eGFP splicing reporter construct was also hampered. To that end, we cloned the gene of a brain-specific ncRNA, i.e., the C/D box snoRNA SNORD115 including flanking regions into the identical intronic location. In contrast to nimtRNAs, a hybridization signal of the expected size for the processed SNORD115 RNA species could be readily observed (Additional file 1: Fig. S2B).

In order to address the discrepancy in canonical processing of a snoRNA, compared to nimtRNAs, we employed an intron-less Pol III reporter containing a single copy of nimtRNA^{Asn} from the *Myo3a* gene. Consistent with our expectations, we observed a fully processed and stable nimtRNA^{Asn} by northern blot analysis (Additional file 1: Fig. S2B), excluding the possibility that nimtRNAs are degraded within the nucleus. In addition to nimtRNA sequences, we also investigated processing of nimtRNAs from their endogenous host gene transcripts. As observed for the eGFP splicing reporter, however, we also failed to detect processed nimtRNAs from the endogenous *DYNC2H1* host gene.

The role of nimtRNAs in pre-mRNA metabolism

To thus investigate alternative functions of nimtRNAs, we inserted either one, two, three, four, or five nimtRNA genes from the *M. musculus Myo3a* intron, i.e., nimtRNA-^{Tyr}, nimtRNA^{Cys}, nimtRNA^{Asn}, nimtRNA^{Ala}, and nimtRNA^{Trp}, into the intron of the Low0-eGFP splicing reporter construct employed above (Fig. 3a). Subsequently, by northern blot and RT-qPCR analysis, we investigated their influence on pre-mRNA splicing of the eGFP host gene. Interestingly, insertion of nimtRNAs into the eGFP reporter gene resulted in a significant increase in the abundance of spliced eGFP mRNA levels, compared to a control lacking the nimtRNA genes (wt; Fig. 3b, c).

Importantly, cells transfected with eGFP constructs containing one (Y), two (YC), three (YCN), four (YCNA), or five nimtRNAs (YCNAW) exhibited a copy number-dependent increase of spliced eGFP mRNA levels of 1.9-fold, 2.9-fold, 3.2-fold, 3.1-fold, and 3.9-fold respectively, as assessed by northern blot and RT-qPCR analysis (Fig. 3b, c). Surprisingly, nimtRNA^{Trp}, which is present in reverse-complementary orientation in the Low0-eGFP splicing reporter (as also observed within the mt genome), also increased reporter mRNA abundance. Normalization of eGFP mRNA levels was performed by employing a co-transfected plasmid (control mRNA), coding for a red fluorescent protein (DsredExpress2). The increase in mRNA levels was accompanied by an increase in eGFP protein level, as assessed by measuring eGFP fluorescence levels, normalized to DsredExpress2 (Additional file 1: Fig. S3A).

In the above experiments, intronic sequences containing single or multiple nimtRNAs were introduced into the intron in addition to the wildtype sequence rather than by substitution. To exclude a potential influence of intron size and/or intron structure on mRNA abundance, an artificial insert (of the same length as YCNAW), containing five nuclear tRNAs, was cloned into the same splicing reporter. Thereby, five nuclear nimtRNA counterparts, i.e., tRNA^{Tyr}, tRNA^{Cys}, tRNA^{Asn}, tRNA^{Ala}, and tRNA^{Trp}, were inserted into the Low0 splicing reporter and analyzed for their effect on host gene splicing. Of note, nuclear tRNAs, although encoding for the same amino acids as nimtRNAs, differ extensively in their sequences from their nimtRNA homologs. In contrast to nimtRNAs, however, the nuclear YCNAW cluster construct resulted in a significant decrease, rather than an increase in mRNA abundance, pointing towards specific sequence or structural features of nimtRNAs (see below) that govern the observed increase in mRNA levels (Fig. 3d).

The above results were also corroborated by introducing the Low0-eGFP splicing reporter into stable cell lines. To that end, the Low0-eGFP splicing reporter was cloned downstream from an EF1α promoter and inserted as a single copy by Flippase

recombination into HEK 293 Flip-In cells [39]. The reporter intron thereby contained either no tRNAs, five nimtRNAs from the *Myo3a* gene, or their nuclear tRNA homologs (see above). By RT-qPCR analysis, we observed an even higher abundance (i.e., 7.9-fold, compared to 3.9-fold in transiently transfected cells) of spliced eGFP mRNA levels in the cell line containing intronic nimtRNAs compared to the cell lines containing either the original intronic sequence or the nuclear tRNAs (Fig. 3e).

NimtRNAs increase mRNA abundance by enhancing splicing efficiency

Next, we wanted to determine whether increased transcription or pre-mRNA processing was responsible for the nimtRNA-mediated increase in spliced eGFP mRNA levels. Thus, we investigated whether unspliced and total transcript levels (i.e., spliced and unspliced levels combined) of the Low0-eGFP splicing reporter were also affected by nimtRNAs. Upon intronic insertion of nimtRNAs into the Low0-eGFP splicing reporter (designated as Low0-YCNAW), by RT-qPCR analysis we observed an approximately 3-fold increase in spliced mRNA abundance as well as in total transcript levels, while pre-mRNA levels remained unchanged (Fig. 3f), consistent with a nimtRNA-mediated increase in splicing.

The first step of spliceosome assembly comprises the recognition of the 5′-splice site by the U1 snRNA [2]. The recognition of weak splice sites, i.e., those displaying low U1 snRNA complementarity, is known to be more dependent on SREs. Thus, we reduced the strength of the 5′-splice site (i.e., U1 snRNA complementarity: H-bond score (HBS) 17.5 > 12.1; designated as Low2-eGFP splicing reporter) and compared unspliced and total transcript levels. Notably, the HBS of this splice donor is still within the range of 12.0 to 20.0, which is observed in 86% of human constitutively spliced exons [40] (Fig. 3g).

As expected, the Low2 wt reporter resulted in an extensive, i.e., 450-fold, decrease in reporter mRNA abundance, compared to the more efficient Low0 wt reporter, while only an about 2-fold reduction in pre-mRNA and total transcript levels was observed. Upon nimtRNA insertion into the Low2 reporter (designated as Low2-YCNAW), spliced mRNA abundance was about 38-fold lower compared to the Low0 wt construct. Thus, insertion of the YCNAW nimtRNA cluster resulted in an about 13-fold increase in spliced mRNA abundance in the inefficient Low2 splicing reporter, hence exhibiting a more pronounced effect on splicing than the efficient Low0 splicing reporter (showing an about 3-fold increase in mRNA abundance). Notably, pre-mRNA and total reporter transcript levels remained unchanged (Fig. 3f). This can be explained by reporter pre-mRNA being significantly more abundant than spliced mRNA.

Single nimtRNAs differently increase host mRNA levels

Analysis of pre-mRNA splicing demonstrated a nimtRNA copy number-dependent increase in mRNA abundance. Hence, to investigate the effect of single nimtRNAs on mRNA abundance, nimtRNAs from the *Myo3a* gene were individually inserted into the Low0-eGFP splicing reporter and assessed for their influence on splicing. In this context, it was observed that nimtRNA^{Tyr} (Y), nimtRNA^{Cys} (C), nimtRNA^{Ala} (A), and nimtRNA^{Trp} (W) significantly increased eGFP mRNA levels by 1.9-fold, 2.8-fold, 2.6-fold, and 2.5-fold, respectively, compared to a scrambled control (Fig. 4a). Interestingly,

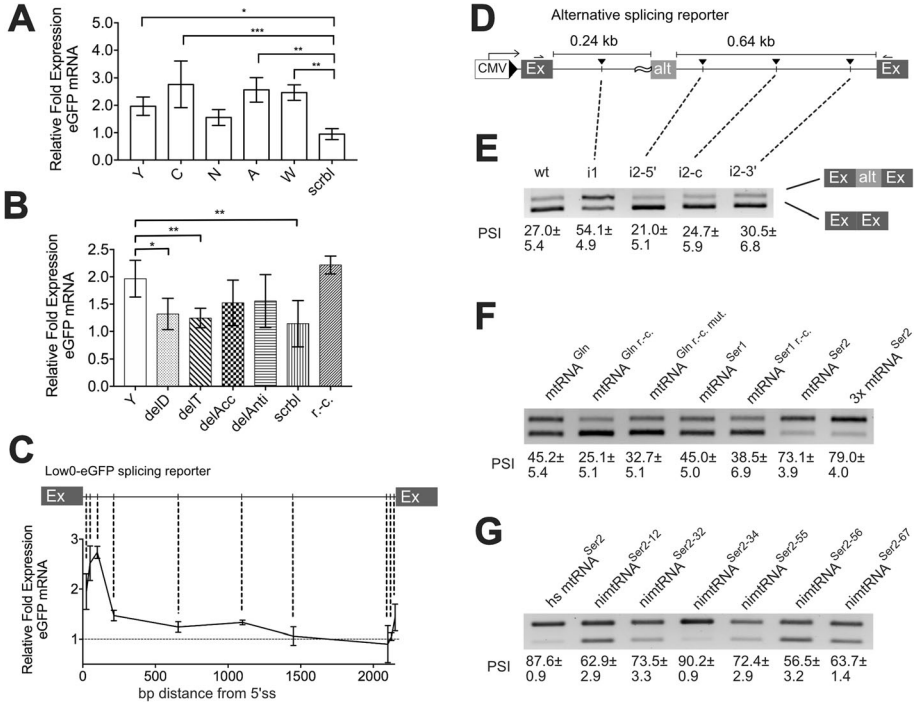


Fig. 4 NimtRNAs increase downstream exon inclusion in a structure- and position-dependent manner. **a** Single nimtRNAs were analyzed for their effects on reporter host gene mRNA abundance compared to a scrambled control (scrbl). **b** Different domains of nimtRNA^{Tyr} were deleted, i.e., the D-arm (delD), the T-arm (delT), the acceptor stem (delAcc), or the anticodon arm (delAnti), respectively. Alternatively, nimtRNA^{Tyr} was exchanged for a scrambled sequence (scrbl), or inserted into the splicing reporter in reverse-complementary orientation (r.-c.). **c** NimtRNA^{Tyr} was integrated into different locations within the intron of the Low0-eGFP splicing reporter. Respective host gene mRNA abundances were assessed by RT-qPCR, error bars represent the standard deviation from three independent experiments. Normalization was performed relative to a co-transfected reporter control, i.e., Dsred2express. **d** The nimtRNA^{Tyr} was inserted into different locations within the introns of the alternative splicing reporter and used for transient transfection experiments. **e** Alternative splicing isoforms were analyzed by RT-PCR and subsequent gel electrophoresis. PSI (percent spliced in) including standard deviation was quantified from three independent sets of experiments. **f**, **g** RT-PCR analysis was performed to detect alternative splicing isoforms of constructs containing different mtRNAs or nimtRNAs within the first intron of the Designer Exon. PSI including standard deviation was quantified from three independent sets of experiments

nimtRNA^{Trp}, which is present in a reverse-complementary orientation in its host intron (i.e., as present within the mt genome), also increased mRNA abundance (see above).

Mutational analysis by deletion of canonical tRNA domains within nimtRNA^{Tyr}, i.e., the D-arm (delD), the T-arm (delT), the acceptor stem (delAcc), or the anticodon arm (delAnti), respectively, resulted in a decrease in mRNA abundance for all mutant versions of nimtRNA^{Tyr}. NimtRNA^{Tyr} was chosen because we observed that the secondary structure of nimtRNA^{Tyr} homologs is well conserved in the human and the mouse genome. The most prominent decrease in splicing was observed upon deletion of the T-arm within nimtRNA^{Tyr} (Fig. 4b). As expected, scrambling the nimtRNA^{Tyr} sequence (designated as scrbl) failed to significantly increase eGFP mRNA levels. Interestingly, in contrast to the scrbl control, the reverse-complementary version of nimtRNA^{Tyr} designated as nimtRNA$^{Tyr\ r.-c.}$, resulted in an increase in eGFP mRNA abundance comparable to that observed using its canonical counterpart.

NimtRNAs affect pre-mRNA splicing dependent on their relative position within an intron

Splicing has been shown to be modulated in a position-dependent manner by splicing regulatory elements (SREs) [3]. Therefore, we investigated the effect of nimtRNA positioning on splice site recognition. Thus, we introduced either the YCNAW nimtRNA cluster or a single nimtRNA^{Tyr} (Y) at different locations within the intron of the Low0-eGFP splicing reporter. We observed the strongest increase in mRNA abundance/splicing when inserting the nimtRNA cluster 200 bp downstream of the 5′-splice site as compared to an insertion in the center of the 2.2-kb-long intron or 200 bp upstream of the 3′-splice site (Additional file 1: Fig. S3B and C). Upon insertion of the single nimtRNA^{Tyr} at different intronic locations, we observed the strongest increase in splicing efficiency for insertions 50 to 100 bp downstream of the 5′-splice site (Fig. 4c). Thus, we conclude that the increase in splicing efficiency upon insertion of a single or multiple nimtRNAs is position-dependent and is therefore comparable to effects observed for bona fide SREs.

NimtRNAs increase alternative exon inclusion in a three-exon splicing reporter

Next, we wanted to investigate whether nimtRNAs are also able to increase exon inclusion in a three-exon splicing reporter exhibiting an alternatively spliced internal exon. The alternative splicing reporter (designated as Designer Exon) consisted of three exons of 126 bp, 82 bp, and 273 bp in length as well as intervening introns of 242 and 637 bp, placed downstream of a CMV promoter (Fig. 4d) [41]. The second exon was designed to contain an inefficient 3′-splice acceptor at the intron1/exon2-border, with its polypyrimidine tract composed of only 50% pyrimidines, thereby reducing inclusion of the alternative exon.

Upon introduction of nimtRNA^{Tyr} (Y) into the first intron, i.e., 88 bp downstream of the 5′-splice site, an increase in PSI (percent spliced in) levels from 27.0 ± 5.4 (wild type) to 54.1 ± 4.9 (with insertion of nimtRNA^{Tyr}) could be observed (Fig. 4e). In contrast, introduction of nimtRNA^{Tyr} into the second intron 100 bp downstream of the alternative exon resulted in a decrease in PSI levels to 21.0 ± 5.1, while insertion into the middle of the intron (i.e., situated 280 bp up- and downstream from the exon borders) or close to the 3′-splice site (i.e., 100 bp upstream of the third exon) resulted in PSI levels of 24.7 ± 5.9 and 30.5 ± 6.8, respectively, comparable to the wt reporter construct lacking nimtRNAs (Fig. 4e).

As stated above, we postulate that nimtRNAs originated from mtRNAs, encoded within the mitochondrial genome. However, in the course of evolution, nuclear nimtRNAs have acquired specific mutations, compared to their mitochondrial ancestors. Thus, to determine whether the mitochondrial ancestors of nimtRNAs, i.e., bona fide mtRNAs, promoted splicing as observed for their nuclear-encoded counterparts, we also analyzed the influence of mtRNAs on alternative splicing. Hence, we introduced bona fide mouse mtRNA^{Gln}, mtRNA^{Ser1}, mtRNA^{Ser2} or the reverse-complementary variant of mtRNA^{Ser1}, designated as mtRNA$^{Ser1r.-c.}$, into intron 1 of the alternative splicing reporter. Indeed, mtRNA^{Gln}, mtRNA^{Ser1}, mtRNA^{Ser2}, and mtRNA$^{Ser1r.-c.}$ increased exon inclusion (Fig. 4f). Furthermore, the reverse-complementary mtRNA$^{Ser1r.-c.}$ showed comparable effects on exon inclusion to a two-exon Low0-eGFP splicing reporter, harboring nimtRNA^{Trp} in reverse-complementary orientation.

In contrast, the reverse-complementary variant of mtRNA^{Gln}, i.e., mtRNA$^{Gln\ r.-c.}$, did not enhance exon inclusion (PSI = 25.1 ± 5.1) (Fig. 4f). Upon closer inspection, mtRNA^{Gln} contained one U-G base pair in each of its stem regions. As a consequence, the reverse-complementary variant mtRNA$^{Gln\ r.-c.}$ exhibits an A-C pair at this position (Additional file 1: Fig. S4). Upon mutation of the respective nucleotides (i.e., A to G or C to U), we observed a partial rescue of alternative exon inclusion (PSI = 32.7 ± 5.1) (Fig. 4f, mtRNA$^{Gln\ r.-c.\ mut.}$). Interestingly, three identical copies of the same nimtRNA, i.e., mtRNA^{Ser2}, resulted in a further substantial increase in alternative exon inclusion (Fig. 4f), as was already observed in the two-exon Low0-eGFP splicing reporter employing multiple, but different, nimtRNAs (see above and Fig. 3c).

NimtRNAs of the same isotype might be derived from different mtRNA founder sequences since we hypothesized that nuclear integration of mtRNAs occurred at several different time points in evolution (see above). Importantly, sequences of nimtRNAs of the same isotype can be influenced to different extents by evolutionary pressure, and thus may differ extensively in their capacity to influence splicing.

Thus, to determine potential differences in the splicing capacity of a single nimtRNA isotype, we investigated different variants of nimtRNA^{Ser2} in the alternative splicing reporter. To this end, all nimtRNA^{Ser2} sequences were aligned by MUSCLE [42] and distances between sequences were estimated by employing MEGA to calculate the maximum likelihood [43]. This analysis resulted in the generation of distinct clusters, from which six candidate nimtRNA^{Ser2} sequences were chosen for splicing analysis in the alternative splicing reporter. In comparison, effects of bona fide human mtRNA^{Ser2} on splicing were also investigated. Interestingly, in the course of these analyses, we observed different stimulatory effects of nimtRNA^{Ser2} variants on alternative exon inclusion, ranging from 59.7 to 90.9 in PSI (Fig. 4g); thereby, bona fide human mtRNA^{Ser2} resulted in a PSI of 87.8. Thus, different nimtRNAs, derived from the same isotype, as well as bona fide mtRNAs, can exert a wide range on exon inclusion within the alternative splicing reporter.

CRISPR/Cas9 mediated partial deletion of endogenous nimtRNA^{Lys} within intron 28 of the *PPFIBP1* gene results in decreased downstream exon 29 inclusion

To analyze the influence of endogenous nimtRNAs on host gene splicing, we targeted nimtRNA^{Lys} located within intron 28 of the *PPFIBP1* gene by a CRISPR/Cas9-based approach (Fig. 5a). The *PPFIBP1* gene comprises 31 exons and contains a single intronic nimtRNA^{Lys}, exhibiting a canonical tRNA-like secondary structure. *PPFIBP1* encodes for the PPFIA binding protein 1 (PPFIBP1), a member of the LAR protein-tyrosine phosphatase-interacting protein family, also designated as liprins and is abundantly expressed in HEK 293T cells. By employing the CRISPOR web tool [44], a sgRNA (single guide RNA) was designed to directly target the T-loop of nimtRNA^{Lys} and was cloned into lentiCRISPRv2 for lentiviral transduction of HEK 293T cells, as described previously [45]. Subsequently, the efficiency of nimtRNA indel formation was confirmed by TIDE analysis [46] (Additional file 1: Fig. S5). We thereby determined an editing efficiency of 94.5%, where approximately 31% of cells harbored deletions between 13 and 17 nts, respectively.

From the bulk of nimtRNA^{Lys}-targeted cells, we screened for single-cell clones in order to obtain defined and extended CRISPR-induced nimtRNA^{Lys} deletions. Indeed,

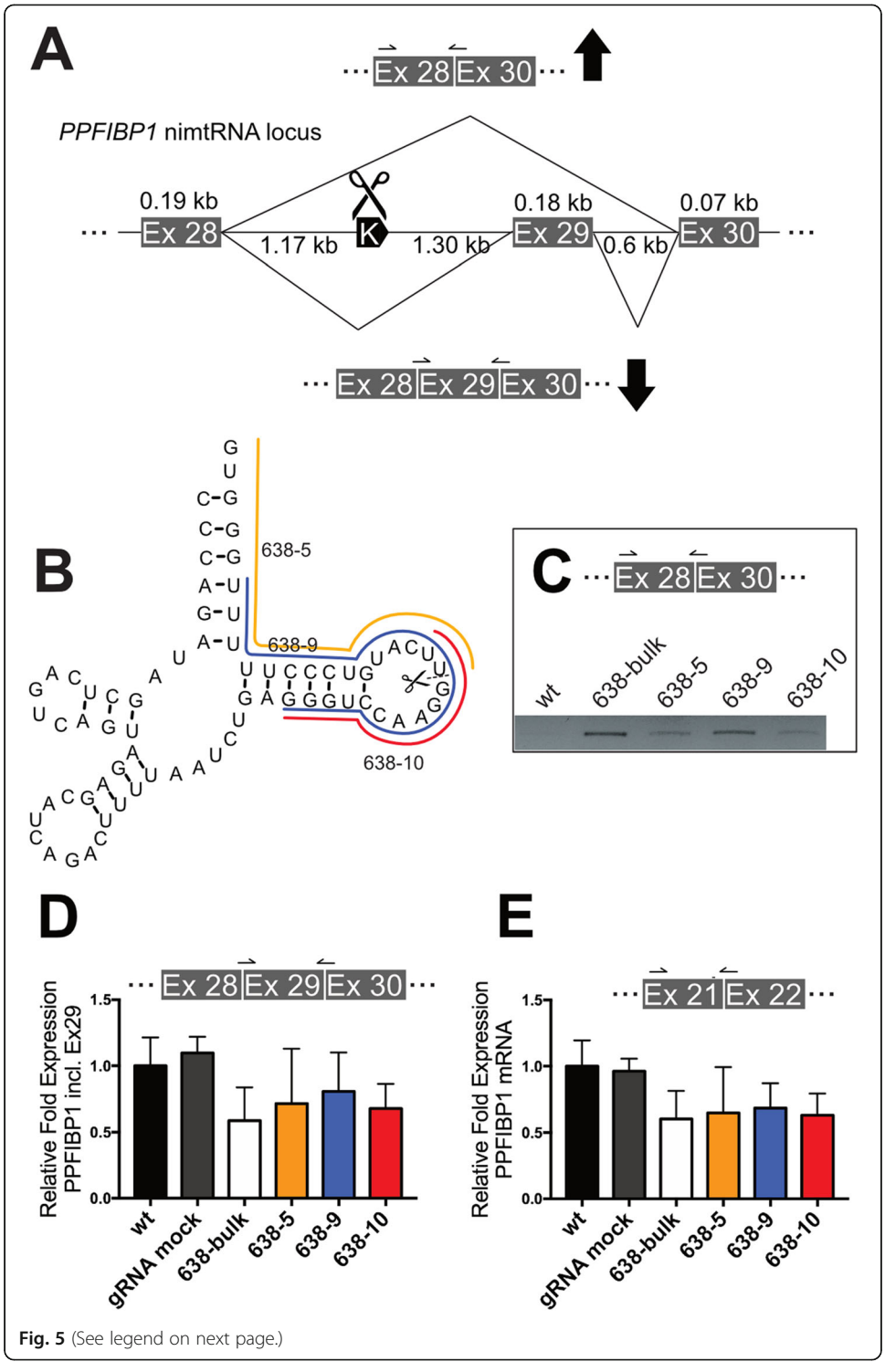


Fig. 5 (See legend on next page.)

(See figure on previous page.)
Fig. 5 A CRISPR-mediated partial deletion of a nimtRNA downregulates downstream exon inclusion of an endogenous host gene. **a** The nimtRNA^{Lys} gene, located within intron 28 of the human *PPFIBP1* gene, was targeted by a CRISPR-mediated approach to analyze the influence of partial deletions within the nimtRNA^{Lys} gene on PPFIBP1 exon 29 inclusion. **b** Single clones of CRISPR-targeted cells were cultured and analyzed by Sanger sequencing for nimtRNA^{Lys} deletions; based on this analysis, three single clones, designated as 638-5, 638-9, and 638-10, indicated in orange, blue, and red, respectively, were selected. **c** Subsequently, by RT-PCR analysis comparing wt to bulk and single clone cells, respectively, the abundance of the PPFIBP1 mRNA transcript lacking exon 29 was assessed, employing primers as indicated. **d** The abundance of PPFIBP1 mRNA transcript harboring exon 29 was determined by RT-qPCR in wt cells and compared to bulk as well as single clone CRISPR-targeted cells employing primers as indicated. Cells targeted by a guideRNA not binding to the *PPFIBP1* gene (gRNA mock) were employed as an additional control. Normalization was performed to GAPDH. **e** In addition to the specific inclusion of exon 29, the general abundance of PPFIBP1 mRNA levels was determined by employing primers binding to exon 21 and 22 upstream of the nimtRNA locus, as indicated. Cells targeted by a guideRNA not binding to the *PPFIBP1* gene (gRNA mock) were employed as an additional control. Normalization was performed to GAPDH. Averages and standard deviations were determined from three independent sets of experiments

through these analyses, we identified three clones, which displayed different larger deletions within the nimtRNA^{Lys} gene, comprising either the T-arm (i.e., 638–10) or the T-arm and the acceptor stem (i.e., 638–5 and 638–9), and exhibiting deletions from 12 to 20 and 25 nt, respectively (Fig. 5b).

By comparing wt cells to bulk or single clones of nimtRNA^{Lys}-targeted cells, similar to our constitutive splicing reporter assay, we analyzed the *PPFIBP1* mRNA for inclusion of exon 29, located downstream of nimtRNA^{Lys}, by employing primers spanning exon/exon borders (Fig. 5c, d). In the course of these analyses, we detected a splicing variant in bulk or single clones of nimtRNA^{Lys}-targeted cells which lacked exon 29, while this variant was absent in wt cells (Fig. 5c).

As for the constitutive splicing reporter (see above), in bulk nimtRNA^{Lys}-targeted cells, a decrease of about 41% in exon 29 inclusion compared to untreated wt cells was observed. In addition, we found a decrease in exon 29 inclusion for all three single clones, compared to wt cells, ranging from 19 to 32% respectively (Fig. 5d).

As an additional control, we investigated *PPFIBP1* exon 29 inclusion in cells with a nimtRNA^{Lys} unrelated sgRNA, targeting an intronic region of the *SYTL4* gene (designated as gRNA mock, Fig. 5d), which resulted in the same levels of exon 29 inclusion as observed for wt cells. The range of standard deviations of exon 29 inclusion levels in nimtRNA^{Lys}-targeted cells might thereby potentially be due to the influence of cellular stress and/or differences in cell confluency in these cells. Likely, cellular stress or varying cell confluences might result, for example, in a high variability in the expression of *trans*-acting protein factors, associated with nimtRNA-mediated splicing increase. Consistent with this hypothesis, we noted the influence of these parameters also in previous transient transfection experiments for our constitutive or alternative splicing reporter assays.

Within the *PPFIBP1* gene, exon 29 is annotated as a constitutive exon and might be essential for proper gene function. By employing a different set of primers, targeting exon 21 and 22 upstream of the nimtRNA locus, we observed a general downregulation of spliced *PPFIBP1* mRNA levels in nimtRNA^{Lys}-targeted cells (Fig. 5e). Again, gRNA mock-treated cells exhibited identical levels of *PPFIBP1* mRNA levels as wt cells. These findings are consistent with CRISPR-induced deletions within the intronic nimtRNA^{Lys} gene decreasing *PPFIBP1* host gene levels by inhibiting exon 29 inclusion.

Splicing-associated proteins bind to a nimtRNA transcript

Previous reports have demonstrated that splicing is modulated by *trans*-acting proteins which bind to SREs located within pre-mRNA transcripts [2]. Interestingly, computational analysis of the nimtRNA^{Tyr} sequence revealed a negative HEXplorer score [40] indicating the presence of potential hnRNP and/or hnRNP-like binding sites (Additional file 1: Fig. S6). Thus, to elucidate which nuclear proteins might bind to nimtRNAs, resulting in the splicing upregulation/exon inclusion, we performed an RNA immunoprecipitation (RIP) assay. This approach utilized a biotinylated T7 transcript containing five nimtRNAs identical to those employed previously (i.e., YCNAW, see above) which was incubated with a nuclear extract generated from HEK 293T cells. We propose that mouse and human nimtRNAs are likely to be recognized by identical *trans*-factors in HEK 293T cells, since we showed that nimtRNA structure, rather than sequence, is responsible for observed effects.

Employing streptavidin beads, proteins associated with nimtRNAs were isolated and subsequently separated by SDS-PAGE (Fig. 6a). Bands which predominantly appeared in the nimtRNA pull-down approach, but not in the control lacking nimtRNAs (designated as scrbl), were excised from the gel and subsequently analyzed by LC-MS. By a STRING protein-protein interaction network analysis of proteins identified by MS, we determined a highly significant network of proteins being bound to the nimtRNA transcript ($p < 1.0e{-}16$). An inherent GO analysis determined "RNA processing" and "RNA splicing" as the biological processes, and "mRNA Splicing-Major Pathway" in the Reactome, enriched significantly (Fig. 6b). The top 10 most abundant nimtRNA transcript interacting proteins are listed in Fig. 6c.

Consistent with our experimental analyses, we observed that binding sites of 24 proteins which function in splicing or exhibit other regulatory roles (see Additional File 6: Table S5 for a complete list) overlap with nimtRNA sequences, either in HepG2 or K652 cell lines, or both, as determined by analysis of ENCODE eCLIP data. Of these, G3BP1 and NSUN2 have a more than 2-fold enrichment of their binding sites in nimtRNAs. Notably, KHDRBS1 was also found to be enriched for nimtRNA binding in this dataset as observed in MS analysis by our pull-down experiments (see above).

Previously, Marnef et al. have demonstrated that canonical mtRNA^{Thr}, but not other mtRNAs, interact with PTBP1 in the cytosol [47]. In contrast to KHDRBS1, we could not retrieve overlaps of PTBP1 binding sites with nimtRNA loci in our analysis (see below). Hence, by an EMSA assay, we also investigated a potential direct interaction of PTBP1 and nimtRNA^{Tyr}. To this end, a transcript of nimtRNA^{Tyr} including an extra 10 nucleotides at the 5′ and 3′ terminus, in order to not resemble processed tRNA ends, was radioactively labeled and incubated with increasing concentrations of PTBP1 protein. However, we did not observe specific binding of PTBP1 to nimtRNA^{Tyr}, which is consistent with reported eCLIP data which provided no evidence of PTBP1 being directly associated with nimtRNAs.

KHDRBS1, also designated as Sam68, was found by MS analysis among the ten most abundant proteins binding to the nimtRNA-containing transcript. Thereby, KHDRBS1 belongs to the STAR (signal transduction and activation of RNA metabolism) protein family and has previously been demonstrated to be associated with several roles in mRNA metabolism, including splice site selection.

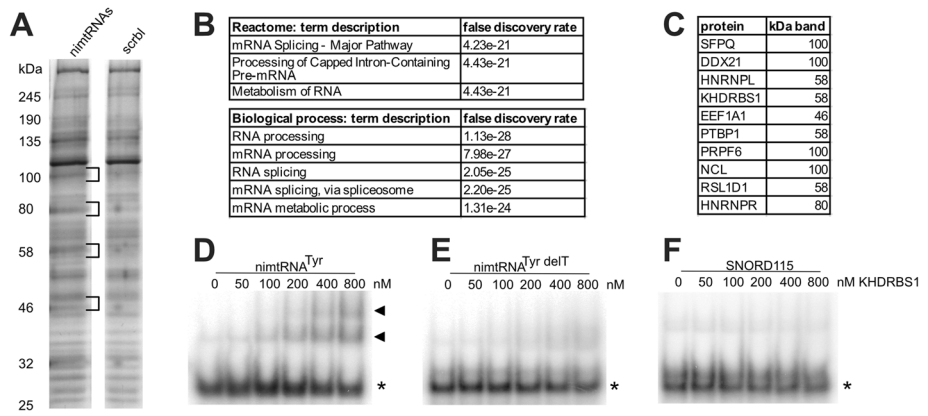


Fig. 6 NimtRNA protein interaction analysis. **a** Biotinylated transcripts containing or lacking nimtRNAs were incubated with nuclear protein extracts. Subsequently, binding proteins were isolated and analyzed by PAGE; differential band patterns were excised and analyzed by MS. **b** GO analysis of nimtRNA transcript interacting proteins. **c** The top 10 nimtRNA transcript interacting proteins ranked by absolute abundance. **d**–**f** An electrophoretic mobility shift assay was performed with increasing concentrations of KHDRBS1 being incubated with a transcript of the nimtRNA^{Tyr} (**d**), nimtRNA$^{Tyr\ delT}$ (**e**), or snoRNA SNORD115 (**f**). The unbound transcript is indicated by an asterisk; the transcript-protein complexes are indicated by triangles

We thus also employed the nimtRNA^{Tyr}-containing transcript in an EMSA analysis by adding increasing amounts of KHDRBS1 protein to the labeled nimtRNA transcript. Indeed, increasing concentrations of KHDRBS1 protein resulted in a mobility shift of the nimtRNA^{Tyr} transcript (Fig. 6d). Interestingly, in this analysis, two distinct bands were observed in the EMSA analysis employing KHDRBS1. As a negative control, the T-arm deletion mutant of nimtRNA^{Tyr} (designated as nimtRNA$^{Tyr\ delT}$), due to its reduced potential to increase splicing efficiency (see above), was employed (Fig. 6e). We observed a significantly reduced binding of KHDRBS1 to the nimtRNA$^{Tyr\ delT}$ mutant transcript compared to the nimtRNA^{Tyr} transcript. As a second negative control, the brain-specific snoRNA SNORD115 was employed in the EMSA assay, but was found to be deficient in binding to KHDRBS1, consistent with KHDRBS1 binding specifically to canonical sequences and/or structural features of nimtRNAs (Fig. 6f).

Discussion

Despite the lack of canonical nimtRNA processing, in particular their intronic location, the partial conservation of their structure, and their association with splicing-related proteins prompted us to investigate the potential role(s) of nimtRNAs in pre-mRNA splicing. By employing well-characterized splicing reporter constructs [3, 41], harboring nimtRNAs within their introns, we could indeed demonstrate that nimtRNAs increase splice site recognition, as previously reported for bona fide intronic splicing regulatory elements, designated as SREs. SREs bind *trans*-acting protein factors, which interact with spliceosomal components during different steps of spliceosomal assembly [48]. Hence, SREs are able to increase splicing of constitutive introns and also promote the inclusion of alternative exons (see below), either by affecting 5′- or 3′-splice site choice [48].

A potential reason for nimtRNAs having remained unidentified for also harboring SREs might be that previous studies on SREs have mainly focused on mini-gene, in vivo pull-down or in vitro SELEX approaches which neglected intron-located RNA

secondary structure elements as major determinants for recognition by protein *trans*-factors (splicing regulatory proteins; SRPs) [49–53]. In addition, in silico RNA secondary or higher-order structure prediction of SREs results in many false-positive structures, as has been shown by the lack of their experimental validation [54].

However, previously some pre-mRNA secondary structures have indeed been demonstrated to significantly affect splicing efficiencies [55] even though SRP binding sites are generally single-stranded [56]. As an example, intron 7 of the *SMN* gene, implicated in the development of spinal muscular atrophy, harbors an intronic SRE element with an intricate RNA secondary structure in addition to specific sequence discriminators [57]. NimtRNAs, functioning as intronic SREs, might thus exhibit a combination of specific sequence elements and structural motifs that are required for proper *trans*-factor recruitment [57]. As has been observed for SREs, we demonstrate that nimtRNAs, located within introns of host genes, are able to increase host gene pre-mRNA splicing *in cis*. In agreement with these findings, *cis*-acting intronic snoRNAs and miRNAs have also been demonstrated to be implicated in regulation of pre-mRNA splicing [5].

By placing nimtRNAs at different positions within the intron of the two-exon splicing reporter, we demonstrate that the nimtRNA-mediated increase in splicing is position-dependent, as was previously reported for a number of SR and hnRNP binding sites [3]. Notably, when placed close to the 5′-splice site but distant from the respective 3′-splice site, nimtRNA^{Tyr} exhibited the most pronounced effect on splicing efficiency in the two-exon splicing reporter in agreement with the predicted negative integral of the nimtRNA^{Tyr} sequence by the HEXplorer profile (Additional file 1: Fig. S4).

In addition to multiple nimtRNAs, also single nimtRNAs increased pre-mRNA splicing efficiency to different extents when located within the intron of the eGFP two-exon splicing reporter. When several nimtRNAs were inserted, an increase in host gene splicing efficiency was positively correlated with the number of intronic nimtRNAs. These findings are consistent with a function of nimtRNAs resembling SREs, which have been shown to cooperate in splicing of a single intron by binding simultaneously to several *trans*-acting splicing factors [49].

Splice site selection is generally regulated by spliceosomal snRNP proteins and auxiliary factors. Thereby, the intrinsic 5′-splice site strength is mainly defined by the complementarity of the 5′-splice site to the U1 snRNA. It has been reported that introns harboring weak splice sites, thereby resulting in a low basic level of splicing, are more subject to splicing regulation by splicing regulatory elements [58, 59], which is corroborated by this study. Upon employing a reporter construct with a weak 5′-splice site, we observed a more pronounced increase in nimtRNA-mediated splicing. The presence of nimtRNAs increased mRNA abundance by about 13-fold for the weak reporter, compared to an increase of approximately 3-fold for the more efficient splicing reporter construct. One explanation for the observed increase in host gene mRNA abundance might be a nimtRNA-mediated increase in host gene transcription. However, we were able to demonstrate that pre-mRNA levels remain largely unchanged when nimtRNAs are placed within the intron of the efficient two-exon splicing reporter, whereas spliced mRNA levels increased significantly (Fig. 3f).

In addition to the Low0-eGFP splicing reporter employed in the experiments described above, we also investigated the effects of nimtRNAs on a three-exon splicing reporter. The internal exon thereby exhibited an inefficient 3′-splice acceptor at the

intron1/exon2-border, thus reducing efficient inclusion of the alternative exon in this reporter (Fig. 4d) [41]. NimtRNA^{Tyr} increased exon inclusion when located in the upstream intron, whereas it decreased exon inclusion when located in the downstream intron close to the 5′-splice site (Fig. 4e). The three-exon splicing reporter employed in our experiments contained a weak 3′-splice site. This resulted in impaired alternative exon inclusion due to inefficient 3′-splice site recognition by the U2 auxiliary factors 1 and 2 (U2AF1 and U2AF2). We show that in this splicing reporter, nimtRNAs are able to significantly increase downstream alternative exon inclusion likely as a result of enhancing 3′-splice site recognition.

To determine sequence or structural requirements for the observed splicing increase within nimtRNAs, we mutated structural domains and sequence motifs within nimtRNA^{Tyr} and introduced these mutated variants into our two-exon splicing reporter assay. In this context, deletions of either tRNA arm, i.e., D-arm, T-arm, anticodon arm, and the acceptor stem of nimtRNA^{Tyr} reduced splicing efficiency to various extents, compared to the non-mutated form of nimtRNA^{Tyr}. These findings are consistent with the requirement of the conservation of canonical tRNA secondary structure for the function of nimtRNAs (see above). Future experiments will attempt to address the question whether in addition specific sequence motifs within nimtRNAs are also required and to what extent the structural requirements are different for nimtRNAs deriving from different mtRNAs. The differences between nimtRNAs and their consensus structures at least hint to some variability.

In addition to nimtRNAs, we also analyzed the effects of bona fide human mtRNAs in the two- and the three-exon splicing reporters (see above). In most cases, nimtRNAs only carry few mutations compared to their mitochondrial tRNA counterparts. As described above, most mutations are found in loop regions or as compensatory base changes in the stem regions of the nimtRNAs. In the three-exon splicing reporter, mtRNA^{Ser2}, lacking the entire D-arm, increased alternative exon inclusion. Also, we observed that different isoforms of nimtRNA^{Ser2}, displaying various base changes compared to human bona fide mtRNA^{Ser2}, promoted alternative exon inclusion to different extents. It is thus tempting to speculate that by changing only a small number of bases within nimtRNAs, the ratio of mRNA splice-isoforms can be modulated.

Surprisingly, we also observed an increase in reported splicing when employing a nimtRNA^{Trp} variant, which is present in the reverse-complementary orientation relative to its host gene. Since bona fide mitochondrial tRNAs are encoded on both strands of the mitochondrial genome both of which are transcribed, several nimtRNAs are present in reverse-complementary orientation within their host genes. Thus, by introducing a single nimtRNA^{Tyr} in a reverse-complementary orientation into the eGFP splicing reporter construct, we observed a comparable stimulatory effect on splicing efficiency as we observed for their canonical counterparts present in the sense-orientation. A plausible explanation for these observations might be that reverse-complementary variants of nimtRNAs also display a canonical cloverleaf structure due to maintaining stem-loop structures (Additional file 1: Fig. S4).

The fact that nimtRNA cloverleaf structures might be a main determinant for the observed splicing effects is in line with results obtained employing mtRNA^{Gln} in sense or antisense orientation in the three-exon splicing reporter. We observed that sense mtRNA^{Gln} increased alternative exon inclusion while in contrast this was not found for

antisense mtRNA$^{Gln\ r.-c}$. We thereby noted that mtRNA^{Gln} exhibits single G-U wobble base pairs within each of its stem regions. Hence, the antisense variant of mtRNA^{Gln} displays A-C pairs in its corresponding stem regions which might potentially interfere with canonical tRNA cloverleaf formation. Indeed, following mutation of the respective nucleotides, i.e., by introducing compensatory base changes (i.e., changing an A to a G or a C to a U), we could partially rescue the effect on alternative exon inclusion. Partial rescue of exon inclusion might, in addition, depend on (short) sequence motifs within nimtRNAs not present in mtRNA$^{Gln\ r.-c}$. It is noteworthy that in the mitochondrial genome both mtDNA strands are transcribed, resulting in the generation of polycistronic transcripts that also contain reverse-complementary variants of mtRNAs. These reverse-complementary mtRNA variants have indeed been suggested to fulfill non-canonical functions within mitochondria [60–62].

Utilizing a pull-down assay followed by MS analysis, we identified specific nuclear proteins which showed high binding affinity to a nimtRNA transcript. These nuclear proteins included Splicing factor, proline- and glutamine-rich (SFPQ), heterogeneous nuclear ribonucleoprotein L (HNRNPL), KH RNA-binding domain containing signal transduction associated 1 (KHDRBS1, also designated as Sam68), and polypyrimidine tract-binding protein 1 (PTBP1), respectively. Since all these proteins are involved in pre-mRNA splicing, we envision that these proteins may form a nimtRNA-associated pre-mRNA splicing network. Whether or not this network is identical in mouse and humans remains to be investigated in the future. In general, splicing machineries are similar and splicing-associated diseases have been successfully explored and treated in different mouse models, including the Spinraza® (Nusinersen) antisense oligonucleotide targeting the intronic splicing silencer ISS-N1 [63]. Furthermore, nimtRNAs can be found in mouse and several other species [33] suggesting a potential common pathway.

The results of the MS analysis are consistent with our analyses of eCLIP data from the ENCODE project which showed a significant enrichment in binding sites for 10 splicing-related genes (24 in total), most notably KHDRBS1, pointing towards a functional involvement of nimtRNAs in regulating splicing efficiency and specificity. Since eCLIP data were derived from two different cell lines (HepG2 and K562), we would not expect a perfect match with our experimental MS data since splicing patterns, and thus splicing regulation, differs substantially between different cell types (see below).

In agreement with our MS analysis, in a gel retardation assay (EMSA), we were able to detect specific binding of KHDRBS1, but not PTBP1 to nimtRNA^{Tyr}. By applying increasing concentrations of KHDRBS1 protein, we observed two distinct band shifts for nimtRNA^{Tyr} consistent with formation of dimers of KHDRBS1, which is in line with previous reports [64]. In contrast, the T-arm deletion mutant exhibited a significant decrease in its affinity to KHDRBS1. The low, but specific affinity of KHDRBS1 for a nimtRNA might indicate that KHDRBS1 is required but not sufficient for nimtRNA-mediated effects on splicing and that additional proteins may also contribute to this process, as also corroborated by our MS analysis.

It is of note that KHDRBS1 has been associated with positive and negative 3′- and 5′-splice site selection as well as with polypyrimidine tract binding [65]. It has been reported to directly interact with U2AF2, which in turn has been shown to associate with both the branchpoint-binding protein SF1 and the 3′-splice site-binding protein U2AF1 and is thus involved in the regulation of splicing. Indeed, the three-exon splicing reporter, employed in our analyses, harbors a weak 3′-splice site at the 5′ terminus

of the alternative exon. Accordingly, an increased inclusion of the alternative exon was observed in the presence of a nimtRNA within the upstream intron.

Similarly, nimtRNAs also increased splicing efficiency in a constitutively spliced intron, since the Low0- and Low2-eGFP splicing reporters also possess a weak 3′-splice site due to a short pyrimidine-rich region and a shorter than canonical spacer between branch-point sequence and 3′-splice site. Hence, KHDRBS1 might play a role in the repression of canonical 3′-splice site recognition in these reporters, by impairing 3′-splice site recognition as previously suggested [66, 67]. This notion is corroborated by our results showing that a more efficient 3′-splice site reduces the potential impact of nimtRNAs on splicing in the constitutive splicing reporter (Additional file 1: Fig. S7).

Based on our experiments employing two- and three-exon splicing reporters, respectively, we propose that nimtRNAs located within introns of their cognate host genes are able to affect host gene splicing patterns. Using a CRISPR/Cas9-based approach, we were able to show for the first time that partial deletions of a single endogenous nimtRNA^{Lys} gene within intron 28 of the *PPFIBP1* gene are able to significantly decrease downstream exon 29 inclusion. The induced deletions from 12 to 25 nt in length were located within the T-arm or the T-arm and the acceptor stem of nimtRNA^{Lys}, respectively, which we also show to be essential for nimtRNA^{Tyr}-mediated splicing effects in our two-exon splicing reporter assay.

CRISPR-targeting of nimtRNA^{Lys} resulted in an increase in the abundance of *PPFIBP1* transcripts lacking exon 29 and a decrease in the abundance of transcripts including exon 29, located downstream from nimtRNA^{Lys}, consistent with a role of nimtRNA^{Lys} in promoting *PPFIBP1* pre-mRNA splicing. Thereby, exon 29 is annotated as a constitutive exon, and thus likely to be essential for *PPFIBP1* protein function. This is corroborated by the observation that exclusion of exon 29 results in a general reduction of *PPFIBP1* mRNA levels (see above) consistent with nimtRNAs acting as ISEs which are known to regulate host gene mRNA levels. Our findings concerning an endogenous host gene, i.e., *PPFIBP1*, recapitulates the splicing effects observed in the two-exon constitutive splicing reporter assay described above, where we show that nimtRNAs increase eGFP reporter mRNA levels and hence eGFP protein synthesis.

Conclusions

By employing splicing reporter constructs as well as investigating an endogenous host gene, our study demonstrates a potential novel function of nimtRNAs, present in introns of host genes in the human genome, in pre-mRNA splicing. Since processing of bona fide mitochondrial tRNAs within mitochondria has been shown to be directly linked to mitochondrial mRNA processing, it is thereby tempting to speculate that nimtRNAs might have acquired a related novel function in processing/splicing of nuclear-encoded pre-mRNAs. Future studies will have to focus on the involvement of nimtRNAs in splicing regulation within all 76 introns of their human host genes as well as on their interaction and regulation by *trans*-acting protein factors.

Material and methods

Cloning and reporters

Cloning of reporter constructs was either performed by classical cloning or PCR mutagenesis approaches. For nimtRNA integration, we digested the Low0/Low2-eGFP

vector with KpnI and NdeI. NimtRNAs were amplified including 10–50 bp up- and downstream by PCR from mouse genomic DNA employing overhanging primers, cleaved with the respective enzyme(s) and ligated into the reporter vectors. The nucl. YCNAW construct was amplified by PCR with overhanging primers from a gene fragment and cloned into the Low0-eGFP reporter by KpnI and NdeI digestion. Mutations and several integrations of nimtRNAs were done by mutagenesis PCR employing the NEB Mutagenesis Kit (NEB). Oligonucleotides and gene fragments were ordered from IDT. Plasmids were transformed into TOP10 *E. coli* (One Shot® TOP10 Chemically Competent or Electrocomp™ *E. coli*), clones were selected, and DNA was extracted using NEB Miniprep Kit and sequenced by Eurofins. Positive clones were cultured; DNA for transfection experiments was extracted employing the Qiagen Midiprep Kit. Primer sequences are provided in the supplementary (Table 1 and 2).

Cells, cell culture, and manipulation

Cell culture experiments were performed with HEK 293T (ATCC® CRL-3216™) and Flip-In™-293 (Invitrogen, #R75007) cells. Cells were cultured in 4.5 g/l glucose and L-glutamine DMEM medium (Gibco) with 100 units/ml penicillin, 100 units/ml streptomycin (Gibco), and 10% heat-inactivated FBS (Gibco) at 37 °C, saturated humidity, and 5% CO_2. Cells were transiently transfected by lipotransfection employing Metafectene (Biontex). A total of 500,000 cells were seeded 24 h prior to transfection in a 6-well dish. 1.5-µg plasmids, i.e., 1 µg of splicing reporter and 0.5 µg DsredExpress2 transfection control, were transfected employing 5 µl Metafectene. Stable transfections were performed by employing the Flip-In system in Flip-In™-293 cells. Low0-eGFP splicing reporters were cloned into an EF1α promoter containing pcDNA5/FRT-derived expression vector. In total, 375 ng of plasmid was co-transfected with 1.125 ng of pOG44 into Flip-In™-293 cells and selected by Hygromycin resistance, as previously described (Invitrogen).

Table 1 List of primers used for classical cloning in this study.

Plasmid	Primer fwd	Primer rev	Entry plasmid
Y	GGGGTACCCCGTTCCG ATATCTTT GTGATTG	GGAATTCCATATGGAA TTCCCACCTTAAGA CCT CTGGTA	Low0-eGFP
YC	GGGGTACCCCGTTCCG ATATCTTT GTGATTG	GGAATTCCATATGGAA TTCCTCTACTTCTA CCG CCGAAA	Low0-eGFP
YCN	GGGGTACCCCGTTCCG ATATCTTT GTGATTG	GGAATTCCATATGGAATTCCAGACCTCAAC TAGATTGGC	Low0-eGFP
YCNA	GGGGTACCCCGTTCCG ATATCTTT GTGATTG	GGAATTCCATATGGAATTCCAACTTCTGATAA GGACTGTAG	Low0-eGFP
(nucl.) YCNAW	GGGGTACCCCGTTCCG ATATCTTT GTGATTG	GGAATTCCATATGGAATTCCGCTGTCATAAGT ACAATAACC	Low0-eGFP/Low2-eGFP
C	GGGGTACCCCTTTTTACCAGAGGT CTTAAGG	GGAATTCCATATGGAA TTCCTCTACTTCTA CCG CCGAAA	Low0-eGFP
N	GGGGTACCCCCTACCGCCATTTTT TTTTTCG	GGAATTCCATATGGAA TTCCAGACCTCAAC TAG A TTGGC	Low0-eGFP
A	GGGGTACCCCGCCAATCT AGTTGAGGTCT	GGAATTCCATATGGAATTCCAACTTCTGATAA GGACTGTAG	Low0-eGFP
W	GGGGTACCCCCTACAGTCCTTA TCAGAAGTT	GGAATTCCATATGGAATTCCGCTGTCATAAGT ACAATAACC	Low0-eGFP

Table 2 List of primers used for mutagenesis PCR cloning in this study

Plasmid	Primer fwd	Primer rev	Template plasmid
Y delD	GTCGAATTGCAAATTCGAAG	CTTAAGACCTCTGGTAAAAAG	Y
Y delT	TAAGACTTCTACCGCCAT	ACCTTCGAATTTGCAATTC	Y
Y delAcc	ttcgaaggtgtagagaaatctctacCTACCGCCATTTTTTTTTTC	tttgcaattcgacatgaatatcacctTCTGGTAAAAAGGGGTAC	Y
Y delAnti	AGGTGTAGAGAAATCTCTAC	CATGAATATCACCTTAAGAC	Y
Y scrbl	agtatttgcggattaacaatgactggtaccactGAGGTCTTAAGGTGGGAATTC	agcattaaggctttaacactcttggttttattaaTATGTTCATTAATCGTTGATTATTCTC	Y
Y r.-c.	tttacagtctaatgcttactcagccattttaccGAGGTCTTAAGGTGGGAATTC	tctaaacacagaggtttaaatcctctttttaccaTATGTTCATTAATCGTTGATTATTCTC	Y
C r.-c.	tttgcaattcgacatgaatatcaccttaagaccCTACCGCCATTTTTTTTTTC	ttcgaaggtgtagagaaatctctactaagacttTCTGGTAAAAAGGGGTAC	C
mt T	TAAACCGGAGATGAAAACCTTTTTCCAAGGACAGAGGTCTTAAGGTGGGAATTC	CAAGACTGGTGTATTAGTTTATACTACAAGGACTATGTTCATTAATCGTTGATTATTCTC	Y
Y25	ctaaacacagaggtttaaatcctctttttaccaCAGTGGCAATGAGAGTGAAG	atttacagtctaatgcttactcagccattttaccTCTTCGGAGAGCTTAAGG	Low0-eGFP
Y50	ctaaacacagaggtttaaatcctctttttaccaGTATCAGCACTTGTGGAG	atttacagtctaatgcttactcagccattttaccTTCTCCTTCACTCTCATTG	Low0-eGFP
Y100	ctaaacacagaggtttaaatcctctttttaccaGGATATTGATGATCTGTAGTGCTACAG	atttacagtctaatgcttactcagccattttaccCAAGGAGCATGGTGCCCC	Low0-eGFP
Y214	ctaaacacagaggtttaaatcctctttttaccaCAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAAC	atttacagtctaatgcttactcagccattttaccTGGGTTGGGGTCTGTGGG	Low0-eGFP
Y657	ctaaacacagaggtttaaatcctctttttaccaGAGCCAATTCCCATACATTATTG	atttacagtctaatgcttactcagccattttaccAAAGGATACCTTTGGACAG	Low0-eGFP
Y1100	ctaaacacagaggtttaaatcctctttttaccaAATAATCTTTAAGCAATCCTC	atttacagtctaatgcttactcagccattttaccGTTTTATTATTTCCAAATTGTTCTC	Low0-eGFP
Y1450	ctaaacacagaggtttaaatcctctttttaccaTGAGGGACAATTGGAGAAGTG	atttacagtctaatgcttactcagccattttaccTATCGCCTCCTCCAGGTC	Low0-eGFP
Y2125	ctaaacacagaggtttaaatcctctttttaccaTATTCACCATTATCGTTTCAG	atttacagtctaatgcttactcagccattttaccTCCCTGCCTAACTCTATTC	Low0-eGFP
Y2150	ctaaacacagaggtttaaatcctctttttaccaCGTTTCAGACCCACCTCC	atttacagtctaatgcttactcagccattttaccATAATGGTGAATATCCCTGCC	Low0-eGFP
i1-Y	ctaaacacagaggtttaaatcctctttttaccaAGTTTCCCAAAATTTTATTTTTGG	atttacagtctaatgcttactcagccattttaccAAAACAGGCTTCCAACAATG	Designer exon
i2–5′-Y	ctaaacacagaggtttaaatcctctttttaccaAACTATAAGGTAGACATTCTTATTC	atttacagtctaatgcttactcagccattttaccAGCTATTAAAAATATTGTTAATGATTC	Designer exon
i2-c-Y	ctaaacacagaggtttaaatcctctttttaccaTCATTTTCATTTCCAGGG	atttacagtctaatgcttactcagccattttaccTAACTGTGCTCAAATTCTAG	Designer exon
i2–3′-Y	ctaaacacagaggtttaaatcctctttttaccaCTGTGCACACGAGTGTAG	atttacagtctaatgcttactcagccattttaccCAAGAAAGAGAGATAACTGGG	Designer exon
mt Ser2 hs	cccccatgtctaacaacatggctttctcaAGTTTCCCAAAATTTTATTTTTGG	catgagttagcagttcttgtgagctttctcAAAACAGGCTTCCAACAATG	Designer exon
nimtRNA Ser2–12	catatattaatataacaatatggctttatcaAGTTTCCCAAAATTTTATTTTTGG	ggggcattattagcagttatcgcatactttctAAAACAGGCTTCCAACAATG	Designer exon

Table 2 List of primers used for mutagenesis PCR cloning in this study *(Continued)*

Plasmid	Primer fwd	Primer rev	Template plasmid
nimtRNA Ser2–32	ccttttgtgtatcatccataacttttctaAGTTTCCCAAAATTTTATTTTTGG	aaaagaatgagcagttttttgttttgtttAAAACAGGCTTCCAACAATG	Designer exon
nimtRNA Ser2–34	gggggtggcagtctctttcatcAGTTTCCCAAAATTTTATTTTTGG	atgttaaaaacatggcttcatcAAAACAGGCTTCCAACAATG	Designer exon
nimtRNA Ser2–55	cccccagaaaccaaactggctctcttgAGTTTCCCAAAATTTTATTTTTGG	tcgtggggttaggtcctcatgcttctctAAAACAGGCTTCCAACAATG	Designer exon
nimtRNA Ser2–56	tgtcaaatgtattagtttattctttcAGTTTCCCAAAATTTTATTTTTGG	tcaaaatttaaaactttgctctttcAAAACAGGCTTCCAACAATG	Designer exon
nimtRNA Ser2–67	ttaacaagaaaggctttttcaAGTTTCCCAAAATTTTATTTTTGG	tcaaagtaggcttttcttttttAAAACAGGCTTCCAACAATG	Designer exon

RNA isolation, reverse transcription, and (q)PCR analysis

Total RNA of cells was isolated 24 h post transfection with TRI reagent (Sigma) following the manufacturer's protocol. In total, 500 ng of total RNA of cells was employed for DNA digestion and subsequent reverse transcription utilizing the SuperScript IV VILO with ezDNase Kit following the manufacturer's protocol. Complete DNase digest was assessed by quantitative PCR. For quantitative PCR analysis, 2 μl of 1:100 diluted cDNA in a 6-μl sample volume with Luna Dye (NEB) was employed according to the manufacturer's protocol. The primers used are listed in Table 3. Quantification was performed employing the ΔΔCt method, normalizing to a co-transfected plasmid containing the DsredExpress2 reporter gene. The respective unaltered reporter construct was employed as reference. Means and standard deviations of the RQ values of at least three individual experiments were calculated. Significance was assessed by one-way ANOVA in GraphPad prism 5 (GraphPad Software, San Diego, CA). RT-PCR analysis of alternative splicing reporters was performed by using 2 μl of 1:100 diluted cDNA in a 20 μl PCR reaction with Pfu Polymerase and primers binding to the first and the last exon (listed in Table 3).

Northern blot analysis

For detection of mRNAs, 10 μg of total RNA was loaded onto a 2% Agarose gel with 2.2 M formaldehyde and subsequently blotted onto Amersham™ Hybond™-N Membranes (Thermo Fisher). The RNA was UV-crosslinked to the membranes at 0.12 kJ using a UV crosslinker (Stratagene, La Jolla, USA). Respective mRNA transcripts were detected with DIG-labeled probes amplified by PCR. For detection of ncRNAs, 10 μg of total RNA was separated on denaturing polyacrylamide gels (8 or 12%, acrylamid:bisacrylamid ≙ 29:1, 7 M urea, 1× TBE) at 150 to 250 V for 3 to 4 h. Subsequently, the gel was stained with ethidium bromide for 10 min. RNA was transferred on Amersham™ Hybond™-N Membranes (GE Healthcare) employing the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad, Vienna, Austria) at 400 mA for 45 min. The RNA was UV-crosslinked to the membranes at 0.12 kJ using a UV crosslinker (Stratagene, La Jolla, USA). Respective transcripts were detected with radioactively labeled oligonucleotides.

Fluorescence measurement

HEK 293T cells were seeded at 40,000 cells per well in a 96-well flat-bottom plate (Greiner) and transfected the next day. Fluorescence was measured in live HEK 293T cells 48 h post transfection in a Clariostar Microplate Reader (BMG Labtech). eGFP was excited at 485 ± 10 nm, dichroic filter at 503 nm, and measured at 525 ± 15 nm with the gain set to 1170. DsredExpress2 was excited at 554 ± 10 nm, dichroic filter at 571.2 nm, and measured at 591 ± 15 with the gain set to 1460. Focal height was determined at 5.2 from top. Scan mode was set to spiral with a scan diameter of 6 mm and 50 flashes per well. Measured values of PBS-transfected cells were subtracted from transfected cells. Ratio of eGFP and Dsredexpress2 was calculated. Mean and standard deviation was calculated from five separate experiments. Statistical significance was determined by one-way ANOVA in GraphPad prism 5 (GraphPad Software, San Diego, CA).

Table 3 List of primers used for RT-(q)PCR analysis

Target	Primer fwd	Primer rev
eGFP mRNA	TGAGGAGGCTTTTTTGGAGG	TTCACTAATCGAATGGATCTGTC
eGFP pre-mRNA	GTAATACGACTCACTATAGGGC	CATCAATATCCCAAGGAGCATG
Beta-Actin	CGTCACCAACTGGGACGACA	CTTCTCGCGGTTGGCCTTGG
DsredExpress2	GTCCTTCCCCGAGGGC	TTCAGCACGCCGTCGCG
GAPDH	CCATGGGGAAGGTGAAGGTC	AGTTAAAAGCAGCCCTGGTGA
Alt. spl. Rep.	AGTGATTCAGAACCGTCAAG	TCCACCACCGTCTTCTTTAG
PPFIBP1 incl ex 29	ccaaagtgaagCCAAAGAAACTT	aatcttccatctgctctaaccg
PPFIBP1 excl ex 29	gttctagagcctcgttttaacg	tgaatcttccatcttcactttgg
PPFIBP1 upstream	gaaacagaaaaagagacagcaga	CTTCTCCTAAGTtttccaaagagt

CRISPR/Cas9 targeting of nimtRNAs

The experimental setup for CRISPR/Cas9 targeting of the endogenous nimtRNA within the gene *PPFIBP1* was designed using CRISPOR [44]. The respective nimtRNA sequence including 50 bp up- and downstream was analyzed using the online tool in order to generate candidate guideRNAs. Only guideRNA sequences directly targeting the respective nimtRNA and without off-targets for 0, 1, or 2 mismatches were considered. The respective primers for gRNA cloning into the lentiCRISPR v2 by Zhang [45] were ordered, following the CRISPOR workflow. Cloning was performed following the protocol by Zhang [45]. Viruses were produced by transfecting HEK 293T cells in a 6-well dish with 400 ng of the respective gRNA construct, 200 ng pSPAX2, and 200 ng VSVg with 5 µl Metafectene following the manufacturer's protocol. Supernatant of the transfected cells was taken after 48 and 72 h for transduction of HEK 293T target cells. Cells were grown and selected with puromycin for 2–3 weeks. Single clones were generated by seeding 0.2–0.5 cells in 96-well plates in DMEM medium supplemented with 10% FBS, 1% pen/strep, and 1% methylcellulose.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were performed as described in [68]. Briefly, T7 transcripts were generated from PCR amplified templates overnight. Transcripts were dephosphorylated by Calf Intestinal Alkaline Phosphatase and subsequently 5′ labeled with [γ-^{32}P]-ATP. In total, 100 fmol of radioactively labeled transcript was incubated with heparin for 1 h at 4 °C and separated by PAGE at 4 °C at 100 V on a native 1× TBE 5% polyacrylamide gel (75:1 Acrylamid:Bisacrylamide).

Biotin-streptavidin pull-down

Transcripts were generated employing the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB, Frankfurt, Germany). Transcripts were biotinylated employing the Pierce™ RNA 3′ End Biotinylation Kit (NEB, Frankfurt, Germany) according to the manufacturer's protocol. Briefly, 50 pmol of transcript was labeled by ligation with a single biotinylated nucleotide at the 3′-terminus and subsequently purified. Labeling efficiency of biotinylated RNA was determined by dot blotting whilst following the description of the Pierce™ Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific, Vienna, Austria).

Proteins binding to the respective transcripts were isolated employing streptavidin magnetic beads (Thermo Scientific, Vienna, Austria) according to the manufacturer's protocol. Briefly, beads were washed and supplemented with RNA Capture buffer. Then, 50 pmol biotin-labeled RNA was added to the beads, followed by an incubation for 30 min at RT with agitation. Protein-RNA binding buffer (Tris pH 7.5 20 mM, NaCl 50 mM, $MgCl_2$ 2 mM, Tween 0.1%(v/v)), 30% glycerol, and 20 μg of nuclear lysate were added to the beads and incubated for 60 min at 4 °C with agitation. RNA-binding protein complexes were collected and washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole) and eluted in 50 mM ammonium acetate. Proteins were separated by SDS-PAGE and subsequently silver stained.

Mass spectrometry

Silver-stained gel bands were excised from SDS-PAGE gels, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with trypsin (Promega) as previously described [69]. Tryptic digests were analyzed using an UltiMate 3000 RSCLnano-HPLC system coupled to a Q Exactive HF mass spectrometer (both Thermo Scientific, Bremen, Germany) equipped with a Nanospray Flex ionization source. The peptides were separated on a homemade fritless fused-silica micro-capillary column (100 μm i.d. × 280 μm o.d. × 20 cm length) packed with 2.4 μm reversed-phase C18 material. Solvents for HPLC were 0.1% formic acid (solvent A) and 0.1% formic acid in 85% acetonitrile (solvent B). The gradient profile was as follows: 0–4 min, 4% B; 4–57 min, 4–35% B; 57–62 min, 35–100% B; and 62–67 min, 100% B. The flow rate was 300 nl/min.

The Q Exactive HF mass spectrometer was operating in the data-dependent mode selecting the top 20 most abundant isotope patterns with charge > 1 from the survey scan with an isolation window of 1.6 mass-to-charge ratio (m/z). Survey full-scan MS spectra were acquired from 300 to 1750 m/z at a resolution of 60,000 with a maximum injection time (IT) of 120 ms, and automatic gain control (AGC) target 1e6. The selected isotope patterns were fragmented by higher-energy collisional dissociation with normalized collision energy of 28 at a resolution of 30,000 with a maximum IT of 120 ms, and AGC target 5e5.

Data analysis was performed using Proteome Discoverer 2.2 (Thermo Scientific) with search engine Sequest. The raw files were searched against the uniprot *Homo sapiens* database. Precursor and fragment mass tolerance was set to 10 ppm and 0.02 Da, respectively, and up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as static modification and oxidation of methionine as variable modification. Acetylation, methionine loss, and methionine loss plus acetylation were set as N-terminal dynamic modification of proteins. Peptide identifications were filtered at 1% false discovery rate. Only proteins identified by at least 2 unique peptides were considered for subsequent analyses. The STRING online tool was used to analyze the proteins thus identified in terms of protein-protein interaction and Gene Ontology. Textmining, experiments, databases, co-expression, neighborhood, gene fusion, and co-occurrence were chosen as active interaction sources. Confidence was set to medium (0.400).

Protein expression and purification

Expression and purification of human recombinant PTBP1 was performed as described in [47]. The expression construct was kindly provided by Douglas Black (University of

California, Los Angeles, US). Human recombinant KHDRBS1 (Sam68; CAT#: TP300263) was ordered from Origene (Rockville, Maryland, USA).

Search for genomic loci of nimtRNA genes

Since there is no tool to accurately annotate MTLs, we tested different annotation strategies on genomic sequences of human and mouse. Annotations were tested either on the nuclear numtDNA sequences reported in Tsuji et al. [25] or the entire nuclear genome as reference. We refer to the two strategies as *NUMT*-based and *genome*-based, respectively. To detect tRNAs, we used the tRNA annotation tool tRNAscan-SE v2.0 [36] in a modified manner, applying the integrated mtRNA search mode (-M option) not to mitochondrial genomes, but to nuclear sequences. Regardless of whether the default (20 bits) or a very low (0–20 bits) cutoff score was used for filtering hits, the same results were returned. In an alternative approach, we applied Infernal v1.1.2 [35] as search engine with specific covariance models (CMs) for each of the 22 mtRNA families taken from MiTFi [18]. These CMs contain information on aberrant mtRNAs in addition to the normal mtRNA sequence and structure consensus which can help to detect MTLs exposed to high selection pressure. All Infernal hits were retained to find also MTLs that are not well conserved. Since we ran Infernal separately with each of the 22 CMs, we obtained overlapping predictions. For each locus, the MTL hit with the highest score was retained. To determine the transcriptional context, e.g., intronic, exonic, and intergenic, we assigned transcript annotations to the MTLs. We defined MTLs as intergenic if they could not be assigned to an annotated transcript. All MTLs located in introns are annotated as nimtRNAs.

Performance evaluation

For each numtDNA, the original mitochondrial sequence is traceable. We therefore can reconstruct the number, types, and order of the mtRNA copies expected within each numtDNA. This synteny information is used to validate the direct MTL annotations obtained with the different analysis strategies. We count each hit as TP if the hit is located inside a numtDNA following the occurrence and order of the given synteny information. Some numtDNAs were copied from mitochondrial sequences that lack mtRNAs. Thus, we counted hits within such numtDNAs as FPs. Hits obtained outside from numtDNAs were also counted as FPs, since we thus far only have evidence for the existence of MTLs within recognizable numtDNAs only they are considered to be true MTLs.

Sequence and structure consensus predictions

Multiple sequence and structure alignments were performed for each type of nimtRNA separately applying LocARNA [70]. The secondary structure predictions of tRNAscan-SE were used as constraints for the alignments.

Measurement of evolutionary conservation

PhyloP (phylogenetic *P* value) scores were assigned to each sequence which has been predicted from multiple genome alignments of mammals. PhyloP scores are available from UCSC [71] and can be used to detect nucleotide substitution rates that are faster or slower than expected under neutral drift in genomic sequences of different species. However, testing the conservation of (parts of) a numtDNA is not trivial. While using PhyloP scores, one has

to take into account that numtDNA, due to their quasi-repetitive nature, may have incurred problems in the genome assemblies and/or may be misaligned. Therefore, a complementary approach that compared the numtDNA to the extant human mitochondrial genome sequence was applied. The observed sequence divergence is in this case a sum of two independent effects: (i) the evolution of the numtDNA since its insertion and (ii) the evolution of mitochondrial genome since the insertion event. It can be expected that the selection pressure on the mitochondrial genome has remained neutral over time t0 because its functionality has been preserved. Since tRNAs are among the most stringently conserved genetic elements, the mitochondrial substitution rate of mtRNAs is smaller than the substitution rate of the mitochondrial proteins. Therefore, the evolutionary distance dt between MTLs or nimtRNA and mtRNA is expected $dt = (sn + st)t0$, while for the numtDNA it is $dp = (sn + sp)t0$, where sn is the neutral substitution rate in the mitochondrial genome. The substitution rates for MTLs or nimtRNA and numtDNA are given by st and sp, respectively. Outliers of this linear regression with unexpectedly large values, $dp - dt$, are then identified as the MTLs or nimtRNAs that have evolved slower than expected, i.e., those that have become subject to stabilizing selection after their insertion into the nuclear genome. Thus, the difference, $dp - dt$, is expected to be a linear function of t0. Since we are not able to calculate substitution rates and t0, we linearly transformed the model with $sn + sp$. The linear transformation leads to a model enabling MTLs or nimtRNAs to be obtained as outliers that are subject to a stronger selection pressure relative to numtDNA. Therefore, the sequence divergences can be used as measurement for the evolutionary sequence conservation. The sequence divergences (Hamming distance) dt and dp were computed by dividing their edit distance to the primordial mitochondrial sequence by their length. The edit distances were obtained by mapping the sequences to the mitochondrial genome. For this purpose, segemehl v0.2.0-418 [72] was applied with a low accuracy of 50%, while seeds with two differences were searched for to enable the mapping of strongly degraded sequences. Cook's distance [73] was applied for the outlier test and was performed in R v3.6.0 using the stats package [74]. An observation with Cook's distance larger than three times the mean Cook's distance was considered to be an outlier. Only numtDNA sequences which are longer than 50 nts were used within this analysis to avoid overestimating shorter sequences.

Determining RBP binding sites of nimtRNAs

To investigate the potential regulatory role of nimtRNAs by interaction with RNA-binding proteins (RBPs), their genomic loci with a list of experimentally validated RBP binding sites were intersected. The latter is readily available from the GENCODE project [75], which hosts a repository for BED files containing binding sites of a large set of RBPs derived from eCLIP experiments. These binding sites have already been quality controlled and show enrichment after normalization against IgG background; for more information on data generation and processing, please refer to https://www.encodeproject.org/eclip/. The genomic coordinates of nimtRNAs were intersected with RBP binding sites on the same strand to derive a list of overlaps by applying the BEDtools suite v2.29.0 [76]. RBPs that bound to each type of nimtRNA were then annotated according to their biological function with information derived from the GeneCards database [77]. The expected coverage of RBP per nucleotide intron was calculated from intersection of the eCLIP dataset with intron annotation (ENSEMBL biomart, hg38, version 98, [78]) for each RBP in the collection. By comparing

this to the RBP coverage of binding sites in nimtRNA, the relative enrichment of RBP binding events in nimtRNAs over background could be calculated.

Data sources

Mitochondrial and nuclear genomes of *Homo sapiens* (assembly hg38) and *Mus musculus* (assembly mm10) were downloaded from NCBI, release 90 [79]. The annotation of numtDNAs was obtained from [25] for the older assemblies mm9 and hg19. The numtDNA coordinates were converted to the latest genome assemblies mm10 and hg38 for mouse and human, respectively, applying the UCSC Liftover utility [71]. PhyloP scores of the multiple alignments of 29 mammalian genomes to hg38 were downloaded from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/phyloP30way/). Transcript annotations were obtained from Ensemble release 96 [78]. RBP interaction sites were downloaded from the ENCODE [80, 81] eCLIP repository [82].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-020-02199-6.

Additional file 1: Figure S1. Bioinformatic approach analysis. Performance evaluation of different MTL annotation strategies. Analysis of conservation densities of MTLs and numtDNAs. **Figure S2.** Intronic nimtRNA and snoRNA processing. The processing of intronic, Pol II- or Pol III-transcribed, plasmid encoded ncRNAs as assessed by northern blot analysis. **Figure S3.** NimtRNA effects on splicing. NimtRNA-mediated splicing increase as assessed by fluorescence. Position-dependent effects of a cluster of nimtRNAs as assessed by RT-qPCR. **Figure S4.** Proposed secondary structures of reverse-complementary nimtRNAs. Proposed secondary structures of nimtRNAs and their reverse-complementary counterparts as determined by tRNAscan or manually. **Figure S5.** TIDE analysis of the CRISPR/Cas9-targeted nimtRNA locus within intron 28 of *PPFIBP1*. CRISPR-induced nimtRNAs deletions within the *PPFIBP1* gene assessed by TIDE analysis. **Figure S6.** HEXplorer profile of nimtRNA^{Tyr} (67 nt). The HEXplorer score of nimtRNA^{Tyr} as determined in silico. **Figure S7.** Intron characteristics affect nimtRNA-mediated splicing regulation. The efficiency of 3′ splice site recognition was increased by mutation to determine its impact on the nimtRNA-mediated splicing increase.

Additional file 2: Table S1. Hg38 scores. MTL and nimtRNA hits within the human genome.

Additional file 3: Table S2. Mm10 scores. MTL and nimtRNA hits within the mouse genome.

Additional file 4: Table S3. MTL and nimtRNA outliers. MTLs and nimtRNAs exhibiting a higher degree of conservation as measured by the Cook's distance.

Additional file 5: Table S4. NimtRNA consensus. Consensus sequence and structure of the different nimtRNA types.

Additional file 6: Table S5. Enrichment RBP binding. Binding sites of proteins enriched in nimtRNA sequences.

Additional file 7. Northern Blots. Full, uncut northern blots.

Additional file 8. ENCODE sources. List of ENCODE sources used for nimtRNA eCLIP meta-analysis.

Additional file 9. Review history.

Acknowledgements

We would like to thank Hubertus Haas and Ludger Hengst for providing materials and Jonathan Vosper for helpful suggestions and proof reading of the manuscript. We would also like to thank Martina Hoelzl, Kristina Singer, Katrin Perfler, and Daniel Schreiber for excellent technical assistance. PFS holds external affiliations with the Institute of Theoretical Chemistry at the University of Vienna, Austria, the Facultad de Ciencias of the Universidad Nacional de Colombia, Bogota, Colombia, and the Santa Fe Institute.

Review history

The review history is available as additional file 9.

Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

A.H. and P.F.S. contributed to the conceptualization; A.H., P.F.S., and A.H. contributed to the methodology; S.M.H., A.M., M.G., L.K., A.H., J.F., S.B., M.P., and L.M. contributed to the investigation; A.H., S.M.H., P.F.S., and A.H. contributed to the writing of the original draft; A.H., S.M.H., L.M., M.M., S.G., and H.S. contributed to the writing and review and editing; A.H. and P.F.S. contributed to the funding acquisition; A.H., P.F.S., H.S., S.G., and H.L. contributed to the resources; A.H. and P.F.S contributed to the supervision. The author(s) read and approved the final manuscript.

Funding
This work was supported by the Austrian Science Fund [FWF ZFP326120 and SFB ZFF044110 to A.H.], the German Research Foundation (DFG) [SFB 1052 and STA 850/27-1 to P.F.S], the German Federal Ministry of Education and Research (BMBF) [de.NBI 031A538A to P.F.S.], and German Research Foundation (DFG) (SCHA909/8-1) and the Jürgen Manchot Foundation [to L.M., and H.S.].

Availability of data and materials
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [83] with the dataset identifier PXD022204 [84]. Experimental eCLIP data and metadata from ENCODE was downloaded via the command "xargs -L 1 curl -O -L < eCLIP_datasets.txt." The input text file for this command is available as supplement to this manuscript (Additional file 8).

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Division of Genomics and RNomics, Biocenter, Medical University of Innsbruck, 6020 Innsbruck, Austria. [2]Helmholtz Institute for Metabolic, Obesity and Vascular Research (HI-MAG) of the Helmholtz Zentrum München at the University of Leipzig and University Hospital Leipzig, Philipp-Rosenthal-Str. 27, 04103 Leipzig, Germany. [3]Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Leipzig University, 04107 Leipzig, Germany. [4]Institute for Virology, Medical Faculty Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany. [5]Division of Molecular Biology, Biocenter, Medical University of Innsbruck, Innsbruck, Austria. [6]Division of Clinical Biochemistry, Protein Micro-Analysis Facility, Biocenter, Medical University of Innsbruck, Innsbruck, Austria. [7]Institute of Pathophysiology, Biocenter, Medical University of Innsbruck, 6020 Innsbruck, Austria. [8]Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany.

Received: 24 March 2020 Accepted: 9 November 2020
Published online: 08 December 2020

References
1. Artamonova II, Gelfand MS. Comparative genomics and evolution of alternative splicing: the pessimists' science. Chem Rev. 2007;107(8):3407–30.
2. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136(4):701–18.
3. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, et al. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA. 2013;19(1):96–102.
4. Shenasa H, Movassat M, Forouzmand E, Hertel KJ. Allosteric regulation of U1 snRNP by splicing regulatory proteins controls spliceosomal assembly. Rna. 2020;26:1389-99.
5. Agranat-Tamir L, Shomron N, Sperling J, Sperling R. Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. Nucleic Acids Res. 2014;42(7):4640–51.
6. Janas MM, Khaled M, Schubert S, Bernstein JG, Golan D, Veguilla RA, et al. Feed-forward microprocessing and splicing activities at a microRNA-containing intron. PLoS Genet. 2011;7(10):e1002330.
7. Ballarino M, Morlando M, Pagano F, Fatica A, Bozzoni I. The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. Mol Cell Biol. 2005;25(13):5396–403.
8. Richard P, Kiss AM, Darzacq X, Kiss T. Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. Mol Cell Biol. 2006;26(7):2540–9.
9. Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J Cell Biol. 2006;173(2):207–18.
10. Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G. Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. Mol Cell Biol. 2005;25(8):3295–304.
11. Lykke-Andersen S, Ardal BK, Hollensen AK, Damgaard CK, Jensen TH. Box C/D snoRNP autoregulation by a cis-acting snoRNA in the NOP56 Pre-mRNA. Molecular Cell. 2018;72(1):99–111. e5.
12. Bachellerie JP, Cavaillé J, Hüttenhofer A. The expanding snoRNA world. Biochimie. 2002;84(8):775–90.
13. Rossmanith W. Of P and Z: mitochondrial tRNA processing enzymes. Biochim Biophys Acta. 2012;1819(9–10):1017–26.
14. Rossmanith W. Localization of human RNase Z isoforms: dual nuclear/mitochondrial targeting of the ELAC2 gene product by alternative translation initiation. PLoS One. 2011;6(4):e19152.
15. Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood A-MJ, et al. The human mitochondrial transcriptome. Cell. 2011;146(4):645–58.
16. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806):457–65.
17. de Bruijn MH, Schreier PH, Eperon IC, Barrell BG, Chen EY, Armstrong PW, et al. A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" loop and stem. Nucleic Acids Res. 1980;8(22):5213–22.

18. Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, et al. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 2012;40(7):2833–45.
19. Suzuki T, Nagao A, Suzuki T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 2011;45:299–329.
20. Gray MW. The Endosymbiont Hypothesis Revisited. In: Wolstenholme DR, Jeon KW, editors. International Review of Cytology. 141: Academic Press; 1992. p. 233–357.
21. Bestwick ML, Shadel GS. Accessorizing the human mitochondrial transcription machinery. Trends Biochem Sci. 2013; 38(6):283–91.
22. Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC. Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum Mutat. 2004;23(2):125–33.
23. Bensasson D, Feldman MW, Petrov DA. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol. 2003;57(3):343–54.
24. Woischnik M, Moraes CT. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res. 2002;12(6):885–93.
25. Tsuji J, Frith MC, Tomii K, Horton P. Mammalian NUMT insertion is non-random. Nucleic Acids Res. 2012;40(18):9073–88.
26. Riba A, Fumagalli MR, Caselle M, Osella M. Genomic duplications shaped current retrotransposon position distribution in human. bioRxiv. 2019:819284. http://biorxiv.org/content/early/2019/10/28/819284.abstract.
27. Mourier T, Hansen AJ, Willerslev E, Arctander P. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol. 2001;18(9):1833–7.
28. Tourmen Y, Baris O, Dessen P, Jacques C, Malthièry Y, Reynier P. Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics. 2002;80(1):71–7.
29. Hazkani-Covo E, Graur D. A comparative analysis of numt evolution in human and chimpanzee. Mol Biol Evol. 2007;24(1):13–8.
30. Chen JM, Chuzhanova N, Stenson PD, Férec C, Cooper DN. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Hum Mutat. 2005;25(2):207–21.
31. Gaziev AI, Shaĭkhaev GO. Nuclear mitochondrial pseudogenes. Mol Biol. 2010;44(3):405–17.
32. Telonis AG, Loher P, Kirino Y, Rigoutsos I. Nuclear and mitochondrial tRNA-lookalikes in the human genome. Front Genet. 2014;5:344.
33. Telonis AG, Kirino Y, Rigoutsos I. Mitochondrial tRNA-lookalikes in nuclear chromosomes: could they be functional? RNA Biol. 2015;12(4):375–80.
34. Piskurek O, Nikaido M, Boeadi, Baba M, Okada N. Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources. Mol Biol Evol 2003;20(10):1659–1668.
35. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5.
36. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.
37. Gstir R, Schafferer S, Scheideler M, Misslinger M, Griehl M, Daschil N, Humpel C, Obermair GJ, Schmuckermair C, Striessnig J, et al. Generation of a neuro-specific microarray reveals novel differentially expressed noncoding RNAs in mouse models for neurodegenerative diseases. Rna. 2014;20:1929-43.
38. Asang C, Hauber I, Schaal H. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic Acids Res. 2008;36(5):1450–63.
39. O'Gorman S, Fox DT, Wahl GM. Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science. 1991;251(4999):1351–5.
40. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res. 2014;42(16):10681–97.
41. Brillen AL, Schoneweis K, Walotka L, Hartmann L, Muller L, Ptok J, et al. Succession of splicing regulatory elements determines cryptic 5ss functionality. Nucleic Acids Res. 2017;45(7):4202–16.
42. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7.
43. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
44. Concordet J-P, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 2018;46(W1):W242–W5.
45. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014; 11(8):783–4.
46. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Research. 2014;42(22):e168-e.
47. Marnef A, Jady BE, Kiss T. Human polypyrimidine tract-binding protein interacts with mitochondrial tRNA(Thr) in the cytosol. Nucleic Acids Res. 2016;44(3):1342–53.
48. Wang Y, Liu J, Huang BO, Xu Y-M, Li J, Huang L-F, et al. Mechanism of alternative splicing and its regulation. Biomedical Reports. 2015;3(2):152–8.
49. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013;499(7457):172–7.
50. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. Nucleic Acids Res. 2011;39(Database issue):D301–D8.
51. Giudice G, Sánchez-Cabo F, Torroja C, Lara-Pezzi E. ATtRACT-a database of RNA-binding proteins and associated motifs. Database (Oxford). 2016;2016:baw035.
52. Paz I, Kosti I, Ares M Jr, Cline M, Mandel-Gutfreund Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res. 2014;42(Web Server issue):W361–W7.
53. Manley JL. SELEX to identify protein-binding sites on RNA. Cold Spring Harb Protoc. 2013;2013(2):156–63.
54. Dasti A, Cid-Samper F, Bechara E, Tartaglia GG. RNA-centric approaches to study RNA-protein interactions in vitro and in silico. Methods. 2020;178:11-8.

55. McManus CJ, Graveley BR. RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev. 2011;21(4):373–9.
56. Afroz T, Cienikova Z, Clery A, Allain FHT. One, two, three, four! How multiple RRMs read the genome sequence. Methods Enzymol. 2015;558:235–78.
57. Singh NN, Lawler MN, Ottesen EW, Upreti D, Kaczynski JR, Singh RN. An intronic structure enabled by a long-distance interaction serves as a novel target for splicing correction in spinal muscular atrophy. Nucleic Acids Res. 2013;41(17):8144–65.
58. Wong MS, Kinney JB, Krainer AR. Quantitative activity profile and context dependence of all human 5′ splice sites. Molecular Cell. 2018;71(6):1012–26. e3.
59. Erkelenz S, Theiss S, Kaisers W, Ptok J, Walotka L, Müller L, et al. Ranking noncanonical 5′ splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. Genome Res. 2018;28(12):1826–40.
60. Seligmann H. Pathogenic mutations in antisense mitochondrial tRNAs. J Theor Biol. 2011;269(1):287–96.
61. Seligmann H. An overlapping genetic code for frameshifted overlapping genes in Drosophila mitochondria: antisense antitermination tRNAs UAR insert serine. J Theor Biol. 2012;298:51–76.
62. Seligmann H. Undetected antisense tRNAs in mitochondrial genomes? Biol Direct. 2010;5:39.
63. Montes M, Sanford BL, Comiskey DF, Chandler DS. RNA splicing and disease: animal models to therapies. Trends Genet. 2019;35(1):68–87.
64. Chen T, Damaj BB, Herrera C, Lasko P, Richard S. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. Mol Cell Biol. 1997;17:5707–18.
65. Bielli P, Busa R, Paronetto MP, Sette C. The RNA-binding protein Sam68 is a multifunctional player in human cancer. Endocr Relat Cancer. 2011;18(4):R91–r102.
66. Romanelli MG, Diani E, Lievens PM. New insights into functional roles of the polypyrimidine tract-binding protein. Int J Mol Sci. 2013;14(11):22906–32.
67. Heiner M, Hui J, Schreiner S, Hung LH, Bindereif A. HnRNP L-mediated regulation of mammalian alternative splicing by interference with splice site recognition. RNA Biol. 2010;7(1):56–64.
68. Rozhdestvensky TS. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. Nucleic Acids Res. 2003;31(3):869–77.
69. Ott HW, Lindner H, Sarg B, Mueller-Holzner E, Abendstein B, Bergant A, et al. Calgranulins in cystic fluid and serum from patients with ovarian carcinomas. Cancer Res. 2003;63(21):7507–14.
70. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. RNA. 2012;18(5):900–14.
71. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006.
72. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol. 2009;5(9):e1000502.
73. Cook RD. Detection of influential observation in linear regression. Technometrics. 1977;19(1):15–8.
74. Team RDC. R: a language and environment for statistical computing. R Foundation for statistical computing. Vienna: R Foundation for Statistical Computing; 2018.
75. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2018;47(D1):D766–D73.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
77. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016;54(1):1.30.1–1.3.
78. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46(D1):D754–D61.
79. Coordinators NR. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2015;44(D1):D7–D19.
80. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018;46(D1):D794–801.
81. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
82. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016;13(6):508–14.
83. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47(D1):D442–d50.
84. Huettenhofer A, Hoser SM. [Hoffmann A, Meindl A, Gamper M, Fallmann J, Bernhart SH, Müller L, Ploner M, Misslinger M, Kremser L, Lindner H, Geley S, Schaal H, Stadler PF] Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing. Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing. PXD022204. https://www.ebi.ac.uk/pride/archive/projects/PXD022204 (2020).

Publisher's Note

3.2 Impairment of cellular and viral SREs contributes to aberrant splicing

As discussed above, the regulation of splice site recognition relies on an intricate network of regulators, which include not only the splice sites themselves but also neighboring splicing regulatory elements. Therefore, nucleotide substitutions including so-called silent mutations, near splice sites can contribute in various ways to changes in gene expression, including aberrant splicing and alterations in splice isoform ratios. Furthermore, viruses employ parts of the hosts' RNA processing machinery to ensure their efficient replication. In the case of HIV-1, the virus fully exploits the hosts' cellular splicing machinery to express all proteins encoded in its 9kb genomic transcript. Successful viral replication requires the balanced expression of all viral proteins. Due to the cap-dependent translation in the eukaryotic host cells, it is necessary to position downstream positioned translation start codons proximal to the cap structure, which is possible by extensive alternative splicing. Similar to human splicing regulation, nucleotide changes in viral genomes can in fact alter or inactivate splicing regulatory elements and thereby cause aberrant splicing. This can, in turn, interfere with viral replication and affect viral fitness, however, disturbance of the viral splicing balance can also potentially be exploited as a drug target.

3.2.1 RHAMM splice isoforms contributing to Multiple Myeloma progression can be predicted based on Single Nucleotide Variants (Chapter 5)

A splicing defect has been found to play a role in Multiple Myeloma, a plasma cell malignancy. RHAMM, the receptor for hyaluronan mediated motility, is a regulator of the motility of malignant B and plasma cells in myeloma. Blood or bone marrow samples of patients with multiple myeloma show it is highly expressed on the surface of malignant cells and furthermore, high ratios of a shorter splice-isoform compared to the full-length transcript in bone marrow samples of newly diagnosed patients correlate with poor survival. With the help of bioinformatics approaches, it might be possible to predict the consequences of nucleotide exchanges that can be used in Multiple Myeloma patient screenings to predict disease progression and adjust the treatment regimen.

This is a project draft that was written during a research stay (01/2019 – 03/2019) at Dana-Farber-Cancer-Institute, Harvard Medical School, Boston, USA.

Introduction

Even though novel treatment approaches and drug discoveries have shifted the treatment paradigms of multiple myeloma (MM), this plasma cell malignancy remains treatable but incurable. Over the last few years, a detailed map of cytogenetic, molecular genetic, and epigenetic lesions that underlie MM has been created. Based on these studies it is clear that MM arises through a progressive acquisition of genetic and epigenetic alterations resulting in a highly heterogeneous malignant cell population in patients. These genetic abnormalities lead to altered tumor cell growth and differentiation in patients. In addition to chromosomal rearrangements and mutations (insertions, deletions, and/or substitutions) in DNA, changes such as alteration in RNA processing, specifically altered pre-RNA splicing events are frequent in MM, and this is an area of great therapeutic interest.

As said, pre-mRNA splicing (AS) is executed in the nucleus by the large macromolecular spliceosome (Matera and Wang 2014), and the efficiency of intron removal is controlled also by *cis*-splicing elements (Cieply and Carstens 2015; De Conti, Baralle, and Buratti 2013). Alterations of splicing elements lead to aberrant splicing of many disease-related genes. These splicing alterations create functionally significant biomarkers and drug targets (Grodecka, Buratti, and Freiberger 2017).

Ongoing studies that monitor alterations in the transcriptome of patients with MM identified hundreds of alternatively spliced genes. Most importantly, RNA-seq analyses of 410 newly diagnosed patients and 18 healthy donor PCs showed a significant impact of alternative splicing on the overall clinical outcome of patients with MM (ASH abstract 2015, 2016). Prior studies of AS identified several genes misspliced in MM patients, including XPB1, MMSET, CD44, HAS1, RHAMM, and Gal-8 (Carrasco et al. 2007; Mimura et al. 2012; Dring et al. 2004; Keats et al. 2005; Keats et al. 2003; Chesi et al. 1998; Maxwell et al. 2004; Adamia et al. 2014; Adamia et al. 2013; Adamia et al. 2005). Studies that report splice variants of these genes also documents the role of these variants in disease biology, and demonstrate that their overexpression in MM clone correlates poor survival of patients. These studies suggest a significant role of splicing mechanisms in MM pathophysiology, and understanding the cause of splicing alterations can provide novel insights into the pathogenesis of MM, leading to the identification of innovative targeted therapy. Here we discuss that recurrent clusters of

mutations (SNPs and somatic genetic variations) detected in the RHAMM gene might contribute to aberrant splicing of RHAMM pre-mRNA. As demonstrated in prior studies RHAMM splicing alteration leads to aberrant RHAMM protein expression. Furthermore, it is reported that RHAMM$^{-\text{splice variant}}$/RHAMM^{-FL} ratio increases with disease burden and predicts poor survival of patients (Maxwell et al. 2004). Moreover, based on bioinformatic analysis here we describe an ASO-based approach to decrease the production of RHAMM splice variant transcripts in MM patients.

Results

Choice of bioinformatics tool to evaluate the impact of SNPs on splicing outcome

It is estimated that around 200 human diseases such as cystic fibrosis, progeria, spinal muscular atrophy, and various types of cancer arise from misregulation of pre-mRNA splicing. This can be caused by single nucleotide changes that affect either splice sites directly by destroying or weakening them or splicing regulators such as splicing regulatory elements that are crucial for the tightly orchestrated splicing process (Wang et al. 2012). Understanding the potential impact single nucleotide polymorphisms (SNPs) execute in splice site selection and exon inclusion, even without altering the coding potential of a sequence, is key to evaluate patient data correctly, especially in the light of finding reliable diagnostic markers for disease progression (Hull et al. 2007). Various bioinformatics approaches have been made in the past, aiming to directly predict the splicing outcome of SNPs to evaluate their functionality (Faber et al. 2011; ElSharawy et al. 2006; Desmet et al. 2009), in this study, however, our approach was different. The known disease-associated splice phenotype RHAMM^{-exon4} (48bp) and the elevated ratio of RHAMM^{-exon4} to RHAMM^{FL} splice isoforms respectively have been shown to be prognostic markers for poor survival in MM patients (Maxwell et al. 2004). Knowing the splicing outcome, the question arises what leads to the exclusion of exon 4 in the mRNA transcript. Therefore, we employed a bioinformatics tool to evaluate SNPs in close proximity or directly located in RHAMM exon 4 for their likelihood to contribute to an aberrant splicing reaction/the skipping of exon 4 by altering splicing regulatory elements or having an impact on splice sites themselves.

We used the “HEXplorer” tool (Erkelenz et al. 2014) (https://www2.hhu.de/rna/html/hexplorer_score.php) for our analysis, as its aim is not restricted to evaluate SNPs, it is rather designed to locate potential SREs in any given sequence and allows the analysis of the impact of mutations in potential SREs. The tool is based on hexamer weights calculated by a RESCUE-type approach (Fairbrother et al. 2002) with hexamer frequencies derived from a dataset containing over 40.0000 canonical human 5’ss sequences and their surrounding that were collected from ENSEMBL (Hartmann 2008). A normalized Z-score was determined for all 4096 possible hexamers discriminating their rather intronic or exonic (ZEI) positioning. The tool then takes overlapping hexamers into account; hence an index nucleotide receives six ZEI scores, one score for each possible position in a hexamer, which are combined to an overall HZEI score per nucleotide. The HZEI scores of each nucleotide are then plotted to generate sequence graphs where positive areas indicate a downstream enhancing segment of a sequence by potential SR protein binding, while negative areas show a likelihood of downstream silencing activities of a sequence, e.g. by hnRNP binding (Erkelenz et al. 2013). Changes in the graph area by mutations/SNPs are considered to impact potential SREs and the binding of trans-acting splicing regulatory RBPs respectively. The tool has recently been tested by various groups (Grodecka et al. 2017; Soukarieh et al. 2016; Grodecka, Buratti, and Freiberger 2017) and was described to have high discriminative power on SRE mutations.

Mapping of SNPs in the region of RHAMM exon 4 reveals a high number of polymorphisms

To overview the occurrence frequency of SNPs located close to RHAMM exon 4, we mapped all ENSEMBL derived reference SNP ID numbers (rs-numbers) located in the exon and 50bp up- and downstream of it since intronic SRE alterations or mutations located in e.g. the polypyrimidine tract (PPT) can also impact the physiological splicing reaction severely (Figure 1). However, since the frequency of SNPs in this region seems to be quite high, it seems obvious that not every single nucleotide polymorphism can contribute to an aberrant splicing reaction. Therefore, we analyzed the SNPs with the HEXplorer tool and choose four to evaluate further.

Figure 1. SNPs annotated in the ENSEMBL database mapped to RHAMM exon 4 and the respective introns 50bp up and downstream. Base numbering for RHAMM (HMMR) transcript ENST00000393915.4 located on chromosome 5 is according to ENSEMBL version GRCh37 (ENSEMBL release 75 GRCh 37.p13). Intronic nucleotides are depicted in lower case letters, exon 4 is marked in green and displayed in upper case letters, SNPs are indicated by bold pink lettering.

Prediction of splicing changing properties of four HEXplorer selected SNPs

Four SNP out of the pool of 58 SNPs located in exon 4 of RHAMM or the adjacent 50bp of both neighboring introns were chosen by HEXplorer graph analysis and subjected to prediction whether they are capable of inducing aberrant splicing of this exon (Figure 2a).

The first SNP that induced a severe change in the HEXplorer graph is rs561052191 located in RHAMM intron 3-4. It is located directly in the polypyrimidine tract of the exon 4 splice acceptor site. The polypyrimidine tract (PPT) is generally characterized as a pyrimidine-enriched upstream sequence element that is located approximately 15-50nt upstream of the 3' end of the respective intron and is followed by the consensus sequence CAG/G (/ = intron-exon border) exhibiting a terminal almost invariant AG-dinucleotide (Will and Luhrmann 2011; Cieply and Carstens 2015). Binding of the U2 auxiliary factor (U2AF) subunits U2AF65 to the PPT and U2AF35 recognizing the splice acceptor sequence (Kralovicova et al. 2015) as well as SF1/mBBP (splicing factor1/mammalian branchpoint binding protein) binding to the BPS leads to the ATP-dependent binding of the U2 spliceosomal subunit which constitutively results in the A complex formation (Berglund et al. 1997; Arias, Lubkin, and Chasin 2015). Interestingly, Y to R mutations in the PPT, as in the case of rs561052191, have been described to be involved in decreased splicing of the respective intron (Vorechovsky 2006; Sebillon et al. 1995). Looking at the HEXplorer graph of the sequence, it becomes evident that the general negative PPT graph area, which is due to the higher frequency of these hexamer combinations in intronic regions, is disrupted by the T>G SNP rs561052191, which is highly likely to contribute to less

U2AF65 binding. This is underlined by the severe change in the HZEI score (∆HZEI = 113.36) which displays the change to a less negative area in the PPT (Figure 2b).
Moving further downstream, another SNP, rs1235517851 seems likely to be involved in the generation of RHAMM^{-exon4} splice isoforms. This intronic SNP is located directly downstream of the AG dinucleotide of the splice acceptor site and although it does not interfere with the CAG intronic consensus sequence, it still decreases the negative graph area depicting the PPT by an ∆HZEI of 50.67. Moreover, it generates an additional AG dinucleotide due to its A>G nucleotide change in a AAA>GCAG sequence, which could lead to competitive binding of U2AF35 which might ultimately result in decreased exon 4 recognition and contributes to the splicing phenotype RHAMM^{-exon4}. The creation of an additional splice acceptor site also weakens the intrinsic strength of the constitutive splice acceptor according to the MaxEnt score algorithm (Yeo and Burge 2004) from 6.76 to 3.96, which also supports the prediction of this SNP being involved in altered splicing of RHAMM exon 4 (Figure 2c).
The exonic SNP rs1175449655 is located directly in the consensus motif of the U1 snRNA binding site, decreasing its intrinsic strength. Splice donor (SD) strength can generally be measured by calculating the complementarity of the RNA tail of the U1 snRNP and the 11-nucleotide long recognition sequence of the SD and taking the H-bonds that can be formed between the two sequences into account, hence, the more H-bonds that can be formed, the higher is the intrinsic strength of a splice donor site. The H-bond score (HBS) ranges between 1.8 (lowest complementarity) and 23.8 (full complementarity) (Freund et al. 2003). The exonic nucleotides that are fully complementary to the U1 snRNA are CAG, however, SNP rs1175449655 shifts sequence to CAA which leads to the decreased intrinsic strength of the splice donor (HBS WT = 17.3 > HBS SNP = 12.6). This drastic reduction in complementarity to the U1 snRNA is highly likely to result in a decreased binding of the U1 snRNP, facilitating the skipping of RHAMM exon 4, which then results in the elevated ratio of RHAMM-exon4/RHAMMFL (Figure 2d).
We also included an SNP in our analysis which we would, according to the HEXplorer graph analysis, determine as not being capable to drastically change the splicing outcome for RHAMM exon 4. The T to C silent mutation of rs767100503 is located in the middle of exon 4. It induces an ∆HZEI of only 26.72, which indicates that it is very

likely to be tolerated (Figure 2e). However, it is important to mention that synonymous mutations should generally be considered in such analysis even though they do not change the underlying amino acid code, they might still be well capable of contributing to aberrant splicing processes. This is for example described by Yamaguchi et. al., (Yamaguchi et al. 2017) where a silent mutation created a functional splice donor site directly upstream of the constitutive SD in a patient with transverse and sigmoid colon cancer, another infamous example is the silent C–T transition in exon 7 of SMN2 that leads to exon 7 exclusion so that SMN2 cannot compensate for the lack of SMN1 in SMA patients (Son et al. 2019).

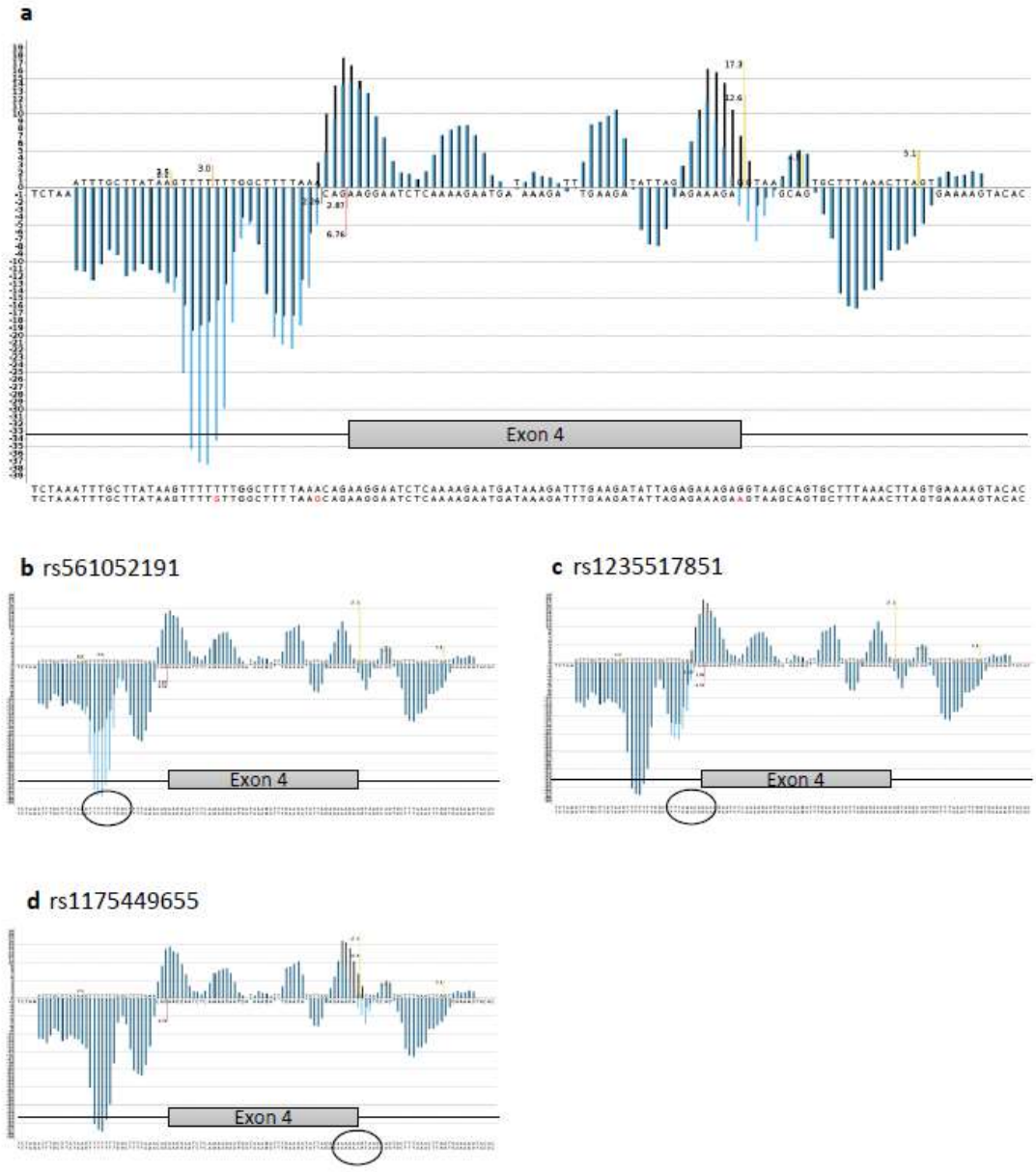


Figure 2. (a) HEXplorer graph of RHAMM exon 4 and parts of the neighboring introns (chromosome 5:162894651 – 5:162894810 in ENSEMBL GRCh 37.p13) containing SNPs rs561052191, rs1235517851, and rs1175449655 that are likely to impact the inclusion of exon 4 in the mRNA transcript. The consensus sequence is depicted in blue while the sequence containing the SNPs is plotted in black. AG dinucleotides are marked with red bars, splice acceptor strength measured as MaxEnt score is indicated beneath. GT dinucleotides are marked by yellow bars and the respective scoring measured as H-bond score (HBS) can be found beneath. (b) HEXplorer graph of SNP rs561052191 resulting in a T to G nucleotide change that decreases the negative plot area of the PPT (△HZEI = 113.36) which indicates a decrease in PPT strength, also resulting in a decreased intrinsic strength of the splice acceptor (MaxEnt score 6.76 > 5.67). This SNP is likely to decrease U2AF binding and subsequently the recognition of the SA, resulting in the loss of the exon. (c) HEXplorer graph of SNP rs1235517851 resulting in an A to G nucleotide change that not only creates an additional weak splice acceptor directly upstream of the constitutional SA but also weakens the intrinsic strength of the latter drastically (MaxEnt score 6.76 > 3.96). With two putative splice acceptor sites in close proximity, the competitional binding of U2AF might lead to steric hindrances and overall decreased recognition of the splice acceptor and the exon in general. (d) HEXplorer graph of SNP rs1175449655 resulting in a G to A nucleotide change directly located in the splice donor site. It decreases the complementarity of the SD to the U1 snRNA which is likely to result in a decreased recognition of the SD by the U1 snRNP which potentially contributes to the exclusion of exon 4 in the mRNA transcript. (e) HEXplorer graph of SNP rs767100503 resulting in a T to C synonymous nucleotide change in the middle of exon 4. It only changes the HZEI score by 26.72 and is highly likely to be tolerated.

Impact of SNPs directly occurring in the splice donor sequence

Single nucleotide polymorphisms are generally capable of affecting the balanced splicing process by disrupting SREs. However, a subset of SNPs, as most of the ones discussed above, has an impact on splice sites themselves and can influence the recognition of the sequences by spliceosomal subunits, especially U1 snRNP and U2snRNP. Here in particular the recognition of the 11-nucleotide long splice donor sequence is prone to be affected by nucleotide changes since the intrinsic strength of a splice donor is highly dependent on the complementarity to the U1 snRNA that forms a duplex upon recognition (Freund et al. 2003). Therefore, we performed a separate analysis of the splice donor sequence of RHAMM exon 4 that can be altered by annotated SNPs located in the 11-nucleotide recognition sequence. As expected, most SNPs induce a change in the complementarity of the sequences, however, it was surprising that some SNPs such as rs752321701 and rs755740711 can indeed increase the complementarity and hence might increase the recognition of the splice donor sequence. Though, SNPs such as rs1175449655 and rs988517256 decrease complementarity and are likely to contribute to aberrant splicing of this exon (Table 1).

Table 1. The impact of SNPs on the intrinsic strength (H-bond score) of RHAMM exon 4 splice donor.

Sequence H-Bond Score

Sequence	HBS	
CAGGTAAGTAT		U1snRNA
gAGGTAAGcAg	17.30	WT
gAaGTAAGcAg	12.60	rs1175449655
gAGGTtAGcAg	13.40	rs988517256
gAGGTAAGTAg	20.80	rs752321701
gAGGTAAGcAa	17.30	rs755740711
gAaGTtAGTAa	9.00	All SNPs

Discussion

The analysis of single nucleotide variants in individuals and the evaluation of their potency to contribute to diseases is a very important step on the way to personalized detection, prevention, and treatment of diseases (Agyeman and Ofori-Asenso 2015). The analysis of the impact of one single nucleotide variation, however, is still a hurdle regarding the vast amount and variety of SNPs reported in various databases (Welter et al. 2014; Bruno et al. 2012; Leslie, O'Donnell, and Johnson 2014). Furthermore, while many genetic variations are only assessed based on their impact on the amino acid sequence, synonymous mutations or mutations that are capable of influencing RNA processing such as the splicing reaction are often out of focus.

In this report, we revisited a well-described disease-associated splice isoform of the receptor for hyaluronan-mediated motility (RHAMM) which is a hyaluronan-binding protein that is expressed on multiple myeloma plasma cells (Crainie et al. 1999). An elevated ratio of RHAMM lacking exon 4 (RHAMM-exon4) to RHAMM full length (RHAMMFL) is associated with poor survival of MM patients and disease progression (Maxwell et al. 2004). We analyzed SNPs in the proximity or directly located in RHAMM exon 4 and used a bioinformatics approach to evaluate their potential to contribute to aberrant splicing and the increase in RHAMM-exon4 splice isoform.

Splicing affecting mutations usually occur as cis-sequence alterations that affect the proper recognition of exon and intron boundaries and the inclusion of exons in the mRNA transcript. While some of them disrupt existing splice sites, others create de novo splice sites or activate cryptic ones. Additionally, mutations can impact splicing

regulatory elements and subsequently the binding of trans-acting splicing regulatory RBPs (Sterne-Weiler and Sanford 2014). In silico analysis tools that aim to evaluate the impact of such splicing altering mutations are of great help in research and also diagnostic settings. Although the tools are based on various algorithms and datasets, many of them provide useful information to set nucleotide change into context of the surrounding sequences and provide predictions whether they can affect splicing ((Grodecka et al. 2017; Soukarieh et al. 2016), also reviewed in (Anna and Monika 2018)). However, it has to be stressed that these analyses are limited to be predictions and still require experimental validation, but they provide starting points for further investigation (Thery et al. 2011). Bioinformatic analyses of sequence variants such as SNPs that are capable of altering the splicing reactions as discussed in this manuscript can lead to research for biomarkers e.g. MM patient screenings since the clinical outcome is already proven. This knowledge can ultimately be used to predict disease burden and to adjust the treatment regimen which provides better care for patients. Additionally, these bioinformatic predictions paired with experimental approaches can reveal the Achilles heel of diseases, pointing out potential targets for new therapeutic approaches such as antisense-oligonucleotide treatment.

In the last years, several RNA therapeutic approaches demonstrated not only the potential of antisense oligonucleotides (ASOs) to specifically target RNA molecules but also that ASOs have the potential to discriminate between SNPs (Stein and Castanotto 2017; Crooke et al. 2018; Levin 2019; Ostergaard et al. 2017; Magner et al. 2017). Consequently, and concerning the influence of the RHAMM-exon4/ RHAMMFL splice isoform ratios in favor of the patient, ASO-based approaches might be an additional therapy opportunity for MM patients with SNPs affecting the RHAMM exon4 splicing. Here, one possible strategy would be to correct splicing, hence RHAMM exon4 inclusion by directing ASOs against RNA sequences containing the SNPs involved in exon4 skipping which would result in ASO-mediated inclusion of RHAMM exon4 and thereby a decreased ratio of RHAMM-exon4/RHAMMFL. The potential of such a strategy is supported by the FDA-approved drugs with a similar underlying molecular mechanism, Eteplirsen, and Nusinersen. Both represent splice switching ASOs (ASOs which mask their sequence to prevent binding of e.g. SR or hnRNP proteins (Havens and Hastings 2016) influencing either exon skipping of exon 51 dystrophin pre-mRNA

(Eteplirsen) or exon 7 inclusion in the survival motor neuron protein 2 mRNA and are therefore promising therapeutic strategies for Duchenne muscular dystrophy (DMD) or spinal muscular atrophy (SMA) patients (Hua et al. 2008; Singh et al. 2017).

Furthermore, a second ASO-based strategy would not aim at correcting the exon4 inclusion but on degrading RHAMM-exon4 mRNAs to increase the expression of the full-length isoform. Here, in contrast to ASOs which correct splicing, ASOs involved in target mRNA degradation (e.g. locked nucleic acid gapmers) recruit RNase H1 after mRNA binding destroying the ASO-bound mRNA (Grunweller and Hartmann 2007; Levin 2019). Such a strategy is also supported by FDA-approved drugs which aim on decreasing the expression of target mRNAs. For example, Mipomersen, an ASO inducing the degradation of the apolipoprotein B mRNA is a therapeutic agent for homozygous hypercholesterolemia (Levin 2019; Raal et al. 2010; Akdim et al. 2010)

Materials and Methods

SNPs were collected from the ENSEMBL database (ENSEMBL release 75 GRCh 37.p13) according to their location in RHAMM (HMMR transcript ENST00000393915.4) on chromosome 5. Results were compared to NCBI dbSNP. Bioinformatic analysis was performed using the HEXplorer algorithm available at https://www2.hhu.de/rna/html/hexplorer_score.php, described in Erkelenz et. al., 2014 (Erkelenz et al. 2014). For the calculation of the intrinsic strength of splice acceptors sites, MaxEntScan::score3ss for human 3' splice sites was used (accessible via http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html, (Yeo and Burge 2004)). Splice donor strength was calculated using the H-bond score web interface version 3.4 (https://www2.hhu.de/rna/html/hbond_score.php, described in (Freund et al. 2003)).

References

1. Matera, A. G.; Wang, Z., A day in the life of the spliceosome. Nature reviews. Molecular cell biology 2014, 15, (2), 108-21.
2. Cieply, B.; Carstens, R. P., Functional roles of alternative splicing factors in human disease. Wiley interdisciplinary reviews. RNA 2015, 6, (3), 311-26.
3. De Conti, L.; Baralle, M.; Buratti, E., Exon and intron definition in pre-mRNA splicing. Wiley interdisciplinary reviews. RNA 2013, 4, (1), 49-60.

4. Grodecka, L.; Buratti, E.; Freiberger, T., Mutations of Pre-mRNA Splicing Regulatory Elements: Are Predictions Moving Forward to Clinical Diagnostics? International journal of molecular sciences 2017, 18, (8).
5. Carrasco, D. R.; Sukhdeo, K.; Protopopova, M.; Sinha, R.; Enos, M.; Carrasco, D. E.; Zheng, M.; Mani, M.; Henderson, J.; Pinkus, G. S.; Munshi, N.; Horner, J.; Ivanova, E. V.; Protopopov, A.; Anderson, K. C.; Tonon, G.; DePinho, R. A., The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. Cancer Cell 2007, 11, (4), 349-60.
6. Mimura, N.; Fulciniti, M.; Gorgun, G.; Tai, Y. T.; Cirstea, D.; Santo, L.; Hu, Y.; Fabre, C.; Minami, J.; Ohguchi, H.; Kiziltepe, T.; Ikeda, H.; Kawano, Y.; French, M.; Blumenthal, M.; Tam, V.; Kertesz, N. L.; Malyankar, U. M.; Hokenson, M.; Pham, T.; Zeng, Q.; Patterson, J. B.; Richardson, P. G.; Munshi, N. C.; Anderson, K. C., Blockade of XBP1 splicing by inhibition of IRE1alpha is a promising therapeutic option in multiple myeloma. Blood 2012, 119, (24), 5772-81.
7. Dring, A. M.; Davies, F. E.; Fenton, J. A.; Roddam, P. L.; Scott, K.; Gonzalez, D.; Rollinson, S.; Rawstron, A. C.; Rees-Unwin, K. S.; Li, C.; Munshi, N. C.; Anderson, K. C.; Morgan, G. J., A global expression-based analysis of the consequences of the t(4;14) translocation in myeloma. Clin Cancer Res 2004, 10, (17), 5692-701.
8. Keats, J. J.; Maxwell, C. A.; Taylor, B. J.; Hendzel, M. J.; Chesi, M.; Bergsagel, P. L.; Larratt, L. M.; Mant, M. J.; Reiman, T.; Belch, A. R.; Pilarski, L. M., Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients. Blood 2005, 105, (10), 4060-9.
9. Keats, J. J.; Reiman, T.; Maxwell, C. A.; Taylor, B. J.; Larratt, L. M.; Mant, M. J.; Belch, A. R.; Pilarski, L. M., In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. Blood 2003, 101, (4), 1520-9.
10. Chesi, M.; Nardini, E.; Lim, R. S.; Smith, K. D.; Kuehl, W. M.; Bergsagel, P. L., The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood 1998, 92, (9), 3025-34.
11. Maxwell, C. A.; Rasmussen, E.; Zhan, F.; Keats, J. J.; Adamia, S.; Strachan, E.; Crainie, M.; Walker, R.; Belch, A. R.; Pilarski, L. M.; Barlogie, B.; Shaughnessy, J., Jr.; Reiman, T., RHAMM expression and isoform balance predict aggressive disease and poor survival in multiple myeloma. Blood 2004, 104, (4), 1151-8.
12. Adamia, S.; Kriangkum, J.; Belch, A. R.; Pilarski, L. M., Aberrant posttranscriptional processing of hyaluronan synthase 1 in malignant transformation and tumor progression. Adv Cancer Res 2014, 123, 67-94.
13. Adamia, S.; Pilarski, P. M.; Belch, A. R.; Pilarski, L. M., Aberrant splicing, hyaluronan synthases and intracellular hyaluronan as drivers of oncogenesis and potential drug targets. Curr Cancer Drug Targets 2013, 13, (4), 347-61.
14. Adamia, S.; Reiman, T.; Crainie, M.; Mant, M. J.; Belch, A. R.; Pilarski, L. M., Intronic splicing of hyaluronan synthase 1 (HAS1): a biologically relevant indicator of poor outcome in multiple myeloma. Blood 2005, 105, (12), 4836-44.
15. Wang, J.; Zhang, J.; Li, K.; Zhao, W.; Cui, Q., SpliceDisease database: linking RNA splicing and disease. Nucleic acids research 2012, 40, (Database issue), D1055-9.
16. Hull, J.; Campino, S.; Rowlands, K.; Chan, M. S.; Copley, R. R.; Taylor, M. S.; Rockett, K.; Elvidge, G.; Keating, B.; Knight, J.; Kwiatkowski, D., Identification of common genetic variation that modulates alternative splicing. PLoS genetics 2007, 3, (6), e99.
17. Faber, K.; Glatting, K. H.; Mueller, P. J.; Risch, A.; Hotz-Wagenblatt, A., Genome-wide prediction of splice-modifying SNPs in human genes using a new analysis pipeline called AASsites. BMC bioinformatics 2011, 12 Suppl 4, S2.
18. ElSharawy, A.; Manaster, C.; Teuber, M.; Rosenstiel, P.; Kwiatkowski, R.; Huse, K.; Platzer, M.; Becker, A.; Nurnberg, P.; Schreiber, S.; Hampe, J., SNPSplicer: systematic analysis of SNP-dependent splicing in genotyped cDNAs. Human mutation 2006, 27, (11), 1129-34.

19. Desmet, F. O.; Hamroun, D.; Lalande, M.; Collod-Beroud, G.; Claustres, M.; Beroud, C., Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research 2009, 37, (9), e67.
20. Erkelenz, S.; Theiss, S.; Otte, M.; Widera, M.; Peter, J. O.; Schaal, H., Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic acids research 2014, 42, (16), 10681-97.
21. Fairbrother, W. G.; Yeh, R. F.; Sharp, P. A.; Burge, C. B., Predictive identification of exonic splicing enhancers in human genes. Science 2002, 297, (5583), 1007-13.
22. Hartmann, L. T., Stephan; Niederacher, Dieter; Schaal, Heiner, Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Frontiers in bioscience : a journal and virtual library 2008, 13, 3252-3272.
23. Erkelenz, S.; Mueller, W. F.; Evans, M. S.; Busch, A.; Schoneweis, K.; Hertel, K. J.; Schaal, H., Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. Rna 2013, 19, (1), 96-102.
24. Grodecka, L.; Hujova, P.; Kramarek, M.; Krsjakova, T.; Kovacova, T.; Vondraskova, K.; Ravcukova, B.; Hrncirova, K.; Soucek, P.; Freiberger, T., Systematic analysis of splicing defects in selected primary immunodeficiencies-related genes. Clinical immunology 2017, 180, 33-44.
25. Soukarieh, O.; Gaildrat, P.; Hamieh, M.; Drouet, A.; Baert-Desurmont, S.; Frebourg, T.; Tosi, M.; Martins, A., Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLoS genetics 2016, 12, (1), e1005756.
26. Will, C. L.; Luhrmann, R., Spliceosome structure and function. Cold Spring Harb Perspect Biol 2011, 3, (7).
27. Kralovicova, J.; Knut, M.; Cross, N. C.; Vorechovsky, I., Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3' splice-site organization and activity of U2AF-related proteins. Nucleic acids research 2015, 43, (7), 3747-63.
28. Berglund, J. A.; Chua, K.; Abovich, N.; Reed, R.; Rosbash, M., The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell 1997, 89, (5), 781-7.
29. Arias, M. A.; Lubkin, A.; Chasin, L. A., Splicing of designer exons informs a biophysical model for exon definition. Rna 2015, 21, (2), 213-29.
30. Vorechovsky, I., Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic acids research 2006, 34, (16), 4630-41.
31. Sebillon, P.; Beldjord, C.; Kaplan, J. C.; Brody, E.; Marie, J., A T to G mutation in the polypyrimidine tract of the second intron of the human beta-globin gene reduces in vitro splicing efficiency: evidence for an increased hnRNP C interaction. Nucleic acids research 1995, 23, (17), 3419-25.
32. Yeo, G.; Burge, C. B., Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 2004, 11, (2-3), 377-94.
33. Freund, M.; Asang, C.; Kammler, S.; Konermann, C.; Krummheuer, J.; Hipp, M.; Meyer, I.; Gierling, W.; Theiss, S.; Preuss, T.; Schindler, D.; Kjems, J.; Schaal, H., A novel approach to describe a U1 snRNA binding site. Nucleic acids research 2003, 31, (23), 6963-75.
34. Yamaguchi, T.; Wakatsuki, T.; Kikuchi, M.; Horiguchi, S. I.; Akagi, K., The silent mutation MLH1 c.543C>T resulting in aberrant splicing can cause Lynch syndrome: a case report. Jpn J Clin Oncol 2017, 47, (6), 576-580.
35. Son, Y. S.; Choi, K.; Lee, H.; Kwon, O.; Jung, K. B.; Cho, S.; Baek, J.; Son, B.; Kang, S. M.; Kang, M.; Yoon, J.; Shen, H.; Lee, S.; Oh, J. H.; Lee, H. A.; Lee, M. O.; Cho, H. S.; Jung, C. R.; Kim, J.; Cho, S.; Son, M. Y., A SMN2 splicing modifier rescues the disease phenotypes in an in vitro human spinal muscular atrophy model. Stem Cells Dev 2019.
36. Agyeman, A. A.; Ofori-Asenso, R., Perspective: Does personalized medicine hold the future for medicine? J Pharm Bioallied Sci 2015, 7, (3), 239-44.

37. Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorff, L.; Parkinson, H., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research 2014, 42, (Database issue), D1001-6.
38. Bruno, A. E.; Li, L.; Kalabus, J. L.; Pan, Y.; Yu, A.; Hu, Z., miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes. BMC genomics 2012, 13, 44.
39. Leslie, R.; O'Donnell, C. J.; Johnson, A. D., GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics 2014, 30, (12), i185-94.
40. Crainie, M.; Belch, A. R.; Mant, M. J.; Pilarski, L. M., Overexpression of the receptor for hyaluronan-mediated motility (RHAMM) characterizes the malignant clone in multiple myeloma: identification of three distinct RHAMM variants. Blood 1999, 93, (5), 1684-96.
41. Sterne-Weiler, T.; Sanford, J. R., Exon identity crisis: disease-causing mutations that disrupt the splicing code. Genome biology 2014, 15, (1), 201.
42. Anna, A.; Monika, G., Splicing mutations in human genetic disorders: examples, detection, and confirmation. Journal of applied genetics 2018, 59, (3), 253-268.
43. Thery, J. C.; Krieger, S.; Gaildrat, P.; Revillion, F.; Buisine, M. P.; Killian, A.; Duponchel, C.; Rousselin, A.; Vaur, D.; Peyrat, J. P.; Berthet, P.; Frebourg, T.; Martins, A.; Hardouin, A.; Tosi, M., Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. Eur J Hum Genet 2011, 19, (10), 1052-8.
44. Stein, C. A.; Castanotto, D., FDA-Approved Oligonucleotide Therapies in 2017. Mol Ther 2017, 25, (5), 1069-1075.
45. Crooke, S. T.; Witztum, J. L.; Bennett, C. F.; Baker, B. F., RNA-Targeted Therapeutics. Cell Metab 2018, 27, (4), 714-739.
46. Levin, A. A., Treating Disease at the RNA Level with Oligonucleotides. N Engl J Med 2019, 380, (1), 57-70.
47. Ostergaard, M. E.; Nichols, J.; Dwight, T. A.; Lima, W.; Jung, M. E.; Swayze, E. E.; Seth, P. P., Fluorinated Nucleotide Modifications Modulate Allele Selectivity of SNP-Targeting Antisense Oligonucleotides. Mol Ther Nucleic Acids 2017, 7, 20-30.
48. Magner, D.; Biala, E.; Lisowiec-Wachnicka, J.; Kierzek, R., Influence of mismatched and bulged nucleotides on SNP-preferential RNase H cleavage of RNA-antisense gapmer heteroduplexes. Sci Rep 2017, 7, (1), 12532.
49. Havens, M. A.; Hastings, M. L., Splice-switching antisense oligonucleotides as therapeutic drugs. Nucleic acids research 2016, 44, (14), 6549-63.
50. Hua, Y.; Vickers, T. A.; Okunola, H. L.; Bennett, C. F.; Krainer, A. R., Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. American journal of human genetics 2008, 82, (4), 834-48.
51. Singh, N. N.; Howell, M. D.; Androphy, E. J.; Singh, R. N., How the discovery of ISS-N1 led to the first medical therapy for spinal muscular atrophy. Gene Ther 2017, 24, (9), 520-526.
52. Grunweller, A.; Hartmann, R. K., Locked nucleic acid oligonucleotides: the next generation of antisense agents? BioDrugs 2007, 21, (4), 235-43.
53. Raal, F. J.; Santos, R. D.; Blom, D. J.; Marais, A. D.; Charng, M. J.; Cromwell, W. C.; Lachmann, R. H.; Gaudet, D.; Tan, J. L.; Chasan-Taber, S.; Tribble, D. L.; Flaim, J. D.; Crooke, S. T., Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. Lancet 2010, 375, (9719), 998-1006.
54. Akdim, F.; Stroes, E. S.; Sijbrands, E. J.; Tribble, D. L.; Trip, M. D.; Jukema, J. W.; Flaim, J. D.; Su, J.; Yu, R.; Baker, B. F.; Wedel, M. K.; Kastelein, J. J., Efficacy and safety of mipomersen, an antisense inhibitor of apolipoprotein B, in hypercholesterolemic subjects receiving stable statin therapy. J Am Coll Cardiol 2010, 55, (15), 1611-8.

3.2.2. Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b. (Chapter 6)

One accessory protein vital for viral replication is Vif, the viral infectivity factor. It poses as a counter actor of the host restriction factor APOBEC3G, and balanced levels are required for efficient viral replication. Here, the noncoding exon 2/2b plays an important role since it contains the Vif start codon between two alternatively used splice donors, namely D2 and D2b. Thus, while exon 2 is non-coding, exon 2b is coding and therefore the use of these splice donors also affects viral proteome diversity. The two splice donor sites are tightly regulated by a dense network of splicing regulatory elements that offer binding sites for splicing regulatory proteins. By mutational analysis, a novel splicing regulatory element, as well as protein binding partners of this regulatory network, were identified.

The following article is published in J Virol. 2017 Jun 26;91(14). pii: e00389-17. doi: 10.1128/JVI.00389-17. by

Brillen AL, Walotka L, Hillebrand F, **Müller L**, Widera M, Theiss S, Schaal H.

A.L.B. and H.S. conceived the study and designed the experiments. A.L.B., L.W., L.M. and M.W. performed cloning, transfection experiments and (q)RT-PCR analyses. A.L.B. performed RNA-pull-down analyses. F.H. performed LNA-related experiments. S.T., A.L.B. and H.S. performed HEXplorer analyses. S.T. provided statistical analyses. A.L.B., S.T. and H.S. wrote the manuscript.

Individual contribution: 10% L.M. performed cloning and experiments shown in Figure 7.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b

Anna-Lena Brillen,[a] Lara Walotka,[a] Frank Hillebrand,[a] Lisa Müller,[a] Marek Widera,[a,b] Stephan Theiss,[c] Heiner Schaal[a]

Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany[a]; University Hospital, University of Duisburg-Essen, Institute of Virology, Essen, Germany[b]; Institute of Clinical Neuroscience and Medical Psychology, University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany[c]

ABSTRACT The HIV-1 accessory protein Vif is essential for viral replication by counteracting the host restriction factor APOBEC3G (A3G), and balanced levels of both proteins are required for efficient viral replication. Noncoding exons 2/2b contain the Vif start codon between their alternatively used splice donors 2 and 2b (D2 and D2b). For *vif* mRNA, intron 1 must be removed while intron 2 must be retained. Thus, splice acceptor 1 (A1) must be activated by U1 snRNP binding to either D2 or D2b, while splicing at D2 or D2b must be prevented. Here, we unravel the complex interactions between previously known and novel components of the splicing regulatory network regulating HIV-1 exon 2/2b inclusion in viral mRNAs. In particular, using RNA pulldown experiments and mass spectrometry analysis, we found members of the heterogeneous nuclear ribonucleoparticle (hnRNP) A/B family binding to a novel splicing regulatory element (SRE), the exonic splicing silencer ESS2b, and the splicing regulatory proteins Tra2/SRSF10 binding to the nearby exonic splicing enhancer ESE2b. Using a minigene reporter, we performed bioinformatics HEXplorer-guided mutational analysis to narrow down SRE motifs affecting splice site selection between D2 and D2b. Eventually, the impacts of these SREs on the viral splicing pattern and protein expression were exhaustively analyzed in viral particle production and replication experiments. Masking of these protein binding sites by use of locked nucleic acids (LNAs) impaired Vif expression and viral replication.

IMPORTANCE Based on our results, we propose a model in which a dense network of SREs regulates *vif* mRNA and protein expression, crucial to maintain viral replication within host cells with varying A3G levels and at different stages of infection. This regulation is maintained by several serine/arginine-rich splicing factors (SRSF) and hnRNPs binding to those elements. Targeting this cluster of SREs with LNAs may lead to the development of novel effective therapeutic strategies.

KEYWORDS HEXplorer score, exon recognition, host restriction factor, human immunodeficiency virus, pre-mRNA processing, splicing regulatory elements

During long terminal repeat (LTR)-driven transcription, over 50 mRNA isoforms emerge by alternative splicing of the HIV-1 precursor mRNA (1, 2). According to their distinct sizes, mRNA isoforms can be divided into three different classes: 2-kb mRNAs (intronless), encoding Tat, Rev, and Nef; intron-containing 4-kb mRNAs, encoding Vif, Vpr, Vpu, and Env; and 9-kb unspliced mRNAs, encoding Gag and Gag-Pol (3). Viral gene expression follows a strict chronological order (4–6). In the early phase, only intronless mRNAs are transported out of the nucleus and translated, whereas intron-

Received 8 March 2017 **Accepted** 18 April 2017

Accepted manuscript posted online 26 April 2017

Citation Brillen A-L, Walotka L, Hillebrand F, Müller L, Widera M, Theiss S, Schaal H. 2017. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. J Virol 91:e00389-17. https://doi.org/10.1128/JVI.00389-17.

Editor Karen L. Beemon, Johns Hopkins University

Address correspondence to Heiner Schaal, schaal@uni-duesseldorf.de.

containing 4-kb and 9-kb mRNAs depend on the accumulation of Rev protein, which facilitates their export into the cytoplasm in the later phase.

Primarily responsible for the vast amount of mRNA isoforms are four splice donor sites (D1 to D4), eight splice acceptor sites (A1 to A7, including the alternative splice acceptors A4 a, b, and c), and several only rarely used sites, like splice donor 2b (D2b) (1–3). Their recognition depends on intrinsic strength, as well as *cis*-acting splicing regulatory elements (SREs) bound by, e.g., serine/arginine-rich (SR) or heterogeneous nuclear ribonucleoparticles (hnRNPs) (7).

Splicing itself is a highly regulated process controlled by several components of the spliceosomal complex. It starts with U1 small nuclear ribonucleic particle (snRNP) binding to the splice donor, followed by U2 snRNP binding to the branch point sequence of the upstream splice acceptor (8). U1 and U2 snRNPs pair in a process named exon definition (9), which is later transformed into an intron definition process (10, 11) in which U1 and U2 snRNPs couple across the intron and thereby initiate the splicing reaction. SR or hnRNPs can support U1 snRNP binding to a splice donor, depending on their exonic or intronic position (12).

Up to this time, many SREs have been identified within the pre-mRNA of HIV-1 (Fig. 1A). Only recently, five novel SREs could be identified using the HEXplorer algorithm (13). This algorithm reflects potentially enhancing and silencing properties of hexamers in the neighborhood of a splice donor by calculating the frequency of hexamer occurrence within introns versus exons. HEXplorer score (HZ_{EI}) profiles along sequences depict exonic enhancing regions as positive and silencing regions as negative values. Furthermore, HEXplorer score differences (ΔHZ_{EI}) between wild-type (WT) and mutant sequences quantitatively reflect the mutation effect on splice-enhancing/silencing properties. Any disruption of a splice site or an SRE can lead to a profound weakening of viral replication (14). Exclusively within HIV-1 exons 2 and 2b, six different SREs have already been described (Fig. 1B). Within exon 2, serine/arginine-rich splicing factor 1 (SRSF1)-dependent exonic splicing enhancers (ESEs) M1 and M2 (15), as well as the SRSF4-dependent ESE-Vif (16), have been shown to activate D2, whereas two G runs suppress exon 2/2b inclusion (16, 17). Furthermore, a novel HEXplorer-identified SRE within exon 2b, ESE2b (previously called $ESE^{5005–5032}$), was shown to activate downstream splice donor usage in minigene analysis (13).

In addition to splice acceptor A1 recognition and removal of the most 5′-proximal intron, use of the downstream splice donor must be prevented to result in the formation of *vif* mRNAs. Downstream splice donor sites D2 and D2b, however, have to be recognized by U1 snRNP to activate A1 but rendered splicing incompetent to maintain the *vif* open reading frame (ORF), whose start codon lies within the downstream intron of D2 (17).

Vif is a low-abundance, 23-kDa small protein that is incorporated into newly assembling virions. Vif counteracts the host restriction factor APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G) (A3G) (18), which is also encapsidated into virus particles and primarily triggers G-to-A hypermutations in the viral genome during reverse transcription (RT). Vif binds to A3G to provoke ubiquitination and proteosomal degradation. Although Vif is absolutely essential for efficient HIV-1 replication in A3G-expressing cells, excessive Vif is deleterious, since massive levels of Vif inhibit proteolytic Gag processing (19).

In the present study, we focused on the functional importance of splicing regulatory elements within exon 2/2b. On the basis of our results, we provide evidence that multiple SREs within exon 2/2b tightly regulate proper *vif* mRNA production. We could underline the functional importance of ESE2b, bound by Tra2 and SRSF10, and the newly discovered exonic splicing silencer 2b (ESS2b), bound by hnRNP A/B proteins, for splice donor use and exon recognition. Point mutations within the SREs predicted via the HEXplorer algorithm, as well as locked nucleic acid (LNA) masking, altered both viral *vif* mRNA and Vif protein amounts by regulating exon 2/2b inclusion and led to a drop in viral particle production.

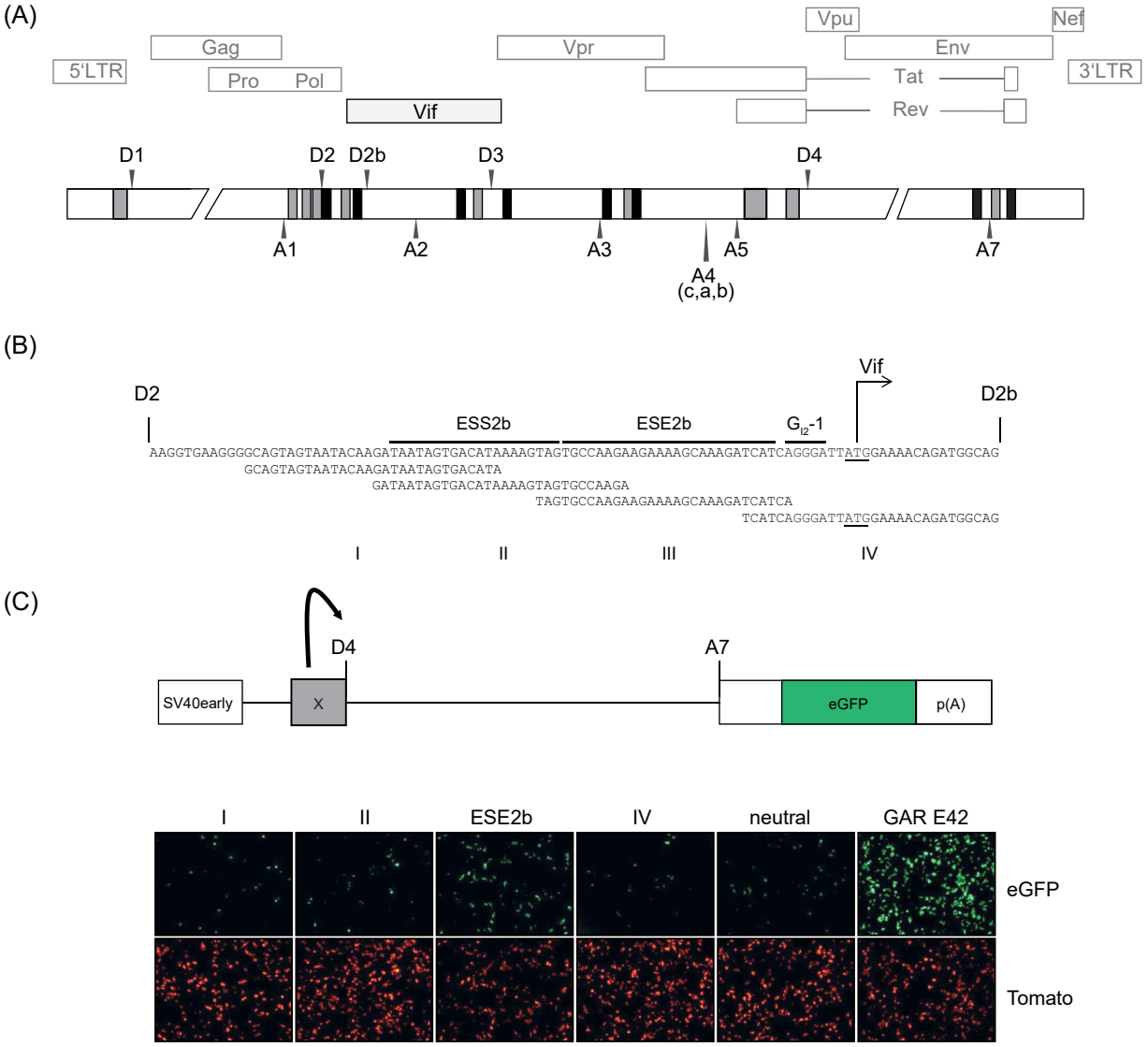


FIG 1 Analysis of SREs in HIV-1 exon 2/2b. (A) Black (silencer) and gray (enhancer) bars represent published SREs. Splice donor sites (D1 to D4), splice acceptor sites (A1 to A7), and protein ORFs are shown. (B) SREs within exon 2/2b and definition of parts I to IV. The translational start codon for Vif is underlined. (C) Fluorescence microscopy analysis of fragments I to IV. (Top) Schematic overview of the single-intron eGFP splicing reporter. Any of sequences I, II, III (ESE2b), IV, neutral $(CCAAACAA)_{5\times}$, and GAR E42 was inserted upstream of D4. (Bottom) HeLa cells were transiently transfected with 1 μg of each construct, together with 1 μg of pCL-dTOM, to monitor transfection efficiency. Twenty-four hours after transfection, fluorescence microscopy was carried out.

(This research was conducted by A.-L. Brillen in partial fulfillment of the requirements for a PhD from Heinrich Heine University, Düsseldorf, Germany, 2017.)

RESULTS

Tra2 and SRSF10 act via ESE2b to activate the downstream splice donor D2b. To understand splice site selection critical for HIV-1 *vif* mRNA formation, we focused on the exonic 2b region downstream of splice donor D2 (Fig. 1). The *vif* start codon is localized upstream of an alternative splice donor, termed D2b, which defines the 3′ end of exon 2b, but needs to be repressed to retain the downstream intronic sequence coding for Vif. Previously, we have shown that D2b is repressed by a conserved G run (G_{I2}-1) located immediately upstream, which is bound by hnRNP F/H (17). As inactivating G_{I2}-1 led to upregulation of the intrinsically rather weak splice donor D2b, we hypothesized that G_{I2}-1 not only represses D2b but might additionally shield an upstream bound SR protein from activating D2b (17). This assumption was further

supported by the observation that, in the presence of multiple exonic SREs, the SRE closest to the splice donor likely dominates splicing decisions (12). Therefore, we tested the region between D2 and D2b for splice donor-enhancing properties and split the region into four overlapping segments, as indicated in Fig. 1B. To test the segments for splice donor-enhancing properties, we used an HIV-1 subgenomic reporter that allows monitoring of SRE-mediated U1 snRNP binding to splice donor D4, forming an enhanced green fluorescent protein (eGFP)-encoding mRNA by splicing to splice acceptor A7 (20–22) (Fig. 1C, top). Following transient transfection, fluorescence microscopy allowed a first rough estimation of enhancing properties in the four exonic 2b segments. We used the sequence CCAAACAA (23) as a splicing-neutral reference and the very strongly enhancing purine-rich SRE HIV-1 GAR E42 fragment as a positive control (GAR contains GAA or GAG repeats [R is A or G]) (20, 22). As expected, fragment IV, covering G_{I2}-1, did not support downstream splice donor use, while ESE2b ($ESE^{5005–5032}$ [13]), contained in fragment III, enhanced D2b use. Neither fragment I nor fragment II led to increased eGFP expression (Fig. 1C, bottom), demonstrating that ESE2b was the only SRE in the 3′ part of exon 2b capable of supporting downstream splice donor usage.

To identify splicing regulatory proteins binding to ESE2b, we made use of the previously published inactivating nucleotide substitutions predicted by the HEXplorer algorithm [$ESE2b^{MUT}$ (ΔHZ_{EI} −267), termed "5015A>T" or "5025A>T (dm)" in reference 13]. Here, regions with positive HEXplorer scores (HZ_{EI}) have been shown to exhibit downstream splice-enhancing properties, and a negative HEXplorer score difference ΔHZ_{EI} means that the mutations render the region less downstream enhancing. We performed RNA affinity purification assays with RNA oligonucleotides containing either the ESE2b or the $ESE2b^{MUT}$ sequence. After coupling to agarose beads, the oligonucleotides were incubated with HeLa cell nuclear extract. After washing and elution, bound proteins were analyzed via mass spectrometry (MS). Besides weak binding to several members of the SR protein family, we found a significant loss of the proteins Tra2α, Tra2β, and SRSF10 in the mutant ESE2b sequence and no significant change in the level of any hnRNP (Table 1) ($P = 0.05$; *t* test).

SREs between D2 and ESE2b are necessary to maintain splicing at D2. To test the impact of ESE2b on D2/D2b splice donor selection, we used a heterologous three-exon minigene splicing reporter (Fig. 2A) previously shown to be suitable to dissect the role of *cis*-acting SREs in splice donor decisions in complex splicing networks (24). Within this splicing reporter, the artificial internal exon was not recognized at all when it was completely composed of splicing-neutral sequences (23) but could be exonized upon replacing neutral sequences by *cis*-acting SREs or by increasing splice donor complementarity to U1 snRNA above an HBond score (HBS) of 15.8 (24). Here, we quantified splice donor complementarity to U1 snRNA by the experimentally derived HBS (range, 1.8 to 23.8) and splice acceptor strength by the MaxEnt score (ME). When we inserted both viral splice donors, D2 (HBS, 10.7) and D2b (HBS, 12.4), into this context of neutral sequences, the exon was not recognized, even though it is bordered by an intrinsically strong splice acceptor (ME, 10.25) (Fig. 2B, lane 1). To recapitulate HIV-1 exon 2 splice site recognition, all known exon 2-localized SREs—ESE-Vif (16), ESE-M1 and -M2 (15), and the GGGG motif (16), as well as ESE2b and G_{I2}-1 (for simplicity, collectively referred to here as ESE2)—were inserted either individually or in combination into the exon at their authentic positions either upstream or downstream of D2 (Fig. 2A).

Replacing corresponding neutral sequences with ESE2 alone comparably activated D2 and D2b (Fig. 2B, cf. lanes 1 and 2). Additionally replacing neutral sequences with ESE2b and G_{I2}-1 switched splice donor selection to almost exclusive D2b rather than D2 use (Fig. 2B, lane 3), indicating that ESE2b not only strongly supported D2b selection, overriding the repressive G_{I2}-1 activity, but at the same time blocked the upstream-localized D2. Even though D2b has higher complementarity to U1 snRNA than D2 (HBS, 12.4 versus 10.7), in the viral context it is rarely used: 0.2% versus 5.3% D2 usage (17).

TABLE 1 Mass spectrometry analysis of ESE2b (average of duplicates)[a]

Gene name	No. of unique peptides	Log_2 difference
TRA2A	**10.5**	**3.97**
TRA2B	**13.5**	**3.20**
SRSF10	**11.5**	**3.16**
SRSF3	**4.5**	**1.39**
SRSF7	**7**	**1.32**
SRSF6	**4**	**0.99**
SRSF4	**4.5**	**0.68**
SRSF9	**12**	**0.45**
SRSF1	**25**	**0.45**
SRSF2; SRSF8	**4**	**−0.06**
SRSF11	**6**	**−0.32**
HNRNPUL2-BSCL2; HNRNPUL2	12.5	1.48956667
HNRNPR	12	1.38463333
HNRNPL	14	0.5345
HNRNPDL	6	0.50626667
HNRNPU	26	0.45483333
HNRNPA2B1	20	0.3558
HNRNPH3	10.5	0.3429
HNRNPM	26	0.32243333
HNRNPA3	17.5	0.2572
HNRNPA1; HNRNPA1L2	21	0.24806667
HNRNPH2	9	0.19285
HNRNPLL; HNRPLL	17	0.14186667
HNRNPK	27	0.1075
HNRNPH1	6.5	0.06496667
HNRNPUL1	17	0.0539
HNRNPF	16	−0.0073
HNRNPC	5	−0.22633333
HNRNPD	6	−0.26476667
HNRNPA0	9.5	−1.01033333

[a]Unique peptides and log_2 differences of SR (boldface) and hnRNPs enriched after RNA affinity purification are shown; log_2 differences of normalized protein intensities were calculated as wild-type minus mutated sequence samples; log_2 differences significant at the 5% level (2 sided) are indicated by shading.

To examine the impact of ESE2b variants on splice donor selection, we tested two ESE2b mutations that reduced its splice-enhancing activity (WT > ΔHZ_{EI}-94 > ΔHZ_{EI}-267). For brevity, we also denote these mutations by their corresponding ΔHZ_{EI} values with respect to the wild-type sequence, which reflect the reduction in splice-enhancing properties (Fig. 2C).

As shown in Fig. 2B, a stepwise switch toward D2 usage occurred when we reduced the ESE2b HEXplorer score by 2-nucleotide (nt) mutations, thus weakening its splice-enhancing activity (Fig. 2B, cf. lanes 3 to 5). This D2b-to-D2 transition occurred with both intact and inactivated G_{I2}-1 (Fig. 2B, cf. lanes 6 to 8), but in the latter case, a larger reduction in ESE2b splice-enhancing activity was required to switch to D2 selection. Thus, splicing occurred at the weaker upstream splice donor, D2, if the combined splice-enhancing properties of ESE2b and G_{I2}-1 did not suffice to move splice donor selection to D2b, located downstream.

So far, however, HIV-1 D2 usage, as in the viral context, could not be mimicked with this minigene, indicating that there may be an additional *cis*-acting element in the viral sequence. Such an SRE, localized between D2 and ESE2b, might act like an "insulator" separating the ESE2 from the ESE2b activity. Therefore, we profiled exon 2b for further enhancing and silencing properties of splice donor neighborhoods. The region located directly downstream of splice donor D2 (Fig. 3A, top left [WT]) includes four consecutive peaks (A, B, C, and D) of the HEXplorer profile. It exhibits predominantly HZ_{EI}-negative areas and is supposed to support the splice donor D2, located upstream.

We then substituted either the whole fragment (WT; A to D) or individual fragments (A, B, C, and D) for neutral sequences of the same length in the minigene reporter. After RT-PCR analysis, it became obvious that, indeed, the region from A through D reversed splice site selection from D2b to the native HIV-1 splice donor, D2, which is more

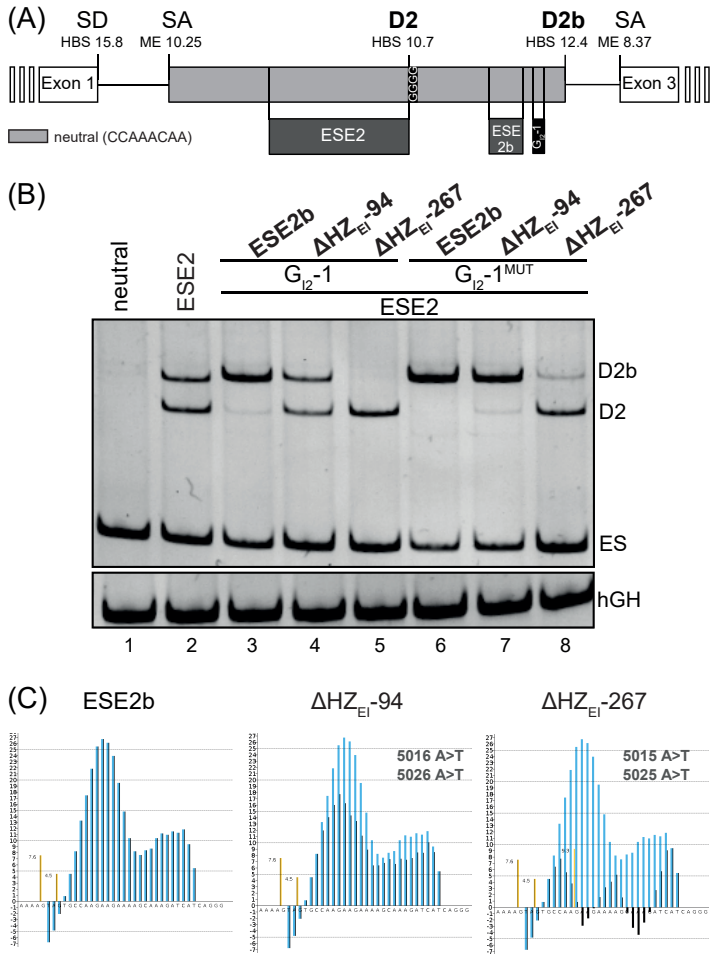


FIG 2 Impact of ESE2b on D2b recognition. (A) Schematic of the three-exon minigene. The middle exon is composed of only neutral CCAAACAA repeats (23), except for D2, D2b, and the depicted SREs. (B) RT-PCR analyses of the splicing pattern of the minigene shown in panel A. HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. RNA isolated from the cells was subjected to RT-PCRs using primer pairs #2648/#2649 and #1224/#1225 (human growth hormone [hGH]). The PCR amplicons were separated on a nondenaturing 10% polyacrylamide gel and stained with ethidium bromide. (C) HZ_{EI} plots of ESE2b and its two mutants ($\Delta HZ_{EI} = -94$ and $\Delta HZ_{EI} = -267$). Black, mutated sequence; blue, wild-type reference; brown, HBond scores for WT; yellow, HBond scores for mutants.

frequently used in the viral context (Fig. 3B, cf. lanes 1 and 2). Further analyses of the individual fragments demonstrated that fragments C and D, rather than fragment A or B, affected splice donor choice (Fig. 3B, lanes 3 to 6). However, as neither fragment C nor D on its own was sufficient to fully induce the splice donor switch, we concluded that the potential SRE spanned both fragments and termed it ESS2b. To examine our hypothesis, we specifically changed ESS2b by HEXplorer-guided mutagenesis in fragment C, D, or both in the context of A to D [Fig. 3A, C^{MUT}, D^{MUT}, and $(C+D)^{MUT}$]. Analysis of the splicing pattern (Fig. 3C) revealed that mutating either C or D led to a partial splice donor switch, whereas simultaneously mutating C and D showed the same splicing phenotype as the neutral sequence (Fig. 3C, cf. lanes 1 and 5). These results demonstrate that ESS2b spans C and D, enhances upstream D2, and represses downstream D2b recognition in the presence of downstream ESE2b. Next, to identify splicing regulatory proteins binding to ESS2b, we again performed RNA affinity purification with WT and mutant sequences as described above. Subsequent MS analysis revealed that, besides hnRNP DL binding, members of the hnRNP A/B family (the hnRNP A/B family includes isoforms A1, A2/B1, A3, and A0) especially were markedly enriched in the WT compared to the mutant sample, whereas, in contrast, no SR protein was significantly enriched ($P = 0.05$; *t* test) (Table 2).

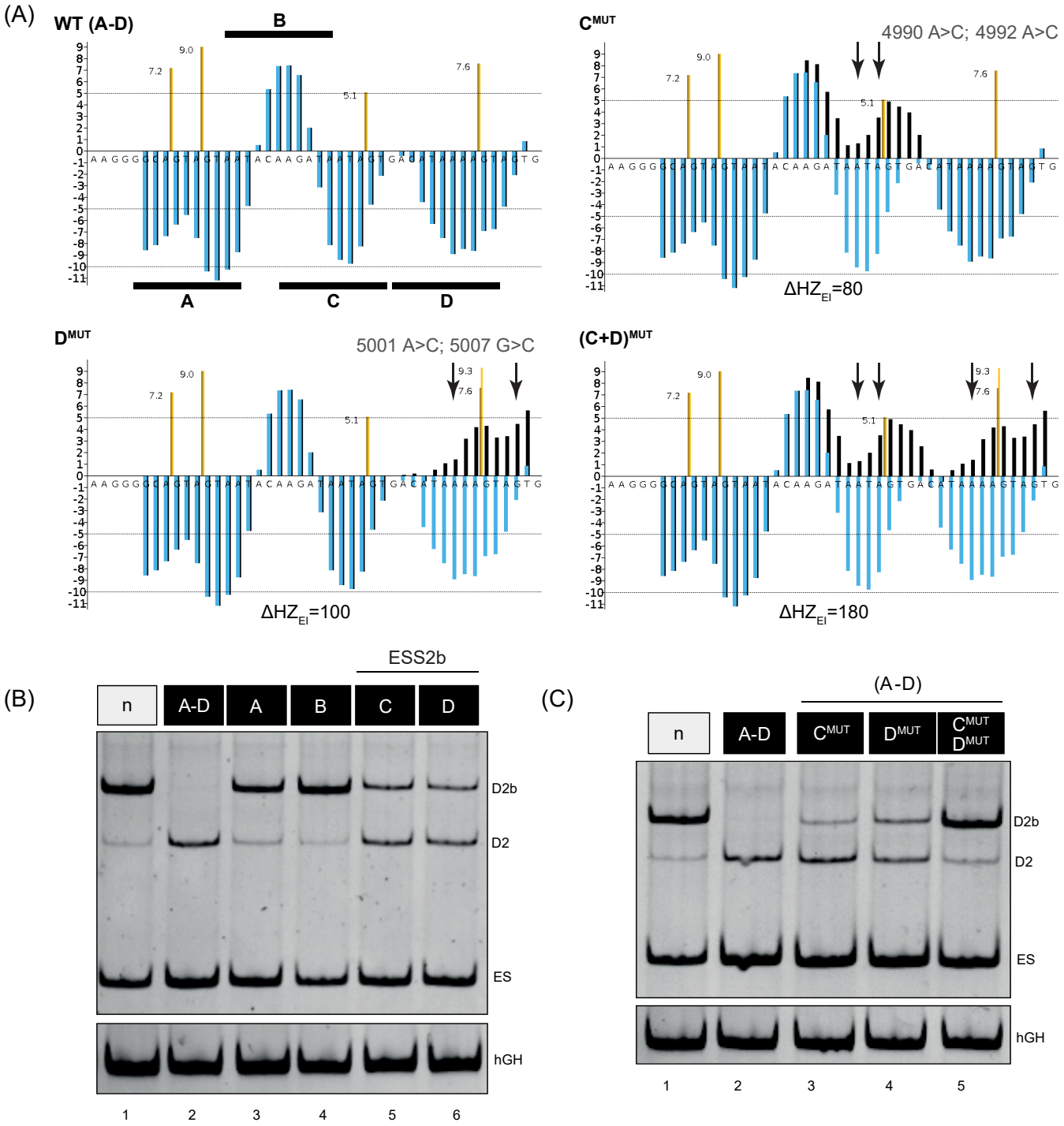


FIG 3 ESS2b, located between D2 and ESE2b, is bound by members of the hnRNP A/B family and counteracts ESE2b. (A) HEXplorer score profiles of segments A to D (black horizontal bars) and mutations of segment C, D, or both, composing ESS2b (black, mutated sequence; blue, wild-type reference; brown, HBond scores for WT; yellow, HBond scores for mutants). The arrows indicate mutated nucleotides. (B and C) Mutational analysis of ESS2b. HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5. Twenty-four hours after transfection, RNA was isolated from the cells and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH).

Taken together, the data show that multiple SREs within exon 2/2b balance splice donor selection in a strictly position-dependent manner.

ESS2b and ESE2b regulate balanced splice donor usage in provirus-transfected cells. To analyze the impacts of ESS2b and ESE2b on viral pre-mRNA splicing, we inserted both of the most promising inactivating mutations, ESE2b^{MUT} ($\Delta HZ_{EI} = -267$)

TABLE 2 Mass spectrometry analysis of ESS2b (average of duplicates)[a]

Gene name	No. of unique peptides	Log_2 difference
HNRNPDL	**6**	**2.11**
HNRNPA1; HNRNPA1L2	**21**	**1.34**
HNRNPA2B1	**20**	**1.13**
HNRNPA3	**17.5**	**1.11**
HNRNPA0	**9.5**	**0.94**
HNRNPUL2	**12.5**	**0.53**
HNRNPH3	**10.5**	**0.41**
HNRNPF	**16**	**0.25**
HNRNPH2	**9**	**0.23**
HNRNPH1	**6.5**	**0.19**
HNRNPR	**12**	**0.16**
HNRNPK	**27**	**0.14**
HNRNPL	**14**	**0.13**
HNRNPD	**6**	**0.12**
HNRNPU	**26**	**0.10**
HNRNPLL; HNRPLL	**17**	**0.05**
HNRNPUL1	**17**	**−0.01**
HNRNPC	**5**	**−0.08**
HNRNPM	**26**	**−0.13**
HNRNPH1; HNRNPH2	**3**	**−0.31**
SRSF2; SRSF8	2	0.38435
SRSF10	11.5	0.37666667
TRA2A	10.5	0.25796667
TRA2B	13.5	0.2346
SRSF1	20.5	0.1378
SRSF4	4.5	0.0756
SRSF9	6	0.05275
SRSF11	6	0.0258
SRSF6	4	0.0229
SRSF7	7	−0.06006667
SRSF3	4.5	−0.31765

[a]Unique peptides and log_2 differences of SR and hnRNPs (boldface) enriched after RNA affinity purification are shown; log_2 differences of normalized protein intensities were calculated as wild-type minus mutated sequence samples; log_2 differences significant at the 5% level (2 sided) are indicated by shading.

and ESS2b^{MUT} [(C+D)^{MUT}], either individually or in combination into pNL4-3 proviral plasmid DNA (GenBank accession no. M19921) (25), with and without the inactivating G_{I2}-1 mutation (17). RNA was isolated 48 h after transfection of HEK293T CD4$^+$ cells, subjected to Northern blot analysis, and detected with an exon 7 probe hybridizing to all viral mRNAs. Mutating ESE2b showed no shift in viral mRNA levels compared to the wild-type proviral clone (Fig. 4A, cf. lanes 1 and 2), whereas, in contrast, inactivating ESS2b caused a strong increase not only in 4-kb *vif* mRNAs, but also in 2-kb mRNAs, which was accompanied by a reduction in 9-kb mRNAs (Fig. 4A, cf. lanes 1 and 3). Interestingly, inactivating mutations of both SREs seem to nearly compensate for each other (Fig. 4A, cf. lanes 1 and 4), suggesting that though there seems to be no obvious effect of mutating ESE2b in viral mRNA distribution at first glance, the two SREs together critically regulate the balance of HIV-1 RNA classes. In agreement with our previous results (17), mutating G_{I2}-1 caused an increased amount of 2-kb and, particularly, of 4-kb *vif* mRNAs, which was comparable to inactivating ESS2b (Fig. 4A, cf. lanes 3 and 5). Inactivation of ESS2b and G_{I2}-1 resulted in an even stronger effect (Fig. 4A, cf. lanes 1, 5, and 7).

Next, to quantitatively measure individual HIV-1 transcript ratios, RT-PCRs were set up with different primer pairs, each normalized to the total amount of all viral mRNAs measured with primers detecting exon 7 (#3387/#3388) (Fig. 4B). Since exon 2 and exon 3 recognition underlie inverse regulation (26–29), we used exon junction primer pairs specifically detecting *vif* and *vpr* or [1.2.5] and [1.3.5] (exon numbers are indicated in square brackets) *nef* mRNAs as two distinct targets for exon 2 versus exon 3 inclusion in viral mRNAs (#3395/#3396 for *vif*, #3397/#3398 for *vpr*, #3395/#4843 for [1.2.5] *nef*, and #3397/#3636 for [1.3.5] *nef*) (Fig. 4B). As expected, inactivation of ESE2b showed no significant change in *vif*, *vpr*, and [1.2.5] and [1.3.5] *nef* mRNA levels (1-way analysis of

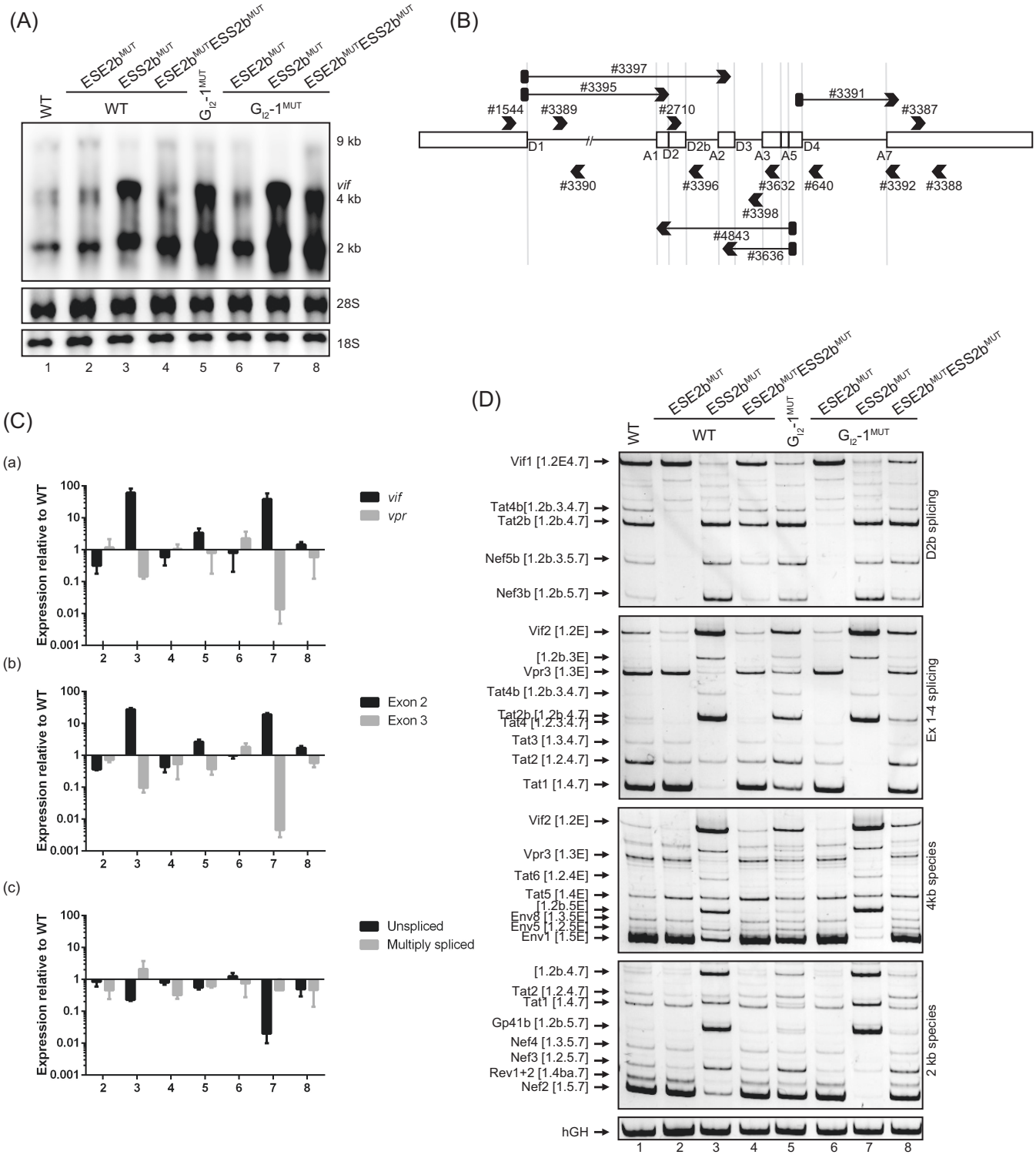


FIG 4 ESE2b and ESS2b cause alterations in proviral pre-mRNA processing. (A) Northern blot analysis of total RNA isolated from HEK293T CD4$^+$ cells transfected with wild-type or mutant pNL4-3. A hybridization probe specifically detecting HIV-1 exon 7 was used. (B) Binding sites of (q)RT-PCR primers. Open boxes indicate exons. Black arrowheads denote primers. Black arrows with filled black rectangles and arrowheads denote exon-exon junction primers. (C) qRT-PCR of total RNA isolated from the same RNA preparation as in panel A to specifically quantitate the levels of *vif* versus *vpr* (a) and [1.2.5] versus [1.3.5.] (b) and multiply spliced versus unspliced (c) mRNA species, displaying exp($-\Delta C_T$) ratios normalized to the wild-type splicing pattern. The bar graphs show means and standard deviations of three replicates. Primer pair #3387/#3388, specifically detecting exon 7, was used for normalization. The following primer pairs were used: *vif*, #3395/#3396; *vpr*, #3397/#3398; [1.2.5], #3395/#4843; [1.3.5], #3397/#3636; multiply spliced, #3391/#3392; and unspliced, #3389/#3390. (D) RT-PCR analysis of RNAs used in panels A and C. The following primer pairs were used: #1544/#3632 (Ex1-4 splicing), #2710/#3392 (D2b splicing), #1544/#3392 (2-kb species), and #1544/#640 (4-kb species). HIV-1 mRNA species according to Purcell and Martin (3) are indicated on the left of each gel image. Exon numbers are indicated in square brackets; those including an E read through D4.

variance [ANOVA] with Dunnett's *post hoc* test), whereas disruption of ESS2b induced a huge upregulation of *vif* and [1.2.5] *nef* ($P < 0.001$; Dunnett's *post hoc* test) and a reduction of *vpr* and [1.3.5] *nef* mRNAs (Fig. 4C, a and b, cf. bars 2 and 3). Inactivation of both SREs resulted in mRNA levels comparable to those of the wild type (Fig. 4C, a and b, bars 4). Likewise, inactivation of G_{I2}-1 led to comparable effects, with an overall higher level of *vif* mRNAs (Fig. 4C, a, bars 5 to 8, *vif*). Furthermore, we measured the levels of unspliced and multiply spliced mRNAs with both intact and inactivated G_{I2}-1 (#3389/#3390 for unspliced; #3391/#3392 for multiply spliced) (Fig. 4B). There was no significant difference from the wild type after disruption of ESE2b (Dunnett's *post hoc* test) but a clear decrease in unspliced mRNAs for inactivating ESS2b (Fig. 4C, c, cf. bars 2 and 3 and bars 6 and 7), which could again be compensated for by additionally mutating ESE2b (Fig. 4C, c, lanes 4 and 8).

To break down what impact the two mutations had on distinct mRNA species, we also performed semiquantitative RT-PCR. In line with minigene analyses and position-dependent effects, inactivating ESE2b revealed a complete loss of D2b use (Fig. 4D, D2b splicing, lane 2, e.g., Tat2b), whereas there was an elevated level of D2b use after inactivating ESS2b (Fig. 4D, D2b splicing, lane 3, e.g., Nef3b) and an upregulation of otherwise low-abundance mRNA species (Fig. 4D, 2-kb species, lane 3, e.g., Gp41b [1.2b.5.7]) (17). Moreover, inactivating ESE2b led to a slight decrease in exon-2-containing transcripts, like *vif2* or *tat2* (Fig. 4D, Ex1-4 splicing, lane 2). A mirror image inverted phenotype occurred after inactivation of ESS2b, where an increased degree of exon 2 inclusion could be observed (*vif2*), entailing a drop in exon 3 inclusion and *vpr* messages (*vpr3*), thereby sustaining their mutually regulated roles in HIV-1 splicing, as shown by quantitative PCR (qPCR) analysis in Fig. 4C (Fig. 4D, Ex1-4 splicing and 4-kb species, lane 3). Comparing overall 2-kb and 4-kb mRNA species in general, only marginal differences from wild-type pNL4-3 could be detected for ESE2b (Fig. 4D, 2 kb and 4 kb, cf. lanes 1 and 2), compatible with Northern blot analysis. As expected, for ESS2b, elevated levels of exon 2 inclusion with a concomitant reduction in mRNAs including exon 3 could be observed (Fig. 4D, 2 kb and 4 kb, cf. lanes 1 and 3). Again, for all detected mRNA species, a splicing pattern comparable to that of wild-type pNL4-3 was observed if both SREs had been mutated (Fig. 4D, cf. lanes 1 and 4). As shown before, inactivation of G_{I2}-1 resulted in enhanced exon 2b inclusion, followed by an increased amount of exon 2-containing transcripts, supporting the exon-bridging function of A1 with respect to D2 and D2b (Fig. 4D, cf. lanes 1 and 5). Additionally mutating ESE2b or ESS2b had no or only minor effects on the splicing patterns (Fig. 4D, cf. lanes 2 and 6 and lanes 3 and 7). In summary, RT-PCR analyses of RNA expressed from proviral clone pNL4-3 confirmed the results of the minigene analyses, revealing ESE2b and ESS2b as essential SREs regulating splice donor usage within exon 2/2b and thus *vif* mRNA processing.

ESE2b and ESS2b are essential for viral infectivity. To test to what extent changes in exon 2/2b inclusion reflect viral protein expression, we performed immunoblot analysis. No obviously different phenotype for the investigated proteins was observed after inactivating ESE2b (Fig. 5A, cf. lanes 1 and 2). In agreement with the data obtained from (q)RT-PCR analysis, a strong increase in the Vif protein level could be observed after inactivating ESS2b (Fig. 5A, cf. lanes 1 and 3). As expected, mutating both SREs brought the Vif protein level back to the wild-type pNL4-3 level (Fig. 5A, cf. lanes 1 and 4). Additionally interrupting G_{I2}-1 enhanced the effect of ESS2b and further increased Vif protein expression (Fig. 5A, lane 7). Moreover, a drop in intracellular p24 Gag levels, as well as in viral capsid within the supernatant, could be observed for the ESS2b mutant with intact or inactivated G_{I2}-1 (Fig. 5A, lanes 3 and 7). Furthermore, we used an antibody directed against the C-terminal domain of Gp41 (Chessie 8 [30]) to examine the presence of the previously described Gp41b isoform (17). In agreement with RT-PCR analyses, Gp41b protein was also enriched after ESS2b mutation (Fig. 5A, cf. lanes 3 and 7).

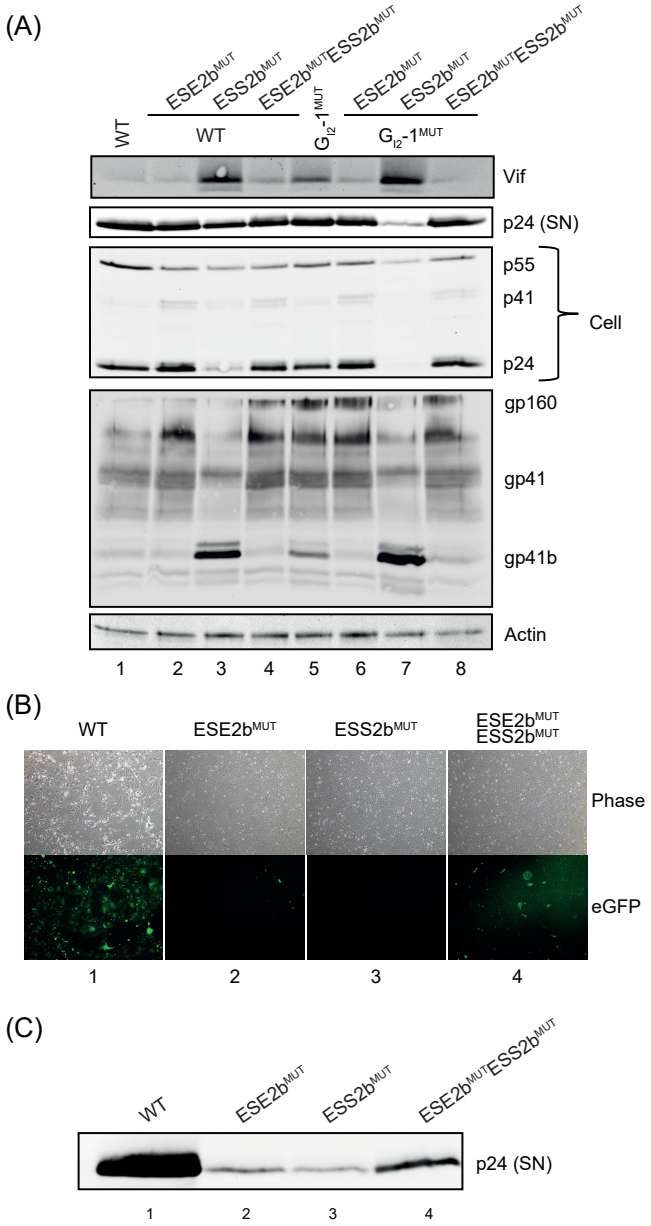


FIG 5 Impairment of proper viral particle production. (A) Immunoblot analysis of proteins of pelleted virions from the supernatants (SN) of transfected cells described in Fig. 4. (B and C) HEK293T cells (2.5×10^5) were transfected with pNL4-3 and mutant proviruses; 48 h posttransfection, the supernatant was collected for infection of GHOST CD4$^+$ cells, an indicator cell line that expresses eGFP after HIV-1 infection. Infection and viral replication were analyzed 48 h postinfection, both by fluorescence microscopy (B) and by p24-gag Western blot analysis (C) of supernatants of the infected GHOST CD4$^+$ cells.

Eventually, we tested whether viral particles within the supernatants harboring either individual or both mutations were still infectious. For this, we used GHOST cells that stably expressed the CD4 receptor and contained an LTR-dependent gene for eGFP. Thus, after successful infection and Tat-mediated transactivation of the LTR promoter, eGFP expression could be easily monitored via fluorescence microscopy. Forty-eight hours postinfection, strong eGFP expression was observed for wild-type pNL4-3, and it was clearly reduced in the ESE2b mutant-infected cells (Fig. 5B, cf. columns 1 and 2). Furthermore, infection with ESS2b mutant viral particles led to complete loss of eGFP expression, which was partially restored in cells infected with

viral particles harboring both mutations (Fig. 5B, cf. columns 3 and 4). p24 levels within the supernatant reflected the observed eGFP expression (Fig. 5C). In summary, the severely altered phenotype of the ESS2b mutant already observed during (q)RT-PCR and Northern and Western blot analyses led to a complete failure of infectiousness. Surprisingly, mutating ESE2b already showed a clear drop in eGFP expression, which was not indicated by the transfection experiments alone. Thus, an already slight imbalance in viral exon 2 splicing could lead to an impairment of proper viral particle production. In viral particles containing both mutations, balance could be restored, at least to some extent.

Masking of ESE2b and ESS2b restricts viral particle production. As was shown previously (26, 27), the use of LNAs can mimic the mutational analysis of SREs within the provirus. The modified antisense oligonucleotides are able to mask any specific sequence, in particular SREs, and thereby inhibit the binding of SR or hnRNPs. We used LNAs targeting either ESE2b or ESS2b and cotransfected them with pNL4-3 (Fig. 6A). Scrambled LNAs not targeting any viral sequence were used as a control. Forty-eight hours posttransfection, RNA and protein were isolated and analyzed for mRNA levels and protein expression. Northern blot analysis revealed distributions of viral mRNA classes when the two SREs were masked by LNAs similar to those obtained by SRE mutation (cf. Fig. 4A and 6B). Here, LNAs targeting ESE2b showed a slight reduction of 4-kb mRNAs, whereas LNAs targeting ESS2b showed a strong increase in 4-kb *vif* mRNA and a decrease in unspliced 9-kb mRNA (Fig. 6B). Furthermore, we examined the levels of both intracellular Gag protein and virus particles released into the supernatant (Fig. 6C). In agreement with the p24 levels detected after virus infection (Fig. 5C), we observed significantly less p24 Gag within both cells and supernatant for both LNAs. Additionally, RT-PCR analysis showed a dramatic loss of exon 2/2b inclusion for LNAs targeting ESE2b (Fig. 6D, e.g., *vif2* and *tat2b*, cf. lanes 1 and 3), followed by an increase in exon 3 inclusion (Fig. 6D, e.g., *vpr3*, cf. lanes 1 and 3). Conversely, splicing shifted toward exon 2 inclusion when LNAs against ESS2b were applied (Fig. 6D, e.g., *vif2* and *tat2*, cf. lanes 1 and 4), while exon 3 inclusion was reduced at the same time (Fig. 6D, e.g., *vpr3*, cf. lanes 1 and 4). Taken together, the data show that masking ESE2b or ESS2b with LNAs showed a phenotype very similar to that in infection experiments and was able to inhibit proper virus particle production.

In summary, thee data obtained in these experiments highlight the existence of a tight cluster of splicing regulatory elements within exon 2/2b that balances viral mRNA and protein production. Inhibiting protein binding to those elements disrupts viral particle production and infectivity.

Multiple SRE sequence variations between HIV-1 subtypes. Aligning the HIV-1 consensus sequences A1 to AE of HIV-1 exon 2/2b using the RIP 3.0 software (https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) showed that sequence variations between viral strains occurred strikingly more often within the regions containing the splicing regulatory elements ESS2b, ESE2b, and G_{I2}-1, while the flanking sequences were mainly conserved (Fig. 7A). The impact of these natural nucleotide variations on splice-enhancing properties was reflected in their HEXplorer profiles. Indeed, HIV strains showed a wide range of ΔHZ_{EI} scores. In order to examine one exemplary naturally occurring variation, we substituted in the minigene reporter the subtype K sequence exhibiting both high ΔHZ_{EI} and an additional deletion of 5 nucleotides within ESS2b. In fact, subtype K experimentally showed a splicing phenotype similar to that of A to D with a slight tendency toward D2 usage (Fig. 7B, left, cf. lanes 1 and 3). The HEXplorer profile of subtype K (Fig. 7B, right, black bars) showed only a minor effect on ESS2b compared to pNL4-3 (blue bars) and weaker ESE2b. Both effects tend to shift splice site selection further toward D2, which is barely visible, since D2 already dominates splicing in pNL4-3. The high SRE sequence variability between HIV-1 subtypes may suggest an equally wide range of splicing regulatory properties that permits adjusting Vif levels to A3G levels in a variety of cellular host environments.

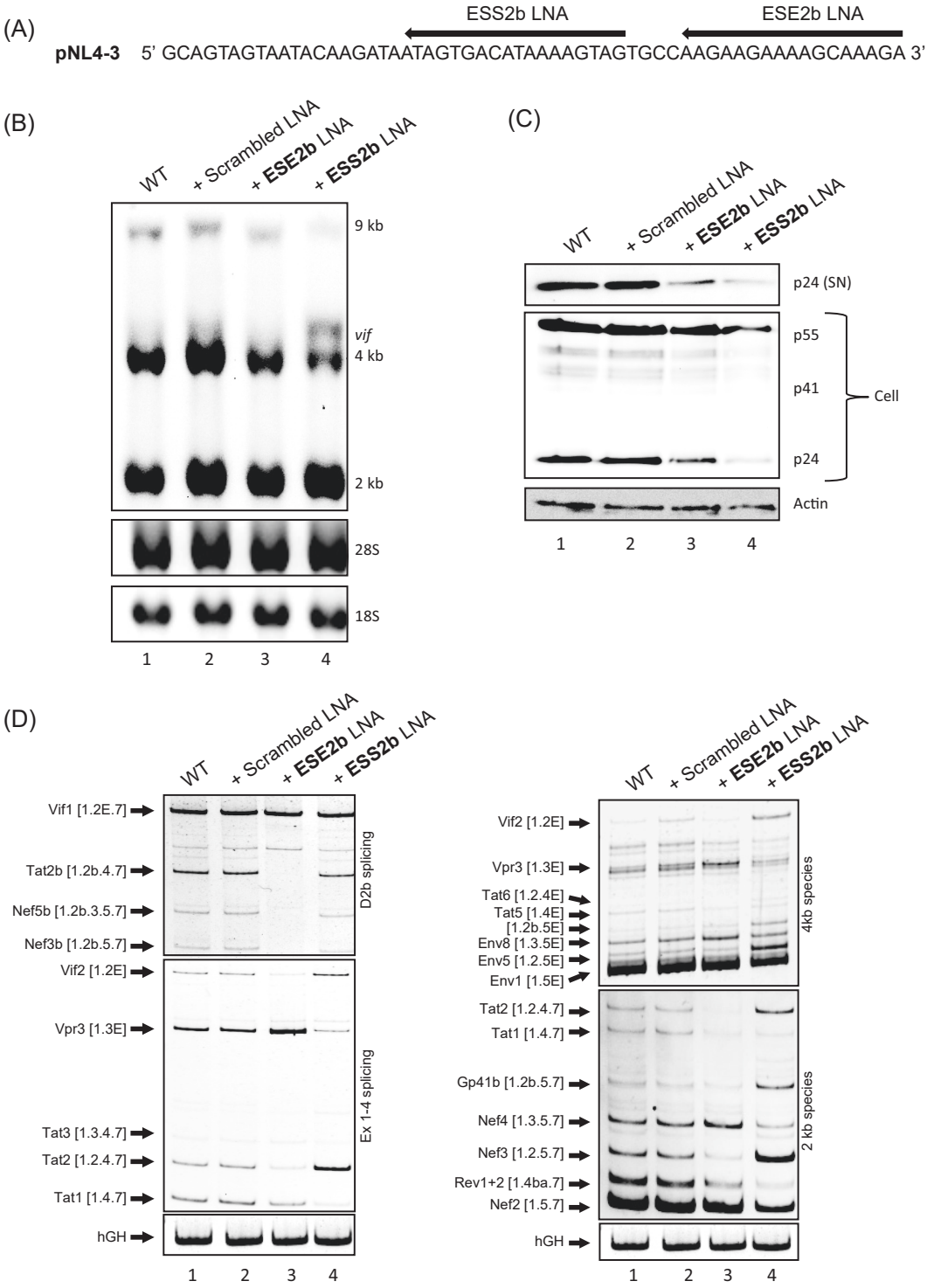


FIG 6 LNA-directed masking of ESE2b and ESS2b mimics the mutational phenotype. (A) Schematic of LNA binding sites. (B) Northern blot analysis of total RNA. HeLa cells were cotransfected with pNL4-3 and either LNAs masking ESE2b or ESS2b or the scrambled LNA. Total RNA was isolated 24 h posttransfection and subjected to Northern blot analysis using an HIV-1 exon 7 probe. (C) Western blot analysis of cellular (Cell) and supernatant (SN) Gag of cotransfected cells from panel B. (D) RT-PCR analysis of different viral mRNA species. The following primer pairs were used: #1544/#3632 (Ex1-4 splicing), #2710/#3392 (D2b splicing), #1544/#3392 (2-kb species), and #1544/#640 (4-kb species). HIV-1 mRNA species are indicated on the left of each gel image according to Purcell and Martin (3). Exon numbers are indicated in square brackets; those including an E read through D4.

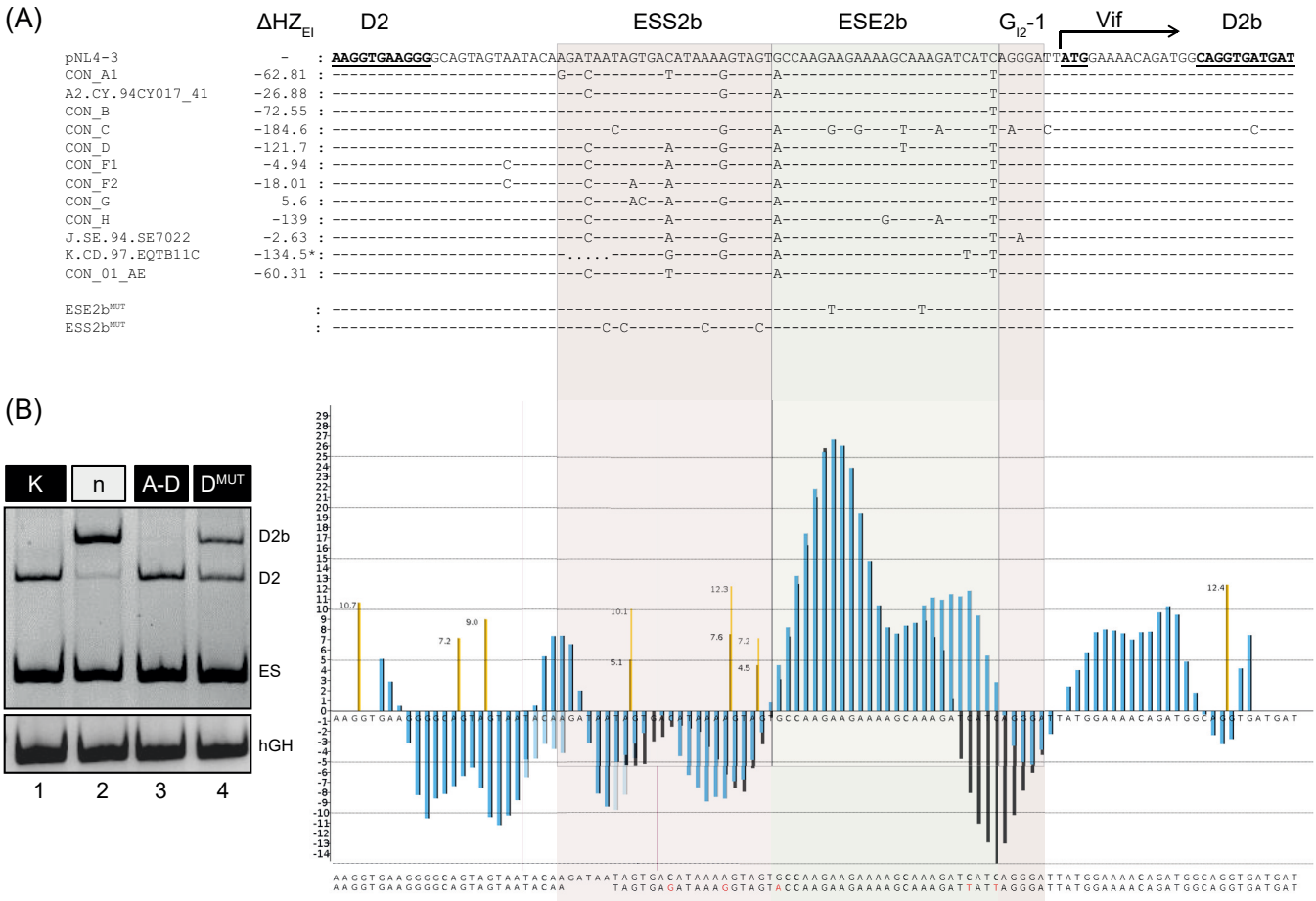


FIG 7 Analysis of SREs within exon 2/2b of different HIV-1 subtypes. (A) pNL4-3-derived HIV-1 exon 2/2b consensus sequences from A1 to AE of the different HIV-1 subtypes, together with their HEXplorer score differences (ΔHZ_{EI}). Conserved sequences are represented by dashes and differences by letters. Regions with SREs are shown with red or green shading. The subtype sequences were analyzed with the RIP 3.0 software (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html). (B) (Left) Splicing patterns of the splicing reporter carrying SRE regions of subtype K (lane 1) and pNL4-3 (lane 3). For reference, lanes 2 and 4, corresponding to the neutral sequence and to D^{MUT}, are also shown. HeLa cells were transiently transfected with 1 μg of each construct and 1 μg of pXGH5; 24 h after transfection, RNA was isolated from the cells and subjected to RT-PCR analysis using primer pairs #2648/#2649 and #1224/#1225 (hGH). (Right) HEXplorer profiles of pNL4-3 and exemplary subtype K containing a 5-nt deletion (between vertical red lines) and several single-nucleotide variations. Blue, HEXplorer profile for pNL4-3; black, subtype K; brown, HBond scores for WT; yellow, HBond scores for mutants.

DISCUSSION

The data presented in this work show that a splicing regulatory network (Fig. 8) regulates HIV-1 exon 2/2b inclusion in viral mRNAs, thus optimizing viral replication via competing actions of several SREs located close to D2 and D2b. In particular, we identified the Tra2/SRSF10-binding site ESE2b and the hnRNP A/B-binding site ESS2b, which could be specifically masked by LNAs. Both SREs contribute to regulating splice donors D2/D2b and splice acceptor A1, as well as *vif* mRNA and protein production.

During alternative splicing, recognition of splice sites is most often facilitated not only by conserved sequence elements, like the splice donor and splice acceptor, but also by RNA secondary structure (2, 31–33) and a multitude of splicing regulatory elements. While splicing patterns of various HIV-1 subtypes are mostly conserved, the frequency of splice site usage can depend on the temperature (2) and the presence of splicing regulatory proteins (14).

Within the noncoding exon 2/2b, six different SREs have already been described. Three elements exist that enhance recognition of splice donor D2 and thereby inclusion of exon 2 in viral mRNAs: ESE-M1 and -M2 (bound by SRSF1) (15) and ESE-Vif (bound by SRSF4) (16). Furthermore, an inhibitory GGGG motif, overlapping the already intrinsically weak D2, inhibits its use and exon 2 inclusion (16), potentially through sterical hindrance of the U1 snRNP. We have previously reported that a G run located down-

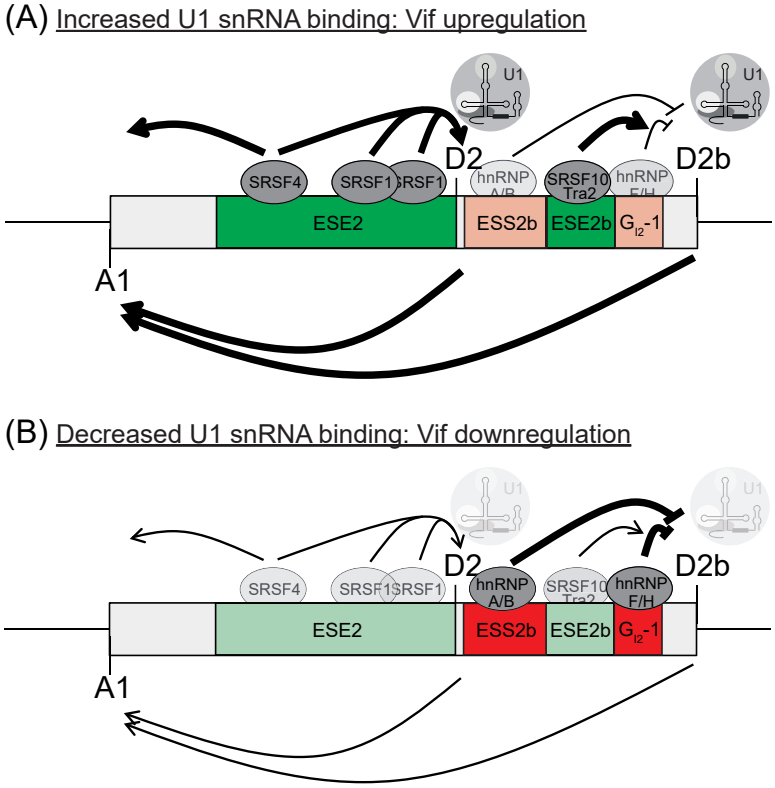


FIG 8 Model for exon 2/2b recognition. Exon 2/2b inclusion and splice donor usage are regulated by a complex network of SREs. (A) SR proteins binding to both ESE2 and ESE2b support U1 snRNP binding at the downstream splice donors D2 and D2b. Exon definition leads to the concomitant upregulation of splice acceptor A1 and to higher *vif* mRNA expression (left-pointing arrows below exon 2/2b). (B) Lower levels of SR proteins, as well as hnRNP binding to sites ESS2b and G_{I2}-1, reduce U1 snRNP binding to D2 and D2b.

stream of exon 2 inhibits the splice donor D2b lying further downstream by binding of hnRNP F/H (G_{I2}-1) (17). Inactivation of this G_{I2}-1 motif led to a strong increase in the use of the otherwise only little used donor D2b. This was attributed to the fact that binding of hnRNP F/H leads to the formation of a "dead-end" complex, meaning that the U1 snRNP binds to the splice donor without actually splicing at this position (12, 34, 35).

Upregulation of D2b use following G_{I2}-1 inactivation indicated that an SR protein binding site could be located within exon 2b. We had previously found an enhancing element located downstream of D2 in a HEXplorer-based screen of total HIV-1 mRNA (13). Continuing analysis of this element here showed that the enhancer ESE2b strongly activates D2b and simultaneously inhibits D2, which is facilitated by binding of SRSF10 and Tra2. Tra2β was previously shown to bind to GA-rich sequence elements (36–39), similar to the sequence of ESE2b. Cloning this element into the minigene indeed led to an excessive splicing phenotype at D2b, which, however, was not observed in a physiological HIV-1 splicing context. During infection, we could confirm by RNA deep sequencing that D2b is only marginally used (0.2%) compared to D2 (5.3%) (17). Here, we resolve this apparent discrepancy between splicing patterns of minigene and infection experiments by identifying a novel SRE located within exon 2b, ESS2b, which counteracts the strong ESE2b effects. By using MS analysis, we show that ESS2b is bound by members of the hnRNP A/B family, which fits earlier studies showing that those proteins bind to sequences that include a TAG motif (40, 41).

It might be surprising that such a multitude of SREs should regulate splice donor selection in a noncoding exon. However, in order to obtain Vif, splice acceptor A1 must be used, and A1 itself seems to require activation by an exon definition complex (42,

43) in which U1 snRNP binding to either D2 or D2b promotes the recognition of the splice acceptor A1, located upstream, by U2 snRNPs. On the other hand, splicing at D2 or D2b prevents Vif expression, which relies on intron 2 retention. This is similar to *env* mRNA processing, where U1 snRNP binding to a splicing-incompetent D4 was needed for splice acceptor A5 activation (44). Thus, the commonly observed larger amounts of both intron-retaining—leading to Vif expression—and exon 2-including mRNAs could be due to increased U1 snRNA complementarity or mutations of neighboring SREs (16, 17, 29, 45).

Only balanced levels of Vif expression contribute to maximal viral replication, while excessive Vif expression is detrimental to viral replication due to perturbation of proteolytic Gag processing (19). On the other hand, excessive splicing at D2 leads to a decrease of unspliced mRNAs and, consequently, a reduction of Gag/Gag-Pol expression levels and a defect in virion production. This effect has also been termed "oversplicing" and is in line with our observation revealing that excessive Vif expression after mutating or masking of ESS2b leads to a reduction of overall unspliced mRNAs and impairment of cellular Gag and viral particles within the supernatant. However, not only excessive Vif levels, but also insufficient amounts are deleterious to viral replication. Vif is essential for counteracting the host cell restriction factor A3G, and an imbalance in the Vif/A3G ratio strongly affects viral replication. It was shown that if restriction pressure is low, lower Vif levels are sufficient to counteract A3G, whereas excessive Vif impedes viral replication ability (19, 46). However, conversely, HIV-1 replicates only in cells with high restriction pressure if sufficient Vif is present (17, 46).

Nomaguchi et al. identified natural single-nucleotide variations within different HIV-1 isolates proximal to HIV-1 A1 (SA1prox) that could be shown to regulate *vif* mRNA and Vif protein expression and were linked to the fact that effective viral replication critically depends on an optimal Vif/A3G ratio (17, 46). Here, we found nucleotide variations predominantly within splicing regulatory elements in exon 2/2b.

Therefore, it is tempting to speculate that the vast number of SREs within exon 2/2b ensures viral replication in cells with different A3G or splicing regulatory protein concentrations, e.g., by a mechanism like mutual evolution (46).

MATERIALS AND METHODS

Single-intron splicing constructs. All eGFP single-intron splicing reporters are based on the well-established HIV-1 glycoprotein/eGFP expression plasmid (20). Insertion of exon 2b parts I to IV was carried out by replacing GAR E42 of SV GAR E42 SD4 Δvpu env eGFP D36G (22) with a PCR product obtained with primer pairs #4200/#4201 (part I), #4202/#4203 (part II), #4204/#4205 (part III), and #4206/#4207 (part IV), respectively. The neutral sequence (23) was inserted 3.5 times as described above with primer pair #4213/#4214.

Three-exon minigenes. The three-exon minigenes are derived from the fibrinogen Bβ minigene pT-Bβ-IVS7 + 1G>T (47, 48). The middle exon was replaced with only splicing-neutral sequences (23) by using a customized synthetic gene from Invitrogen and inserted into pT-Bβ-IVS7 + 1G>T via EcoNI/Bpu10I. HIV-1-derived splice donors D2 and D2b were inserted with PCR products resulting from primer pair #4793/#4794. ESE-Vif, -M1, and -M2 were inserted by PCR with primer pair #4853/#2620. Fragments of HIV-1 exon 2/2b were added at their authentic positions relative to D2 or D2b, respectively, by using primer pairs #4795/#2620 (ESE2b and G_{I2}-1), #4798/#2620 (ΔHZ_{EI}-94 and G_{I2}-1), #5318/#2620 (ΔHZ_{EI}-267 and G_{I2}-1), #4796/#2620 (ESE2b and G_{I2}-1^{MUT}), #5319/#2620 (ΔHZ_{EI}-94 and G_{I2}-1^{MUT}), #5317/#2620 (ΔHZ_{EI}-267 and G_{I2}-1^{MUT}), #5251/#2620 (ESS2b WT [A to D]), #5337/#2620 (ESS2b part A), #5339/#2620 (ESS2b part B), #5341/#2620 (ESS2b part C), and #5343/#2620 (ESS2b part D). The fragment of HIV-1 subtype K was added at its authentic position flanked by D2 and D2b using primer pair #5712/#5713. HEXplorer-guided mutations of ESS2b were inserted via PCR products resulting from primer pairs #5392/#2620 (C^{MUT}), #5393/#2620 (D^{MUT}), and #5394/#2620 [$(C+D)^{MUT}$].

Proviral plasmids. pNL4-3 $ESE2b^{MUT}$ proviral DNA was generated by the overlapping-PCR technique using primers #5549/#4773 and #5553/#5550, pNL4-3 $ESS2b^{MUT}$ using primer pairs #5547/#4773 and #5553/#5548, and pNL4-3 $ESE2b^{MUT}$ $ESS2b^{MUT}$ using primer pairs #5551/#4773 and #5553/#5552. pNL4-3 G_{I2}-1^{MUT} has been described previously (17) and was used as a template instead of pNL4-3, using the primer pairs described above to generate double or triple mutations.

Expression plasmids. pXGH5 (49) was cotransfected to monitor transfection efficiency. pCL-dTOM was cotransfected to detect the transfection efficiency of each sample in fluorescence microscopy analysis. The plasmid expresses the fluorescent protein Tomato and was kindly provided by H. Hanenberg.

Oligonucleotides. All the oligonucleotides used were obtained from Metabion GmbH (Planegg, Germany) (Table 3). RNase-free high-performance liquid chromatography (HPLC)-purified LNAs were purchased from Exiqon (Denmark).

Cell culture and transfection. HeLa, HEK293T ($CD4^+$), or GHOST (3) $CXCR4^+$ cells (50) were cultured in Dulbecco's high-glucose modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum and 50 μg/ml penicillin-streptomycin (Invitrogen). For transient transfection, 2×10^5 cells per six-well plate were used. Transient-transfection experiments were performed using TransIT-LT1 transfection reagent (Mirus Bio LLC) according to the manufacturer's instructions. LNA transfection was performed as described previously 26.

RNA isolation and RT-PCR. Twenty-four or 48 h posttransfection, total RNA was isolated by using acid guanidinium thiocyanate-phenol-chloroform (51). For semiquantitative and quantitative RT-PCR analyses, RNA was reverse transcribed by using Superscript III reverse transcriptase (Invitrogen) and oligo(dT) primers (Invitrogen) and amplified using the primer pairs depicted in Fig. 4B.

Northern blotting. Three micrograms of total RNA isolated by using acid guanidinium thiocyanate-phenol-chloroform (51) was separated on a denaturing 1% agarose gel and then capillary blotted onto a positively charged nylon membrane. Hybridization was carried out using a digoxigenin (DIG)-labeled HIV-1 exon 7 PCR amplicon (#3387/#3388) as previously described (17).

Protein isolation and Western blotting. Protein samples were heated to 95°C for 10 min and loaded onto SDS-PAGE gels for Western blot analysis. The samples were transferred to a nitrocellulose membrane; probed with primary and secondary antibodies (sheep antibody against HIV-1 p24 CA; Aalto), mouse monoclonal antibody specific for HIV-1 Vif (ab66643; Abcam), mouse anti-gp41 (Chessie 8 [30]), and mouse anti-β-actin monoclonal antibody (A5316; Sigma-Aldrich); and developed with ECL chemiluminescence reagent (GE Healthcare).

RNA affinity purification assay. RNA oligonucleotides (3,000 pmol) for either a WT or mutant version of ESE2b and ESS2b, respectively, were covalently coupled to adipic acid dihydrazide agarose beads (Sigma). A 60% HeLa cell nuclear extract (Cilbiotech) was added to the immobilized RNAs. After five stringent washing steps with buffer D containing different concentrations of KCl (0.1, 0.25, 0.5, 0.25, and 0.1 M KCl, together with 20 mM HEPES-KOH [pH 7.9], 5% [vol/vol] glycerol, 0.2 M EDTA, 0.5 mM dithiothreitol, 0.4 M $MgCl_2$), the precipitated proteins were eluted in protein sample buffer. Samples were sent to the Molecular Proteomics Laboratory, BMFZ, Heinrich Heine University, Düsseldorf, Germany, for MS analysis as described in detail previously (24).

HBond score. The HBS measures splice donor strength by its complementarity to U1 snRNA, combining experimental evidence with a computational hydrogen bond weight model.

The HBond score algorithm models hydrogen bond formation at individual positions, as well as nucleotide interdependence beyond nearest-neighbor relationships. It also takes positions +7 and +8 fully into account, as experiments have confirmed U1 snRNA duplex dependency on these nucleotides. The hydrogen bond pattern between a splice donor site and all 11 nt of the free 5′ end of U1 snRNA is translated into a numerical HBond score in the range 1.8 to 23.8, with CAG/GTAAGTAT corresponding to an HBS of 23.8. The HBond score is available through the Web interface (http://www.uni-duesseldorf.de/rna/html/hbond_score.php).

HEXplorer score. In a RESCUE-type approach, the HZ_{EI} is based on different hexamer occurrences in exonic and intronic sequences in the neighborhood of splice donors, and it has been successfully used as a basis for the identification of exonic splicing regulatory elements (13, 24).

Different hexamer frequencies up- and downstream of splice donors are first translated into Z-scores for all 4,096 hexamers. For any index nucleotide in a genomic sequence, its HZ_{EI} is then calculated as the average hexamer Z-score of all six hexamers overlapping with this index nucleotide. This algorithm permits the plotting of HEXplorer score profiles along genomic sequences, and they reflect splice-enhancing or -silencing properties in the neighborhood of a splice donor: HEXplorer score positive regions support downstream splice donors and repress upstream ones, and HZ_{EI} negative regions do the opposite. HEXplorer score profiles of wild-type and mutant sequences were calculated using the Web interface (https://www2.hhu.de/rna/html/hexplorer_score.php).

qPCR statistics. In qPCR experiments, expression levels relative to the WT were calculated as $\exp(-\Delta C_T)$ (threshold cycle) ratios. The bar graphs show means and standard deviations of three replicates. Statistical significance was determined separately for each sample (*vif*, *vpr*, exon 2, exon 3, unspliced, and multiply spliced) by 1-way ANOVA followed by Dunnett's *post hoc* test correcting for multiple comparisons.

ACKNOWLEDGMENTS

We thank Björn Wefers for excellent technical assistance. The following reagents were obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: Chessie 8 from George Lewis and GHOST (3) $CXCR4^+$ cells from Vineet N. KewalRamani and Dan R. Littman.

Funding was provided by the Deutsche Forschungsgemeinschaft (DFG) (SCHA 909/8-1) and Jürgen Manchot Stiftung (to A.-L.B., L.W., and H.S.) and by Stiftung für AIDS-Forschung, Düsseldorf (to H.S.).

A.-L.B. and H.S. conceived the study and designed the experiments. A.-L.B., L.W., L.M., and M.W. performed cloning, transfection experiments, and (q)RT-PCR analyses. A.-L.B. performed RNA pulldown analyses. F.H. performed LNA-related experiments. S.T.,

TABLE 3 Sequences of primers used for cloning and (q)RT-PCR analyses and of LNAs

Primer or LNA	Sequence
Primers	
Cloning	
#2620	GATCCCGGGAAAGATTTGTTGTCACATACAGAAG
#4200	AATTCGCAGTAGTAATACAAGATAATAGTGACATAGAGCT
#4201	CTATGTCACTATTATCTTGTATTACTACTGCG
#4202	AATTCGATAATAGTGACATAAAAGTAGTGCCAAGAGAGCT
#4203	CTCTTGGCACTACTTTTATGTCACTATTATCG
#4204	AATTCTAGTGCCAAGAAGAAAAGCAAAGATCATCAGAGCT
#4205	CTGATGATCTTTGCTTTTCTTCTTGGCACTAG
#4206	AATTCTCATCAGGGATTATGGAAAACAGATGGCAGGAGCT
#4207	CCTGCCATCTGTTTTCCATAATCCCTGATGAG
#4213	AATTCCCAAACAACCAAACAACCAAACAACCAAACGAGCT
#4214	CGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGG
#4773	TGGATGCTTCCAGGGCTC
#4793	AACAAACCGGTAAGGTGAAGGGTCTAGACCAAACAACCAAACAAC
#4794	AACAGCGTACGTTGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGTTGTTTGGGTCGACATCATCACCTGGCGGCCGCTTGTTTG
#4795	AAGGGGCTAGCCCAAGAAGAAAAGCAAAGATCATCAGGGATTATGGAAAACAGGCGGCCGCCAGGT
#4796	AAGGGGCTAGCCCAAGAAGAAAAGCAAAGATCATCCGCGATTATGGAAAACAGGCGGCCGCCAGGT
#4798	AAGGGGCTAGCCCAAGATGAAAAGCAATGATCATCAGGGATTATGGAAAACAGGCGGCCGCCAGGT
#4853	AACAACCTTAGGGGACAGCAGAGATCCAGTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGGGACCCAAGGTGAAG
#5251	AAGGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAGATCATCA
#5317	AAGGGGCTAGCCCAAGTAGAAAAGCATAGATCATCCGCGATTATGGAAAACAGGCGGC
#5318	AAGGGGCTAGCCCAAGTAGAAAAGCATAGATCATCAGGGATTATGGAAAACA
#5319	AAGGGGCTAGCCCAAGATGAAAAGCAAGATCATCCGCGATTATGGAAAACAGGCGG
#5337	AAGGGGCTAGCGCAGTAGTAATACACCAAACAACCAAACAACCAAACAACCAAGAAGAAAAGCAAAGATCATCA
#5339	AAGGGGCTAGCCAACCAAACAAAATACAAGATAACCAAACAACCAAACAACCAAGAAGAAAAGCAAAGATCATCA
#5341	AAGGGGCTAGCCCAAACAACCAAACAAAGATAATAGTGACCCAAACAACCAAGAAGAAAAGCAAAGATCATCA
#5343	AAGGGGCTAGCCCAAACAACCAAACAACCAAACAAGTGACATAAAAGTAGTGCCAAGAA
#5392	AAGGGGCTAGCGCAGTAGTAATACAAGATACTCGTGACATAAAAGTAGTGCCAAGAA
#5393	AAGGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATACAAGTACTGCCAAGAAGAAAAGCAAAGATCAT
#5547	CTCGTGACATACAAGTACTGCCAAGAAGAAAAGCAAAGATCAT
#5548	GTACTTGTATGTCACGAGTATCTTGTATTACTACTGCCCCTT
#5549	TAGAAAAGCATAGATCATCAGGGATTATGGAAAAC
#5550	ATGCTTTTCTACTTGGCACTACTTTTATGTCACT
#5551	CTCGTGACATACAAGTACTGCCAAGTAGAAAAGCATAGATCATCAGGGATTATGGAAAAC
#5552	ATGCTTTTCTACTTGGCAGTACTTGTATGTCACGAGTATCTTGTATTACTACTGCCCCTT
#5553	CTGGCAGAAAACAGGGAGATT
#5712	GGGCTAGCGCAGTAGTAATACAATAGTGAGATAAAGGTAGTACCAAGAAGAAAAGCAAAGAT
#5713	CCTGGCGGCCGCCCATCTGTTTTCCATAATCCCTAATAATCTTTGCTTTTCTTCTTGG
(q)RT-PCR	
#640	CAATACTACTTCTTGTGGGTTGG
#1224	TCTTCCAGCCTCCCATCAGCGTTTGG
#1225	CAACAGAAATCCAACCTAGAGCTGCT
#1544	CTTGAAAGCGAAAGTAAAGC
#2648	AGTGATTCAGAACCGTCAAG
#2649	TCCACCACCGTCTTCTTTAG
#2710	GGGGGGATCGATAATTAAGGAGTTTATATGGAAACCCTTAAAGGTAAAGGGGCAGTAGTAATACAA
#3387	TTGCTCAATGCCACAGCCAT
#3388	TTTGACCACTTGCCACCCAT
#3389	TTCTTCAGAGCAGACCAGAGC
#3390	GCTGCCAAAGAGTGATCTGA
#3391	TCTATCAAAGCAACCCACCTC
#3392	CGTCCCAGATAAGTGCTAAGG
#3395	GGCGACTGGGACAGCA
#3396	CCTGTCTACTTGCCACAC
#3397	CGGCGACTGAATCTGCTAT
#3398	CCTAACACTAGGCAAAGGTG
#3632	TGGATGCTTCCAGGGCTC
#3636	CCGCTTCTTCCTTGTTATGTC
#4843	CCGCTTCTTCCTTTCCAGAGG
LNAs	
Scrambled	GACGCGTCCTTACGCG
ESE2b	TCTTTGCTTTTCTTCT
ESS2b	CTACTTTTATGTCACTAT

A.-L.B., and H.S. performed HEXplorer analyses. S.T. provided statistical analyses. A.-L.B., S.T., and H.S. wrote the manuscript.

REFERENCES

1. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res 40:10345–10355. https://doi.org/10.1093/nar/gks753.
2. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 splicing by using next-generation sequencing. J Virol 91:e02515–16. https://doi.org/10.1128/JVI.02515-16.
3. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365–6378.
4. Kim SY, Byrn R, Groopman J, Baltimore D. 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J Virol 63:3708–3713.
5. Klotman ME, Kim S, Buchbinder A, DeRossi A, Baltimore D, Wong Staal F. 1991. Kinetics of expression of multiply spliced RNA in early human immunodeficiency virus type 1 infection of lymphocytes and monocytes. Proc Natl Acad Sci U S A 88:5011–5015. (Erratum, 89:1148, 1992.) https://doi.org/10.1073/pnas.89.3.1148b.
6. Mohammadi P, Desfarges S, Bartha I, Joos B, Zangger N, Munoz M, Gunthard HF, Beerenwinkel N, Telenti A, Ciuffi A. 2013. 24 hours in the life of HIV-1 in a T cell line. PLoS Pathog 9:e1003161. https://doi.org/10.1371/journal.ppat.1003161.
7. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2:a006916. https://doi.org/10.1101/cshperspect.a006916.
8. Will CL, Luhrmann R. 2011. Spliceosome structure and function. Cold Spring Harb Perspect Biol 3:a003707. https://doi.org/10.1101/cshperspect.a003707.
9. Berget SM. 1995. Exon recognition in vertebrate splicing. J Biol Chem 270:2411–2414. https://doi.org/10.1074/jbc.270.6.2411.
10. Romfo CM, Alvarez CJ, van Heeckeren WJ, Webb CJ, Wise JA. 2000. Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol Cell Biol 20:7955–7970. https://doi.org/10.1128/MCB.20.21.7955-7970.2000.
11. Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 102:16176–16181. https://doi.org/10.1073/pnas.0508489102.
12. Erkelenz S, Mueller WF, Evans MS, Busch A, Schoneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA 19:96–102. https://doi.org/10.1261/rna.037044.112.
13. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res 42:10681–10697. https://doi.org/10.1093/nar/gku736.
14. Stoltzfus CM. 2009. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res 74:1–40. https://doi.org/10.1016/S0065-3527(09)74001-1.
15. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. 2006. The strength of the HIV-1 3′ splice sites affects Rev function. Retrovirology 3:89. https://doi.org/10.1186/1742-4690-3-89.
16. Exline CM, Feng Z, Stoltzfus CM. 2008. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. J Virol 82:3921–3931. https://doi.org/10.1128/JVI.01558-07.
17. Widera M, Erkelenz S, Hillebrand F, Krikoni A, Widera D, Kaisers W, Deenen R, Gombert M, Dellen R, Pfeiffer T, Kaltschmidt B, Munk C, Bosch V, Kohrer K, Schaal H. 2013. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. J Virol 87:2707–2720. https://doi.org/10.1128/JVI.02755-12.
18. Sheehy AM, Gaddis NC, Choi JD, Malim MH. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646–650. https://doi.org/10.1038/nature00939.
19. Akari H, Fujita M, Kao S, Khan MA, Shehu-Xhilaga M, Adachi A, Strebel K. 2004. High level expression of human immunodeficiency virus type-1 Vif inhibits viral infectivity by modulating proteolytic processing of the Gag precursor at the p2/nucleocapsid processing site. J Biol Chem 279:12355–12362. https://doi.org/10.1074/jbc.M312426200.
20. Kammler S, Leurs C, Freund M, Krummheuer J, Seidel K, Tange TO, Lund MK, Kjems J, Scheid A, Schaal H. 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. https://doi.org/10.1017/S1355838201001212.
21. Freund M, Asang C, Kammler S, Konermann C, Krummheuer J, Hipp M, Meyer I, Gierling W, Theiss S, Preuss T, Schindler D, Kjems J, Schaal H. 2003. A novel approach to describe a U1 snRNA binding site. Nucleic Acids Res 31:6963–6975. https://doi.org/10.1093/nar/gkg901.
22. Caputi M, Freund M, Kammler S, Asang C, Schaal H. 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. https://doi.org/10.1128/JVI.78.12.6517-6526.2004.
23. Zhang XH, Arias MA, Ke S, Chasin LA. 2009. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. RNA 15:367–376. https://doi.org/10.1261/rna.1498509.
24. Brillen AL, Schoneweis K, Walotka L, Hartmann L, Muller L, Ptok J, Kaisers W, Poschmann G, Stuhler K, Buratti E, Theiss S, Schaal H. 2017. Succession of splicing regulatory elements determines cryptic 5′ss functionality. Nucleic Acids Res 45:4202–4216. https://doi.org/10.1093/nar/gkw1317.
25. Adachi A, Gendelman HE, Koenig S, Folks T, Willey R, Rabson A, Martin MA. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol 59:284–291.
26. Widera M, Hillebrand F, Erkelenz S, Vasudevan AA, Munk C, Schaal H. 2014. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. Retrovirology 11:72. https://doi.org/10.1186/s12977-014-0072-1.
27. Erkelenz S, Hillebrand F, Widera M, Theiss S, Fayyaz A, Degrandi D, Pfeffer K, Schaal H. 2015. Balanced splicing at the Tat-specific HIV-1 3′ss A3 is critical for HIV-1 replication. Retrovirology 12:29. https://doi.org/10.1186/s12977-015-0154-8.
28. Madsen JM, Stoltzfus CM. 2005. An exonic splicing silencer downstream of the 3′ splice site A2 is required for efficient human immunodeficiency virus type 1 replication. J Virol 79:10478–10486. https://doi.org/10.1128/JVI.79.16.10478-10486.2005.
29. Mandal D, Feng Z, Stoltzfus CM. 2010. Excessive RNA splicing and inhibition of HIV-1 replication induced by modified U1 small nuclear RNAs. J Virol 84:12790–12800. https://doi.org/10.1128/JVI.01257-10.
30. Abacioglu YH, Fouts TR, Laman JD, Claassen E, Pincus SH, Moore JP, Roby CA, Kamin Lewis R, Lewis GK. 1994. Epitope mapping and topology of baculovirus-expressed HIV-1 gp160 determined with a panel of murine monoclonal antibodies. AIDS Res Hum Retroviruses 10:371–381. https://doi.org/10.1089/aid.1994.10.371.
31. Abbink TE, Berkhout B. 2008. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. J Virol 82:3090–3098. https://doi.org/10.1128/JVI.01479-07.
32. Zychlinski D, Erkelenz S, Melhorn V, Baum C, Schaal H, Bohne J. 2009. Limited complementarity between U1 snRNA and a retroviral 5′ splice site permits its attenuation via RNA secondary structure. Nucleic Acids Res 37:7429–7440. https://doi.org/10.1093/nar/gkp694.
33. Mueller N, van Bel N, Berkhout B, Das AT. 2014. HIV-1 splicing at the major splice donor site is restricted by RNA structure. Virology 468–470C:609–620. https://doi.org/10.1016/j.virol.2014.09.018.
34. Domsic JK, Wang Y, Mayeda A, Krainer AR, Stoltzfus CM. 2003. Human immunodeficiency virus type 1 hnRNP A/B-dependent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts. Mol Cell Biol 23:8762–8772. https://doi.org/10.1128/MCB.23.23.8762-8772.2003.
35. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol 15:183–191. https://doi.org/10.1038/nsmb.1375.

36. Tacke R, Tohyama M, Ogawa S, Manley JL. 1998. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell 93:139–148. https://doi.org/10.1016/S0092-8674(00)81153-8.
37. Grellscheid S, Dalgliesh C, Storbeck M, Best A, Liu Y, Jakubik M, Mende Y, Ehrmann I, Curk T, Rossbach K, Bourgeois CF, Stevenin J, Grellscheid D, Jackson MS, Wirth B, Elliott DJ. 2011. Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. PLoS Genet 7:e1002390. https://doi.org/10.1371/journal.pgen.1002390.
38. Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, Unzai S, Inoue M, Harada T, Watanabe S, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A, Sugano S, Guntert P, Yokoyama S, Muto Y. 2011. Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. Nucleic Acids Res 39:1538–1553. https://doi.org/10.1093/nar/gkq854.
39. Erkelenz S, Poschmann G, Theiss S, Stefanski A, Hillebrand F, Otte M, Stuhler K, Schaal H. 2013. Tra2-mediated recognition of HIV-1 5′ splice site D3 as a key factor in the processing of vpr mRNA. J Virol 87:2721–2734. https://doi.org/10.1128/JVI.02756-12.
40. Burd CG, Dreyfuss G. 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J 13:1197–1204.
41. Kajita Y, Nakayama J, Aizawa M, Ishikawa F. 1995. The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. J Biol Chem 270:22167–22175.
42. Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol 10:84–94. https://doi.org/10.1128/MCB.10.1.84.
43. Hoffman BE, Grabowski PJ. 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. https://doi.org/10.1101/gad.6.12b.2554.
44. Asang C, Hauber I, Schaal H. 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. https://doi.org/10.1093/nar/gkm1147.
45. Madsen JM, Stoltzfus CM. 2006. A suboptimal 5′ splice site downstream of HIV-1 splice site A1 is required for unspliced viral mRNA accumulation and efficient virus replication. Retrovirology 3:10. https://doi.org/10.1186/1742-4690-3-10.
46. Nomaguchi M, Doi N, Sakai Y, Ode H, Iwatani Y, Ueno T, Matsumoto Y, Miyazaki Y, Masuda T, Adachi A. 2016. Natural single-nucleotide variations in the HIV-1 genomic SA1prox region can alter viral replication ability by regulating Vif expression levels. J Virol 90:4563–4578. https://doi.org/10.1128/JVI.02939-15.
47. Spena S, Duga S, Asselta R, Malcovati M, Peyvandi F, Tenchini ML. 2002. Congenital afibrinogenemia: first identification of splicing mutations in the fibrinogen Bbeta-chain gene causing activation of cryptic splice sites. Blood 100:4478–4484. https://doi.org/10.1182/blood-2002-06-1647.
48. Spena S, Tenchini ML, Buratti E. 2006. Cryptic splice site usage in exon 7 of the human fibrinogen B beta-chain gene is regulated by a naturally silent SF2/ASF binding site within this exon. RNA 12:948–958. https://doi.org/10.1261/rna.2269306.
49. Selden RF, Howie KB, Rowe ME, Goodman HM, Moore DD. 1986. Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Mol Cell Biol 6:3173–3179. https://doi.org/10.1128/MCB.6.9.3173.
50. Morner A, Bjorndal A, Albert J, Kewalramani VN, Littman DR, Inoue R, Thorstensson R, Fenyo EM, Bjorling E. 1999. Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. J Virol 73:2343–2349.
51. Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159.

3.2.3. Altered HIV-1 mRNA splicing due to drug resistance-associated mutations in exon 2/2b (Chapter 7)

HIV-1 infection is still not sterilely curable to this day, although the development and availability of antiviral therapies contribute significantly to improving quality of life and increasing life expectancy. Country-specific studies and data from the World Health Organization suggest that overall, both the number of HIV-1 patients receiving highly active antiretroviral therapy as well as the duration of treatment increased over the past years. This, however, might contribute to increased incidents of drug resistance, especially in non-adherent patients or in areas with limited access to drugs and regular clinical monitoring. Unfortunately, the underlying molecular mechanism of drug resistance-associated mutations is often poorly understood. Since HIV-1 highly relies on a balanced splicing reaction, nucleotide substitutions during viral replication can contribute to changes in the splicing outcome and thus, to changes in viral replication which might add to the effect of drug escape mutations. To elucidate the effect of two drug resistance-associated mutations on the splicing outcome, bioinformatic predictions were combined with transfection and infection experiments.

The following unpublished article is written by

Müller, L*, Moskorz, M*, Brillen, ALB, Hillebrand, F, Ostermann, PNO, Kiel, N, Ptok, J, Timm, J, Lübke, N, Schaal, H

* contributed equally

Contributions:

A.L.B., N.L. and H.S. conceived the study. H.S., A.L.B., J.T., N.L. W.M. and L.M. designed the experiments. L.M. W.M. F.H. P.N.O. N.K. performed cloning, transfection experiments, infection experiments and RT-PCR analyses. J.P. performed bioinformatics analysis. L.M. and H.S. wrote the manuscript.

Individual contribution: 30% L.M. performed infection experiments shown in Figure 2. L.M. designed and cloned the constructs and performed the experiments shown in Figure 3. L.M. wrote the first draft of the manuscript.

Altered HIV-1 mRNA splicing due to drug resistance-associated mutations in exon 2/2b

Abstract

The underlying molecular mechanism and their general effect on the replication capacity of HIV 1 drug resistance-associated mutations are often poorly understood. To elucidate the effect of two such mutations located in a region with a high density of spicing regulatory elements on the HIV-1 splicing outcome, bioinformatic predictions were combined with transfection and infection experiments. Results show that the previously described R263K drug resistance-associated integrase mutation has additionally a severe effect on the ESE2b splicing regulatory element (SRE) in exon 2b which causes loss of SD2b recognition. This is confirmed by an R263R silent mutation with a similar predicted effect on the exon 2b SRE. In contrast, a V260I mutation and its silent counterpart with a lower effect on ESS2b did not exhibit any differences in the splicing pattern. Because HIV-1 relies heavily on a balanced splicing reaction, changes in the splicing outcome may contribute to changes in viral replication and might add to the effect of escape mutations to antiviral drugs. Thus, a pure protein addressing classification of mutations is insufficient.

Introduction

Human immunodeficiency virus 1 (HIV-1) infection is still not sterile curable to this day, although several promising results have at least demonstrated viral clearance from infected hosts in cell cultures and small animal models using genome-editing approaches [1]. Nevertheless, the development and current availability of antiviral treatment is already making a significant contribution to improving quality of life and increasing life expectancy [2, 3]. Country-specific studies and data from the World Health Organization and UNAIDS respectively suggest that overall, both the number of HIV-1 patients receiving highly active antiretroviral therapy (HAART) as well as the duration of treatment increased over the past years [4-6]. This, however, might contribute to increased incidents of drug resistance, especially in non-adherent patients or in areas with limited access to drugs and regular clinical monitoring [7, 8].

On a molecular level, HIV-1 highly relies on host factors for replication, in particular the cellular splicing machinery [9]. Upon infection and reverse transcription of the viral RNA

into the 9.7 kb proviral DNA, the assembled pre-integration complex (PIC) facilitates nuclear incorporation and subsequent integration of HIV-1 genome into the host's chromosome, which is then transcribed into a full-length precursor mRNA (pre-mRNA) [10, 11]. For the production of the complete range of mRNAs needed for balanced expression of all viral proteins and functional replication, transcripts undergo extensive alternative splicing (AS) which yields approximately 50 different mRNAs species Depending on their size, the resulting mRNAs are classified into full-length 9 kb, intron-containing 4 kb and intron-less 2 kb mRNAs. Four major splice donor (SD) sites (SD1 to SD4) in combination with eight splice acceptor (SA) sites (SA1, SA2, SA3, SA4a,b,c, SA5, and SA7) contribute to the generation of the majority of mRNA isoforms, while the rare use of additional splice sites such as splice donor SD2b or splice acceptors SA4d or SA5b further enlarge the transcript repertoire [12-17].

In general, their recognition depends on the intrinsic strength of the splice sites, which can be scored by various algorithms such as the HBond-score (HBS) [18] or MaxEnt scan [19]. A subset of HIV-1 3’ SA exhibit lower intrinsic strength scores compared to cellular 3’ SA sites [20], hence they are likely more influenced by nearby splicing regulatory elements [21]. Additionally, the order of intrinsic splice site strengths does not correlate with the observed levels of mRNAs, which implies that cis-acting splicing regulatory elements add an additional layer of regulation and dominate splice site selection of HIV-1 mRNAs [22].

The compact HIV-1 genome contains numerous cis-acting regulatory RNA elements that can influence a variety of processes essential for viral replication including genomic RNA packaging, pre-mRNA processing, polyadenylation, and nuclear RNA export [17, 23]. The sequence elements and changes thereof can subsequently influence the overall replication capacity of viruses [24]. In addition to altering the amino acid sequence, missense mutations, but of course also silent mutations in SREs, can affect viral replication by altering the binding of splicing regulatory proteins. This could then ultimately lead to altered splice site recognition and severely disrupt viral replication by altering the ratio of alternatively spliced mRNA isoforms required for effective replication [14]. Some of these Sanger or Next Generation Sequencing detected mutations exhibit the potential to alter the viral splicing patterns and thus

change the viral proteome [25]. Such mutations might ultimately be associated with reduced sensitivity to antiviral agents.

Here, we analyze two antiviral drug resistance-associated mutations, V260I and R263K located within the C-terminal domain of Integrase. The HIV Integrase is encoded in the pol gene and promotes the integration of the provirus into the host's genome. The splice sites SA1, SD2, and SD2b are located within this distal region of pol. Among others, the splicing regulatory elements ESE-Vif, ESS2b, and ESE2b modulate their usage [17, 26, 27].

The bioinformatics tool used for the prediction, the HEXplorer tool, is one of various publicly available algorithms to analyse SREs and the consequences of mutations and has already been employed to identify SREs within HIV-1 [26, 27]. It is based on differential hexamer frequencies and has the particularity of analysing a sequences' potential to act a SREs in a sliding window [28]. The two mutations are of particular interest since bioinformatics analysis via the HEXplorer tool indicated a certain potential to alter the splicing outcome. The tool evaluates sequence stretches on their probability to act as SREs by posing as binding sites for either SR-or hnRNP proteins [28, 29]. The consequences of nucleotide changes are given as differences in hexamer frequencies, calculated as ΔHZEI. Thus, higher changes in the HZEI score indicate a higher potential to alter the binding potential of SREs so that the chosen V260I mutation with an ΔHZEI of -37.79 compared to the parental sequence is predicted to have only a minor impact on splice site recognition compared to R263K with an ΔHZEI of -132.16. Despite the two chosen mutations being located outside of the catalytic core of the Integrase, they are described as secondary resistance associated mutations [30]. Additionally, R263K was recently described to decrease HIV integration and to be associated with therapy failure using Dolutegravir in HIV-1 treated patients [31-33].

The aim of this study was the evaluation of the influence of the V260I and R263K mutation on the alternative splicing of HIV-1 mRNA.

Results:

In the C-terminal domain of the integrase coding region, i.e the 3' end of HIV-1 pol, both mutations, V260I and R263K, are located within ESS2b and ESE2b (Figure 1A),

and were analyzed for their potential to impact splice site recognition of exon 2/2b using the HEXplorer algorithm [24]. The 3' end of HIV-1 pol is of particular interest since it is interspersed with a dense network of splicing regulatory elements that has been shown to mediate the use of splice acceptor SA1 (MaxEnt 6.41) [34] and splice donors SD2 (HBond score 10.7) and SD2b (HBond score 12.4) (Figure 1B) [22]. HEXplorer plots show the predicted impact of the two secondary resistance-associated mutations on the ESE2b SRE. V260I displays a change by an ΔHZEI of -32.79 suggesting a potential only minor addition to the negative regulation of SD2b and a slightly increased enhancement of SD2. Mutation R263K, however, largely alters the HEXplorer plot by an ΔHZEI of -132.16, and thus, this mutation is predicted to increase SD2 use while the support of SD2b should be drastically decreased (Figure 1C).

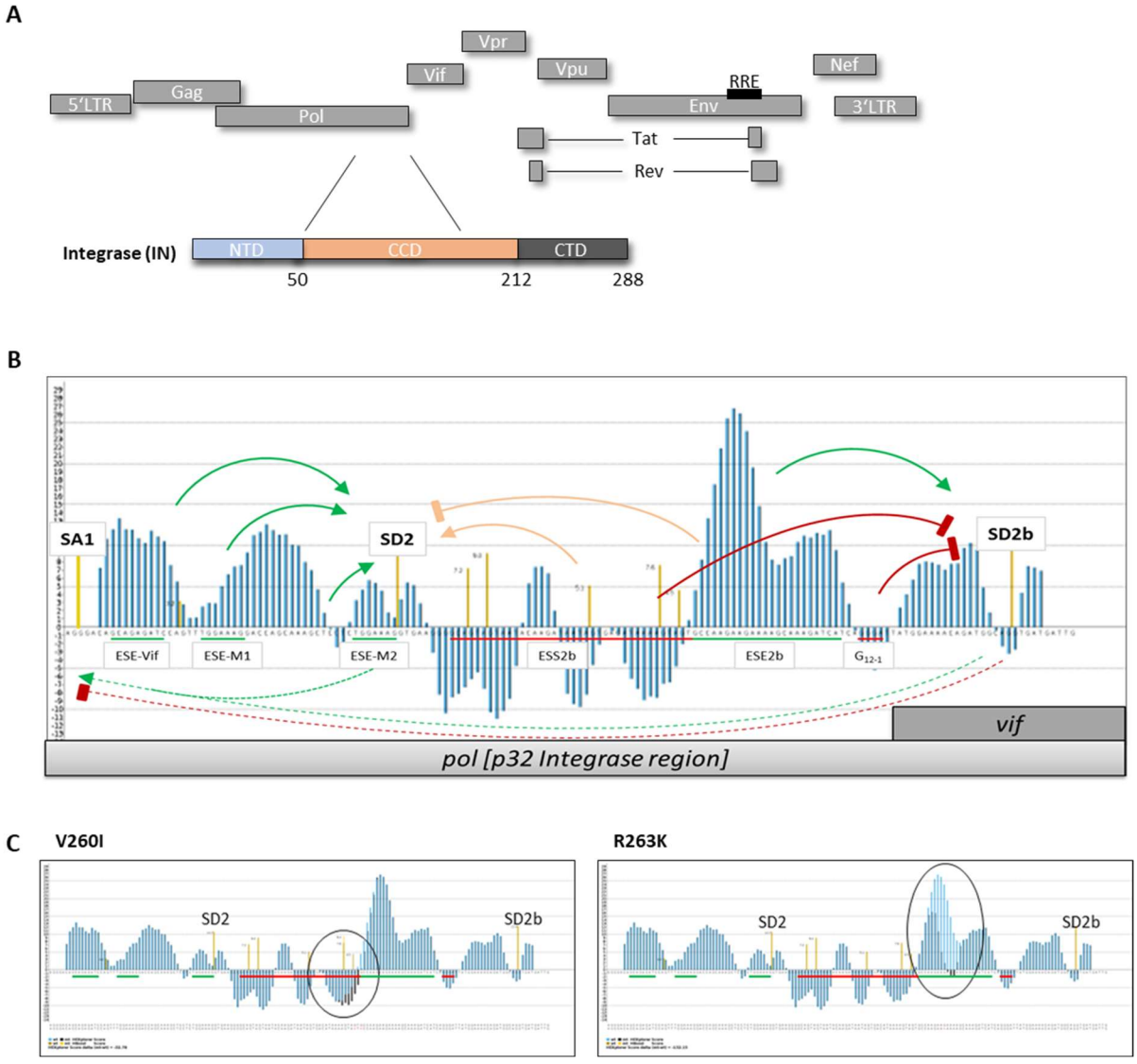


Figure 1: Organization and regulation of the HIV-1 genome.

A *The positions of the HIV-1 open reading frames are shown as grey boxes. Ribosomal frame-shifting allows the production of the viral enzymes Protease (PR), Reverse Transcriptase (RT) and Integrase (IN), which are also cleavage products of the Gag-Pol polyprotein. The 288 amino acids (32kDa) HIV-1 IN mediates the integration of the provirus into the host genome and contains three domains i) N-terminal domain (NTD, AS 1-50), ii) catalytic core domain (CCD, AS 50-212) and iii) C-terminal domain (CTD, AS 212-288).* ***B*** *Integrase coding region is interspersed with well-characterized SREs that mediate the use of splice acceptor SA1 (MaxEnt 6.41) and both splice donors, SD2 (HBond score 10.7) and SD2b (HBond score 12.4). Positive regulation is marked in green, negative regulation in red and intermediate regulation in orange.* ***C*** *Impact of two secondary resistance-associated mutations on ESE2b depicted by HEXplorer plots. V260I (∆HZEI = -32.79), potentially slightly adding to the negative regulation of SD2b but increasing the enhancement of SD2. Mutation R263K (∆HZEI = -132.16) is likely to increase SD2 at the expense of SD2b.*

Using site-directed mutagenesis, the selected nucleotide exchanges were inserted into the pNL4-3 proviral plasmid coding for an HIV-1 laboratory wild-type strain. This permits the analysis and comparison of the splicing pattern after both, transfection and infection. Different splicing patterns were amplified after transfection of HEK293T cells or infection of PM1 cells which is a commonly used T-cell line in HIV-1 research. Therefore, particular primer sets were used either amplifying transcripts of the 2 kb class or transcripts using splice donors D2 and D2b (#2710/#3392). For monitoring transfection efficiency, a plasmid coding for hGH (human growth hormone) was co-transfected and expressed transcripts amplified using primer pair #1224/#1225. To monitor the infection, a short region of HIV-1 exon 7 was amplified (#3387/#3388). Significant differences were only found in the analysis of the SD2/SD2b splicing pattern, both in comparison to the parental NL4-3 and between the mutants themselves. In the transfection experiments, the V260I variant showed a slight increase in Tat2b message which was in contrast to the HEXplorer prediction of a slight decrease in SD2b recognition. A similar pattern could be observed upon infection with the V260I virus (Figure 2). In the transfection experiment, the R263K mutant showed a reduced recognition of splice donor SD2b. This pattern was even more pronounced upon infection of PM1 cells with the R263K virus. Recognition of SD2b was completely abolished, which has a particular effect on the processing of Tat and Nef RNA species (Figure 2).

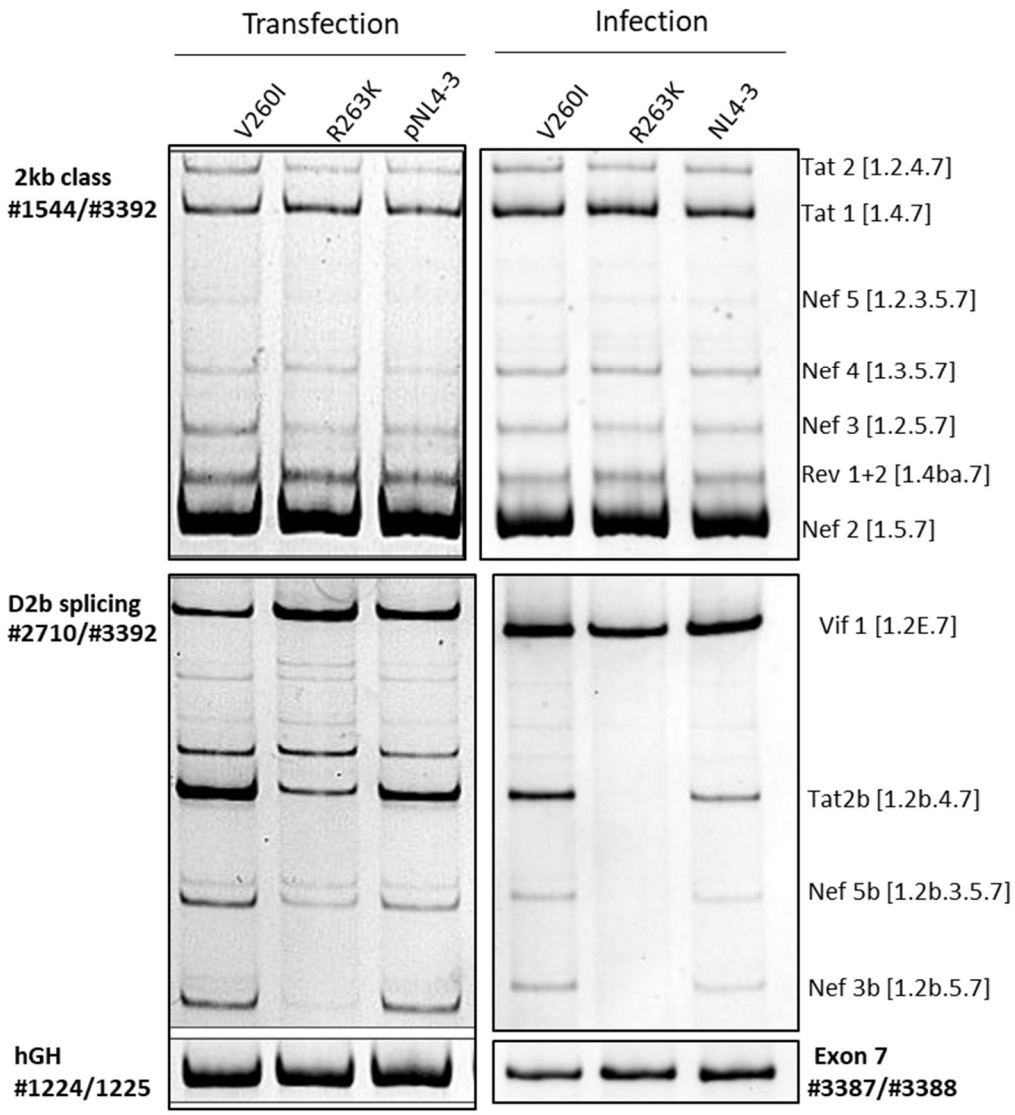


Figure 2: Effects of secondary drug resistance-associated mutations on the HIV-1 splicing pattern

Left panel: Proviral plasmids were used for transfection as wildtype (pNL4-3) or carrying either the V260I or R263K mutation. 2.5 × 10^5 Hek 293T cells were transiently transfected with 1µg of each construct together with 1µg of pXGH5 (hGH) to monitor transfection efficiency. Twenty-four hours after transfection, RNA was isolated and subjected to RT-PCR analysis using different primer pairs (#1544/#3392 for 2kb class, #2710/#3392 for D2b splicing, #1224/#1225 for hGH). PCR products were separated by a 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide. Right panel: 1.0 x 10^6 PM1 cells were infected with either wildtype or mutant virus (MOI 0.05). RNA was harvested 72h post-infection, RT-PCR was performed with the same primer pairs as for transfection except for the control (#3387/#3388 for Exon 7). Again, PCR products were separated by a 10% non-denaturing polyacrylamide gel electrophoresis and stained with ethidium bromide. The main

difference between the splicing patterns can be seen for R263K D2b splicing where the Tat 2b, Nef 5b, and Nef 3b messages are lost upon infection.

To dissect whether the mutations' effect on replication and splicing is dependent on the altered coding potential of the sequence by the missense mutations, the HEXplorer tool was used to design silent versions of the selected mutations with the same in silico predicted potential to change splicing behavior. For the missense mutation V260I with an ΔHZEI of -37.79, a silent version V260V with an ΔHZEI of -39.0 was generated. For the missense mutation R263K with a ΔHZEI of -132.16, a silent version R263R with a slightly lower ΔHZEI of -107.8 was generated (Figure 3A, left panel). For further analysis, the mutations were also inserted into subgenomic splicing reporters. The 3-exon-minigene splicing reporter consists of two exons flanking an amendable middle exon. It carries the naturally occurring splicing regulatory elements of exon 2 as well as the two splice donors SD2 and SD2b. In between the two splice donors, the missense mutations V260I and R263K and the silent mutations V260V and R263R as well as the parental sequence were inserted (Figure 3A). Transfection and subsequent RT-PCR analysis of the splicing pattern, in particular recognition of SD2 and SD2b, revealed differences in the splicing pattern between the mutations and compared to the parental NL4-3 sequence. There was no difference in SD2 and SD2b recognition upon insertion of the V260I or V260V silent mutation, however, a slight increase in SD2b recognition could be recognized compared to the NL4-3 sequence for both versions of the mutation. For R263K, a similar pattern was observed as previously seen in the infection experiment (Figure 2) with SD2b recognition fully diminished. Interestingly, the silent version R263R with a similar but even slightly lower ΔHZEI of -107.8 showed only a slight decrease in SD2b recognition compared to the parental NL4-3 sequence (Figure 3B).

To further analyze the silent mutation R263R, the mutation was inserted into the pNL4-3 proviral plasmid by site-directed mutagenesis and used for infection. Upon infection of PM1 cells and RNA isolation 72h post-infection, RT-PCR analysis was carried out using the previously described primers #2710/3392 to analyze the SD2/SD2b usage. As expected from the HEXplorer prediction, the silent R263R mutation led to a similar splicing pattern as R263K, albeit, in contrast to the missense version R263K a slight band for the most abundant mRNA species Tat2b was still present (Figure 3C).

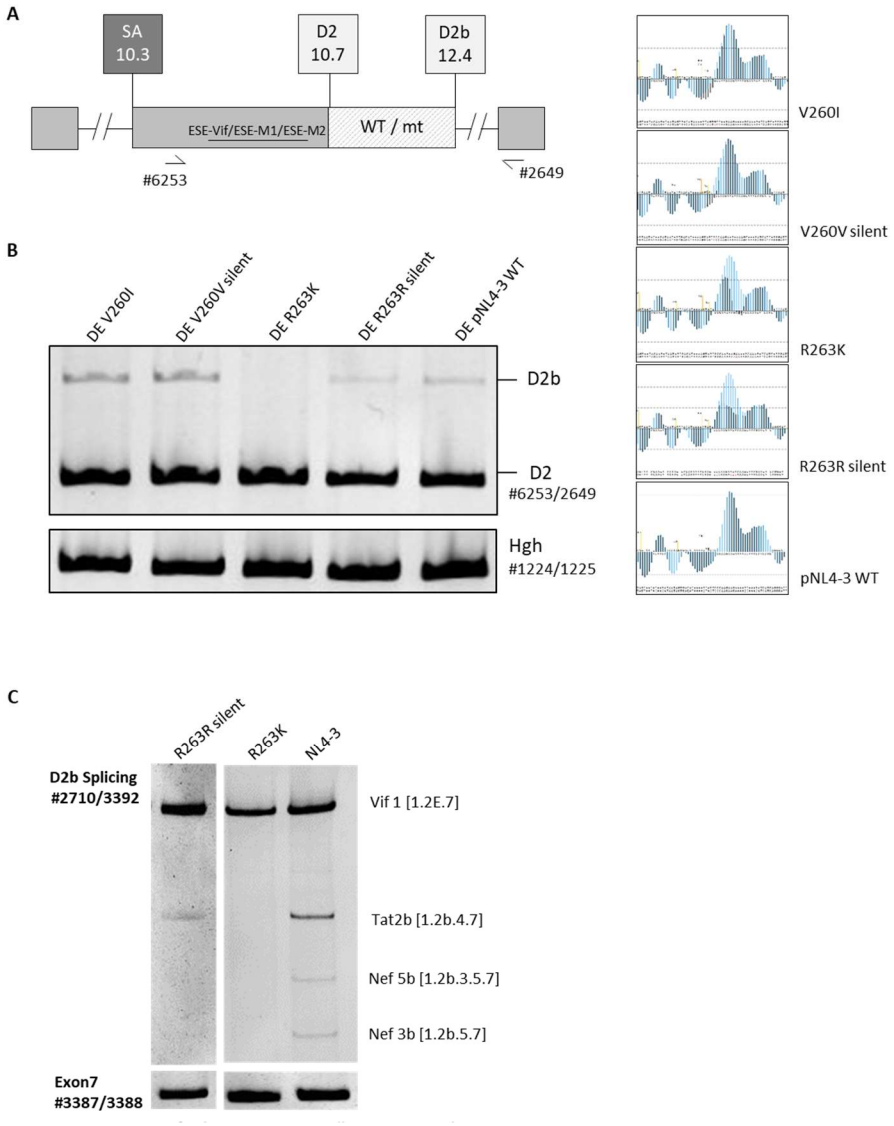


Figure 3: Silent variant of R263K with a comparable effect on SREs.

A *Schematic overview of the three-exon minigene splicing reporter containing the HIV-1 native splice donors SD2 and SD2b. Parental or mutated sequences are inserted between the two splice donors. The right panel shows the respective HEXplorer plots of the inserted sequences.* ***B*** *2.5x10^5 HeLa cells were transiently transfected with the splicing reporter plasmids, RNA was harvested 24h post-transfection. Hgh was co-transfected as a control. RT-PCR samples were run on a non-denaturing 10% polyacrylamide gel.* ***C*** *HIV-1 proviral plasmids carrying either the wild-type, the missense, or the silent mutation were prepared and used for infection of PM1 cells with an MOI of 0.05. RNA was harvested 72h post-infection and the splicing pattern was analyzed via RT-PCR on a 10% non-denaturing PAA gel.*

Discussion

By infection and transfection experiments and the employment of splicing reporters, this work demonstrates that resistance-associated HIV-1 Integrase mutations can highly influence the viral splicing pattern.

We report about two secondary resistance-associated mutations, V260I and R263K located within the C-terminal domain of Integrase [30]. Both mutations were selected for further analysis since they are bioinformatically predicted to affect the splicing pattern by disrupting ESE2b and thus, potentially contribute to an imbalance in HIV-1 mRNA transcripts (Figure 1C).

The SREs mostly affected by these mutations are the recently described ESS2b and ESE2b, which regulate the recognition of splice donors 2 and 2b as well as the closely located splice acceptor A1 (Figure 1B). Both splice donors located within leader exon 2 are generally used infrequently, recognition and use of D2 however exceed D2b [26, 27]. Exon 2/2b recognition particularly contributes to the generation of Vif mRNA species, a counter actor of the host restriction factor APOBEC3G [35], as well as minor Tat mRNA species including Tat2b. A balanced regulation of Vif expression is crucial for viral replication as both, too high or too low amount of Vif can have detrimental effects [36]. However, at exon 2, the Vif mRNA overlaps with the integrase reading frame. Therefore, the SREs and potential sequence changes due to mutations in this region can have a major impact on both Vif levels and integrase function. Both contribute greatly to efficient virus replication [34, 37].

Transfection and infection experiments revealed that in particular, the R263K mutation, located within ESE2b contributed to a splicing imbalance where mRNA transcripts employing SD2b are fully diminished (Figure 2). This pattern was reproducible with a silent version of the mutation that had a similar potential to alter ESE2b (Figure 3). R263K was of particular interest since it is not only described to be the most commonly selected resistance mutation selected under therapy with the second-generation integrase strand-transfer inhibitor (INSTI) Dolutegravir (DTG) [36], it was also recently described to contribute to low-level resistance in patients receiving first-line treatment with the DTG [31]. Reports show that R263K reduces DTG susceptibility by

approximately 2-fold [39]. However, the molecular mechanism underlying the drug resistance caused by this mutation is yet to be fully understood. A possible explanation of the molecular mechanism of resistance could be an interaction of the non-transferred strand with the transferred viral DNA strand, that results in a change in the positioning of the end of the viral DNA and thus could affect the interaction between the INSTIs and the DNA [38]. The general effect of the mutation on the viral replication has been described to decrease in integration capacity [32]. Our results reveal that furthermore, the R263K, as well as a silent version of the mutation, can contribute to a highly imbalanced splicing behavior at SD2/SD2b.

The data imply that splicing analysis may play an important part in resistance profiling and consequently may contribute to a better understanding of resistance mechanisms regarding integrase mutations. Furthermore, since HIV-1 is generally dependent on a tightly regulated splicing balance and the mRNA isoforms generated, targeting the splicing process has been discussed as a potential approach to disrupting viral replication [39]. One approach along this line is the use of locked nucleic acids (LNA), antisense oligonucleotides forming a particularly stable Watson-Crick base pairing with the RNA through an additional methylene bridge in the ribose sugar. We have recently been able to report that the gymnotic, without transfection reagent, application of LNA mixmers to the cell culture medium of HIV-1 infected cells induces the degradation of viral mRNA carrying the target sequence of the LNAs, and thus inhibit viral replication [40]. The characterization of possible new targets of antiviral agents is an important branch of research, particularly in the light of increased drug resistance.

Material and Methods

Proviral plasmids.

The proviral plasmids pNL4-3 V260I, R263K, and R263R were generated by an overlapping-PCR technique (PCR1 with primers a+b, PCR2 with primers c+d, PCR 3 with PCR 1 and 2 as a template using primer a+d) using their respective primer pairs (Table 1).

Three-exon minigenes.

The three-exon minigenes are derived from the fibrinogen Bß-minigene pT-Bß-IVS7ß 1G>T and were previously described [26]. The middle exon carries HIV-1-derived SD2 and SD2b as well as splicing regulatory elements ESE-Vif, -M1, and -M2. The respective wild-type and mutated fragments of HIV-1 exon 2/2b were added at their authentic position between SD2 or SD2b, by using primer pairs #5941/#6040 (V260I), #5941/#6042 (V260V silent), #5941/#6041 (R263K) and #5941/#6043 (R263R silent). All oligonucleotides used were obtained from Metabion GmbH.

Expression plasmids.

A plasmid encoding for the human growth hormone hGH (pXGH5) [42] was co-transfected to monitor transfection efficiency.

Cell culture, infection, and transfection.

HeLa and HEK293T cells were cultured in Dulbecco's high-glucose modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and 50 g/ml penicillin-streptomycin. PM1 cells were cultured in Roswell Park Memorial Institute Medium (RPMI) supplemented with 10% fetal calf serum and 50 g/ml penicillin-streptomycin. Cells were propagated twice a week. For transient transfection, 2x105 cells per well were plated in six-well plates. Transient-transfection experiments were performed using Mirus TransIT-LT1 transfection reagent according to the manufacturer's instructions. For infection experiments, PM1 cells were adjusted to 1x106 cells per ml and inoculated with virus stock at an MOI of 0.05 for 6h in 1ml medium before they were washed with 5ml PBS and kept in 2ml RPMI at 37°C.

RNA isolation and RT-PCR.

Either 24h post-transfection or 72h post-infection, total cellular RNA was isolated by using acid guanidinium thiocyanate-phenol-chloroform. RNA was reverse transcribed by using Superscript III reverse transcriptase (Invitrogen) and oligo(dT) primers (Invitrogen) for semiquantitative RT-PCR with the denoted primer pairs.

Bioinformatic tools.

The HBond score was calculated via http://www.uni-duesseldorf .de/rna/html/hbond_score.php. The MaxEnt scan algorithm can be accessed via http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html and the HEXplorer tool is available under https://www2.hhu.de/rna/html/hexplorer_score.php.

Table 1: Primer sequences:

RT-PCR:

#1554 CTTGAAAGCGAAAGTAAAGC

#2710 AAGGGGCAGTAGTAATACAA

#3392 CGTCCCAGATAAGTGCTAAGG

#3632 TGGATGCTTCCAGGGCTC

#1224 TCTTCCAGCCTCCCATCAGCGTTTGG

#1225 CAACAGAAATCCAACCTAGAGCTGCT

#2649: TCCACCACCGTCTTCTTTAG

#6253 CAACCAAACAACCTTAGGGGA

Cloning

a) Proviral plasmids

#3632 (d) TGGATGCTTCCAGGGCTC

#5553 (a) CTGGCAGAAAACAGGGAGATT

V260I as (b) CTTTGCTTTTCTTCTTGGTATTACTTTTATGTCACTATTATC

V260I s (c) GTGACATAAAAGTAATACCAAGAAGAAAAGCAAAGATCATCAG

R263K as (b) GATGATCTTTGCTTTTTTTCTTGGCACTACTTTTATG

R263K s (c) GTAGTGCCAAGAAAAAAAGCAAAGATCATCAGG

R263R s (c) GTAGTGCCAAGACGGAAAGCAAAGATCATCAGG

R263R as (b) GATGATCTTTGCTTTCCGTCTTGGCACTACTTTTATG

b) Three-exon-minigene

#6040

GGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAATACCAAGAAGAAAAGCAAA
GATCAT

#6041

GGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAAAAAAGCAAA
GATCATCAGGGATTAT

#6042

GGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTCCCAAGAAGAAAAGCAAA
GATCAT

#6043
GGGGCTAGCGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGACGGAAAGCAAA
GATCATCAGGGATTAT
#5941 GGACAGTGGCTGACAGT

Author Contribution:

N.L., A.L.B. and H.S. conceived the study. N.L., A.L.B., H.S., W.M., J.T. and L.M. designed the experiments. L.M. W.M. F.H. P.N.O. N.K. performed cloning, transfection experiments, infection experiments and RT-PCR analyses. J.P. performed bioinformatics analysis. L.M. and H.S. wrote the manuscript.

References:

1. Buchholz F, Hauber J. Antiviral therapy of persistent viral infection using genome editing. Curr Opin Virol 2016; 20: 85-91.
2. Archin NM, Sung JM, Garrido C, Soriano-Sarabia N, Margolis DM. Eradicating HIV-1 infection: seeking to clear a persistent pathogen. Nat Rev Microbiol 2014; 12(11): 750-64.
3. Shang HT, Ding JW, Yu SY, Wu T, Zhang QL, Liang FJ. Progress and challenges in the use of latent HIV-1 reactivating agents. Acta Pharmacol Sin 2015; 36(8): 908-16.
4. Montaner JS, Lima VD, Harrigan PR, et al. Expansion of HAART coverage is associated with sustained decreases in HIV/AIDS morbidity, mortality and HIV transmission: the "HIV Treatment as Prevention" experience in a Canadian setting. PloS one 2014; 9(2): e87872.
5. Bradley H, Mattson CL, Beer L, Huang P, Shouse RL, Medical Monitoring P. Increased antiretroviral therapy prescription and HIV viral suppression among persons receiving clinical care for HIV infection. Aids 2016; 30(13): 2117-24.
6. Maskew M, Bor J, MacLeod W, Carmona S, Sherman GG, Fox MP. Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. Lancet HIV 2019; 6(11): e760-e8.
7. Keita A, Sereme Y, Pillet S, et al. Impact of HIV-1 primary drug resistance on the efficacy of a first-line antiretroviral regimen in the blood of newly diagnosed individuals in Bamako, Mali. J Antimicrob Chemother 2019; 74(1): 165-71.
8. Nachega JB, Marconi VC, van Zyl GU, et al. HIV treatment adherence, drug resistance, virologic failure: evolving concepts. Infect Disord Drug Targets 2011; 11(2): 167-74.
9. Sertznig H, Hillebrand F, Erkelenz S, Schaal H, Widera M. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. Virology 2018; 516: 176-88.
10. Jayappa KD, Ao Z, Yao X. The HIV-1 passage from cytoplasm to nucleus: the process involving a complex exchange between the components of HIV-1 and cellular machinery to access nucleus and successful integration. Int J Biochem Mol Biol 2012; 3(1): 70-85.
11. Liu RD, Wu J, Shao R, Xue YH. Mechanism and factors that control HIV-1 transcription and latency activation. J Zhejiang Univ Sci B 2014; 15(5): 455-65.
12. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med 2012; 2(2): a006916.

13. Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. Journal of virology 1993; 67(11): 6365-78.

14. Takata MA, Soll SJ, Emery A, Blanco-Melo D, Swanstrom R, Bieniasz PD. Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. PLoS Pathog 2018; 14(1): e1006824.

15. Ocwieja KE, Sherrill-Mix S, Mukherjee R, et al. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic acids research 2012; 40(20): 10345-55.

16. Nguyen Quang N, Goudey S, Segeral E, et al. Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. Retrovirology 2020; 17(1): 25.

17. Emery A, Zhou S, Pollom E, Swanstrom R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. Journal of virology 2017; 91(6).

18. Freund M, Asang C, Kammler S, et al. A novel approach to describe a U1 snRNA binding site. Nucleic acids research 2003; 31(23): 6963-75.

19. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 2004; 11(2-3): 377-94.

20. Stoltzfus CM, Madsen JM. Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. Curr HIV Res 2006; 4(1): 43-55.

21. Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. The strength of the HIV-1 3' splice sites affects Rev function. Retrovirology 2006; 3: 89.

22. Asang C, Hauber I, Schaal H. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. Nucleic acids research 2008; 36(5): 1450-63.

23. Ostermann PN, Ritchie A, Ptok J, Schaal H. Let it go: HIV-1 cis-acting repressive sequences. Journal of virology 2021.

24. Wong RW, Balachandran A, Ostrowski MA, Cochrane A. Digoxin suppresses HIV-1 replication by altering viral RNA processing. PLoS Pathog 2013; 9(3): e1003241.

25. Arias A, Lopez P, Sanchez R, Yamamura Y, Rivera-Amill V. Sanger and Next Generation Sequencing Approaches to Evaluate HIV-1 Virus in Blood Compartments. Int J Environ Res Public Health 2018; 15(8).

26. Brillen AL, Walotka L, Hillebrand F, et al. Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b. Journal of virology 2017; 91(14).

27. Widera M, Erkelenz S, Hillebrand F, et al. An intronic G run within HIV-1 intron 2 is critical for splicing regulation of vif mRNA. Journal of virology 2013; 87(5): 2707-20.

28. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic acids research 2014; 42(16): 10681-97.

29. Erkelenz S, Mueller WF, Evans MS, et al. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. Rna 2013; 19(1): 96-102.

30. Ceccherini-Silberstein F, Malet I, D'Arrigo R, Antinori A, Marcelin AG, Perno CF. Characterization and structural analysis of HIV-1 integrase conservation. AIDS reviews 2009; 11(1): 17-29.

31. Lubke N, Jensen B, Huttig F, et al. Failure of Dolutegravir First-Line ART with Selection of Virus Carrying R263K and G118R. N Engl J Med 2019; 381(9): 887-9.

32. Mesplede T, Leng J, Pham HT, et al. The R263K Dolutegravir Resistance-Associated Substitution Progressively Decreases HIV-1 Integration. mBio 2017; 8(2).

33. Ahmed N, Flavell S, Ferns B, et al. Development of the R263K Mutation to Dolutegravir in an HIV-1 Subtype D Virus Harboring 3 Class-Drug Resistance. Open Forum Infect Dis 2019; 6(1): ofy329.

34. Nomaguchi M, Doi N, Sakai Y, et al. Natural Single-Nucleotide Variations in the HIV-1 Genomic SA1prox Region Can Alter Viral Replication Ability by Regulating Vif Expression Levels. Journal of virology 2016; 90(9): 4563-78.

35. Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Molecular cell 2003; 12(3): 591-601.

36. Mandal D, Exline CM, Feng Z, Stoltzfus CM. Regulation of Vif mRNA splicing by human immunodeficiency virus type 1 requires 5' splice site D2 and an exonic splicing enhancer to counteract cellular restriction factor APOBEC3G. Journal of virology 2009; 83(12): 6067-78.

37. Exline CM, Feng Z, Stoltzfus CM. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. Journal of virology 2008; 82(8): 3921-31.

38. Smith SJ, Zhao XZ, Passos DO, Lyumkis D, Burke TR, Jr., Hughes SH. Integrase Strand Transfer Inhibitors Are Effective Anti-HIV Drugs. Viruses 2021; 13(2).

39. Rhee SY, Grant PM, Tzou PL, et al. A systematic review of the genetic mechanisms of dolutegravir resistance. J Antimicrob Chemother 2019; 74(11): 3135-49.

40. Dlamini Z, Hull R. Can the HIV-1 splicing machinery be targeted for drug discovery? HIV AIDS (Auckl) 2017; 9: 63-75.

41. Hillebrand F, Ostermann PN, Muller L, et al. Gymnotic Delivery of LNA Mixmers Targeting Viral SREs Induces HIV-1 mRNA Degradation. International journal of molecular sciences 2019; 20(5).

42. Selden RF, Howie KB, Rowe ME, Goodman HM, Moore DD. Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Molecular and cellular biology 1986; 6(9): 3173-9.

3.2.4. Gymnotic Delivery of LNA Mixmers Targeting Viral SREs Induces HIV-1 mRNA Degradation (Chapter 8)

The understanding of underlying molecular mechanisms of emerging drug resistance in HIV-1 patients can lead ways for research to exploit certain pathways to disrupt viral replication by drug interference. Since HIV-1 highly relies on a balanced splicing regulation, it is evident that this process needs to be discussed as a novel therapeutic drug target. However, since the virus entirely uses the hosts' cellular splicing machinery, no components of the spliceosome itself can be targeted without causing significant side effects. As discussed above, regulation of viral alternative splice site usage depends on splicing regulatory elements, and changes thereof can heavily influence RNA expression and viral replication. Therefore, masking these elements by transfecting locked nucleic acid mixmers, antisense oligonucleotides that carry an additional methylene bridge in the ribose backbone and form a particularly stable bond with the RNA, can highly affect the viral splicing pattern and thus, replication. Furthermore, it was shown that delivering these locked nucleic acid mixmers gymnotically, hence without a transfection reagent, can mediated RNA degradation, which results in abrogation of viral replication in HIV-1 infected T-cells.

The following article is published in Int J Mol Sci. 2019 Mar 3;20(5). pii: E1088. doi: 10.3390/ijms20051088. by

Hillebrand F, Ostermann PN, **Müller L**, Degrandi D, Erkelenz S, Widera M, Pfeffer K, Schaal H.

Contributions:
F.H. and H.S. conceived the study and designed the experiments. S.E. and M.W. were involved in study design. F.H. and P.N.O. performed the NL4-3 infection and gymnotic LNA experiments. F.H., P.N.O. and L.M. performed transfection, RT-PCR, northern blot and immunoblot analyses. D.D. and K.P. performed the confocal laser scanning analysis. F.H., L.M. and H.S. wrote the manuscript.
Individual contribution: 10% L.M. performed reproductions of experiments shown in Figure 4 and 5. L.M. edited the manuscript draft.

International Journal of
Molecular Sciences

MDPI

Article

Gymnotic Delivery of LNA Mixmers Targeting Viral SREs Induces HIV-1 mRNA Degradation

Frank Hillebrand [1], Philipp Niklas Ostermann [1], Lisa Müller [1,†], Daniel Degrandi [2], Steffen Erkelenz [1,‡], Marek Widera [1,§], Klaus Pfeffer [2] and Heiner Schaal [1,*]

[1] Institute of Virology, Medical Faculty, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany; Frank.Hillebrand@med.uni-duesseldorf.de (F.H.); Philipp.Ostermann@uni-duesseldorf.de (P.N.O.); Lisa_Muller@dfci.harvard.edu (L.M.); Steffen.Erkelenz@gmx.de (S.E.); Marek.Widera@uni-due.de (M.W.)

[2] Institute of Medical Microbiology and Hospital Hygiene, Medical Faculty, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany; Daniel.Degrandi@uni-duesseldorf.de (D.D.); Klaus.Pfeffer@med.uni-duesseldorf.de (K.P.)

* Correspondence: schaal@uni-duesseldorf.de; Tel.: +49-(0)211-81-12393

† Current address: Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA.

‡ Current address: Institute for Genetics and Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50931 Cologne, Germany.

§ Current address: Institute for Virology, University of Duisburg-Essen, University Hospital Essen, 45147 Essen, Germany.

Received: 31 January 2019; Accepted: 25 February 2019; Published: 3 March 2019

Abstract: Transcription of the HIV-1 provirus generates a viral pre-mRNA, which is alternatively spliced into more than 50 HIV-1 mRNAs encoding all viral proteins. Regulation of viral alternative splice site usage includes the presence of splicing regulatory elements (SREs) which can dramatically impact RNA expression and HIV-1 replication when mutated. Recently, we were able to show that two viral SREs, G_{I3}-2 and ESE_{tat}, are important players in the generation of viral *vif*, *vpr* and *tat* mRNAs. Furthermore, we demonstrated that masking these SREs by transfected locked nucleic acid (LNA) mixmers affect the viral splicing pattern and viral particle production. With regard to the development of future therapeutic LNA mixmer-based antiretroviral approaches, we delivered the G_{I3}-2 and the ESE_{tat} LNA mixmers "nakedly", without the use of transfection reagents (gymnosis) into HIV-1 infected cells. Surprisingly, we observed that gymnotically-delivered LNA mixmers accumulated in the cytoplasm, and seemed to co-localize with GW bodies and induced degradation of mRNAs containing their LNA target sequence. The G_{I3}-2 and the ESE_{tat} LNA-mediated RNA degradation resulted in abrogation of viral replication in HIV-1 infected Jurkat and PM1 cells as well as in PBMCs.

Keywords: antisense oligonucleotides; locked nucleic acids; splicing regulatory elements; mRNA degradation; human immunodeficiency virus type 1 (HIV-1)

1. Introduction

According to the World Health Organization (WHO) in 2017, 36.9 million people were living with the human immunodeficiency virus type 1 (HIV-1) and globally only 21.7 million people received antiretroviral therapy (ART) which combines drugs targeting crucial steps of the HIV-1 replication cycle. However, although ART successfully reduces viral replication to a level below the detection limit, and thus saves the lives of many HIV-infected individuals, it is not curative, and non-adherent patients are especially at risk of developing multidrug-resistant viruses.

Furthermore, life-long ART is often accompanied by multiple adverse side-effects (e.g., lipodystrophy, insulin resistance, dyslipidemia, chronic inflammation) [1,2]. Thus, the identification of additional and alternative targets within the viral life cycle for antiviral drug development is desirable.

As a member of the family of retroviruses (retroviridae) the (+) RNA genome of HIV-1 is reverse transcribed, imported into the nucleus and integrated into the host cellular genome. After transcription the viral pre-mRNA undergoes extensive alternative splicing leading to viral mRNA transcripts essential for all viral proteins. The viral mRNAs are grouped into three viral RNA classes according to their size: The spliced and intron-less 2 kb class (including *tat*, *rev* and *nef* mRNAs), the spliced but intron-containing 4 kb class (including *vif, vpr, vpu* and *env* mRNAs), and the unspliced 9 kb class which serves as mRNA for Gag/Pol and as the viral genome [3]. Generation of more than 50 alternatively spliced viral mRNAs is regulated by differential usage of at least five viral splice donor sites (5′-splice sites) and eight viral splice acceptor sites (3′-splice sites) as well as the presence of several viral exonic and intronic splicing regulatory elements (SREs) [4–8]. Here, the SREs are bound by family members of the serine- and arginine-rich phosphoproteins (SR proteins) or heterogeneous nuclear ribonucleoproteins (hnRNPs), which positively or negatively influence viral splice site selection depending on their position relative to them [5,8,9]. For efficient replication a balanced generation of all viral mRNAs is crucial. Therefore, disruption of the viral splicing process, e.g., by preventing binding of splicing regulatory proteins to their RNA target seems to be a promising approach to impair viral replication. Indeed, as shown in several mutational analyses of viral SREs, interference with the SREs' function not only dramatically impacts viral splicing or RNA expression, but also influences HIV-1 particle production [4,10–16].

Masking viral sequences, e.g., to prevent protein binding, however, has been tried already in the late 1980s when two groups provided evidence that HIV-1 replication can be reduced by adding DNA-antisense oligonucleotides (ASOs), complementary to HIV-1 RNA sequences, to cell culture medium. At that time, unfortunately, very high ASO concentrations were necessary in order to impair HIV-1 RNA expression [17,18], hampering the dissemination of this approach. Later on, mostly by transfection or cell-free in vitro experiments, it was demonstrated that masking various HIV-1 sequences such as the HIV-1 dimerization initiation site, the trans-activation response element (TAR) element, the major splice donor 1 or the viral guanine-adenine-rich (GAR) SRE by ASOs or modified U7 snRNAs, also interfered with viral RNA expression [19–22]. The applied ASOs, however, have to exhibit obligatory characteristics: (i) Efficient masking and specific binding of the target sequence, (ii) low toxicity range, and (iii) an increased stability against endo- and exonucleases.

In the past, several ASOs exhibiting these characteristics have been generated. These include, e.g., 2′-*O*-methyl (2′-OMe) and 2′-*O*-methoxyethyl (2′-MOE) ASOs, as well as phosphorodiamidate morpholinos (PMOs). All these ASOs exhibit either a modified phosphate-backbone (e.g., phosphorothioate (PS) backbone), and modifications at the 2′ sugar position within their ribose moiety or a morpholine ring instead of the furanose ring (PMOs). Furthermore, because these mentioned ASOs can influence alternative splice site recognition when targeted against SREs or splice sites, they are also applied as so-called splice-switching oligonucleotides (SSOs) [23,24]. Indeed, already two FDA-approved antisense oligonucleotides were developed targeting either the exon 51 of the dystrophin pre-mRNA for exon skipping (a PMO ASO, Eteplirsen/Exondys 51), or an intronic SRE inducing exon 7 inclusion in the survival motor neuron protein 1 and 2 mRNAs (a 2′-MOE ASO, Nusinersen/Spinraza). Both ASO strategies are promising therapeutics for patients suffering from Duchenne muscular dystrophy (DMD) or spinal muscular atrophy (SMA) [24–27].

In addition to the aforementioned ASOs, locked nucleic acids (LNAs) are also promising and commonly used ASOs, which can either be used as SSOs to induce steric blocks (LNA mixmers, a combination of LNA and DNA residues), or to induce degradation of target mRNAs (LNA gapmers, with LNAs at the 5′- and 3′-end, and a DNA strand in the center of the ASO). Due to an extra methylene bridge between the 2′-O- and the 4′-C-atoms of the ribose moiety, LNAs are locked in the

ideal conformation for Watson–Crick binding, resulting in an increased melting temperature and an extreme antisense-target-duplex stability [28,29].

Furthermore, it was shown that LNAs can be administered without the use of transfection reagents in a process termed "gymnosis" and that human primary T-cells are susceptible to gymnotic LNA delivery [30–32]. For gymnotic cellular uptake a phosphorothioate- (PS) backbone is essential and therefore, the PS modification must be present in at least 75% of all phosphate linkage in ASOs like 2′-OMe, 2′-MOE or LNAs [33]. ASOs containing a PS-backbone enter the cells via a combination of adsorptive and fluid-phase endocytosis and later accumulate in multivesicular bodies, late endosomes and lysosomes. However, how LNAs escape these membrane vehicles to fulfil their purpose is poorly understood, and is still under investigation [34–37]. LNAs also have an increased stability against endo- and exonucleases, and are in a low toxicity range, hence displaying all characteristics for in vivo usage. Moreover, LNA compounds displaying good pharmacokinetics and –dynamics have been used as "naked" phosphorothioate-modified oligonucleotides, and have been successfully tested in many animal model systems (e.g., mice, rats, monkeys, chimpanzees) and humans [28,29]. Miravirsen for example, an LNA compound developed by the pharmaceutical company Santaris Pharma A/S, is currently in phase II of clinical trials, and masks the liver-specific micro-RNA miR-122. This LNA-based drug is a therapeutic agent for hepatitis C virus (HCV) infection because miR-122 serves as a crucial host factor for HCV [24,38,39].

Recently, we have shown that both splicing regulatory elements G_{I3}-2 and ESE_{tat} play a major role in the generation of viral mRNA species such as *vif*, *vpr* and *tat* mRNAs. Furthermore, we demonstrated that mutating these elements by site-directed mutagenesis or masking these elements by co-transfecting host cells with the proviral DNA and the respective LNA mixmer successfully interfered with viral pre-mRNA splicing and viral replication [13,15]. With regard to the development of an alternative antiretroviral therapy we gymnotically-delivered the G_{I3}-2 and ESE_{tat} LNA mixmers into HIV-1 infected cells and observed that "nakedly" delivered LNA mixmers localize within the cytoplasm, and induce degradation of viral mRNA species containing their target sequence rather than influencing recognition of adjacent splice sites. Consequently, gymnotically-delivered LNA mixmers efficiently block viral replication in HIV 1 infected Jurkat and PM1 cells as well as in PBMCs demonstrating their antiretroviral potential.

2. Results

2.1. Gymnotically-Delivered LNA Mixmers Binding the SREs G_{I3}-2 and ESE_{tat} Specifically Induce Degradation of Their Target mRNAs

Both viral splicing regulatory elements (SREs), G_{I3}-2 and ESE_{tat} (Figure 1a), localized within HIV-1 intron 3 and downstream of the viral SA3 respectively, are involved in regulating HIV-1 splice site usage which is essential for the generation of *tat* as well as *vpr* and *vif* mRNA species. In previous studies we were able to show that individual delivery of either locked nucleic acid (LNA) mixmer, masking the G_{I3}-2 or the ESE_{tat} element (Figure 1a), by transfection induced changes within the viral splicing pattern comparable to their mutational inactivation. Likewise, they also efficiently interfered with HIV-1 replication and RNA expression [13,15].

To test if the G_{I3}-2 and ESE_{tat} LNAs (Figure 1a) also affect the splicing pattern when delivered "nakedly" (via gymnosis) we added the LNAs (3 μM) to the culture medium of HIV-1 infected (laboratory strain NL4-3, MOI: 0.005) Jurkat cells without the use of any transfection reagent. To be able to compare the effect of the gymnotically-delivered LNA mixmers on the viral mRNA splicing pattern, we also transfected HeLa cells with both LNA mixmers (80 nM) as described in [13,15]. Twenty-four hours after LNA addition we isolated total RNA and performed RT-PCR analysis and investigated the viral splicing pattern.

After gymnotic delivery of the G_{I3}-2 or the ESE_{tat} LNA mixmers to HIV-1 infected Jurkat cells, both LNAs specifically induced a decrease only in their targeted transcript isoforms: for the G_{I3}-2 LNA mixmer, the *vpr*3 mRNA (Figure 1b, cf. lanes 1 and 2 [upper panels]) and for the ESE_{tat} LNA mixmer the *vpr3*, *tat1*, *tat2* and *tat3* mRNAs (Figure 1b cf. lanes 4 and 5 [upper panels]). This result

was in contrast to the previously described effect [13,15] of both delivered LNAs by transfection, which induced, comparable to the corresponding mutational inactivation of the G_{I3}-2 or ESE_{tat} SREs, an increase of *vpr* mRNA and reduced *tat* mRNA level (Figure 1b, cf. lanes 10 and 11 and lanes 13 and 14 [lower panels]; [13,15]). Because the *vpr3* mRNA contains the LNA target sequence for both LNA mixmers whereas the *tat* mRNAs only contain the LNA target sequence for the ESE_{tat} LNA (Figure 1b, cf. target vs. non-target) the impact of both gymnotically-delivered LNA mixmers on the viral mRNA expression indicated an LNA mixmer-induced degradation of their target mRNAs rather than affecting pre-mRNA splicing (Figure 1b, cf. target vs. non-target). A mismatch (MM) G_{I3}-2 LNA, which neither affected the expression of the *vpr3* mRNA nor the *tat* mRNAs (Figure 1b. cf. lanes 1 and 2 with lanes 7 and 8 [upper panels]) demonstrated its sequence-specificity.

LNA mixmers whereas the *tat* mRNAs only contain the LNA target sequence for the ESE_{tat} LNA (Figure 1b, cf. target vs. non-target) the impact of both gymnotically-delivered LNA mixmers on the viral mRNA expression indicated an LNA mixmer-induced degradation of their target mRNAs rather than affecting pre-mRNA splicing (Figure 1b, cf. target vs. non-target). A mismatch (MM) G_{I3}-2 LNA, which neither affected the expression of the *vpr3* mRNA nor the *tat* mRNAs (Figure 1b. cf. lanes 1 and 2 with lanes 7 and 8 [upper panels]) demonstrated its sequence-specificity.

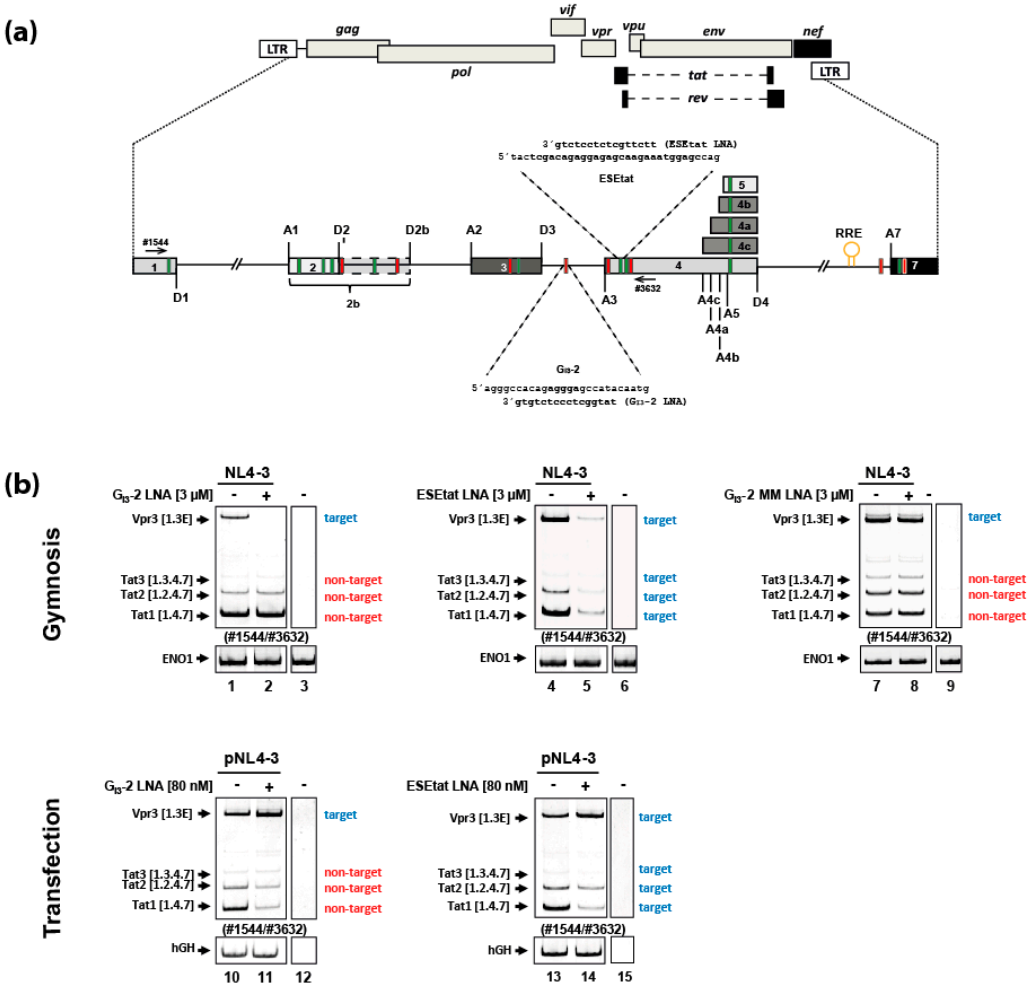


Figure 1. Gymnotically-delivered LNA mixmers targeting viral splicing regulatory elements (SREs) induce degradation of viral mRNAs containing their target sequence. (**a**) The diagram illustrates the HIV-1 proviral genome including open reading frames (ORFs) of viral proteins, long terminal repeats (LTRs), viral splice donor (D) and acceptor (A) sites as well as the viral splicing regulatory elements (enhancing SREs: Green; silencing SREs: Red). The Rev responsive element (RRE) for the Rev-dependent export of the intron-containing viral mRNA species is indicated by an open circle within the last HIV-1 intron. The primer binding sites (#1544, #3632; see Figure 1a) are also shown. The sequences as well as their target sequences of the G_{I3}-2 and ESE_{tat} LNAs are indicated. (**b**) Upper panel: Jurkat cells were infected with NL4-3 (MOI: 0.005) and the LNA mixmers were gymnotically-delivered by adding them at a concentration of 3 µM to the cell culture medium 6 h post infection. Cellular ENO1 was detected to monitor RT-PCR efficiency. Lower panel: HeLa cells were co-transfected with pNL4-3 and either the G_{I3}-2 or ESE_{tat} LNA mixmer. To monitor transfection and RT-PCR efficiencies, a plasmid expressing human growth hormone 1 (hGH) was co-transfected and hGH mRNA was detected using specific primers. Blue: Viral LNA target mRNAs; red: Viral LNA non-target mRNAs. HIV-1 mRNA species are indicated to the left of the gel images according to the nomenclature published previously [3]. Lanes 3, 6, 9, 12 and 15 represent mock samples running on the same gel. MM: Mismatch.

2.2. A Gymnotically-Delivered SRSF6 Exon/Junction LNA Mixmer Induces Degradation of the SRSF6 mRNA

To rule out that the LNA-mediated mode of action on RNA expression was only specific for LNAs directed against HIV-1 RNA, we tested an additional LNA mixmer targeting the cellular serine/arginine-rich splicing factor 6 (SRSF6) mRNA. Here, to analyze the impact of the LNA mixmer (SRSF6 D3 LNA) on SRSF6 splice donor usage, the LNA was complementary designed to mask the

splice donor 3 (D3) sequence of the SRSF6 pre-mRNA (Figure 2a). Furthermore, the LNA mixmer was 5′-labeled with a 6-Carboxyfluorescein (6-FAM™) modification to allow analysis of its cellular localization after delivery by using confocal laser scanning microscopy.

Int. J. Mol. Sci. 2019, 20, x FOR PEER REVIEW 6 of 18

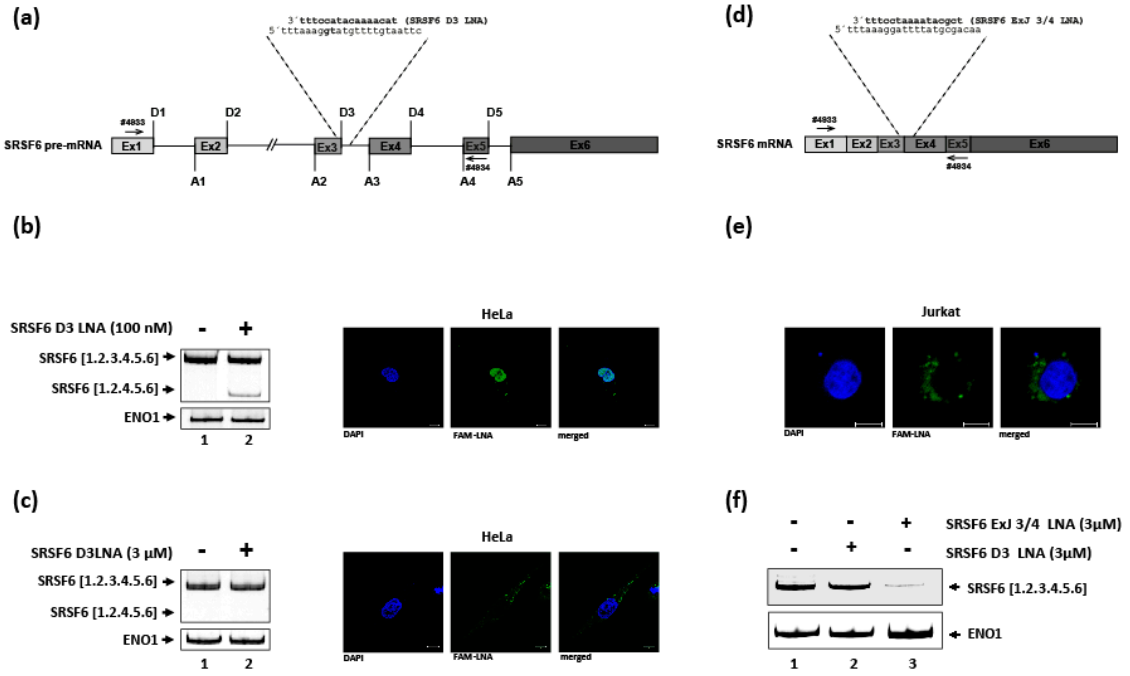


Figure 2. Influence of 6-FAM™-SRSF6 LNA mixmers on SRSF6 expression and their cellular localization. (**a**) Schematic drawing of the SRSF6 pre-mRNA with the indicated 6-FAM™-SRSF6 D3 LNA sequence as well as its target sequence and primer binding sites. (**b**) HeLa cells were transfected with the 6-FAM™-SRSF6 D3 LNA and after 24 h either total RNA was isolated to perform a RT-PCR (left panel), or the cells were fixed and the LNAs cellular localization was analyzed using confocal laser scanning microscopy (right panel; white bar: 10 µm). (**c**) The 6-FAM™-SRSF6 D3 LNA (3 µM) was gymnotically-delivered into HeLa cells, and after 48 h incubation, either total RNA was isolated (RT-PCR, left panel), or the cells were analyzed via confocal laser scanning microscopy (right panel; white bar: 10 µm). (**d**) Schematic drawing of the SRSF6 mRNA with the indicated 6-FAM™-SRSF6 ExJ 3/4 LNA sequence as well as its target sequence and primer binding sites. (**e**) Cellular localization of the 6-FAM™-SRSF6 ExJ 3/4 LNA in Jurkat cells after gymnotic delivery (white bar: 5 µm). (**f**) RT-PCR analysis after gymnotic delivery of the 6-FAM™-SRSF6 D3 and 6-FAM™-SRSF6 ExJ 3/4 LNA into Jurkat cells. Total RNA was isolated 48 h after LNA delivery.

2.3. The FAM-Labeled SRSF6 ExJ 3/4 LNA Mixmer Co-Localizes with GW-182 in HeLa and Jurkat Cells

First, as reference, we delivered the SRSF6 D3 LNA mixmer by transfecting HeLa cells followed by RT-PCR analysis. As expected, masking splice donor 3 competed with U1 snRNA binding, and thus led to SRSF6 exon 3 skipping (Figure 2b, cf. lanes 1 and 2). Moreover, a predominantly nuclear localization of the SRSF6 D3 LNA mixmer could be observed (Figure 2b, right panel).

Next, we wanted to analyze whether the "nakedly" delivered LNA mixmers and cytoplasmic GW-bodies known to contain RNA degrading enzymes co-localize. The reasons for addressing this question were first that glycine-tryptophan protein of 182 kDa (GW-182, GW-body marker) co-localized with gymnotically-delivered LNA gapmers targeting the ApoB or Bcl-2 mRNA in Huh-7 and HT 1080 cells [30,34]. Secondly, the gymnotically-delivered LNA mixmers used here also accumulated within the cytoplasm, and induced RNA degradation in HeLa and Jurkat cells which led to the assumption that the LNA mixmers might also co-localize with GW-bodies.

Therefore, we delivered the FAM-SRSF6 ExJ 3/4 LNA mixmer gymnotically, incubated HeLa or Jurkat cells for 48 h and after permeabilization and fixation, incubated the cells with a GW-182 antibody, and performed confocal laser scanning microscopy. Indeed, in both cell types, cytoplasmic co-localization of the LNA mixmer with GW-182 was observed to some extent (Figure 3, merged and co-localization, arrows). Although we detected several GW-bodies not co-localizing with the FAM-SRSF6 ExJ 3/4 LNA mixmer in HeLa cells (Figure 3, upper panel, red dots), co-localization was observed at cytoplasmic foci where the LNA mixmer accumulated within the cytoplasm (Figure 3, upper panel, FAM-LNA, merged and co-localization).

In contrast, when the SRSF6 D3 LNA was delivered gymnotically, skipping of exon 3 could not be detected (Figure 2c, cf. lanes 1 and 2). Furthermore, when analyzing the intracellular localization of the "nakedly" delivered SRSF6 D3 LNA mixmer, we detected fluorescence signals exclusively in the cytoplasm within a speckled distribution (Figure 2c, right panel). Because splicing takes place in the nucleus, this result plausibly explains why we could not detect any effects on SRSF6 exon 3 inclusion after gymnotic LNA delivery. Furthermore, LNA-mediated RNA degradation of the SRSF6 mRNA was expected not to be detectable, since the intronic portion of the LNA target sequence, the majority of the sequence of splice donor 3, is removed during nuclear splicing (Figure 2a).

To confirm this hypothesis, we designed an exon-exon-junction 5′ 6-FAM™ labeled SRSF6 LNA mixmer targeting the same 3′ end of exon 3, but instead of the downstream intronic sequence, it contains the 5′ end of the downstream exon 4 sequence which remains in the cytoplasmic mRNA after intron removal (SRSF6 ExJ 3/4 LNA, Figure 2d). Since we delivered the HIV targeting LNAs to HIV-1 infected Jurkat cells (Figure 1), we now gymnotically-delivered both SRSF6 LNA mixmers into this T-cell line. In line with the results obtained for HeLa cells, we again observed a cytoplasmic localization with a speckled distribution for both LNA mixmers (Figure 2e, and data not shown) confirming that at least within 48 h, LNA mixmers fail to enter the nucleus when gymnotically-delivered into HeLa or Jurkat cells.

Furthermore, as expected due to their cytoplasmic distribution, only gymnotically-delivered SRSF6 ExJ 3/4 LNA mixmer effectively induced SRSF6 mRNA degradation (Figure 2f, lane 3), substantiating our hypothesis that cytoplasmic-localized LNA mixmers induce degradation of mRNAs containing their target sequence.

2.3. The FAM-Labeled SRSF6 ExJ 3/4 LNA Mixmer Co-Localizes with GW-182 in HeLa and Jurkat Cells

Next, we wanted to analyze whether the "nakedly" delivered LNA mixmers and cytoplasmic GW-bodies known to contain RNA degrading enzymes co-localize. The reasons for addressing this question were first that glycine-tryptophan protein of 182 kDa (GW-182, GW-body marker) co-localized with gymnotically-delivered LNA gapmers targeting the ApoB or Bcl-2 mRNA in Huh-7 and HT 1080 cells [30,34]. Secondly, the gymnotically-delivered LNA mixmers used here also accumulated within the cytoplasm, and induced RNA degradation in HeLa and Jurkat cells which led to the assumption that the LNA mixmers might also co-localize with GW-bodies.

Therefore, we delivered the FAM-SRSF6 ExJ 3/4 LNA mixmer gymnotically, incubated HeLa or Jurkat cells for 48 h and after permeabilization and fixation, incubated the cells with a GW-182 antibody, and performed confocal laser scanning microscopy. Indeed, in both cell types, cytoplasmic co-localization of the LNA mixmer with GW-182 was observed to some extent (Figure 3, merged and co-localization, arrows). Although we detected several GW-bodies not co-localizing with the FAM-SRSF6 ExJ 3/4 LNA mixmer in HeLa cells (Figure 3, upper panel, red dots), co-localization was observed at cytoplasmic foci where the LNA mixmer accumulated within the cytoplasm (Figure 3, upper panel, FAM-LNA, merged and co-localization).

Co-localization could also be observed in Jurkat cells (Figure 3, lower panel, co-localization), though co-localization studies were more difficult due to their spherical shape and the unfavorable nucleus to cytoplasm ratio, which might explain why some foci appear to be in the nucleus.

In conclusion, these results suggest that enzymes localizing within GW-bodies might be involved in LNA mixmer-mediated degradation of mRNAs containing the LNA binding sequences.

Co-localization could also be observed in Jurkat cells (Figure3, lower panel, co-localization), though co-localization studies were more difficult due to their spherical shape and the unfavorable nucleus to cytoplasm ratio, which might explain why some foci appear to be in the nucleus.

In conclusion, these results suggest that enzymes localizing within GW-bodies might be involved in LNA mixmer-mediated degradation of mRNAs containing the LNA binding sequences.

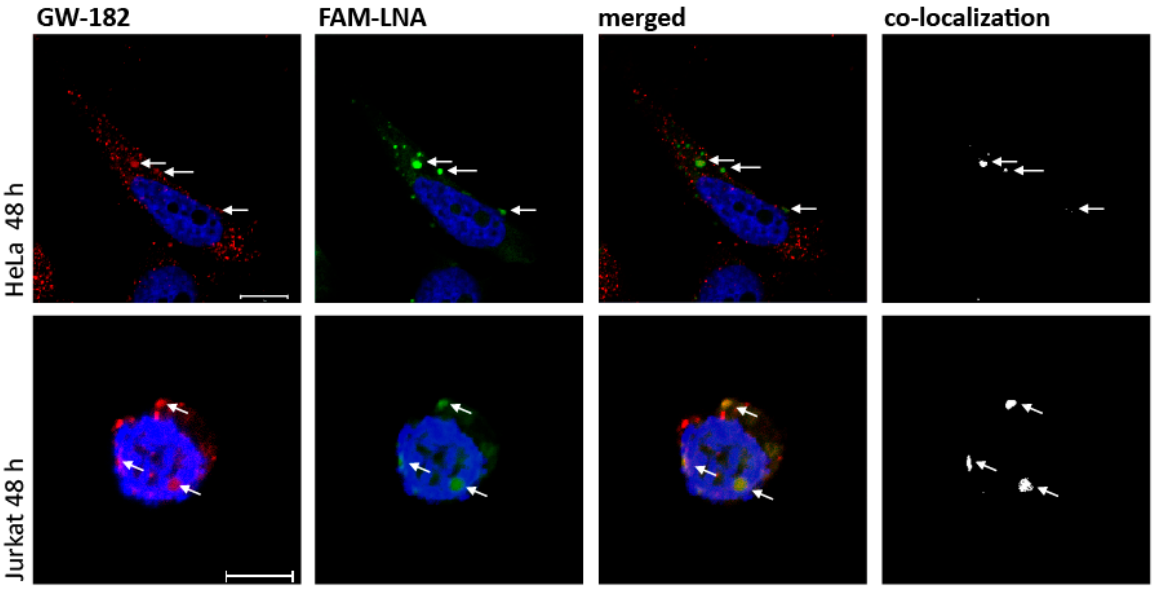


Figure 3. Co-localization of gymnotically-delivered 6-FAM™-SRSF6 ExJ 3/4 LNA mixmer with GW-182. HeLa or Jurkat cells were incubated with the 6-FAM™-SRSF6 ExJ 3/4 LNA, 48 h after gymnotic LNA delivery cells were fixed, permeabilized and incubated with a GW-182 antibody to detect GW-bodies. To monitor co-localization of the 6-FAM™-SRSF6 ExJ 3/4 LNA with GW-182, confocal laser scanning microscopy was performed. Red: GW-182; green: 6-FAM™-SRSF6 ExJ 3/4 LNA; white: Co-localization channel defining co-localized fluorescence. White bars: HeLa cells (10 µm), Jurkat cells (5 µm).

2.4. Gymnotic Delivery of both LNA Mixmers, G_{I3}-2 and ESE_{tat}, Efficiently Interferes with Viral RNA Expression and HIV-1 Replication in Infected T-Cells

So far, our results demonstrated that gymnotically-delivered LNA mixmers induced degradation of mRNAs containing their target sequence within the cytoplasm, suggesting that the gymnotic delivery of the G_{I3}-2 and ESE_{tat} LNA mixmers might interfere with HIV-1 replication.

Therefore, we next analyzed the impact of the gymnotically-delivered LNA mixmers on viral RNA expression and HIV-1 particle production. For this we infected different T-cell lines (Jurkat and PM1 cells) as well as primary T-cells (peripheral blood mononuclear cell (PBMC)) from a healthy donor, with the laboratory HIV-1 strain NL4-3 (MOI: 0.005). Six hours after infection we added the LNA mixmers in increasing concentrations (0.5, 1, 2 and 3 µM) to the cell culture medium. As an internal control we also applied the non SRSF6 RNA degrading SRSF6 D3 LNA (cf. Figure 2e,f) as this should not impair SRSF6 expression and thus not HIV-1 replication [15]. To analyze viral replication, we gymnotically-delivered LNA mixmers to HIV-1 infected cells and incubated the cells for 6 days. To measure the influence of the LNA mixmers on viral RNA expression we performed northern blot analysis using an HIV-1 exon 7 DIG-labeled probe. We performed immunoblot analysis detecting viral p24 Capsid (CA) protein (cellular and supernatant) to measure viral replication.

As shown in Figure 4, the "nakedly" delivered G_{I3}-2 (Figure 4a) and ESE_{tat} (Figure 4b) LNA mixmers dramatically interfered with viral RNA expression and viral p24 CA expression in Jurkat and PM1 cells as well as in PBMCs. Both LNA mixmers, at a concentration of 1 or 2 µM, reduced the amount of all three viral RNA classes and p24 CA protein levels in Jurkat cells and in PBMCs (Figure 4a,b, cf. lane 1 with lanes 3–5 and lane 13 with lanes 15–17). The LNA-mediated antiviral effect was even stronger in PM1 cells affecting viral RNA expression and replication already at a concentration of 0.5 µM (Figure 4a,b, cf. lane 7 with lanes 8–11). On the contrary, when the SRSF6 D3 LNA was gymnotically-delivered at the highest concentration (3 µM) we could not observe effects neither on viral RNA expression nor HIV-1 replication (Figure 4c), indicating that not the LNA mixmers *per se* affect viral replication but the specific binding of the G_{I3}-2 and ESE_{tat} LNA mixmers respectively to their target sequences.

mixmers *per se* affect viral replication but the specific binding of the G_{I3}-2 and ESE_{tat} LNA mixmers respectively to their target sequences.

In summary, both gymnotically-delivered LNA mixmers, G_{I3}-2 and ESE_{tat}, strongly interfered with HIV-1 RNA expression and replication in a low micromolar scale, which seems likely to be the result of the LNA-induced degradation of viral mRNA, demonstrating that unassisted delivered LNA mixmers efficiently interfere with HIV-1 replication in all host cells tested in this study.

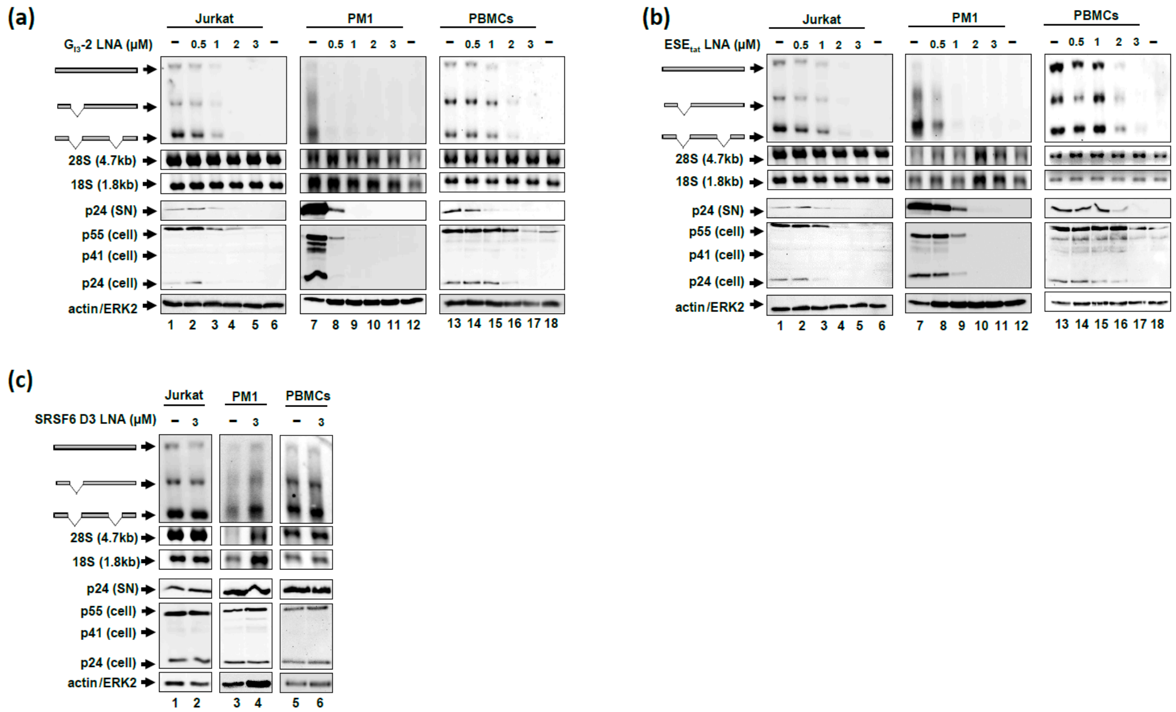


Figure 4. Interference of "naked" delivered G_{I3}-2 and ESE_{tat} LNA mixmers with HIV-1 replication. Six hours after infection (NL4-3, MOI 0.005) of Jurkat cells, PM1 cells or PBMCs with the laboratory strain NL4-3 the G_{I3}-2 (a) and ESE_{tat} (b) LNA mixmers were added to the medium (0.5, 1, 2 and 3 µM). The "naked" delivery of the SRSF6 D3 LNA (c, 3 µM) served as the control. Total RNA and protein as well as virions from the supernatant were harvested six days post-delivery. The three viral mRNA classes were detected performing northern blot analysis using a DIG-labeled HIV-1 exon 7 probe. Cellular and supernatant p24 CA protein levels were determined by immunoblot analysis using an anti-p24 CA antibody. Detection of actin (Jurkat cells and PBMCs) or ERK2 (PM1 cells) served as loading control. In a and b lanes 6, 12 and 18 represent mock samples.

2.5. The Antiretroviral Effect of the G_{I3}-2 and ESE_{tat} LNA Mixmers as Well as a Cocktail Consisting of Both LNAs Lasts up to Nearly 14 Days in Jurkat Cells and in PBMCs

Because the unassisted delivery of G_{I3}-2 and ESE_{tat} LNA mixmers displayed a strong antiretroviral effect, we next wanted to analyze for how long this inhibitory effect may last. For this, we set up a 14-day kinetic infecting Jurkat cells (NL4-3, MOI: 0.005) at day 1, and incubated the cells either with the G_{I3}-2 or ESE_{tat} LNA mixmers at a concentration of 3 µM 6 h post infection. At day 6, 10 and 14 total RNA and protein were isolated and viral RNA expression and replication was monitored performing northern and immunoblot analysis.

As shown in Figure 5a, after 6 and 10 days a clear impact on viral RNA and p24 CA protein expression for both LNAs could be observed (Figure 5a, 6 days and 10 days) and even after 14 days a slight influence was still visible (Figure 5a, 14 days), indicating that the LNA-mediated antiretroviral effect of the G_{I3}-2 and ESE_{tat} LNA mixmer lasts at least 10 days.

For this analysis we infected Jurkat cells and PBMCs from a healthy donor (NL4-3, MOI: 0.005) and repeated the kinetic using 1.5 µM of each LNA mixmer.

As before, we observed a comparable interference with viral RNA and p24 CA protein expression after 6 and 10 days (Figure 5b, 6 days and 10 days). Furthermore, in PBMCs even 14 days after LNA cocktail application, viral replication was severely impaired (Figure 5b, PBMCs, 14 days) demonstrating the potential of LNA mixmers as antiretroviral compound.

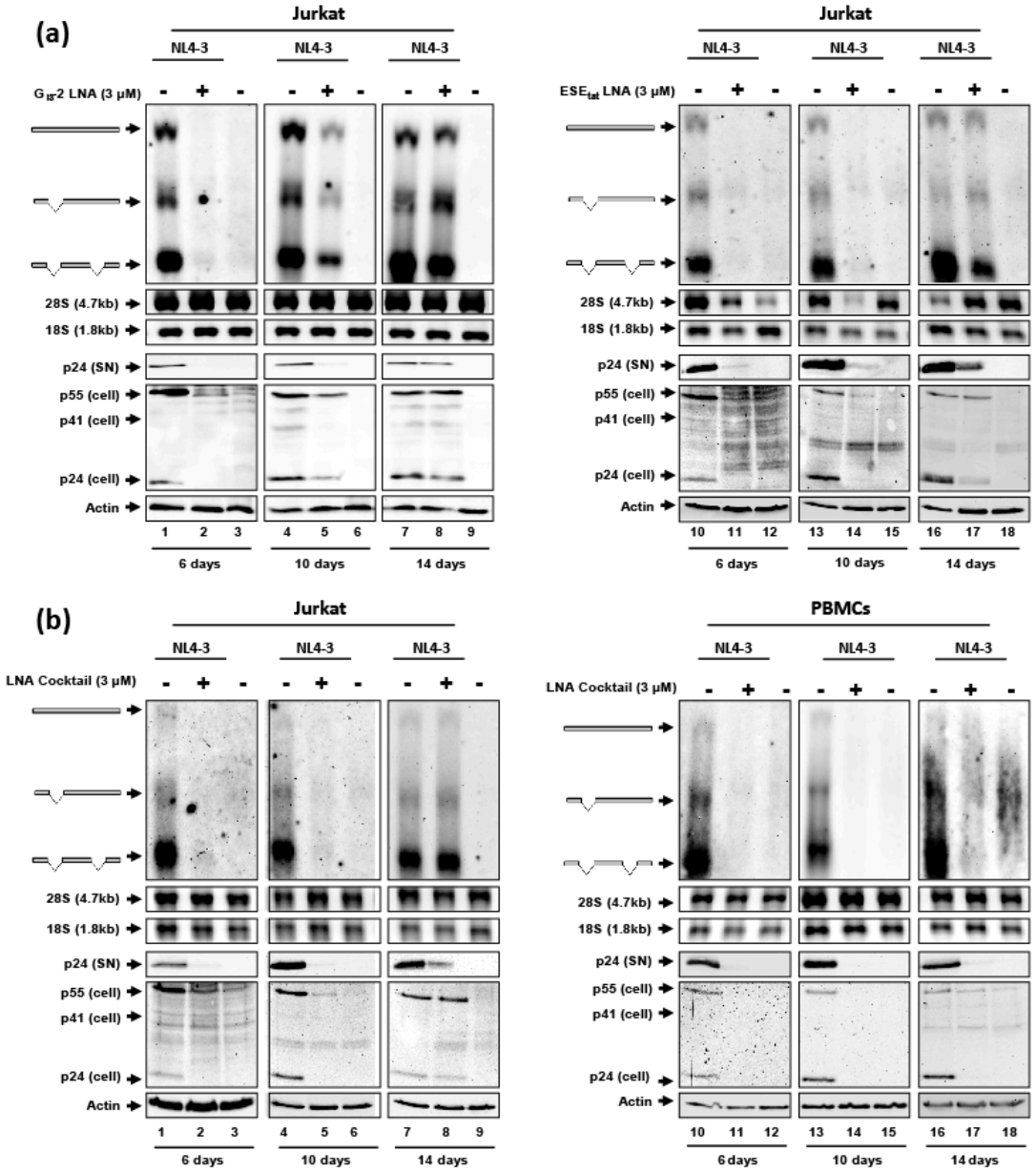


Figure 5. Antiretroviral effect of the gymnotically-delivered G_{15}-2 and ESE_{tat} LNA mixmers. Jurkat and PBMCs were infected with the laboratory HIV-1 strain NL4-3 (MOI: 0.005). Six hours post infection 3 µM of the G_{15}-2 or ESE_{tat} LNA (**a**) or both combined as an LNA cocktail (**b**, 1.5 µM of each LNA) were added to the cell culture medium. After 6 days, 10 days or 14 days total RNA and proteins, as well as virions from the supernatant were harvested. Viral mRNA classes were detected performing northern blot analysis using a DIG-labeled HIV-1 exon 7 probe. p24 CA protein levels were determined by immunoblot analysis using an anti-p24 CA antibody. Lanes 3, 6, 9, 12, 15 and 18 represent mock samples.

In addition, because ART combines substances with different effects on the HIV-1 replication cycle to minimize viral escape mutants, we finally tested if combining both LNAs targeting different HIV-1 sequences to a LNA mixmer cocktail also displays a strong inhibitory effect on viral replication. For this analysis we infected Jurkat cells and PBMCs from a healthy donor (NL4-3, MOI: 0.005) and repeated the kinetic using 1.5 µM of each LNA mixmer.

As before, we observed a comparable interference with viral RNA and p24 CA protein expression after 6 and 10 days (Figure 5b, 6 days and 10 days). Furthermore, in PBMCs even 14 days after LNA

cocktail application, viral replication was severely impaired (Figure 5b, PBMCs, 14 days) demonstrating the potential of LNA mixmers as antiretroviral compound.

3. Discussion

In this study we demonstrated that LNA mixmers delivered in the absence of any transfection reagent (gymnosis), targeting the HIV-1 splicing regulatory elements (SREs) G_{I3}-2 and ESE_{tat} induced degradation of HIV-1 mRNA species containing their target sequence in HIV-1 infected T-cells. As a consequence, HIV-1 RNA expression and HIV-1 replication was affected, underlining the concept to use LNA mixmers as potential therapeutic compounds for the development of an antiretroviral therapy.

Since the G_{I3}-2 and ESE_{tat} LNA mixmers as well as the SRSF6 D3 LNA mixmer localized within the nucleus and clearly influenced viral and cellular splicing after transfection (Figures 1b and 2b) [13,15], it was surprising to find that not only LNA gapmers (LNAs at the 5′- and 3′ end and a DNA strand in the center of the antisense oligonucleotide (ASO)), but also LNA mixmers (mixed combination of LNA and DNA residues within the ASO), normally known to induce steric blocks, localize in the cytoplasm, and are able to induce RNA degradation after gymnotic delivery (Figures 1b and 2f). This result suggested that the LNA mixmers intracellular localization (cytoplasmic vs. nuclear), and its effect on RNA expression (splice switching vs. RNA degradation), seems to depend on the mode of delivery. Although nuclear localization of gymnotically-delivered LNA gap/mixmers has been described [31,32,40], a predominantly cytoplasmic distribution after "naked" delivery of LNA gap/mixmers or other ASOs with phosphorothioate (PS) backbone seems to be the more prominent observation [30,34,41,42]. It was shown that chemical modifications of ASOs (e.g., PS-backbone, 2′-ribose-modification), as well as their sequences, can lead to interactions with intracellular proteins influencing the ASOs' intracellular distribution and pharmacological actions [42–44]. Furthermore, the stress induced response complex (SIRC) mediates the translocation of ASOs, siRNAs or miRNAs to the nucleus, and could be induced either by transfection or other chemicals e.g., arsenite (As III) [45]. However, it was assumed that gymnotic delivery, contrary to lipofection, may not be a significant stressor to trigger SIRC-mediated nuclear transfer, resulting in a cytoplasmic localization of ASOs [45], explaining why the LNA mixmers used in this study accumulate in the cytoplasm and induce RNA degradation instead of affecting splicing within the nucleus. Indeed, we were not able to detect any ESE_{tat} LNA-mixmer-induced RNA degradation of viral target mRNAs after gymnotic delivery, when the cells were treated with transfection reagents at any time during the course of the experiment, excluding synergistic effects of combining both delivery methods (data not shown).

At the moment we are not aware of any off-target effects caused by the LNA mixmers used in this study. However, due to the fact that chemical modifications or sequences of ASOs and hence LNAs can e.g., lead to interactions with intracellular proteins, and thereby induce unspecific effects [42–44], further investigations are required to exclude such off-target effects for the G_{I3}-2, ESE_{tat} and SRSF6 LNA mixmers.

So far, degradation of target mRNAs after gymnotic LNA delivery is only described for LNA gapmers whereas LNA mixmers are found to induce steric blocks and prevent protein binding, e.g., to induce switches within the splicing pattern [28,29]. Since LNA gapmers are known to induce RNase H1-mediated RNA degradation, which requires a gap of 7 to 10 neighboring deoxynucleotides within the LNA gapmer for noteworthy RNase H1 activity [46,47], it is highly unlikely that RNase H1 plays any role in an LNA-mixmer-induced RNA degradation. However, data obtained by Castanotto and colleagues [34] suggest that besides RNase H1, other proteins also involved in the cellular mRNA silencing machinery can be responsible for LNA gapmer-mediated RNA degradation.

The authors observed that gymnotically-delivered LNA gapmers displayed an identical cytoplasmic distribution as siRNAs delivered via transfection reagents, and that they co-localized with glycine-tryptophan protein of 182 kDa (GW-182), a finding we also observed when using LNA mixmers (Figure 3). Furthermore, they found that argonaute-2 (Ago-2), a protein involved in RNA interference

(RNAi) which also interacts with GW-182, can bind LNA gapmers via its piwi/argonaute/zwille (PAZ) domain. In addition, they showed that after silencing the intracellular Ago-2 expression, the LNA gapmers' function was impaired, indicating an involvement of Ago-2 in the LNA gapmers' mode of action. Although the endonucleolytic cleavage activity of Ago-2 seems not to be involved in the LNA-gapmer-mediated RNA degradation, Ago-2 apparently functions as an escort protein, and is also a component of stress-induced response complex (SIRC) [34,45,48,49]. Therefore, and in agreement with the postulated hypothesis by Castanotto et al. [34], our gymnotically-delivered LNAs, targeting SREs or exon junctions, either as LNA gapmers or LNA mixmers, might be taken up by the cell via adsorptive and fluid-phase endocytosis. After membrane trafficking and transiting the late endosomal membrane, the LNAs bind to target mRNAs and induce translational steric blocks, and this results in the formation of GW-bodies. This, in a similar way described for cellular miRNAs, might be followed by the LNA-mediated RNA degradation, including recruitment of exo/endo-ribonucleases (e.g., XRN1), decapping enzymes (e.g., DCP1 and 2), and deadenylase-complexes (e.g., CCR4–NOT complex) [50–52]. However, it needs further and extended investigations to unravel the mechanism responsible for cytoplasmic LNA gapmer/mixmer-mediated RNA degradation.

Any new approach either affecting viral splicing, or inducing degradation of viral mRNA in order to suppress HIV-1 replication would be desirable. Both strategies result in an imbalance of HIV-1 RNA expression, and consequently interfere with HIV-1 replication. With regard to further development of such an anti-HIV-1 strategy, identification of additional viral RNA sequences as targets for LNAs would be beneficial. Since the LNA mixmers tested here accumulate within the cytoplasm and induce RNA degradation after gymnotic delivery, new LNA target sequences also available and accessible within the cytoplasm, are suitable. Furthermore, these viral sequences should be highly conserved among HIV-1 groups and subtypes. Several studies demonstrated that HIV-1 RNA sequences involved in viral gene expression and viral particle production are promising targets for ASOs. These include the viral primer binding site, the viral dimerization site, the viral major splice donor 1, the gag start codon, as well as the guanosine-adenosine-rich (GAR) splicing regulatory elements (SRE) and the trans-activation response element (TAR) [19–22]. Nevertheless, all these sequences were targeted via ASOs in cell-free in vitro or transfection experiments and therefore, it is of great interest to analyze if targeting these sequences with LNA mixmers also induce viral mRNA degradation, and are also suitable targets upon gymnotic delivery.

To further follow an LNA-based antiretroviral strategy, a combination of LNAs, which are most effective in inhibiting viral replication, would be an additional next step. In both, Jurkat cells and PBMCs, the G_{I3}-2 and ESE_{tat} LNA-cocktail was very effective in inhibiting the viral replication (Figure 5b), indicating the potential of combining LNAs. The combination of different compounds interfering with different steps of the HIV-1 replication is successful with regard to viral escape in antiretroviral therapy (ART). Likewise, most of the aforementioned HIV-1 RNA sequences, as well as HIV-1 SREs and splice sites are highly conserved among different HIV subtypes. This brings along the advantage that an LNA cocktail, consisting of three to five efficient LNAs targeting different viral RNA sequences, may hamper the emergence of escape mutants. Although this has to be confirmed experimentally, an efficient LNA cocktail might be an alternative strategy for multi-drug resistant viruses and would be a promising addition or alternative to currently administered ART regimes.

4. Materials and Methods

4.1. Oligonucleotides

All locked nucleic acid (LNAs) antisense oligonucleotides used in this study were LNA/DNA mixmers of 16 nucleotide length with a phosphorothioate (PS) backbone. LNA mixmers were obtained from Qiagen/Exiqon (Hilden, Germany/Vedbaek, Denmark; see Table 1; red: Mismatch). Furthermore, the LNA content/positions within the LNA/DNA mixmers were also defined by Qiagen/Exiqon. Used Primer pairs were ordered from Metabion GmbH (see Table 2).

Table 1. LNA/DNA mixmer oligonucleotides used in this work.

LNA Oligonucleotide	Sequence	Design-ID (Exiqon)	Cat. No Qiagen
G_{I3}-2 [13]	TATGGCTCCCTCTGTG	164610	YCO0073444
ESE_{tat} [15]	TTCTTGCTCTCCTCTG	256589	YCO0073445
SRSF6 D3 (5′-end modiefied with 6-FAM™)	TACAAAACATACCTTT	319384	-
SRSF6 ExJ 3/4 (5′-end modiefied with 6-FAM™)	TCGCATAAAATCCTTT	548164	-
G_{I3}-1-MM-control	TTTGGCTCACTCCGTG	240758	-

Table 2. DNA oligonucleotides used for semi-quantitative RT-PCR.

mRNA Type	Primer No.	Primer Sequence
HIV-1 exon1-4 mRNAs	#1544 (exon1) #3632 (exon4)	5′ CTTGAAAGCGAAAGTAAAGC 3′ 5′ TGGATGCTTCCAGGGCTC 3′
HIV-1 exon7	#3387 #3388	5′ TTGCTCAATG CCACAGCCAT 3′ 5′ TTTGACCACT TGCCACCCAT 3′
SRSF6 mRNA	#4933 #4934	5′ GAGTTCGAGGACTCCCG 3′ 5′ TCTACTGCGGCTGCTCCT 3′
ENO1 mRNA	#4907 #4908	5′ CTGTGCCCAGTGGTGCT 3′ 5′ GACCTGAAGAACTCGGAGG 3′
hGH mRNA	#1224 #1225	5′ TCTTCCAGCCTCCCATCAGCGTTTGG 3′ 5′ CAACAGAAATCCAACCTAGAGCTGCT 3′

4.2. Cell Culture, Preparation of Virus Stocks, Infection Experiments and Gymnotic LNA Delivery

Jurkat and PM1 cells as well as PBMCs were cultured in RPMI1640 GlutaMax medium (Invitrogen, Carlsbad, Carlifornia, USA) containing 10% (*v*/*v*) fetal calf serum and 50 µg/mL each of both penicillin and streptomycin. PBMCs were additionally activated using 25 U/mL IL-2 (Roche, Basel, Switzerland). HeLa cells were maintained in Dulbecco's high glucose modified Eagle's medium (Invitrogen), supplemented with 10% (*v*/*v*) fetal calf serum, 50 µg/mL of penicillin and streptomycin each (Invitrogen).

Virus stocks were prepared as described in [13]. Multiplicity of infection (MOI) was determined by calculating the Tissue Culture Infection Dose of 50% (TCID50). 5×10^5 Jurkat and PM1 cells or 1×10^6 PBMCs were infected with the HIV-1 laboratory strain NL4-3 (MOI: 0.005) and, after six hours, cells were centrifuged, washed with PBS (Invitrogen) and resuspended in RPMI1640 GlutaMax medium (Invitrogen), containing only 10% (*v*/*v*) fetal calf serum. LNAs were then added to the cell culture medium at the indicated end concentration, respectively (gymnotic delivery). Jurkat cells, PM1 cells, and PBMCs were either incubated for 24 h, 48 h or 6, 10 or 14 days. The cell culture medium of PBMCs was supplemented with 25 U/mL IL-2 (Roche) every 3 days.

For gymnotic LNA delivery using HeLa cells, 1×10^4 cells per well (six-well plate) were cultured and maintained in Dulbecco's high glucose modified Eagle's medium (Invitrogen) supplemented only with 10% (*v*/*v*) fetal calf serum. LNAs were then added to the cell culture medium at the indicated end concentration and the cells were incubated for 24 h, 48 h or 6 days.

4.3. Transfection of LNA Antisense Oligonucleotides

2.5×10^5 HeLa cells per well were plated in a six-well plate and cultured in Opti-MEM medium containing 5% fetal calf serum. For LNA transfection, 80 nM (G_{I3}-2 or ESE_{tat} LNA) or 100 nM (SRSF6 D3 LNA) of the LNAs were mixed with 4 µl Lipofectamine 2000 (Invitrogen) and Opti-MEN medium as described in [13,15,16].

4.4. RNA and Protein Isolation

Cells were centrifuged and washed with PBS (Invitrogen). Total RNA was isolated using acid guanidinium thiocyanate-phenol-chloroform as described in [53]. For protein isolation, cells were lysed in RIPA buffer (25 mM Tris·HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, protease inhibitor cocktail (Roche)).

Furthermore, supernatants of infected cells were collected by sucrose centrifugation at 50,000× *g* for 1 h to quantify viral release by immunoblot analysis.

4.5. Immunoblot Analysis

Separation of proteins was performed by 15% SDS polyacrylamide gel electrophoresis. Subsequently, proteins were transferred to a nitrocellulose membrane (pore size, 0.45 µm; Protran, GE Healthcare, Chicago, Illinois, USA), and subjected to immunoblotting procedure. Membranes were probed with the respective primary antibodies: Sheep anti-p24 CA antibody (Aalto Bioreagents Ltd., Dublin, Ireland), mouse anti-β actin (Sigma-Aldrich, St. Louis, Missouri, USA) antibody, or anti-ERK2 antibody (Santa Cruz Biotechnology, Dallas, Texas, USA). After incubation with secondary antibodies (HRP-conjugated anti-mouse superclonal antibody (Invitrogen); HRP-conjugated anti-sheep antibody (Aalto Bioreagents Ltd.) the membrane was developed with ECL enhanced chemiluminescent reagent (GE Healthcare, Chicago, Illinois, USA).

4.6. RT-PCR-Analysis

For reverse transcription 2 µg of total RNA was subjected to cDNA synthesis. cDNA synthesis was performed for 1 h at 50 °C and 15 min at 72 °C by using 200 U Superscript III RNAse H Reverse Transcriptase (Invitrogen), 7.5 pmol oligo(dT)$_{12\text{-}18}$ (Roche) as primer, 20 U of RNAsin (Promega, Fitchburg, Wisconsin, USA) and 10 mM of each deoxynucleoside triphosphate (Qiagen, Hilden, Germany). For semi-quantitative analysis cDNA was used as a template for PCR reactions. To detect the different mRNAs the following primer pairs were used: HIV-1 viral mRNAs (#1544/#3632); SRSF6 mRNA (#4933/#4934); ENO1 mRNA (#4907/#4908) or hGH mRNA (#1224/#1225). PCR products were separated on 10% nondenaturing polyacrylamide gels, stained with ethidium bromide, and visualized with an Imager (INTAS science imaging, Göttingen, Germany).

4.7. Northern Blot Analysis

Total-RNA was harvested from HIV-1 infected cells as described above and separated by gel electrophoresis. Subsequently, the RNA was capillary blotted overnight onto a nylon membrane, UV cross-linked and pre-hybridized with DIG Easy Hyb hybridization solution (Roche) for 2 h at 55 °C. The specific DIG-labeled probe (HIV-1 exon 7; #3387/#3388) was hybridized at 55 °C overnight. Finally, the membrane was washed, blocked and probed with anti-Digoxigenin-AP Fab fragments (Roche) followed by detection of the RNA bands using CDP-Star for chemiluminescent reactions (Roche).

4.8. Confocal Laser Scanning Microscopy

4.8.1. Intracellular Localization of 6-FAM™-Labeled LNAs

To investigate the intracellular distribution of 6-FAM™-labeled LNAs after transfection, 1×10^4 HeLa were seeded on cover slips in 24 well plates and were transfected with the FAM-SRSF6 D3 LNA as described above.

For monitoring the 6-FAM™-LNA localization within cells after gymnotic LNA delivery, 3 µM FAM-SRSF6 D3 or FAM-SRSF6 ExJ 3/4 LNA were added to the cell culture medium of 1×10^4 HeLa or 5×10^5 Jurkat cells. After 24 h (transfection) or 48 h (gymnotic delivery) of incubation, cells were washed three times with PBS, fixed with 4% paraformaldehyde (*w*/*v*) in PBS (10 min, RT), washed again two times with PBS and incubated with DAPI (1:5000 in PBS) to stain the nuclei. Cells were then

washed two additional times with PBS. Cover slips with HeLa cells were fixed on glass slides with FluoromountG (Southern Biotech, Birmingham, Alabama, USA). The intracellular LNA localization was analyzed using a LSM780 confocal microscope (Zeiss, Oberkochen, Germany). Image analyses and processing was performed with the ZEN software (Zeiss).

4.8.2. Co-Localization of Gymnotically-Delivered 6-FAM™-Labeled LNAs with GW-182

1×10^4 HeLa cells seeded on cover slips or 5×10^5 Jurkat cells were incubated with 3 µM FAM-SRSF6 ExJ 3/4 LNA for 48 h. Cells were fixed with 4% paraformaldehyde (w/v) in PBS (10 min, RT), permeabilized with 0.02% saponin (v/v) in PBS (15 min, RT) and blocked with 1% BSA (v/v), 0.002% saponin (v/v) in PBS for 60 min.

For staining GW-182, cells were incubated with a monoclonal mouse anti-GW-182 antibody (Abcam, Cambridge, UK) diluted 1:200 in 1% BSA (v/v), 0.0002% saponin (v/v) in PBS. After 2 h the antibody-solution was removed, cells were washed three times with 0.0002% saponin (v/v) in PBS for 5 min and incubated with Cy3 conjugated AffiniPure anti-mouse IgG diluted 1:300 1% BSA (v/v), 0.0002% saponin (v/v) in PBS for 45 min. This was followed by two times washing with 0.0002% saponin (v/v) in PBS for 5 min, staining of nuclei with DAPI (Invitrogen) (1:5000 in PBS) for 3 min and two additional washing steps with PBS. Co-localization of the FAM-SRSF6 ExJ 3/4 LNA with GW-182 was analyzed (including the co-localization channel to define co-localized fluorescence) using the ZEN software (Zeiss).

5. Conclusions

As demonstrated by the two FDA-approved antisense oligonucleotide-therapeutics for Duchenne muscular dystrophy (DMD) and spinal muscular atrophy (SMA), Eteplirsen and Nusinersen [24,25,27], we now have the tools to approach "undruggable" targets by specifically affecting RNA expression, and thereby expression of proteins. This includes viruses and thereby viral infections where RNA expression can be specifically affected utilizing the ASO technology in order to inhibit viral replication.

Author Contributions: F.H. and H.S. conceived the study and designed the experiments. S.E. and M.W. were involved in study design. F.H. and P.N.O. performed the NL4-3 infection and gymnotic LNA experiments. F.H., P.N.O. and L.M. performed transfection, RT-PCR, northern blot and immunoblot analyses. D.D. and K.P. performed the confocal laser scanning analysis. F.H., L.M. and H.S. wrote the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG) (SCHA909/8-1), the Forschungskommission of the Heinrich-Heine-University, Düsseldorf (to H.S. and F.H) the Heinz-Ansmann Stiftung für AIDS-Forschung, Düsseldorf (to H.S. and F.H.) and the Jürgen Manchot Stiftung (to L.M., P.N.O and H.S.).

Acknowledgments: We thank Björn Wefers for technical assistance. We thank Jed Cohan for proofreading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

2′-MOE	2′-*O*-methoxyethyl
2′-OMe	2′-*O*-methyl
6-FAM™	6-Carboxyfluorescein
Ago-2	protein argonaute-2
ApoB	Apolipoprotein B
ART	antiretroviral therapy
ASO	antisense oligonucleotide
Bcl-2	Apoptosis regulator Bcl-2

CA	capsid
CCR4–NOT	carbon catabolite repressor 4- negative on TATA
D3	splice donor 3
DCP1/2	mRNA-decapping enzyme 1/2
DMD	Duchenne muscular dystrophy
ERK2	mitogen-activated protein kinase 1
ESEtat	exonic splicing enhancer tat
ExJ 3/4	exon junction exon3/exon4
FDA	U.S. Food and Drug Administration
Gapmer	LNAs at the 5′- and 3′ end and a DNA strand in the center of the ASO
GAR	guanosine-adenosine-rich exonic splicing enhancer
GI3-2	second G-run within HIV-1 intron 3
GW-182	glycine-tryptophan protein of 182 kDa
GW-body	cytoplasmic foci containing enzymes involved in RNA degradation and translational repression
HCV	hepatitis C virus
HIV-1	human immunodeficiency virus type 1
hnRNP	Heterogeneous nuclear ribonucleoproteins
IL-2	Interleukin-2
LNA	Locked nucleic acid
miRNA	microRNA
Mixmer	mixed combination of LNA and DNA residues within the ASO
PAZ domain	Piwi/Argonaute/Zwille domain
PBMC	peripheral blood mononuclear cell
PMO	phosphorodiamidate morpholinos
PS	phosphorothioate
RNAi	RNA interference
RRE	Rev responsive element
SA3	splice acceptor 3
SIRC	stress-induced response complex
siRNA	small interfering RNA
SMA	spinal muscular atrophy
SR	serine and arginine-rich protein
SRE	splicing regulatory element
SRSF6	serine/arginine-rich splicing factor 6
SSO	splice-switching oligonucleotide
TAR	trans-activation response element
U1 snRNA	U1 small nuclear ribonucleic acid
WHO	World Health Organization
XRN1	5′-3′ exoribonuclease 1

References

1. Pennings, P.S. HIV Drug Resistance: Problems and Perspectives. *Infect. Dis. Rep.* **2013**, *5* (Suppl. 1), e5. [CrossRef]
2. Loucif, H.; Gouard, S.; Dagenais-Lussier, X.; Murira, A.; Stager, S.; Tremblay, C.; Van Grevenynghe, J. Deciphering natural control of HIV-1: A valuable strategy to achieve antiretroviral therapy termination. *Cytokine Growth Factor Rev.* **2018**, *40*, 90–98. [CrossRef] [PubMed]
3. Purcell, D.F.; Martin, M.A. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* **1993**, *67*, 6365–6378. [PubMed]
4. Widera, M.; Erkelenz, S.; Hillebrand, F.; Krikoni, A.; Widera, D.; Kaisers, W.; Deenen, R.; Gombert, M.; Dellen, R.; Pfeiffer, T.; et al. An intronic G-run within HIV-1 intron 2 is critical for splicing regulation of vif-mRNA. *J. Virol.* **2013**, *87*, 2707–2720. [CrossRef] [PubMed]
5. Stoltzfus, C.M. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. *Adv. Virus Res.* **2009**, *74*, 1–40. [CrossRef] [PubMed]

6. Ocwieja, K.E.; Sherrill-Mix, S.; Mukherjee, R.; Custers-Allen, R.; David, P.; Brown, M.; Wang, S.; Link, D.R.; Olson, J.; Travers, K.; et al. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* **2012**, *40*, 10345–10355. [CrossRef] [PubMed]
7. Emery, A.; Zhou, S.; Pollom, E.; Swanstrom, R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* **2017**, *91*, e02515-16. [CrossRef] [PubMed]
8. Sertznig, H.; Hillebrand, F.; Erkelenz, S.; Schaal, H.; Widera, M. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology* **2018**, *516*, 176–188. [CrossRef] [PubMed]
9. Erkelenz, S.; Mueller, W.F.; Evans, M.S.; Busch, A.; Schoneweis, K.; Hertel, K.J.; Schaal, H. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* **2012**, *19*, 96–102. [CrossRef] [PubMed]
10. Madsen, J.M.; Stoltzfus, C.M. An exonic splicing silencer downstream of the 3′ splice site A2 is required for efficient human immunodeficiency virus type 1 replication. *J. Virol.* **2005**, *79*, 10478–10486. [CrossRef] [PubMed]
11. Exline, C.M.; Feng, Z.; Stoltzfus, C.M. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. *J. Virol.* **2008**, *82*, 3921–3931. [CrossRef] [PubMed]
12. Asang, C.; Hauber, I.; Schaal, H. Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic Acids Res.* **2008**, *36*, 1450–1463. [CrossRef] [PubMed]
13. Widera, M.; Hillebrand, F.; Erkelenz, S.; Vasudevan, A.; Munk, C.; Schaal, H. A functional conserved intronic G run in HIV-1 intron 3 is critical to counteract APOBEC3G-mediated host restriction. *Retrovirology* **2014**, *11*, 72. [CrossRef] [PubMed]
14. Erkelenz, S.; Poschmann, G.; Theiss, S.; Stefanski, A.; Hillebrand, F.; Otte, M.; Stuhler, K.; Schaal, H. Tra2-mediated recognition of HIV-1 5′ss D3 as a key factor in processing vpr-mRNA. *J. Virol.* **2013**, *87*, 721–2734. [CrossRef] [PubMed]
15. Erkelenz, S.; Hillebrand, F.; Widera, M.; Theiss, S.; Fayyaz, A.; Degrandi, D.; Pfeffer, K.; Schaal, H. Balanced splicing at the Tat-specific HIV-1 3′ss A3 is critical for HIV-1 replication. *Retrovirology* **2015**, *12*, 29. [CrossRef] [PubMed]
16. Brillen, A.L.; Walotka, L.; Hillebrand, F.; Muller, L.; Widera, M.; Theiss, S.; Schaal, H. Analysis of competing HIV-1 splice donor sites uncovers a tight cluster of splicing regulatory elements within exon 2/2b. *J. Virol.* **2017**, *91*, e00389-17. [CrossRef] [PubMed]
17. Goodchild, J.; Agrawal, S.; Civeira, M.P.; Sarin, P.S.; Sun, D.; Zamecnik, P.C. Inhibition of human immunodeficiency virus replication by antisense oligodeoxynucleotides. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 5507–5511. [CrossRef] [PubMed]
18. Matsukura, M.; Zon, G.; Shinozuka, K.; Robert-Guroff, M.; Shimada, T.; Stein, C.A.; Mitsuya, H.; Wong-Staal, F.; Cohen, J.S.; Broder, S. Regulation of viral expression of human immunodeficiency virus in vitro by an antisense phosphorothioate oligodeoxynucleotide against rev (art/trs) in chronically infected cells. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 4244–4248. [CrossRef] [PubMed]
19. Arzumanov, A.; Walsh, A.P.; Rajwanshi, V.K.; Kumar, R.; Wengel, J.; Gait, M.J. Inhibition of HIV-1 Tat-dependent trans activation by steric block chimeric 2′-O-methyl/LNA oligoribonucleotides. *Biochemistry* **2001**, *40*, 14645–14654. [CrossRef] [PubMed]
20. Lebars, I.; Richard, T.; Di Primo, C.; Toulme, J.J. LNA derivatives of a kissing aptamer targeted to the trans-activating responsive RNA element of HIV-1. *Blood Cells Mol. Dis.* **2007**, *38*, 204–209. [CrossRef] [PubMed]
21. Jakobsen, M.R.; Haasnoot, J.; Wengel, J.; Berkhout, B.; Kjems, J. Efficient inhibition of HIV-1 expression by LNA modified antisense oligonucleotides and DNAzymes targeted to functionally selected binding sites. *Retrovirology* **2007**, *4*, 29. [CrossRef] [PubMed]
22. Asparuhova, M.B.; Marti, G.; Liu, S.; Serhan, F.; Trono, D.; Schumperli, D. Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. *J. Gene Med.* **2007**, *9*, 323–334. [CrossRef] [PubMed]
23. Havens, M.A.; Hastings, M.L. Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res.* **2016**, *44*, 6549–6563. [CrossRef] [PubMed]

24. Crooke, S.T.; Witztum, J.L.; Bennett, C.F.; Baker, B.F. RNA-Targeted Therapeutics. *Cell Metab.* **2018**, *27*, 714–739. [CrossRef] [PubMed]
25. Hua, Y.; Vickers, T.A.; Okunola, H.L.; Bennett, C.F.; Krainer, A.R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet.* **2008**, *82*, 834–848. [CrossRef] [PubMed]
26. Singh, N.N.; Howell, M.D.; Androphy, E.J.; Singh, R.N. How the discovery of ISS-N1 led to the first medical therapy for spinal muscular atrophy. *Gene Ther.* **2017**, *24*, 520–526. [CrossRef] [PubMed]
27. Stein, C.A.; Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther. J. Am. Soc. Gene Ther.* **2017**, *25*, 1069–1075. [CrossRef] [PubMed]
28. Grunweller, A.; Hartmann, R.K. Locked nucleic acid oligonucleotides: The next generation of antisense agents? *Biodrugs Clin. Immunother. Biopharm. Gene Ther.* **2007**, *21*, 235–243. [CrossRef]
29. Lundin, K.E.; Hojland, T.; Hansen, B.R.; Persson, R.; Bramsen, J.B.; Kjems, J.; Koch, T.; Wengel, J.; Smith, C.I. Biological activity and biotechnological aspects of locked nucleic acids. *Adv. Genet.* **2013**, *82*, 47–107. [CrossRef] [PubMed]
30. Stein, C.A.; Hansen, J.B.; Lai, J.; Wu, S.; Voskresenskiy, A.; Hog, A.; Worm, J.; Hedtjarn, M.; Souleimanian, N.; Miller, P.; et al. Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic Acids Res.* **2010**, *38*, e3. [CrossRef] [PubMed]
31. Zhang, Y.; Qu, Z.; Kim, S.; Shi, V.; Liao, B.; Kraft, P.; Bandaru, R.; Wu, Y.; Greenberger, L.M.; Horak, I.D. Down-modulation of cancer targets using locked nucleic acid (LNA)-based antisense oligonucleotides without transfection. *Gene Ther.* **2011**, *18*, 326–333. [CrossRef] [PubMed]
32. Fazil, M.H.; Ong, S.T.; Chalasani, M.L.; Low, J.H.; Kizhakeyil, A.; Mamidi, A.; Lim, C.F.; Wright, G.D.; Lakshminarayanan, R.; Kelleher, D.; et al. GapmeR cellular internalization by macropinocytosis induces sequence-specific gene silencing in human primary T-cells. *Sci. Rep.* **2016**, *6*, 37721. [CrossRef] [PubMed]
33. Souleimanian, N.; Deleavey, G.F.; Soifer, H.; Wang, S.; Tiemann, K.; Damha, M.J.; Stein, C.A. Antisense 2′-Deoxy, 2′-Fluroarabino Nucleic Acids (2′F-ANAs) Oligonucleotides: In Vitro Gymnotic Silencers of Gene Expression Whose Potency Is Enhanced by Fatty Acids. *Mol. Ther. Nucleic Acids* **2012**, *1*, e43. [CrossRef] [PubMed]
34. Castanotto, D.; Lin, M.; Kowolik, C.; Wang, L.; Ren, X.Q.; Soifer, H.S.; Koch, T.; Hansen, B.R.; Oerum, H.; Armstrong, B.; et al. A cytoplasmic pathway for gapmer antisense oligonucleotide-mediated gene silencing in mammalian cells. *Nucleic Acids Res.* **2015**, *43*, 9350–9361. [CrossRef] [PubMed]
35. Akhtar, S.; Juliano, R.L. Cellular uptake and intracellular fate of antisense oligonucleotides. *Trends Cell Biol.* **1992**, *2*, 139–144. [CrossRef]
36. Koller, E.; Vincent, T.M.; Chappell, A.; De, S.; Manoharan, M.; Bennett, C.F. Mechanisms of single-stranded phosphorothioate modified antisense oligonucleotide accumulation in hepatocytes. *Nucleic Acids Res.* **2011**, *39*, 4795–4807. [CrossRef] [PubMed]
37. Juliano, R.L. Intracellular Trafficking and Endosomal Release of Oligonucleotides: What We Know and What We Don't. *Nucleic Acid Ther.* **2018**, *28*, 166–177. [CrossRef] [PubMed]
38. Janssen, H.L.; Reesink, H.W.; Lawitz, E.J.; Zeuzem, S.; Rodriguez-Torres, M.; Patel, K.; van der Meer, A.J.; Patick, A.K.; Chen, A.; Zhou, Y.; et al. Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* **2013**, *368*, 1685–1694. [CrossRef] [PubMed]
39. Ottosen, S.; Parsley, T.B.; Yang, L.; Zeh, K.; van Doorn, L.J.; van der Veer, E.; Raney, A.K.; Hodges, M.R.; Patick, A.K. In vitro antiviral activity and preclinical and clinical resistance profile of miravirsen, a novel anti-hepatitis C virus therapeutic targeting the human factor miR-122. *Antimicrob. Agents Chemother.* **2015**, *59*, 599–608. [CrossRef] [PubMed]
40. Morelli, E.; Biamonte, L.; Federico, C.; Amodio, N.; Di Martino, M.T.; Gallo Cantafio, M.E.; Manzoni, M.; Scionti, F.; Samur, M.K.; Gulla, A.; et al. Therapeutic vulnerability of multiple myeloma to MIR17PTi, a first-in-class inhibitor of pri-mir-17-92. *Blood* **2018**, *132*, 1050–1063. [CrossRef] [PubMed]
41. Liang, X.H.; Shen, W.; Sun, H.; Prakash, T.P.; Crooke, S.T. TCP1 complex proteins interact with phosphorothioate oligonucleotides and can co-localize in oligonucleotide-induced nuclear bodies in mammalian cells. *Nucleic Acids Res.* **2014**, *42*, 7819–7832. [CrossRef] [PubMed]
42. Bailey, J.K.; Shen, W.; Liang, X.H.; Crooke, S.T. Nucleic acid binding proteins affect the subcellular distribution of phosphorothioate antisense oligonucleotides. *Nucleic Acids Res.* **2017**, *45*, 10649–10671. [CrossRef] [PubMed]

43. Liang, X.H.; Sun, H.; Shen, W.; Crooke, S.T. Identification and characterization of intracellular proteins that bind oligonucleotides with phosphorothioate linkages. *Nucleic Acids Res.* **2015**, *43*, 2927–2945. [CrossRef] [PubMed]
44. Vickers, T.A.; Crooke, S.T. Development of a Quantitative BRET Affinity Assay for Nucleic Acid-Protein Interactions. *PLoS ONE* **2016**, *11*, e0161930. [CrossRef] [PubMed]
45. Castanotto, D.; Zhang, X.; Alluin, J.; Zhang, X.; Ruger, J.; Armstrong, B.; Rossi, J.; Riggs, A.; Stein, C.A. A stress-induced response complex (SIRC) shuttles miRNAs, siRNAs, and oligonucleotides to the nucleus. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5756–E5765. [CrossRef] [PubMed]
46. Kurreck, J.; Wyszko, E.; Gillen, C.; Erdmann, V.A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **2002**, *30*, 1911–1918. [CrossRef] [PubMed]
47. Frieden, M.; Christensen, S.M.; Mikkelsen, N.D.; Rosenbohm, C.; Thrue, C.A.; Westergaard, M.; Hansen, H.F.; Orum, H.; Koch, T. Expanding the design horizon of antisense oligonucleotides with alpha-L-LNA. *Nucleic Acids Res.* **2003**, *31*, 6365–6372. [CrossRef] [PubMed]
48. Jakymiw, A.; Lian, S.; Eystathioy, T.; Li, S.; Satoh, M.; Hamel, J.C.; Fritzler, M.J.; Chan, E.K. Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell Biol.* **2005**, *7*, 1267–1274. [CrossRef] [PubMed]
49. Siomi, H.; Siomi, M.C. RISC hitches onto endosome trafficking. *Nat. Cell Biol.* **2009**, *11*, 1049–1051. [CrossRef] [PubMed]
50. Eulalio, A.; Behm-Ansmant, I.; Izaurralde, E. P bodies: At the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 9–22. [CrossRef] [PubMed]
51. Eulalio, A.; Huntzinger, E.; Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. *Cell* **2008**, *132*, 9–14. [CrossRef] [PubMed]
52. Wu, L.; Belasco, J.G. Let me count the ways: Mechanisms of gene regulation by miRNAs and siRNAs. *Mol. Cell* **2008**, *29*, 1–7. [CrossRef] [PubMed]
53. Chomczynski, P.; Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **1987**, *162*, 156–159. [CrossRef]

3.3. Viral pathogenesis and immunological features of SARS-CoV-2

In December 2019, an unknown lung disease was registered by authorities in China's Wuhan province. In March 2020, after the novel coronavirus SARS-CoV-2 was identified as the trigger of the disease, the World Health Organization (WHO) declared the spread of the virus a pandemic. In a global effort, scientists shifted their research towards the understanding of the viruses' peculiar features in order to accelerate the understanding of clinical symptoms, reliable diagnostics, treatment options, and vaccine development. In this context, the knowledge acquired by then on how to deal with viruses that pose a serious threat to humans was also used to help combat the pandemic.

The zoonotic beta-coronavirus is genetically similar to severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), which caused outbreaks in 2002/2003 and 2012 respectively. However, SARS-CoV-2 and the COVID-19 disease it causes have several unique characteristics, including a highly variable incubation period, a wider range of symptoms, and viral replication starting in the upper respiratory tract, which contributes to the rapid spread of the virus in the population. While the clinical course of COVID-19 is usually mild, often with flu-like symptoms, patients can develop a severe course of infection. The elderly population is primarily at risk, as adults over 65 years of age account for the majority of hospitalizations, however, other risk factors including chronic illness such as diabetes or obesity are known.

With only limited treatment options available, their evaluation and development as well as the generation of prophylactic vaccinations were, and still are, the focus of global research in order to ease the burden of the pandemic.

3.3.1. SARS-CoV-2 targets neurons of 3D human brain organoids (Chapter 9)

While the main symptoms of COVID-19 present as respiratory illnesses, early reports indicate that also neurological symptoms are associated with the infection, suggesting an influence of SARS-CoV-2 on the central nervous system. To elucidate this step during infection, 3D human brain organoids were used as a model system and inoculated with an infectious SARS-CoV-2 isolate that was generated from a patient's swab. In this system, SARS-CoV-2 preferably targeted neurons of the brain organoids where an altered distribution of Tau from the neuronal axons to soma, hyperphosphorylation, and apparent neuronal death were seen. These data add to the better understanding of the underlying mechanism behind neurological symptoms upon SARS-CoV-2 infection and emphasize that brain organoids could aid as model systems to elucidate the pathologies of COVID-19.

The following article is published in EMBO J. 2020 Oct 15;39(20):e106230. doi: 10.15252/embj.2020106230. by

Ramani A*, **Müller L***, Ostermann PN*, Gabriel E, Abida-Islam P, Müller-Schiffmann A, Mariappan A, Goureau O, Gruell H, Walker A, Andrée M, Hauka S, Houwaart T, Dilthey A, Wohlgemuth K, Omran H, Klein F, Wieczorek D, Adams O, Timm J, Korth C, Schaal H#, Gopalakrishnan J#.

* shared first authors

shared senior authors

Contributions:

AR and JG conceptualize the project. AR designed coordinated and conducted experiments. LM and PNO performed virology works. HS supervised virology. EG, PA-I, AM-S and AM provided technical support. AM-S and CK purified antibodies. OG provided iPS cells. HG, DW and FK provided convalescent serum. KW and HO provided airway cells. AW, MA, SH, TH, AD, OA and JT involved in sequencing and viral strains. JG and HS supervised the work. AR and JG wrote the manuscript.

Individual contribution: 20% L.M. established and performed infection experiments as basis for all figures.

Article

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

SARS-CoV-2 targets neurons of 3D human brain organoids

Anand Ramani[1,†], Lisa Müller[2,†], Philipp N Ostermann[2,†], Elke Gabriel[1], Pranty Abida-Islam[1], Andreas Müller-Schiffmann[3], Aruljothi Mariappan[1], Olivier Goureau[4], Henning Gruell[5], Andreas Walker[2], Marcel Andrée[2], Sandra Hauka[2], Torsten Houwaart[6], Alexander Dilthey[6], Kai Wohlgemuth[7], Heymut Omran[7], Florian Klein[5,8,9], Dagmar Wieczorek[1], Ortwin Adams[2], Jörg Timm[2], Carsten Korth[3], Heiner Schaal[2,*] & Jay Gopalakrishnan[1,**]

Abstract

COVID-19 pandemic caused by SARS-CoV-2 infection is a public health emergency. COVID-19 typically exhibits respiratory illness. Unexpectedly, emerging clinical reports indicate that neurological symptoms continue to rise, suggesting detrimental effects of SARS-CoV-2 on the central nervous system (CNS). Here, we show that a Düsseldorf isolate of SARS-CoV-2 enters 3D human brain organoids within 2 days of exposure. We identified that SARS-CoV-2 preferably targets neurons of brain organoids. Imaging neurons of organoids reveal that SARS-CoV-2 exposure is associated with altered distribution of Tau from axons to soma, hyperphosphorylation, and apparent neuronal death. Our studies, therefore, provide initial insights into the potential neurotoxic effect of SARS-CoV-2 and emphasize that brain organoids could model CNS pathologies of COVID-19.

Keywords brain organoids; cell death; neurons; SARS-CoV-2; Tau pathology
Subject Categories Microbiology, Virology & Host Pathogen Interaction; Molecular Biology of Disease; Neuroscience
DOI 10.15252/embj.2020106230 | Received 13 July 2020 | Revised 28 August 2020 | Accepted 31 August 2020 | Published online 23 September 2020
The EMBO Journal (2020) 39: e106230

Introduction

The novel coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is spreading worldwide, and the outbreak continues to rise, posing a severe emergency (Worl Health Organization, 2020) . Understanding the biology of the current COVID-19 pandemic is a high priority for combatting it efficiently. Thus, it is essential to gain initial insights into the infection mechanisms of SARS-CoV-2, including its target cell types and tropism, to contain its short- and long-term effects on human health. Furthermore, it is vital to establish an experimental system that could allow designing measures on how to stop viral replication and protect human health rapidly. However, practical problems associated with the isolation and handling of highly infective viral strains and lack of reliable *in vitro* human model systems that can efficiently model COVID-19 hamper these efforts.

Clinical symptoms of COVID-19 patients include upper respiratory tract infection with fever, dry cough, and dyspnea, indicating that the respiratory tract is the first target (Yang *et al*, 2020b). However, emerging case reports showed that patients infected with SARS-CoV-2 suffered a sudden and complete loss of the olfactory function, stroke, and other severe neurological symptoms (Chen *et al*, 2020; Helms *et al*, 2020; Poyiadji *et al*, 2020; Sedaghat & Karimi, 2020; Virani *et al*, 2020). All of these indicate that SARS-CoV-2 could infect the central nervous system (CNS) and is therefore neurotropic (Baig *et al*, 2020; Conde Cardona *et al*, 2020; De Felice *et al*, 2020). Earlier studies show that SARS-CoV target the brains of mice, and since the different coronaviruses share a similar structure, it is likely that SARS-CoV-2 exhibits the same infection mechanism and possibly invades into the brain (McCray *et al*, 2007). Indeed, a clinical report detected the presence of viral RNA in autopsy of brain samples (Puelles *et al*, 2020). Furthermore, a postmortem brain MRI analysis has identified the presence of hemorrhagic and encephalopathy syndromes suggesting that

1 Institute of Human Genetics, University Hospital Düsseldorf, Heinrich-Heine-Universität, Düsseldorf, Germany
2 Institute of Virology, Medical Faculty, University Hospital Düsseldorf, Heinrich-Heine-Universität, Düsseldorf, Germany
3 Institute of Neuropathology, University Hospital Düsseldorf, Heinrich-Heine-Universität, Düsseldorf, Germany
4 Institut de la Vision, Sorbonne Université, INSERM, CNRS, Paris, France
5 Institute of Virology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Cologne, Germany
6 Institute of Medical Microbiology and Hygiene, University Hospital, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
7 Department of General Pediatrics, University Children's Hospital Muenster, Muenster, Germany
8 German Center for Infection Research (DZIF), partner site Bonn-Cologne, Cologne, Germany
9 Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany
*Corresponding author. Tel: +49 211 811 2393; E-mail: schaal@uni-duesseldorf.de
**Corresponding author. Tel: +49 211 8111561; E-mail: jay.gopalakrishnan@hhu.de
†These authors contributed equally to this work

SARS-CoV-2 infection could cause neuronal stress and inflammations (Coolen *et al*, 2020). Thus, at this point, it is of utmost priority to test whether SARS-CoV-2 directly infects human neurons and productively replicates in the CNS.

To investigate the potential neurotropism of SARS-CoV-2, it is essential to employ a suitable *in vitro* human model system that recapitulates the physiological effects of SARS-CoV-2 infection. In this regard, the recently emerged human brain organoids that closely parallel the complex neural epithelium exhibiting a wide diversity of cell types could serve as a suitable model system to test the neurotoxic effects of SARS-CoV-2. Induced pluripotent stem cells (iPSCs)-derived human brain organoids have revealed useful insights into human brain development and helped to model a variety of neurological disorders(Lancaster *et al*, 2013; Gabriel *et al*, 2016; Birey *et al*, 2017; Gabriel & Gopalakrishnan, 2017; Xiang *et al*, 2017; Goranci-Buzhala *et al*, 2020). Notably, others and our work using brain organoids have revealed unprecedented insights into infection mechanisms, target cell types, and the toxicity effects of the Zika virus (ZIKV) during the recent ZIKV epidemic (Cugola *et al*, 2016; Qian *et al*, 2016; Gabriel *et al*, 2017). These studies validate organoids as a tool for studying not only genetic but also environmental hazards to the human brain.

Here, we report that SARS-CoV-2 readily targets neurons of 3D human brain organoids. Neurons invaded with SARS-CoV-2 at the cortical area display altered distribution of Tau, Tau hyperphosphorylation, and apparent neuronal death. Moreover, we show that although SARS-CoV-2 can readily target brain organoids, SARS-CoV-2 does not appear to efficiently replicate, suggesting that the CNS may not support the active replication of SARS-CoV-2.

Results

Isolation of an infectious SARS-CoV-2 virus

We isolated SARS-CoV-2 (SARS-CoV-2 NRW-42) from a nasopharyngeal and oropharyngeal swab specimen of an infected patient admitted to our university hospital, University of Düsseldorf (see Materials and Methods section for culturing and propagation). To investigate whether SARS-CoV-2 replicates in inoculated African green monkey kidney cells (Vero CCL-81), we performed real-time quantitative polymerase chain reaction (qPCR) analysis with cell culture supernatant. The amount of SARS-CoV-2 RNA drastically increased from 0-dpi until 3-dpi (Appendix Fig S1A). Next, we analyzed the infectivity of generated SARS-CoV-2 particles by propagating virus-containing supernatant to yet uninfected Vero cells. We confirmed the infection of new Vero cells by the emergence of virus-induced cytopathic effects (CPEs) and an increase in SARS-CoV-2 RNA over 4-dpi. The sequence (access number PRJNA627229 at the European Nucleotide Archive and the Sample accession number for NRW-42 which is SRS6522060) showed only eight nucleotide exchanges compared to SARS-CoV-2 Wuhan-Hu-1 isolate.

Isolation and validation of COVID-19 convalescent serum to detect SARS-CoV-2 infection

As of April 1, 2020, we could not procure commercial antibodies that can specifically determine SARS-CoV-2 infection. Therefore, we isolated COVID-19 convalescent serum and tested if they can specifically recognize SARS-CoV-2 infections in our experiments. We obtained blood samples of four independent individuals who recently recovered from COVID-19 (AB1, AB2, AB3, and AB4). Testing them in an enzyme-linked immunosorbent assay (ELISA) that used the SARS-CoV-2 S1 domain of the spike protein as an antigen revealed that, except for AB2, the rest of the convalescent serum contained SARS-CoV-2-specific IgG (Appendix Fig S1B). We then affinity-purified the convalescent serum against the full length ORF of SARS-CoV-2-N (see Materials and Methods section). In Western blots, which used extracts of brain organoids and Vero cells exposed to SARS-CoV-2, the antibodies affinity-purified from convalescent serum specifically recognized a signal similar to the size of the nucleoprotein of SARS-CoV-2. The recombinant SARS-CoV-2-N serves as a positive control in this experiment (Appendix Fig S1C). The convalescent serum AB4 also specifically recognized SARS-CoV-2-infected Vero cells. To further validate the specificity of the AB4, we performed co-immunostaining with a mouse monoclonal anti-SARS-CoV-2 S and a polyclonal anti-SARS-CoV-2 NP. As expected, all of these antibodies recognized only the SARS-CoV-2-infected Vero cells (Appendix Fig S2A). Similarly, AB4 could specifically recognize somas of SARS-CoV-2-positive cells in SARS-CoV-2 exposed brain organoids which were further labeled by the monoclonal anti-SARS-CoV-2 S antibody (Appendix Fig S2B). In Western blots that used SARS-CoV-2-exposed organoid extracts, both AB4 and mouse monoclonal antibodies recognized protein bands around 50 and 180 kDs, sizes similar to the nucleoprotein and uncleaved spike proteins Together, these experiments validate that AB4 detects SARS-CoV-2 infection (Appendix Fig S2C).

SARS-CoV-2 targets neurons of human brain organoids

Before we infected our 3D human brain organoids with the new SARS-CoV-2 NRW-42 isolate, we first tested if our experimental conditions are suitable to infect the well-studied ciliated human respiratory epithelial cells (hRECs), an apparent target for the SARS-CoV-2 (Lamers *et al*, 2020). We noticed that SARS-CoV-2 readily targets hRECs within 2 days of virus exposure (Fig 1A). We then tested if SARS-CoV-2 could infect 3D human brain organoids. To do this, we adapted our previously described protocol and differentiated brain organoids from two different iPSC lines (Donor 1, IMR90 and Donor 2, Crx-iPS; Gabriel *et al*, 2017). In brief, we started with 10,000 iPSCs and induced differentiation into neural epithelium directly using SB431542 and dorsomorphin, the TGF beta and BMP4 inhibitors, respectively. Our differentiation condition did not also include an exogenous addition of retinoic acid, which could activate retinoic acid receptors (RAR) and induce an aberrant neuronal differentiation (Janesick *et al*, 2015; Gabriel *et al*, 2016, 2017; Gabriel & Gopalakrishnan, 2017). As this method skips embryoid bodies formation, it reduces the heterogeneity in organoid sizes simultaneously avoiding the formation of mesoderm and endoderm, which are not required for ectodermal differentiation at early stages of differentiation (Streit *et al*, 2000). As described before, organoids exhibit their specific neuronal cell types, which are spatially restricted. The ventricular zone (VZ) harbors proliferating neural progenitors cells (NPCs) that display typically elongated nuclei which align to form a lumen, a neural tube-like structure. Cortical neurons are positioned basally to the VZ, forming a cortical plate

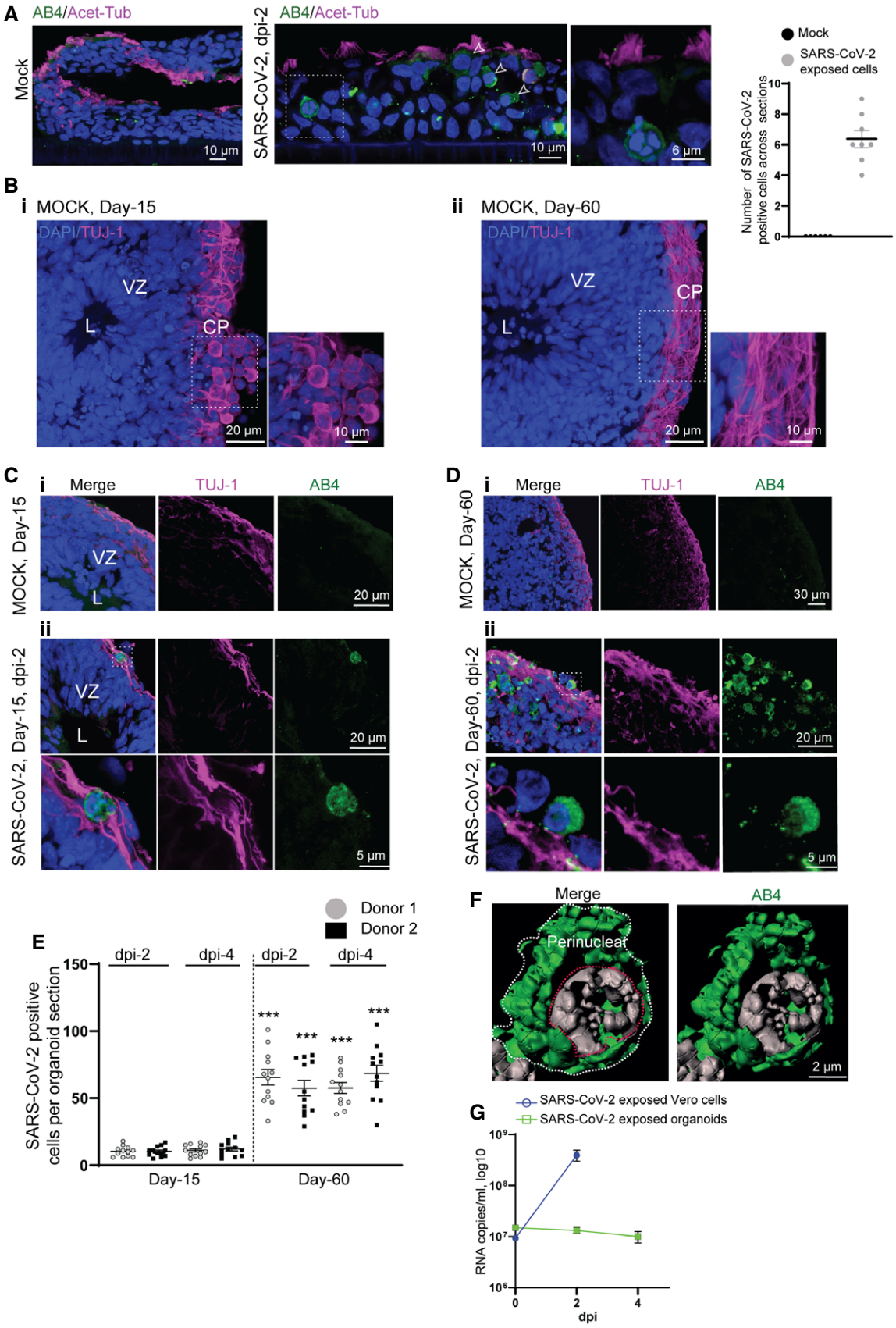

A
AB4/Acet-Tub
Mock
SARS-CoV-2, dpi-2
10 µm
6 µm
Mock
SARS-CoV-2 exposed cells
Number of SARS-CoV-2 positive cells across sections
B
i MOCK, Day-15
ii MOCK, Day-60
DAPI/TUJ-1
VZ
L
CP
20 µm
10 µm
C
i Merge TUJ-1 AB4
MOCK, Day-15
SARS-CoV-2, Day-15, dpi-2
20 µm
5 µm
D
i Merge TUJ-1 AB4
MOCK, Day-60
SARS-CoV-2, Day-60, dpi-2
30 µm
20 µm
5 µm
E
Donor 1
Donor 2
dpi-2
dpi-4
SARS-CoV-2 positive cells per organoid section
Day-15
Day-60
***
F
Merge
AB4
Perinuclear
2 µm
G
SARS-CoV-2 exposed Vero cells
SARS-CoV-2 exposed organoids
RNA copies/ml, log10
dpi

Figure 1.

Figure 1. SARS-CoV-2 targets the cortical region of human brain organoids.

A A positive control experiment. SARS-CoV-2 readily targets ciliated human respiratory epithelial cells (hRECs). Acetylated α-tubulin labels cilia. Arrows point SARS-CoV-2-positive cells labeled by AB4 (green). Figures display scale bars. Bar diagram at right quantifies frequencies of SARS-CoV-2-positive cells in hRECs. At least six hREC sections from three (n = 3) independent samples were examined. Data presented as mean ± SEM.

B Mock organoids of two age groups Day-15 (i) and-60 (ii) display typical cytoarchitecture of brain organoids. L, lumen, VZ, ventricular Zone is containing compact and palisade-like elongated nuclei of neural progenitor cells (NPCs, blue) and CP, a cortical plate containing TUJ-1-positive neurons (magenta). Note a distinct difference TUJ-1 labeling pattern between younger (Day-15) and older (Day-60) brain organoid. Figures display scale bars. Representative images from eight organoids cultured in at least three independent batches (n = 3) derived from donor-1 (IMR90) iPSC line.

C Compared to mock organoids (i), SARS-CoV-2-exposed Day-15 organoids display SARS-CoV-2-positive cells (AB4, green) in their outer periphery, a region of the cortical plate (ii) that is specified by TUJ-1-positive neurons (magenta). L, the lumen of a VZ, the inner area of an organoid where NPCs are located, is free from SARS-CoV-2-positive cells. Magnified region (dotted while box) is given below. At least 10 organoids from five different batches (n = 5) are tested. Figures display scale bars.

D SARS-CoV-2-exposed Day-60 organoids. Compared to Day 15 organoids and mock (i), Day-60 organoids display an increased number of SARS-CoV-2-positive cells (AB4, green) in their cortical plate that is specified by TUJ-1-positive neurons (magenta) (ii). Magnified region (dotted while box) is given below, showing the perinuclear location of SARS-CoV-2 in cortical neurons. At least 10 organoids from five different batches (n = 5) are tested. Figures display scale bars.

E The bar diagram quantifies frequencies of SARS-CoV-2-positive cells in different brain organoid sections derived from two donor iPSC lines (IMR90 and Crx-iPS, see Materials and Methods). Please note that each point represents one organoid section. SARS-CoV-2 shows an enhanced tropism for Day-60 organoids. Note, comparative statistics are shown between different age groups and respective days post-infection (dpi) of organoids, and the significance is given as Asterisks in Day-60 groups. There is no significant difference in SARS-CoV-2-positive cells between 2- and 4-dpi within each age groups. At least twelve organoids sections from four (n = 4) independent batches, from each donor and day post-infections (dpi), were analyzed. One-way ANOVA, followed by Tukey's multiple comparisons test, ***$P < 0.001$. Data presented as mean ± SEM.

F Subcellular localization SARS-CoV-2 in cortical neurons. High-resolution imaging and deconvolution show perinuclear localization of SARS-CoV-2. SARS-CoV-2 (AB4, green) and nucleus (gray). Figures display scale bars. Representative images from at least 200 cells are examined. White line surrounds perinuclear border, and red line encircles the nucleus.

G Determination of viral progeny. Supernatants of SARS-CoV-2 exposed Vero cells, and brain organoids were analyzed for viral RNA assessed by qRT–PCR. While an increase in viral RNA was detected in the supernatants of Vero cells, no apparent increase was identified in brain organoid supernatants. Data are obtained from five technical replicates from four (n = 4) independent batches of organoids. Data presented as mean ± SEM.

(Fig 1B) (Lancaster & Knoblich, 2014; Giandomenico & Lancaster, 2017; Gopalakrishnan, 2019).

We exposed at least two different age groups of organoids (Day-15 and Day-60) to SARS-CoV-2 (TCID50/ml of 50 which is equivalent to 17.5 PFU/organoid, see Materials and Methods section for details) and analyzed after 2 and 4 days post-infection (dpi). First, we began analyzing Day-15 organoids, a developmental stage used to study ZIKV infections (Gabriel *et al*, 2017). At this developmental stage, organoids mostly constitute actively proliferating NPCs at the VZs and a primitive cortical plate containing fewer early neurons (Fig 1B). Testing the target cell types of SARS-CoV-2 in these organoids revealed that SARS-CoV-2 could mostly target the cortical plate specified by pan-neuronal marker TUJ-1 that is spatially distinct from the VZ (Fig 1C and Appendix Fig S2D). To exclude the possibility that the virus may have a limited capacity of diffusion to target NPCs at the inner part of the intact 3D organoids, we directly exposed NPCs' 2D cultures to SARS-CoV-2. Compared to 2D cortical neuronal cultures, NPCs cultures displayed only fewer cells positive for SARS-CoV-2. These findings indicate that SARS-CoV-2 has a preferred tropism to neurons, as reported recently (preprint: Mesci *et al*, 2020; preprint: Song *et al*, 2020; Yang *et al*, 2020a) (Appendix Fig S3A). This is indeed in striking contrast to ZIKV, which directly targets NPCs present at the inner region of brain organoids and triggers them to prematurely differentiate into neurons leading to congenital microcephaly (Cugola *et al*, 2016; Qian *et al*, 2016; Gabriel *et al*, 2017).

Analyzing the cortical regions of Day-60 organoids revealed that the number of SARS-CoV-2-positive cells was significantly higher than in Day-15 organoids. This suggests that SARS-CoV-2 prefers relatively mature neuronal cell types present in older organoids (Fig. 1D and E). Day-60 organoids indeed displayed signs of maturation as judged by more MAP2-positive neurons, S100β-positive astrocytes, and fewer Iba-1-positive microglial cells (Appendix Fig S3B–D). Importantly, the perinuclear localization of SARS-CoV-2 in somas of cortical neurons is similar to the virus's localization pattern in Vero cells, indicating that SARS-CoV-2 can enter into neuronal cells of brain organoids (Fig 1F). Turning our analysis to the later time point of infection (dpi-4 and dpi-6) revealed no apparent increase in SARS-CoV-2-positive cells although dpi-6 organoids exhibited a slightly compromised integrity (Fig 1E and Appendix Fig S4A). Corroborating to this, we could not detect an increase in viral RNA in the supernatants between 2- and 4-dpi (Fig 1G). In contrast to brain organoids, SARS-CoV-2 productively infects vascular, kidney, and gut organoids (Lamers *et al*, 2020; Monteil *et al*, 2020; Zhou *et al*, 2020). Notably, angiotensin-converting enzyme 2 (ACE-2), an entry receptor of SARS-CoV-2, is highly expressed in these organoid types. Testing the ACE-2 expression at the mRNA level via a qRT–PCR revealed that both iPSCs-derived brain organoids and neurons exhibited ~12.5- and 50-fold lesser than human respiratory epithelial cells (hREC), which served as a positive control (Appendix Fig S4B). Our Western blots using anti-ACE2 antibodies recognized ACE2 in organoid extracts only at higher exposure conditions (Appendix Fig S4C).

Since SARS-CoV-2 appears to preferably target neurons, we wondered if SARS-CoV-2 could productively replicate when exposed to an abundant number of mature neurons. To test this, we cultured organotypic slices of 60-day-old organoids, an alternative organoid culturing method that enhances neuronal maturation and viability. These cultures exhibit neuronal outgrowths as long-range axonal fibers expressing mature neuronal markers of MAP2, Tau, synapsin-1, and PSD95 (Gabriel *et al*, 2016; Giandomenico *et al*, 2019; Goranci-Buzhala *et al*, 2020). After directly exposing these slices to SARS-CoV-2, we detected the virus localized at the cell bodies of the neurons which are labeled by MAP2 and Tau (Appendix Fig S5A and B). We noticed only a slight increase in SARS-CoV-2 RNA within 2 days of viral exposure (Appendix Fig S5C). These experiments demonstrate that SARS-CoV-2 enters neurons of brain organoids but does not actively replicate.

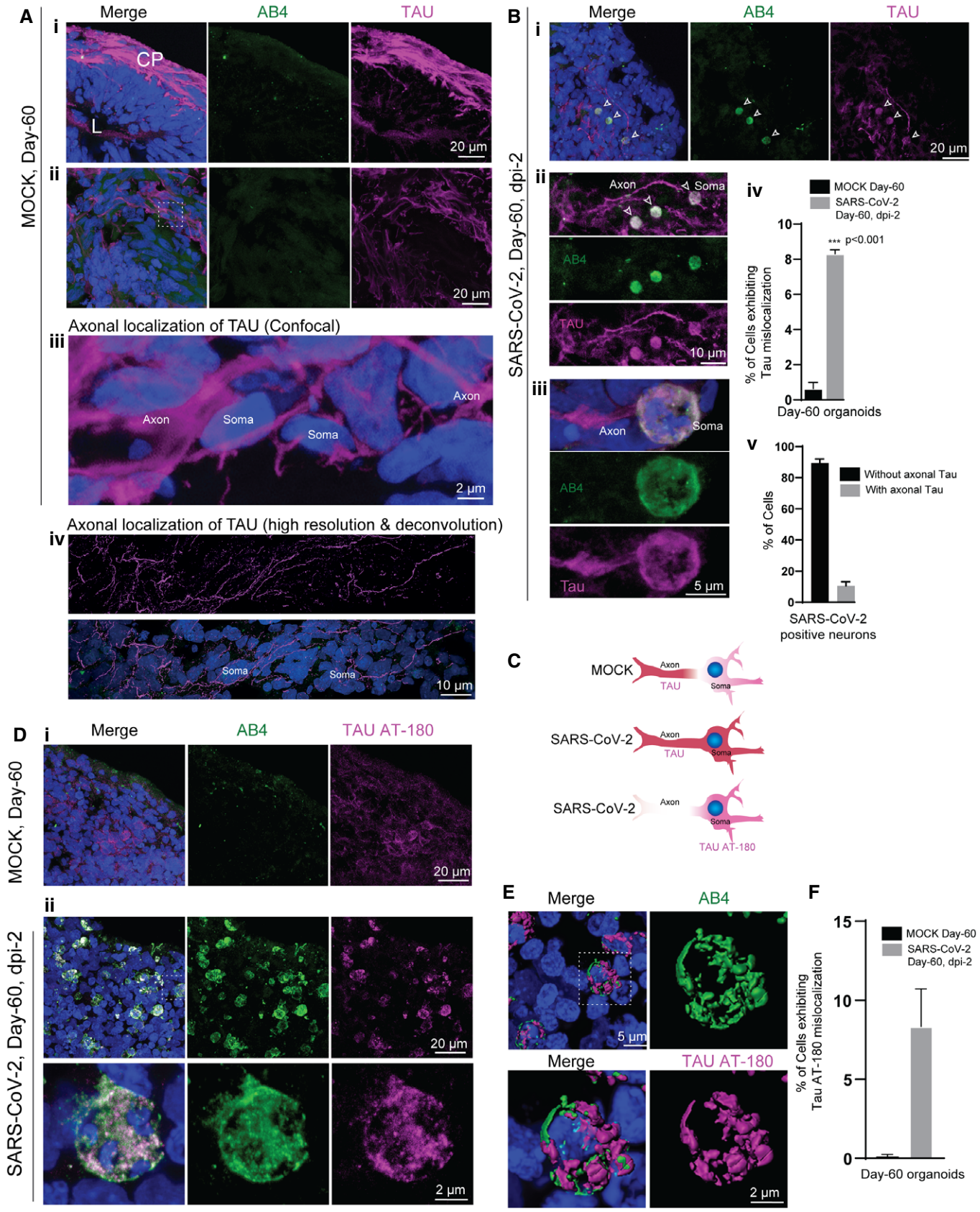

A
Merge
AB4
TAU
MOCK, Day-60
i
CP
L
20 µm
ii
20 µm
Axonal localization of TAU (Confocal)
iii
Axon
Soma
Soma
Axon
2 µm
Axonal localization of TAU (high resolution & deconvolution)
iv
Soma
Soma
10 µm
B
Merge
AB4
TAU
SARS-CoV-2, Day-60, dpi-2
i
20 µm
ii
Axon
Soma
AB4
TAU
10 µm
iii
Axon
Soma
AB4
Tau
5 µm
iv
MOCK Day-60
SARS-CoV-2 Day-60, dpi-2
*** p<0.001
% of Cells exhibiting Tau mislocalization
Day-60 organoids
v
Without axonal Tau
With axonal Tau
% of Cells
SARS-CoV-2 positive neurons
C
MOCK
Axon
TAU
Soma
SARS-CoV-2
Axon
TAU
Soma
SARS-CoV-2
Axon
Soma
TAU AT-180
D
Merge
AB4
TAU AT-180
MOCK, Day-60
i
20 µm
ii
SARS-CoV-2, Day-60, dpi-2
20 µm
2 µm
E
Merge
AB4
5 µm
Merge
TAU AT-180
2 µm
F
MOCK Day-60
SARS-CoV-2 Day-60, dpi-2
% of Cells exhibiting Tau AT-180 mislocalization
Day-60 organoids

Figure 2.

Figure 2. SARS-CoV-2 deregulates of Tau in cortical neurons.

A Tau immunoreactivity (magenta) specifies the cortical plate (CP) surrounding the lumen (L) (i). Selected optic sections at high magnification (ii and iii) and high-resolution imaging (iv) show Tau localization only in axons of cortical neurons. Note the somas of neurons are free from Tau protein. At least eight organoids from four different batches (n = 4) are tested. Figures display scale bars.

B Tau localization in SARS-CoV-2-positive neurons (AB4, green) in selected optic sections (i). Note, in contrast, to control groups, SARS-CoV-2-exposed organoids display mislocalized Tau (magenta) majorly into the somas of neurons (arrowheads). Selected confocal slices are shown to distinguish Tau mislocalization into neuronal soma (arrowheads). At high magnification, neuronal soma is further specified by the perinuclear localization of SARS-CoV-2 (green) (ii and iii). Bar diagrams at right quantifies the percentage of neurons (Mock and SARS-CoV-2 exposed) exhibiting mislocalized Tau (iv) and the fraction of SARS-CoV-2-positive neurons exhibiting Tau-positive axons spanning different cortical areas (v). For statistics, at least 300 cells from six organoids from four different batches (n = 4) were tested. Figures display scale bars. Unpaired t-test with Welch's correction, ***$P < 0.001$. Data presented as mean ± SD.

C Schematic cartoon of differential Tau distribution in mock compared to SARS-CoV-2-positive neurons. In mock, Tau is sorted mainly to axons. In SARS-CoV-2-positive neurons, Tau is missorted to the soma (determined by Pan-Tau antibody). Furthermore, phosphorylated Tau (at T231) majorly localizes in the soma (bottom panel, determined by Tau AT-180 antibody, see below).

D In contrast to controls (i), Tau AT180 antibody (magenta) that specifically recognizes the phosphorylated Threonine 231 of Tau protein distinctly localizes at the somas of SARS-CoV-2-positive neurons (AB4, green) (ii). At least four organoids from two different batches (n = 2) are tested. Figures display scale bars.

E Co-localization of SARS-CoV-2 (AB4, green) and phosphorylated Tau protein (magenta) at somas of cortical neurons revealed by high-resolution imaging and deconvolution. Representative images from at least 300 cells examined. Figures display scale bars.

F The bar diagram quantifies the fraction of Tau AT180-positive neurons that co-localize with SARS-CoV-2-positive neurons. For statistics, at least 250 cells from four organoids and two independent batches (n = 2) were examined. Unpaired t-test with Welch's correction. Data presented as mean ± SEM.

SARS-CoV-2-positive neurons reveal aberrant Tau localization

Next, we identified that the SARS-CoV-2-positive region of the cortical plate is further substantiated by Tau, a microtubule-associated protein that stabilizes neuronal microtubules and promotes axonal growth (Fig 2) (Wang & Mandelkow, 2016). Tau dysfunction is implicated in Alzheimer's disease (AD) and other Tauopathies. Post-translational modifications in Tau, in particular, phosphorylations, modulate the ability of Tau to bind and assemble microtubules. In Tauopathies, Tau is aberrantly phosphorylated (hyperphosphorylation; Cho & Johnson, 2004; Cohen *et al*, 2011; Castellani & Perry, 2019). A recent report showed that herpes simplex virus type 1 can induce AD-like effects, including hyperphosphorylation of Tau in 3D human brain-like tissue model (Cairns *et al*, 2020). This prompted us to investigate if SARS-CoV-2 has a consequence upon its entry into neurons.

Under physiological conditions, Tau is mainly an axonal protein that localizes at the axons of mature neurons (Fig 2Ai–iv). Applying high-resolution imaging followed by deconvolution, we could visualize Tau's localization (as probed by a Pan-Tau antibody Tau5A6) exclusively in axons of the cortical neurons (Fig 2Av). The term Tau "missorting" is used when Tau protein is mislocalized into a cell soma and is observed at the early stages of Tau pathology (Zempel & Mandelkow, 2014).

Compared to control organoids where Tau normally localizes in axons, SARS-CoV-2-positive neurons exhibited an altered Tau localization pattern, although it was challenging to visualize mislocalization of Tau in 3D tissues. Nevertheless, using selected confocal sections, we could image an altered Tau localization in SARS-CoV-2-positive neurons. In particular, we identified an enhanced level of Tau into the somas of the SARS-CoV-2-positive neurons. Importantly, we could visualize fractions of these neurons still contained Tau and TUJ-1 in their axons, indicating that these neurons are still viable (Fig 2B and C and Appendix Fig S6A).

During the pathogenesis of AD and other Tauopathies, Tau also gets hyperphosphorylated at multiple sites. Sequential phosphorylation at different sites ultimately leads to hyperphosphorylation of Tau (Castellani & Perry, 2019). Phosphorylation of Threonine 231 (T231) is one of the first events in the cascade of phosphorylation, and it regulates the microtubule binding. Still, it is also implicated in disease progression such as detachment of Tau from axonal microtubules (Sengupta *et al*, 1998; Augustinack *et al*, 2002a,b; Luna-Munoz *et al*, 2007; Alonso *et al*, 2010; Frost *et al*, 2015). More precisely, we found that compared to control organoids, early Tau phosphorylation marker AT180 recognizes pT231Tau localized at the soma of the SARS-CoV-2-positive neurons (Fig 2D–F). Imaging the neurons for additional phosphorylated Tau using AT8 antibodies (specific for S202 and T205 of Tau) and p396 (specific for S396 of Tau) revealed that unlike pT231Tau, these phospho-species were restricted to the axons and did not mislocalize to the soma of SARS-CoV-2-positive neurons (Appendix Fig S6B–E). In summary, these results demonstrate the aberrant localization of Tau pT231Tau in SARS-CoV-2-positive neurons suggesting the potential neuronal stress reactions upon virus entry.

SARS-CoV-2 induces neuronal cell death

Phosphorylation of Tau at T231 allows for isomerization of the following proline residue into distinct *cis*- and *trans*-conformations by the propyl-isomerase PIN1 (Lu *et al*, 1999). Cis-pT231Tau is acutely produced by neurons after traumatic brain injury, leading to disruption of the axonal microtubule network and apoptosis (Nakamura *et al*, 2012; Kondo *et al*, 2015). Analyzing the nuclei of SARS-CoV-2-positive cells (Fig 3A), we realized that they are highly condensed or fragmented exhibiting a strong reaction to 4′,6-diamidino-2-phenylindole (DAPI) that labels nuclei, a feature quite frequently observed in dead cells. To test neuronal cell death as a consequence of SARS-CoV-2 infection, we stained the SARS-CoV-2-exposed samples with terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) that detects fragmented DNA in dead cells (Darzynkiewicz *et al*, 2008). Compared to un-exposed control organoids, we identified an overall increase in TUNEL-positive cells in SARS-CoV-2-exposed organoids suggesting that virus exposure has caused cell death within 2-dpi (Fig 3B). Staining for SARS-CoV-2-positive cells revealed that most of the virus-positive cells were TUNEL-positive. Besides, we also noticed that some SARS-CoV-2-positive cells were also positive for caspase-3, a protease that specifies programmed cell death (Fig 3C). Interestingly, a fraction of

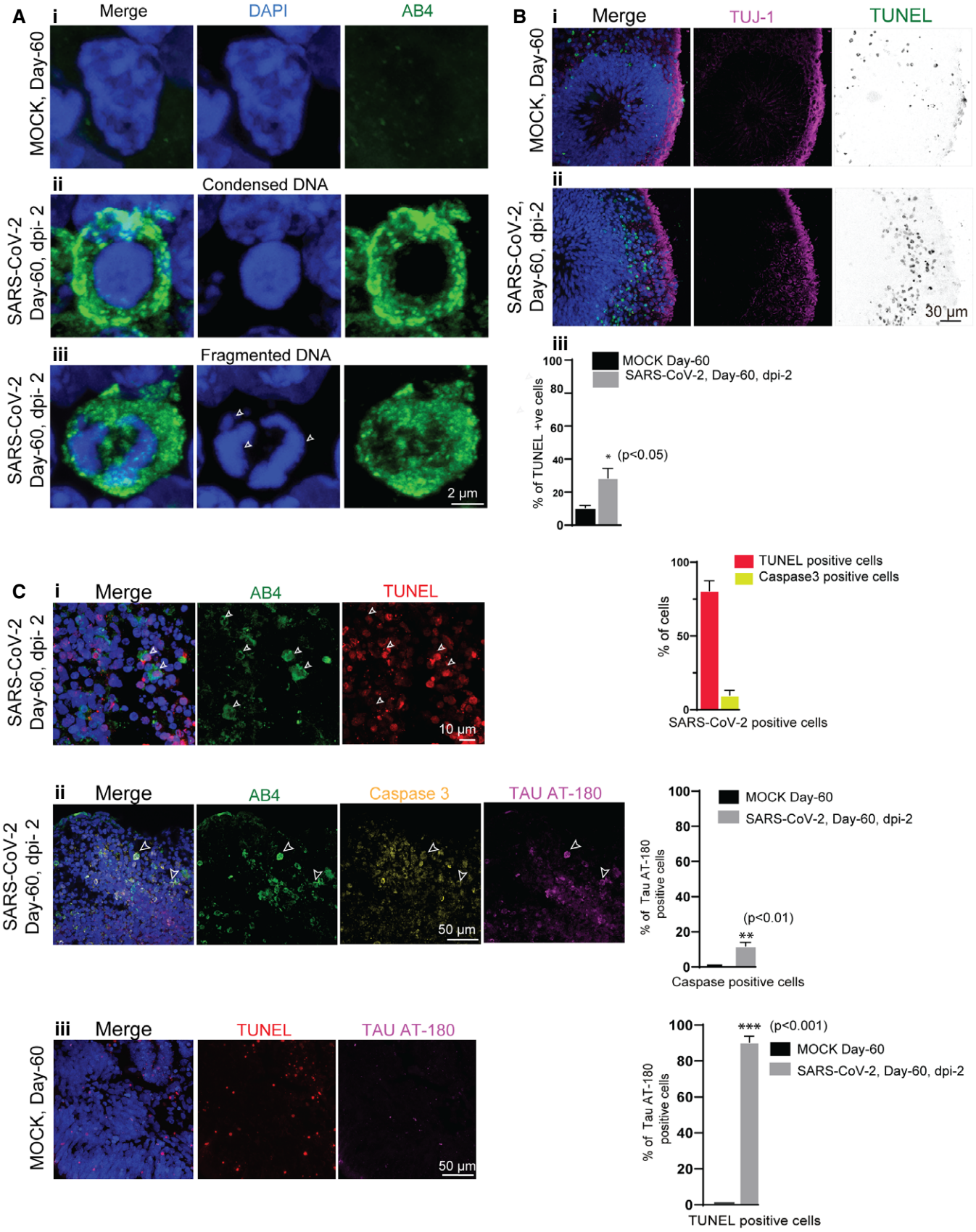

A
i
Merge
DAPI
AB4
MOCK, Day-60
ii
Condensed DNA
SARS-CoV-2 Day-60, dpi- 2
iii
Fragmented DNA
SARS-CoV-2 Day-60, dpi- 2
2 µm
B
i
Merge
TUJ-1
TUNEL
MOCK, Day-60
ii
SARS-CoV-2, Day-60, dpi-2
30 µm
iii
MOCK Day-60
SARS-CoV-2, Day-60, dpi-2
% of TUNEL +ve cells
100
80
60
40
20
0
* (p<0.05)
C
i
Merge
AB4
TUNEL
SARS-CoV-2 Day-60, dpi- 2
10 µm
TUNEL positive cells
Caspase3 positive cells
% of cells
SARS-CoV-2 positive cells
ii
Merge
AB4
Caspase 3
TAU AT-180
SARS-CoV-2 Day-60, dpi- 2
50 µm
MOCK Day-60
SARS-CoV-2, Day-60, dpi-2
% of Tau AT-180 positive cells
(p<0.01)
**
Caspase positive cells
iii
Merge
TUNEL
TAU AT-180
MOCK, Day-60
50 µm
*** (p<0.001)
MOCK Day-60
SARS-CoV-2, Day-60, dpi-2
% of Tau AT-180 positive cells
TUNEL positive cells

Figure 3.

Figure 3. SARS-CoV-2 induces of neuronal death.

A Cells from mock organoids display a healthy nucleus labeled by DAPI (blue) (i). SARS-CoV-2-positive cells (green) display condensed (middle panel, ii) and fragmented DNA (bottom panel, iii, arrows). At least 75 cells from two (n = 2) independent batches of organoids were examined. Figures display scale bars.

B Compared to mock organoids, (i) SARS-CoV-2-exposed organoids (ii) display increased TUNEL-positive cells (displayed as inverted LUT) at the cortical plate that is specified by TUJ-1 (magenta). At least four organoids from two (n = 2) independent batches of organoids were examined. Figures display scale bars. The bar diagram below quantifies the frequencies of TUNEL-positive cells between mock and SARS-CoV-2-exposed organoids. Four organoids from two (n = 2) independent batches were examined. Unpaired t-test, $^{*}P < 0.05$. Data presented as mean ± SD.

C Most of the SARS-CoV-2-positive cells (AB4, green) are TUNEL-positive (i) and some of the SARS-CoV-2-positive cells are caspase-positive (ii). Caspase-positive cells in SARS-CoV-2-exposed organoids display pT231Tau localization at the cell soma specified by AT-180 (ii), which are not observed in mock organoids (iii). Arrowheads point SARS-CoV-2-positive cells (AB4, green) that are also positive for TUNNEL (red), caspase 3 (yellow), and Tau AT-180 (magenta). Figures display scale bars. Bar diagrams at right quantifies proportions of TUNEL and caspase-positive cells among SARS-CoV-2-positive cells. The second graph below quantifies the proportions of pT231Tau-positive cells among caspase-positive cells between control and virus exposed groups. At least 400 cells from four organoids and two independent (n = 2) batches were examined. Unpaired t-test, $^{**}P < 0.01$. Data presented as mean ± SEM. TUNEL-positive cells in control un-exposed organoids do not contain pT231Tau (iii). Figures display scale bars. Bar diagrams at right quantifies proportions of pT231Tau-positive cells among TUNNEL-positive cells between control and virus exposed groups. At least 350 cells from 4 organoids and two independent (n = 2) batches were examined. Unpaired t-test, $^{***}P < 0.001$. Data presented as mean ± SEM.

caspase-positive cells displayed pT231Tau localization at the cell soma. Furthermore, TUNEL-positive cells in un-exposed control organoids (which could be after programmed cell death) did not contain pT231Tau suggesting that this different Tau phosphorylation pattern is associated with SARS-CoV-2 entry (Fig 3Ciii). Thus, it appears that Tau is aberrantly phosphorylated in response to the viral-induced stress, which may elicit further cell death programs that remains to be elucidated.

Discussion

So far, the possible direct effect of SARS-CoV-2 on the CNS has been debated but not experimentally demonstrated (Baig *et al*, 2020; Conde Cardona *et al*, 2020:718; Coolen *et al*, 2020; Helms *et al*, 2020; Poyiadji *et al*, 2020). Thus, it was essential to examine whether SARS-CoV-2 can directly target human neurons and whether this leads to productive infection. In contrast to vascular, kidney, and intestinal organoids (Lamers *et al*, 2020; Monteil *et al*, 2020; Zhou *et al*, 2020), brain organoids does not appear to strongly support the active replication of SARS-CoV-2 at least until 6-dpi. There are several reasons for this. Firstly, the developmental stages of brain organoids used in this work may not contain the full complement of SARS-CoV-2's host cell replication factors. As an example, efficient replication of SARS-CoV requires ACE-2 (Li *et al*, 2003) whose expression appears to be relatively low in brain organoids (Appendix Fig S4B). Next, in brain organoids, post-mitotic neurons seem to be susceptible for SARS-CoV-2 that may not be permissive (Fig 1, Appendix Figs S3 and S5). Finally, brain organoids are simplified reductionist models and lack an additional cell type that can influence viral replication such as blood–brain barrier, vasculature, and mature glial cells, including microglia. To our surprise, we did notice the appearance of fewer Iba-1 and S100β-positive cells in our organoids pointing toward the need of further engineering of our differentiation conditions which could lead to the differentiation of mature microglia and astrocytes (Appendix Fig S3C and D). Thus, future experiments using aged organoids and bioengineered organoids with SARS-CoV-2 replication factors are required to conclude if brain organoids can support productive infection of SARS-CoV-2.

ACE-2 is an entry receptor for SARS-CoV and efficient replication of SARS-CoV (SARS outbreak in the year 2003) and also depends on the expression level of ACE-2 (Li *et al*, 2003; Hoffmann *et al*, 2020). Curiously, SARS-CoV could only infect the brain of transgenic mice expressing an elevated level of human ACE-2 but not non-transgenic mice. This key finding suggests that the neurotropism of SARS-CoV, to some extent, depends on the expression level of human ACE-2 in the brain (McCray *et al*, 2007). Using our 3D human brain organoid system, we unexpectedly find that although these organoids express low level of ACE-2, the human neurons are indeed a target for SARS-CoV-2. This finding offers a couple of possibilities. First, even a basal level of ACE2 expression is sufficient for viral entry into the neurons. Second, the presence of yet unknown neuron-specific viral entry factors has to be elucidated. It is indeed intriguing that even a low level of ACE-2 is sufficient for the viral entry, and this could explain why SARS-CoV-2 has a broad spectrum of target organs and cell types (Puelles *et al*, 2020).

Detection of Tau phosphorylation at T231 in SARS-CoV-2-positive neurons is remarkable as it could trigger a cascade of downstream effects that finally could initiate neuronal stress and toxicity. Intriguingly, there is growing evidence that viral infections, particularly herpes simplex virus type I (HSV-1), is a potential causative agent leading to Alzheimer's disease (AD). Indeed a recent work demonstrated infection of 3D human brain-like tissue model with HSV-1 and showed that the HSV-1 infection is sufficient to elicit AD-like effects, including hyperphosphorylation of Tau. Early Tau phosphorylation at T231 could be reversible (Castellani & Perry, 2019). However, phosphorylation events observed in conjunction with apparent neuronal cell death suggest that SARS-CoV-2 has potential detrimental effects on neurons at least in our organoid test system (Fig 3). Future biochemical experiments dissecting the ratio of soluble and sarkosyl-stable Tau extracted from SARS-CoV-2-positive neurons are required to obtain insights into the cause and effect of potential Tau pathology and neuronal death. Although we observe Tau abnormalities in SARS-CoV-2-positive neurons, we could not conclude whether the observed effect is directly caused by the virus or an effect due to neuronal stress, which warrants future investigations.

In conclusion, COVID-19 research has taken center stage in biomedical research. It is noteworthy that three coronavirus epidemics have occurred within the last two decades, and thus, the future zoonotic coronavirus outbreak is not unexpected. With the advent of emerging human organoid research, which did not exist 20 years ago, we should be able to model the current SARS-CoV-2

infections and sufficiently prepare us for the future. Recent works utilizing kidney, gut, and liver organoids have already revealed insights into the infection mechanisms (Lamers *et al*, 2020; Monteil *et al*, 2020; Yang *et al*, 2020a; Zhou *et al*, 2020). Adding to them is the current work that establishes brain organoids as a test system for SARS-CoV-2 infection and provides indications for potential neurotoxic effects of SARS-CoV-2. Since organoids are an experimentally tractable human *in vitro* system and convenient to culture as well as to infect, organoid systems may serve well as a test-bed to screen for anti-SARS-CoV-2 agents. The presented work only provides initial insights into primitive brain-like tissues and requires further experiments to dissect viral replication mechanisms and whether there are ACE2 independent pathways for viral entry. It is important to note that although the virus seems to preferably target neurons, future experiments are required to test if the virus can have extended access across the entire organoids. Advanced experiments utilizing a mature state of brain organoids, bioengineered organoids, and orthogonal experiments with complementary *in vivo* experimental models are assured to dissect the neuropathology of SARS-CoV-2.

Materials and Methods

Clinical specimens

For the isolation of infectious SARS-CoV-2 particles, nasopharyngeal and oropharyngeal swab specimens from one individual with positive qRT–PCR results for SARS-CoV-2 infection were used. The swab specimen was transported in a viral cultivation medium and stored at 4°C overnight. Freezing at −20°C was found to interfere with the infectivity of viral particles. Before the inoculation of susceptible cells, 500 µl maintenance medium (Dulbecco's Modified Eagle Medium (Thermo Fisher), 2% fetal calf serum (PAN Biotech), 100 U/ml penicillin, and 100 µg/ml streptomycin (Gibco) were added to the swab specimen. To get rid of major impurities, samples were briefly centrifuged (3,000 *g*; 60 s) and the supernatant was transferred to new vials.

Human respiratory epithelial cells and culturing

To obtain respiratory epithelia, a MedScand Cytobrush Plus GT (Cooper Surgical, Trumbull, USA) with a gentle-touch tip was rinsed with isotonic saline before use. Afterward, the brush was inserted into the inferior nasal meatus followed by rotatory and linear motions against the medial and superior side. Isolated cells were transferred into a 15-ml centrifuge tube (Corning Incorporated, New York, USA) with 5 ml pre-warmed RPMI 1640 medium containing 2% Antibiotic-Antimycotic 100× (Gibco® Life Technology, Grand Island, USA). The brushes were vigorously shaken several times within the tube, and cells were pelletized by centrifugation at 900 rpm for 5 min at room temperature. hRECs were re-suspended in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12, Gibco® Life Technology, New York, USA) supplemented with 2% UltroserTM G Serum Substitute (Pall Corporation, Port Washington, USA) and 2% Antibiotic-Antimycotic 100 x and seeded on T-25 or T-75 rat-tail collagen-coated tissue flasks (Greiner Bio-One, Kremsmünster, Austria), according to the pellet size, respectively, and incubated at 37°C, 5% CO_2.

To reduce the risk of contamination, the medium was replaced after 24 h, and the flasks were then integrated into the regular feeding procedure (exchange of medium every 48–72 h). After 1 week, the concentration of Antibiotic-Antimycotic was reduced to 1%. Reaching confluency of 90%, the collagen layer was digested by incubating with 200 U/ml collagenase type IV (Worthington Biochemical Company, New Jersey, USA) for 30–60 min, followed by several washing steps with DMEM/F-12 supplemented with 1% Antibiotic-Antimycotic. To reduce the number of fibroblasts, the pellet was re-suspended in 7 ml DMEM/F12 supplemented with 2% UltroserTM G, seeded on tissue culture treated T-25 flasks (Corning Incorporated, New York, USA) and incubated for 1 h at 37°C, 5% CO_2. The cells were then separated by incubating with Trypsin-EDTA 0.05% for 5 min before the reaction was stopped with FBS followed by centrifugation at 900 rpm for 5 min at room temperature.

After re-suspending in PneumaCultTM-Ex Medium (STEMCELLTM Technologies, Vancouver, Canada), 4×10^5 cells/ml were seeded on collagen-coated 6.5 mm Transwell®, 0.4 µm pore Polyester membrane inserts (Corning Incorporated, New York, USA) with 250 µl medium on the apical side and 500 µl on the basolateral side, respectively. Before airlift, after 3–5 days, depending on cell confluency, PneumaCultTM-Ex Medium was replaced every day at the apical and basolateral side. To perform airlift, the medium on the apical side was carefully removed, whereas the basolateral medium was exchanged with PneumaCultTM-ALI Medium. The airlifted inserts were then integrated into the regular feeding procedure and incubated at 37°C, 5% CO_2. A fully differentiated pseudostratified epithelium is expected 15–30 days after airlift and resembles human airway epithelium (*in vivo*) with respect to function and morphology.

Inoculation of Vero cells

In compliance with the German committee's decision on biological agents (ABAS) of the Federal Institute for Occupational Safety and Health, all experimental studies involving infectious SARS-CoV-2 were performed within the biosafety level 3 (P3) facility at the University Hospital Düsseldorf. To isolate SARS-CoV-2 from a clinical specimen, 2.5×10^5 Vero cells (ATCC-CCL-81, obtained from LGC Standards) were seeded into T-25 cell culture flasks in maintenance medium and cultured at 37°C in a humidified cell culture incubator. The following day, SARS-CoV-2 inoculum was prepared by diluting 200 µl of a clinical specimen with 800 µl maintenance medium. The medium was removed from Vero cells, and 1 ml inoculum (1 ml of maintenance medium for control Vero cells) was added onto the Vero cell monolayer. Vero cells were incubated for 1 h on a laboratory shaker at 37°C in a humidified incubator. Afterward, 4 ml of maintenance medium was added. To monitor viral replication, 100 µl of supernatant was directly harvested as the first sample (0 h post-inoculation) and every 24 h for 4 days post-inoculation. Additionally, cells were imaged by light microscopy.

Real-time qPCR analysis for quantification of SARS-CoV-2 RNA copies per ml

For extraction, 100 µl cell culture supernatant was incubated with 400 µl AVL buffer (viral lysis buffer used for purifying viral nucleic

acids; cat No. 19073, Qiagen, Hilden Germany) for 10 min at RT and mixed with 400 µl 100% ethanol. RNA extraction was performed with 200 µl cell culture mix using the EZ1 Virus Mini Kit v2. (cat. no. 955134, Qiagen, Hilden, Germany) following the manufacturer's instructions. A total of 60 µl were eluted from the 200 µl starting material. 5 µl of the eluate was tested in qRT–PCR using the real-time TaqMan®-technique. A 113 base pair amplicon in the E-gene of SARS-Cov-2 was amplified and detected, as described by Corman *et al* (2020) with minor modifications. The thermal protocol described has been shortened to 40 cycles of 95° C. We used the LightMix® Modular SARS and Wuhan CoV E-gene (Cat.-No. 53-0776-96) and the LightMix® Modular EAV RNA Extraction Control. We used the AgPath-ID® One-Step RT–PCR Kit (Applied Biosystems, Cat. No. 4387391). RT–PCR was performed with an ABI 7500 FAST sequence detector system (PE Applied Biosystems, Weiterstadt, Germany). As a DNA-standard, a plasmid (pEX-A128-nCoV2019-E-gene) that encompasses the amplified region was created and serially diluted after purification. The software constructed a standard graph of the CT values obtained from serial dilutions of the standard. The CT values of the unknown samples are plotted on the standard curves, and the number of SARS-CoV-2 RNA copies was calculated.

For gene expression analysis of ACE2, quantitative RT–PCR analysis was performed by using qPCR MasterMix (PrimerDesign Ltd) and fluorescence emission was monitored by LightCycler 1.5 (Roche). For normalization, primers #5163 (5′ CCA CTC CTC CAC CTT TGA 3′) and #5164 (5′ ACC CTG TTG CTG TAG CCA 3′) were used monitoring cellular GAPDH expression. Expression was then calculated as $2^{(-\Delta C_t)}$.

Propagation of infectious SARS-CoV-2 particles

For propagation of infectious SARS-CoV-2 particles from Vero cell culture supernatant, 2.5×10^5 Vero cells were seeded into T-25 cell culture flasks in maintenance medium and incubated at 37°C in a humidified cell culture incubator. The next day, the supernatant of inoculated Vero cells at day four post-inoculation (see above) was diluted with maintenance medium (1:2, 1:10, 1:100, 1:1,000) in a total volume of 5 ml and added to the cells, which were incubated for 4 days at 37°C.

Determining SARS-CoV-2 viral titer by TCID50 assay in 96-well plates with Vero cells

For determination of viral titer in TCID50/ml, 5×10^3 Vero cells were seeded in the first 10 columns of 96-well plate in 100 µl maintenance medium and incubated at 37°C in a humidified cell culture incubator for 24 h. In a new 96-well plate, 180 µl maintenance medium was added to all wells of the first 10 columns. For serial dilutions of the virus stock, 20 µl of the stock solution was added to the wells of the first column. Then, 20 µl of the first dilution was transferred to the wells of the next column to obtain 10-fold serial dilutions up to 10^{-9}. The tenth column of the 96-well plate serves as a control. After exchanging the medium of the previously prepared Vero cell plate with 100 µl fresh maintenance medium, 100 µl of each virus dilution was transferred to the Vero cell plate. After incubation at 37°C for 4 days, microscopic inspection of the plate was used to monitor cytopathic effects (CPEs) in the form of detached cells. TCID50/ml was determined as:

$$\mathrm{TCID}_{50/\mathrm{ml}} = \frac{D_S^{(N/R+0.5)} D_0 \times 1{,}000}{D_S \times V}$$

D_S = dilution factor of consecutive dilutions (10); N = total number of wells showing CPE; R = replicates per dilution (8); D_0 = dilution factor of the first dilution (10); V = volume per well in µl (200 µl).

To estimate MOI, we first calculated the viral titer as TCID50/ml of our generated SARS-CoV-2 by an end-point dilution assay as previously described (Flint *et al*, 2015). In brief, based on induced cytopathic effects, we calculated the TCID50/ml using the above formula based on the Spearman-Karber method (Ramakrishnan, 2016). To further confirm this calculation with respect to the novelty of this formula, we also applied the commonly used Reed and Muench method (Lei *et al*, 2020). Both of these methods resulted in a TCID50/ml of 5,000 that we then used to calculate the PFU/ml. Applying poisson distribution, we estimated that the amount of infectious viral particles per ml (PFU/ml) in our stock is 3,500 PFU/ml.

In the context of our infection experiments, we provided 5 µl virus stock per organoid. According to our calculation, the 5 µl volume of SARS-CoV-2 stock contains approximately 17.5 PFUs. Having then estimated the number of viable cells after disintegrating organoids (an average of 100,000 for Day 15 and 200,000 for day 60), we could determine the multiplicity of infection (MOI). Considering 17.5 PFUs, our estimated MOI is 1.8×10^{-4} and 8.8×10^{-5} for day 15 and day 60, respectively. Importantly, we found that such a low viral load is sufficient for our studies.

SARS-CoV-2 infection

All experiments including SARS-CoV-2 infections were performed in a P3 safety laboratory (see above). Neurons and brain organoids were tested and found free from mycoplasma contamination using the mycoplasma kit (Minevera, Cat. No. 11-1050). For viral exposure, 15- and 60-day-old organoids were transferred from spinner flasks into low-adherent 12 well plates. Each well contained one organoid in 2 ml differentiation medium and added with SARS-CoV-2 and was incubated as stationary suspension culture. To exclude that the observed effects were not induced by SARS-CoV-2, the control organoids (control, uninfected) were treated with supernatants of non-infected Vero cells.

Generation of convalescent serum, ELISA validation, and affinity purification of SARS-CoV-2-N specific antibodies

AB1 and AB2 were obtained 23 and 16 days after the diagnosis of SARS-CoV-2 infection. AB3 and AB4 were obtained 27 and 28 days after the diagnosis of SARS-CoV-2 infection (by PCR). Blood samples were drawn directly into serum collection tubes and spun for 15 min at 1,450 *g*. After centrifugation, the clear supernatant was aliquoted and stored at −80°C. ELISA was performed using semi-quantitative SARS-CoV-2-IgA and SARS-CoV-2-IgG ELISAs that detect binding against the recombinant S1 domain of the SARS-CoV-2 spike protein (Euroimmun, Lübeck, Germany).

The full length ORF of SARS-CoV-2-N was amplified from the vector pUC57-2019-nCoV-N (GeneScript) with the primers:

5′-aaaaaagtcgacatgtctgataatggacccc-3′ and 5′-aaaaaaggatccttaggcctgagttgagtc-3′ and ligated into the expression vector pET15b via XhoI and BamHI allowing expression of SARS-CoV-2-N with an N-terminal His_6-tag. The correct sequence was validated by sequencing. pET15b-SARS-CoV-2 was then heat-shock transformed into BL21 (DE3)-Rosetta2-pLysS bacteria, plated on LB plates containing 50 μg/ml carbenicillin and 34 μg/ml chloramphenicol. Following over night incubation at 37°C, a single colony was transferred into 20 ml of 2YT medium (1.6% Bacto tryptone, 1% Yeast extract, 0.5% NaCl) and bacteria were grown over night at 37°C. This preculture was used to inoculate a main culture of 1 l 2YT at the next morning. Expression of the nucleoprotein was induced at an OD_{600} of 0.8 with 1 mM IPTG and continued over night at 30°C. Bacteria were harvested by centrifugation at 4°C for 20 min at 5,000 *g* and the pellets were frozen at −80°C.

One pellet corresponding to 250 ml of culture was thawed and re-suspended in 25 ml lysis buffer (PBS, 100 μg/ml lysozyme, 20 μl DNase, 20 mM $MgSO_4$, 1 mM DTT) and incubated for 1 h at RT. Following a protocol from (Schlager *et al*, 2012), 1% of SDS was added to the lysate which was then transferred to 2 ml Eppendorf tubes and sonicated five times for 2 min in a cooled Misonix S-4000 water bath sonicator, applying an amplitude of 100%. The samples were then placed on ice for 30 min in order to precipitate excess of SDS. The chilled lysates were centrifuged at 4°C for 20 min at 20,000 g and the supernatant was passed through a 0.45 μm syringe cellulose-acetate filter (VWR).

For purification of SARS-CoV-2 N, the lysate was passed over a 3 ml Ni-NTA agarose column (Qiagen) equilibrated with PBS, 0.1% Sarkosyl (w/v), 1 mM DTT by gravity flow. The column was then washed with 30 ml WB1 (PBS, 0.1% Sarkosyl (w/v), 5 mM imidazole, 1 mM DTT), and 30 ml of WB2 (PBS, 0.1% Sarkosyl (w/v), 20 mM imidazole, 1 mM DTT). Bound proteins were eluted with EB (PBS, 0.1% Sarkosyl (w/v), 200 mM imidazole, 1 mM DTT) into 1 ml aliquots. Samples were analyzed by SDS–PAGE regarding purity (by Coomassie stain) and identity (by Western Blot using the anti-SARS-CoV-2-nucleocapsid mouse IgG antibody clone #6F10 from BioVision (#A2060) at a dilution of 1:10,000). The total yield was around 100 mg/l culture, and purity was estimated to be > 95% (including < 5% of degradation products).

The eluate was then directly coupled to activated NHS-sepharose (GE Healthcare) at 1 mg protein/1 ml of NHS-beads. Incubation was carried out over night at 4°C. At the next day, the NHS-sepharose was first blocked with 100 mM of TRIS–HCl pH 8 for 3 h at RT and then washed five times each with 100 mM TRIS pH8 and 100 mM NaAc pH 4.5, 150 mM NaCl. One microliter of SARS-CoV-2-N-NHS-sepharose was then incubated with 2 ml of AB4 serum in a 5 ml tube (Sarstedt) at 4°C over night. At the next day, the beads were filled into a 3 ml column and washed with 10 ml of PBS. Bound antibodies were then diluted with 100 mM of glycine pH 2.5 in 100 μl steps. Eluates were immediately neutralized with 40 mM of unbuffered TRIS base and spectrometrically quantified by using a NanoDrop. The total yield was around 250 μg of SARS-CoV-2-N specific antibodies from 2 ml of serum.

Generation of iPSCs-derived cortical neurons

We differentiated iPSCs into NPCs using STEMdiff Neural Induction Medium (Stem cell technologies, USA). Five days later, the formed neurospheres were collected and cultured on poly-L-ornithine (PLO)-/laminin coated dishes. Seven days later, using a neural rosette selection medium (Stem cell technologies, USA), we re-cultured neural rosettes to generate NPCs. NPCs were differentiated into cortical neurons as described previously (*37*). Briefly, NPCs were seeded on poly-L-ornithine (PLO)-/laminin coated coverslips. Forty-eight hours later, NPCs were switched to cortical neuronal differentiation medium consisting of BrainPhys basal medium(*38*) supplemented with 1× B27 (without vitamin A, Thermo Scientific, USA), 1× N2 (Thermo Scientific, USA), 20 ng/ml BDNF (PeproTech, USA), 20 ng/ml GDNF (PeproTech, USA), 20 ng/ml NT3, 1 μM cAMP (Sigma, USA), and 0.2 μM ascorbic acid (Sigma, USA). Fresh medium was added every 2–3 days.

Generation of iPSCs-derived brain organoids and outgrowths

Organoids were generated from two different iPS cell lines, namely IMR90 (Donor 1, Miltenyi, 130-096-726) and Crx-iPS (Donor 2) as described previously (Gagliardi *et al*, 2018). We adapted previously described protocol to differentiate iPSCs into brain organoids described earlier (Lancaster *et al*, 2013; Gabriel *et al*, 2016). Five-day-old neurospheres were harvested and embedded in matrigel (Corning, USA) drops. Differentiation medium mixture of DMEM/F12 and Neural Basal Medium (in 1:1 ratio), supplemented with 1:200 N2, 1:100 L-glutamine, 1:100 B27 w/o vitamin A, 100 U/ml penicillin, 100 μg/ml streptomycin, 23 μM insulin (Sigma-Aldrich), 0.05 mM MEM non-essential amino acids (NAA), and 0.05 mM β-mercaptoethanol (Life Technologies) was used to differentiate the matrigel embedded droplets in suspension culture. After 4 days of culturing, embedded neurospheres were transferred to spinner flasks (IBS, Integra biosciences) containing the same differentiation medium supplemented with 0.5 μmol dorsomorphin (Sigma-Aldrich, USA).

Organoids with outgrowing neurons were generated as described previously using 60-day-old organoids (Gabriel *et al*, 2016). Organoids were sliced and plated onto coverslips previously coated with poly-L-ornithine and laminin. The slices were grown in organoid medium for 14 days until the extended neural structures were observed under a stereomicroscope.

Western blot

The gel electrophoretic separation of proteins was performed under denaturing conditions in the presence of SDS in a non-continuous gel system, which consisted of a 5% stacking gel and 10% resolving gel, which was then transferred to nitrocellulose membranes. Once the transfer was finished, the membrane was soaked into 5% milk in TRIS–HCl-based buffer (TBST) for a minimum of 30 min at RT. After incubating with primary antibodies overnight at 4°C, the blots were treated with secondary antibodies at RT for 1 h. Super Signal West Pico or Femto Chemiluminescent substrates (Pierce) were used for detection. Antibody dilutions for Western blots: human convalescent serum AB4 (1:400), polyclonal rabbit SARS-CoV-2 (1:500, Biozol, GTX-GTX632604), monoclonal mouse anti-SARS-CoV-2 (1:500, GeneTex, GTX635679), and mouse anti-GAPDH (1:20,000, Proteintech, 60004-I-Ig). Secondary antibodies goat anti-mouse IgG (H+L) HRP (1:5,000, 31430, Thermo Fisher Scientific), goat anti-rabbit IgG (H+L) HRP (1:5,000, 31466, Invitrogen), anti-human secondary antibodies conjugated to HRP (1:5,000, Thermo Scientific).

Immunofluorescence and confocal microscopy

For light microscopy analysis, monolayer cells (Vero and aspics-derived neurons) were fixed for 10 min. Brain organoids were fixed for 30 min. We used 4% paraformaldehyde/PBS as a fixative (Gabriel *et al*, 2016). Organoids were incubated in 30% sucrose overnight at 4°C, embedded in Tissue-Tek O.C.T. compound (Sakura, Netherlands). Organoids were cryofrozen at −80°C before sectioning into 10–15 μm thin slices using Cryostat Leica CM3050 S. Thin sections and cells were permeabilized with a buffer containing 0.5% Triton X100 for 10 min. Specimens were blocked with 0.5% fish gelatin/PBS for 1 hr, both at room temperature. For SOX2 staining, antigen retrieval was required. For this, sections were treated with repeated heating (microwave) in odium citrate buffer (10 mM Sodium citrate, 0.05% Tween 20, pH 6.0) and were applied before permeabilization and blocking.

We used different antibodies as follows: human convalescent serum (AB1 to AB3 (1:40), AB4, 1:40 or 1:50 or 1:100), rabbit anti-TUJ-1 (1:400, Sigma-Aldrich, T2200), monoclonal mouse anti-phospho-Tau AT180 (1:100, Thermo Scientific, MN1040), polyclonal rabbit anti-phospho-Tau (S396), (1:100, Thermo Fisher), monoclonal mouse anti-SARS-CoV-2 (1:200, GeneTex, GTX635679), rabbit anti-SARS-CoV-2 NP (1:200, Biozol, GTX-GTX632604), rabbit anti S-100 beta (1:100, Abcam, ab52642), rabbit anti Iba-1 (1:100, Abcam, ab178846), rabbit anti-MAP2 (1:100, Proteintech, Cat# 17490-1-AP), mouse anti-Pan-Tau (1:100, DSHB, 5A6). Specimens with primary antibodies were incubated overnight at 4°C. For secondary antibodies, donkey anti-rabbit IgG (H+L) secondary antibody, Alexa Fluor 594 (Thermo Scientific, Cat# A21207), donkey anti-rabbit IgG (H+L) secondary antibody, Alexa Fluor 647 (Thermo Scientific, Cat# A31573), donkey anti-rabbit IgG (H+L) secondary antibody, HRP (Thermo Scientific, Cat# A16023), goat anti-mouse IgG (H+L) secondary antibody, Alexa Fluor 488 (Thermo Scientific, Cat# A28175), Goat anti-mouse IgG (H+L) secondary antibody, Alexa Fluor 594 (Thermo Scientific, Cat# A-11032), Goat anti-mouse IgG (H+L) secondary antibody, Alexa Fluor 647 (Thermo Scientific, Cat# A-21236), Alexa Fluor Dyes conjugated either with goat/donkey anti-mouse, anti-human, or anti-rabbit (1:500 or 1:1,000, molecular probes, Invitrogen) was used. For DNA staining, DAPI at a concentration of 1 μg/ml (Thermo Scientific, Cat# 32670) was used, and the coverslips were mounted using Mowiol (Carl Roth, Germany). The raw images were collected using a Leica SP8 confocal system (Leica microsystems, Germany) and processed with the help of Adobe Photoshop (Adobe Systems, USA). For deconvolution, the captured image files were processed using ZEN software (2.3, SP1, black, 64 bit, release version 14.0.0.0; ZEISS, Oberkochen, Germany) for 3D reconstruction and deconvolution. After deconvolution, files were imported into Fiji and further processed using Image J, Adobe Photoshop CC 2018, and Adobe Illustrator CC 2018. For 3D surface and volume rendering, raw image files were processed using Imaris (64× version 7.7.1).

TUNEL assay

Apoptotic cells were detected by using DeadEnd™ Fluorometric TUNEL System (Promega, G3250, USA) according to the manufacturer's protocol.

Ethical approval and patient samples

Serum samples AB1 and AB2 were obtained under a protocol approved by the ethical committee, medical faculty, University Hospital Düsseldorf, Heinrich-Heine-University (study number 5350). Serum samples AB3 and AB4 were obtained under a protocol approved by the Institutional Review Board of the University of Cologne (protocol 16-054). Human respiratory epithelial cells (hREC) were obtained by nasal brush biopsy from healthy control individuals. The study was endorsed by the local ethical committee at the University of Münster, and each patient gave written informed consent (Study number, 2015-104-f-S, Flimmerepithel) and 2020-274-f-S (COVID-19). Trained physicians from the Department of General Pediatrics, University Hospital of Münster, performed biopsies.

Statistical analysis

The statistical analyses were performed using GraphPad Prism version 8. All experiments were performed at least in triplicates, and the statistical significance of each dataset was analyzed using Student's *t*-test followed by Welche's correction and non-parametric one-way ANOVA followed by Tukey's *post hoc* test. For immunofluorescence-based experiments, we randomized the samples to avoid any bias. The values are expressed as mean ± SD or SEM. Independent experiments have been represented by "*n*".

Data availability

No amenable data sets.

Expanded View for this article is available online.

Acknowledgements
We want to thank Dr. Boris Görg for offering generous support with their microscope facility. We want to thank Ms Gladiola Goranci and Nazlican Altnisk for their excellent technical assistance. This work was financially supported by a grant from Fritz-Thyssen Foundation and Foundation and by the Jürgen Manchot Foundation. We would like to thank the diagnostics department of the Institute of Virology, University Hospital Düsseldorf. Open access funding enabled and organized by Projekt DEAL.

Author contributions
AR and JG conceptualize the project. AR designed, coordinated and conducted experiments. LM and PNO performed virology works. HS supervised virology. EG, PA-I, AM-S and AM provided technical support. AM-S and CK purified antibodies. OG provided iPS cells. HG, DW and FK provided convalescent serum. KW and HO provided airway cells. AW, MA, SH, TH, AD, OA and JT involved in sequencing and viral strains. JG and HS supervised the work. AR and JG wrote the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References

Alonso AD, Di Clerico J, Li B, Corbo CP, Alaniz ME, Grundke-Iqbal I, Iqbal K (2010) Phosphorylation of tau at Thr212, Thr231, and

Ser262 combined causes neurodegeneration. *J Biol Chem* 285: 30851 – 30860

Augustinack JC, Sanders JL, Tsai LH, Hyman BT (2002a) Colocalization and fluorescence resonance energy transfer between cdk5 and AT8 suggests a close association in pre-neurofibrillary tangles and neurofibrillary tangles. *J Neuropathol Exp Neurol* 61: 557 – 564

Augustinack JC, Schneider A, Mandelkow EM, Hyman BT (2002b) Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathol* 103: 26 – 35

Baig AM, Khaleeq A, Ali U, Syeda H (2020) Evidence of the COVID-19 virus targeting the CNS: tissue distribution, host-virus interaction, and proposed neurotropic mechanisms. *ACS Chem Neurosci* 11: 995 – 998

Birey F, Andersen J, Makinson CD, Islam S, Wei W, Huber N, Fan HC, Metzler KRC, Panagiotakos G, Thom N *et al* (2017) Assembly of functionally integrated human forebrain spheroids. *Nature* 545: 54 – 59

Cairns DM, Rouleau N, Parker RN, Walsh KG, Gehrke L, Kaplan DL (2020) A 3D human brain-like tissue model of herpes-induced Alzheimer's disease. *Sci Adv* 6: eaay8828

Castellani RJ, Perry G (2019) Tau biology, tauopathy, traumatic brain injury, and diagnostic challenges. *J Alzheimers Dis* 67: 447 – 467

Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, Ma K, Xu D, Yu H, Wang H *et al* (2020) Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ* 368: m1091

Cho JH, Johnson GV (2004) Primed phosphorylation of tau at Thr231 by glycogen synthase kinase 3beta (GSK3beta) plays a critical role in regulating tau's ability to bind and stabilize microtubules. *J Neurochem* 88: 349 – 358

Cohen TJ, Guo JL, Hurtado DE, Kwong LK, Mills IP, Trojanowski JQ, Lee VM (2011) The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nat Commun* 2: 252

Conde Cardona G, Quintana Pajaro LD, Quintero Marzola ID, Ramos Villegas Y, Moscote Salazar LR (2020) Neurotropism of SARS-CoV 2: mechanisms and manifestations. *J Neurol Sci* 412: 116824

Coolen T, Lolli V, Sadeghi N, Rovai A, Trotta N, Taccone FS, Creteur J, Henrard S, Goffard JC, De Witte O *et al* (2020) Early postmortem brain MRI findings in COVID-19 non-survivors. *Neurology* https://doi.org/10.1212/WNL.0000000000010116

Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brunink S, Schneider J, Schmidt ML *et al* (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 25: 2000045

Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JL, Guimaraes KP, Benazzato C, Almeida N, Pignatari GC, Romero S *et al* (2016) The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* 534: 267 – 271

Darzynkiewicz Z, Galkowski D, Zhao H (2008) Analysis of apoptosis by cytometry using TUNEL assay. *Methods* 44: 250 – 254

De Felice FG, Tovar-Moll F, Moll J, Munoz DP, Ferreira ST (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the central nervous system. *Trends Neurosci* 43: 355 – 357

Flint SJ, Racaniello VR, Rall GF, Skalka AM, Enquist LW (2015) *Principles of virology*. Washington, DC: ASM Press

Frost B, Gotz J, Feany MB (2015) Connecting the dots between tau dysfunction and neurodegeneration. *Trends Cell Biol* 25: 46 – 53

Gabriel E, Wason A, Ramani A, Gooi LM, Keller P, Pozniakovsky A, Poser I, Noack F, Telugu NS, Calegari F *et al* (2016) CPAP promotes timely cilium disassembly to maintain neural progenitor pool. *EMBO J* 35: 803 – 819

Gabriel E, Gopalakrishnan J (2017) Generation of iPSC-derived human brain organoids to model early neurodevelopmental disorders. *J Vis Exp* 122: 55372

Gabriel E, Ramani A, Karow U, Gottardo M, Natarajan K, Gooi LM, Goranci-Buzhala G, Krut O, Peters F, Nikolic M *et al* (2017) Recent zika virus isolates induce premature differentiation of neural progenitors in human brain organoids. *Cell Stem Cell* 20: 397 – 406.

Gagliardi G, Ben M'Barek K, Chaffiol A, Slembrouck-Brec A, Conart JB, Nanteau C, Rabesandratana O, Sahel JA, Duebel J, Orieux G *et al* (2018) Characterization and transplantation of CD73-positive photoreceptors isolated from human iPSC-derived retinal organoids. *Stem Cell Reports* 11: 665 – 680

Giandomenico SL, Lancaster MA (2017) Probing human brain evolution and development in organoids. *Curr Opin Cell Biol* 44: 36 – 43

Giandomenico SL, Mierau SB, Gibbons GM, Wenger LMD, Masullo L, Sit T, Sutcliffe M, Boulanger J, Tripodi M, Derivery E *et al* (2019) Cerebral organoids at the air-liquid interface generate diverse nerve tracts with functional output. *Nat Neurosci* 22: 669 – 679

Gopalakrishnan J (2019) The emergence of stem cell-based brain organoids: trends and challenges phosphorylates. *BioEssays* 41: e1900011

Goranci-Buzhala G, Mariappan A, Gabriel E, Ramani A, Ricci-Vitiani L, Buccarelli M, D'Alessandris QG, Pallini R, Gopalakrishnan J (2020) Rapid and efficient invasion assay of glioblastoma in human brain organoids. *Cell Rep* 31: 107738

Helms J, Kremer S, Merdji H, Clere-Jehl R, Schenck M, Kummerlen C, Collange O, Boulay C, Fafi-Kremer S, Ohana M *et al* (2020) Neurologic features in severe SARS-CoV-2 infection. *N Engl J Med* 382: 2268 – 2270

Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A *et al* (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181: 271 – 280.e8

Janesick A, Wu SC, Blumberg B (2015) Retinoic acid signaling and neuronal differentiation. *Cell Mol Life Sci* 72: 1559 – 1576

Kondo A, Shahpasand K, Mannix R, Qiu J, Moncaster J, Chen CH, Yao Y, Lin YM, Driver JA, Sun Y *et al* (2015) Antibody against early driver of neurodegeneration cis P-tau blocks brain injury and tauopathy. *Nature* 523: 431 – 436

Lamers MM, Beumer J, van dVJ, Knoops K, Puschhof J, Breugem TI, Ravelli RBG, Paul vSJ, Mykytyn AZ, Duimel HQ *et al* (2020) SARS-CoV-2 productively infects human gut enterocytes. *Science* 369: 50 – 54

Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA (2013) Cerebral organoids model human brain development and microcephaly. *Nature* 501: 373 – 379

Lancaster MA, Knoblich JA (2014) Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* 345: 1247125

Lei C, Yang J, Hu J, Sun X (2020) On the calculation of TCID50 for quantitation of virus infectivity. *Virol Sin*. https://doi.org/10.1007/s12250-020-00230-5

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC *et al* (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426: 450 – 454

Lu PJ, Wulf G, Zhou XZ, Davies P, Lu KP (1999) The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* 399: 784 – 788

Luna-Munoz J, Chavez-Macias L, Garcia-Sierra F, Mena R (2007) Earliest stages of tau conformational changes are related to the appearance of a

sequence of specific phospho-dependent tau epitopes in Alzheimer's disease. *J Alzheimers Dis* 12: 365 – 375

McCray PB Jr, Pewe L, Wohlford-Lenane C, Hickey M, Manzel L, Shi L, Netland J, Jia HP, Halabi C, Sigmund CD *et al* (2007) Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *J Virol* 81: 813 – 821

Mesci P, Macia A, Saleh A, Martin-Sancho L, Yin X, Snethlage C, Avansini S, Chanda SK, Muotri A (2020) Sofosbuvir protects human brain organoids against SARS-CoV-2. *BioRxiv* https://doi.org/10.1101/2020.05.30.125856 [PREPRINT]

Monteil V, Kwon H, Prado P, Hagelkruys A, Wimmer RA, Stahl M, Leopoldi A, Garreta E, Hurtado Del Pozo C, Prosper F *et al* (2020) Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* 181: 905 – 913.e7

Nakamura K, Greenwood A, Binder L, Bigio EH, Denial S, Nicholson L, Zhou XZ, Lu KP (2012) Proline isomer-specific antibodies reveal the early pathogenic tau conformation in Alzheimer's disease. *Cell* 149: 232 – 244

Poyiadji N, Shahin G, Noujaim D, Stone M, Patel S, Griffith B (2020) COVID-19-associated acute hemorrhagic necrotizing encephalopathy: CT and MRI features. *Radiology* 296: E119 – E120

Puelles VG, Lutgehetmann M, Lindenmeyer MT, Sperhake JP, Wong MN, Allweiss L, Chilla S, Heinemann A, Wanner N, Liu S *et al* (2020) Multiorgan and renal tropism of SARS-CoV-2. *N Engl J Med* 383: 590 – 592

Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, Yao B, Hamersky GR, Jacob F, Zhong C *et al* (2016) Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* 165: 1238 – 1254

Ramakrishnan MA (2016) Determination of 50% endpoint titer using a simple formula. *World J Virol* 5: 85 – 86

Schlager B, Straessle A, Hafen E (2012) Use of anionic denaturing detergents to purify insoluble proteins after overexpression. *BMC Biotechnol* 12: 95

Sedaghat Z, Karimi N (2020) Guillain Barre syndrome associated with COVID-19 infection: a case report. *J Clin Neurosci* 76: 233 – 235

Sengupta A, Kabat J, Novak M, Wu Q, Grundke-Iqbal I, Iqbal K (1998) Phosphorylation of tau at both Thr 231 and Ser 262 is required for maximal inhibition of its binding to microtubules. *Arch Biochem Biophys* 357: 299 – 309

Song E, Zhang C, Israelow B, Lu P, Weizman OE, Liu F, Dai Y, Szigeti-Buck K, Yasumoto Y, Wang G *et al* (2020) Neuroinvasive potential of SARS-CoV-2 revealed in a human brain organoid model. *BioRxiv* https://doi.org/10.1101/2020.06.25.169946 [PREPRINT]

Streit A, Berliner AJ, Papanayotou C, Sirulnik A, Stern CD (2000) Initiation of neural induction by FGF signalling before gastrulation. *Nature* 406: 74 – 78

Virani A, Rabold E, Hanson T, Haag A, Elrufay R, Cheema T, Balaan M, Bhanot N (2020) Guillain-Barre Syndrome associated with SARS-CoV-2 infection. *IDCases* 20: e00771

Wang Y, Mandelkow E (2016) Tau in physiology and pathology. *Nat Rev Neurosci* 17: 5 – 21

Worl Health Organization (2020) Naming the coronavirus disease (COVID-19) and the virus that causes it. https://wwwwhoint/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

Xiang Y, Tanaka Y, Patterson B, Kang YJ, Govindaiah G, Roselaar N, Cakir B, Kim KY, Lombroso AP, Hwang SM *et al* (2017) Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell* 21: 383 – 398.e7

Yang L, Han Y, Nilsson-Payant BE, Gupta V, Wang P, Duan X, Tang X, Zhu J, Zhao Z, Jaffre F *et al* (2020a) A human pluripotent stem cell-based platform to study SARS-CoV-2 tropism and model virus infection in human cells and organoids. *Cell Stem Cell* 27: 125 – 136.e7

Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M *et al* (2020b) Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 8: 475 – 481

Zempel H, Mandelkow E (2014) Lost after translation: missorting of Tau protein and consequences for Alzheimer disease. *Trends Neurosci* 37: 721 – 32

Zhou J, Li C, Liu X, Chiu MC, Zhao X, Wang D, Wei Y, Lee A, Zhang AJ, Chu H *et al* (2020) Infection of bat and human intestinal organoids by SARS-CoV-2. *Nat Med* 26: 1077 – 1083

3.3.2. Sensitivity of anti-SARS-CoV-2 serological assays in a high-prevalence setting (Chapter 10)

At the beginning of the pandemic, there were no uniform protective measures in place, such as mandatory masks or distance regulations to contain the spread of the virus. In addition, testing for citizens were not easily available, so that in particular asymptomatic infections often went undetected. Various commercial and in-house methods have been rapidly developed to better assess the populations' seroprevalence and to better characterize the seroprotection of convalescent by antibody titers. Here, four commercial serological tests from EUROIMMUN, DiaSorin, Abbott, and Roche as well as an in-house immunofluorescence and neutralization assay using replication-competent SARS-CoV-2 were developed and assessed for their capability to identify SARS-CoV-2 seropositive individuals in a high-prevalence setting. While the commercial assays showed comparable results, the in-house tests and in particular, the neutralization test, the gold standard in the assessment of seroprotection, was found to be the most sensitive assay. Furthermore, this study revealed that commercial SARS-CoV-2 spike protein-based assays correlate better with the neutralization titer than nucleoprotein-based assays.

The following article is published in Eur J Clin Microbiol Infect Dis. 2021 May;40(5):1063-1071. doi: 10.1007/s10096-021-04169-7. by

Müller L*, Ostermann PN*, Walker A, Wienemann T, Mertens A, Adams O, Andree M, Hauka S, Lübke N, Keitel V, Drexler I, Di Cristanziano V, Hermsen DF, Kaiser R, Boege F, Klein F, Schaal H, Timm J, Senff T.
* contributed equally

Contributions:
Study design: FB, FK, HS, JT, and TS; conducting experiments and acquiring data: LM, PNO, RK; VDC, DFH, and TS; analyzing data: LM, PNO, AW, and TS; sample collection: TW, AM, VK, ID, and TS; development and implementation of serological testing: OA, SH, MA, and NL; TS wrote the manuscript with support from NL, AW, LM, PNO, and JT. All authors read and critically revised the manuscript.
Individual contribution: 20% L.M. established and performed the neutralization test.

European Journal of Clinical Microbiology & Infectious Diseases (2021) 40:1063–1071
https://doi.org/10.1007/s10096-021-04169-7

ORIGINAL ARTICLE

Sensitivity of anti-SARS-CoV-2 serological assays in a high-prevalence setting

Lisa Müller[1] · Philipp Niklas Ostermann[1] · Andreas Walker[1] · Tobias Wienemann[2] · Alexander Mertens[3] · Ortwin Adams[1] · Marcel Andree[1] · Sandra Hauka[1] · Nadine Lübke[1] · Verena Keitel[3] · Ingo Drexler[1] · Veronica Di Cristanziano[4] · Derik Franz Hermsen[5] · Rolf Kaiser[4] · Friedrich Boege[5] · Florian Klein[4] · Heiner Schaal[1] · Jörg Timm[1] · Tina Senff[1]

Received: 29 September 2020 / Accepted: 17 January 2021 / Published online: 3 February 2021

Abstract

Evaluation and power of seroprevalence studies depend on the performed serological assays. The aim of this study was to assess four commercial serological tests from EUROIMMUN, DiaSorin, Abbott, and Roche as well as an in-house immunofluorescence and neutralization test for their capability to identify SARS-CoV-2 seropositive individuals in a high-prevalence setting. Therefore, 42 social and working contacts of a German super-spreader were tested. Consistent with a high-prevalence setting, 26 of 42 were SARS-CoV-2 seropositive by neutralization test (NT), and immunofluorescence test (IFT) confirmed 23 of these 26 positive test results (NT 61.9% and IFT 54.8% seroprevalence). Four commercial assays detected anti-SARS-CoV-2 antibodies in 33.3-40.5% individuals. Besides an overall discrepancy between the NT and the commercial assays regarding their sensitivity, this study revealed that commercial SARS-CoV-2 spike-based assays are better to predict the neutralization titer than nucleoprotein-based assays are.

Keywords SARS-CoV-2 · COVID-19 · Serology · Neutralizing antibodies · Immunofluorescence test · Seroprevalence

Lisa Müller and Philipp Niklas Ostermann contributed equally to this work. Author order was determined on the basis of seniority.

Highlights
- Forty-two contacts of a COVID-19 index patient with 26 individuals showing neutralizing antibodies
- Neutralization test suggests low sensitivity of commercial anti-SARS-CoV-2 tests.
- Higher sensitivity of nucleocapsid-restricted assays
- Stronger correlation of spike-based assays with neutralization capacity

Author order was determined on the basis of seniority.

✉ Jörg Timm
timm@uni-duesseldorf.de

✉ Tina Senff
senff@hhu.de

[1] Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[2] Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, 40255 Düsseldorf, Germany

[3] Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, 40255 Düsseldorf, Germany

[4] Institute of Virology, Faculty of Medicine and University Hospital of Cologne, University of Cologne, 50935 Cologne, Germany

[5] Institute of Clinical Chemistry and Laboratory Diagnostics, Medical Faculty, University Düsseldorf, 40255 Düsseldorf, Germany

Introduction

In December 2019, a new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in China and its pandemic spread resulted in more than 30 million infected people according to the World Health Organization [1–3]. Sensitive serological SARS-CoV-2 assays are of great importance for seroprevalence studies and retrospective diagnosis of SARS-CoV-2 infections and aide in estimating prevalence and incidence [4]. Additionally, these assays are necessary to identify donors for convalescent plasma therapy and to determine antibody titers to assess induced immunity after vaccination [5]. Here, we assess and compare different commercial serological tests as well as an in-house neutralization and immunofluorescence test (IFT) in the context of a SARS-CoV-2 high-prevalence setting.

Material and methods

Patients

Serum samples from 42 non-randomized volunteers from the same local area were collected on April 9, 2020. Individuals had direct or indirect contact to a German index patient with a PCR-confirmed SARS-CoV-2 infection and hospitalization on April 24. Previous SARS-CoV-2 PCR testing and symptoms and their onset were queried. Due to the small cohort size, a classification of the severity of symptoms in PCR-confirmed cases was not performed. Health authorities tested 26 of 42 by PCR before sample collection on April 9 with 8 of 26 being SARS-CoV-2 PCR positive.

Ethical statement

The study was approved by the local ethics committee (study number: 5350). Written informed consent was given from each included individual.

Commercially available anti-SARS-CoV-2 test systems

Samples were tested for anti-SARS-CoV-2 antibodies with four commercially available test systems: EUROIMMUN (EI), Roche, Abbott, and DiaSorin, recognizing either SARS-CoV-2 virus spike (S) protein or nucleocapsid (N) antibodies. Euroimmun ELISA directed against the S1 domain of the spike protein detecting IgA and IgG was performed on the Euroimmune Analyzer I-2P according to manufacturer's instructions. OD ratio ≥1.1 for IgA and IgG was considered positive, ≤0.8 as negative, and ≥ 0.8 ≤ 1.1 as borderline. Upper detection limits were OD ratio = 7 for IgA and OD ratio = 10 for IgG. IgG antibodies against S1/S2 domains of SARS-CoV-2 spike were detected through chemiluminescent immunoassay (CLIA) from DiaSorin on a LIAISONX. SARS-CoV-2 S1/S2 IgG antibody concentrations are given as arbitrary units (AU/ml). Samples <12 AU/ml were interpreted as negative, 12-15 AU/ml as borderline, and ≥ 15 AU/ml as positive. The Elecsys® anti-SARS-CoV-2 electrochemiluminescence immunoassay (ECLIA) from Roche was performed on a cobas e801 immunoassay analyzer for the detection of antibodies (including IgG) against SARS-CoV-2 N antigen. Cut-off was based on the measurement of two calculators, and the result was given as signal sample to cutoff (COI). COI <1.0 is negative for anti-SARS-CoV-2 antibodies and COI ≥1.0 is considered positive. The SARS-CoV-2 IgG chemiluminescent microparticle immunoassay (CMIA) from Abbott performed on an ARCHITECT i2000 SR detects IgG antibodies to N. The relation of chemiluminescent RLU and the calibrator is given as the calculated index (S/C). An index (S/C) ≥1.4 is considered positive and <1.4 as negative.

Cell culture

Vero cells (ATCC-CCL-81 obtained from LGC Standards) were grown in Dulbecco's modified essential medium (DMEM) with 1% penicillin and streptomycin (Gibco, 100 U/ml penicillin and 100 μg/ml streptomycin) and 2% fetal calf serum (FCS, PAN Biotech) and were cultured in a 5% CO_2 humified atmosphere at 37 °C.

SARS-CoV-2 virus isolate

For the neutralization test, SARS-CoV-2 isolate NRW-42 was used [6]. The complete sequence of this virus isolate is online (EPI_ISL_425126). There is a single-nucleotide exchange in the spike ORF between the Wuhan-Hu1 reference sequence and the NRW-42 sequence used for NT and IFT. The isolate carries a A>G mutation within the Spike gene at nucleotide position 23,403 which is located within the S1 domain, but outside of the RBD or RBM motif respectively. There is no nucleotide exchange in the nucleocapsid ORF. Unfortunately, antigenic identity of sequences used in the commercial tests is unavailable.

Neutralization test

To detect SARS-CoV-2-neutralizing antibodies, a modified neutralization test was performed [7]. Sera were heat inactivated for 30 min at 56 °C and briefly centrifuged. Initial 1:5 dilutions were prepared in duplicate per patient followed by twofold serial dilutions performed in 50-μl volume with DMEM (1% penicillin and streptomycin, 2% FCS). A total of 50 μl of SARS-CoV-2 stock dilution (final conc. $TCID_{50}$ of 50) was added to the sera dilutions, control sera, and virus only controls (no serum added). Cell-free plates

were pre-incubated at 37 °C for 1 h. Afterwards, 100 μl of cell suspension containing 7×10^4/ml Vero cells was added to samples and cell growth controls. Plates were incubated for 4 days. By microscopic inspection, the titer of neutralizing antibodies was determined as the highest serum dilution without a cytopathic effect (CPE). The reciprocal of the serum dilution is given as the NT titer. A neutralization titer of ≥20 was considered positive. Samples from three individuals with documented coronavirus HCoV-229E, HCoV-OC43, and HCoV-NL63 infections served as controls for cross reactivity (NT titer = 0).

Immunofluorescence test

Vero cells were seeded at a density of 10^4 cells per well into a 48-well plate. After 24 h, cells were infected with SARS-CoV-2 NRW-42 isolate ($TCID_{50}$ of 50) except for controls. At 2 dpi, fixation was performed with ice-cold methanol for 20 min at −20 °C. Subsequently, cells were washed three times with PBS, permeabilized with 0.5% Triton X-100 in PBS for 20 min, and washed again three times. Sera were diluted 1:40 in PBS containing 5% FCS. Cells were incubated in 200-μl patient serum, for 2 h at room temperature. Two SARS-CoV-2-infected wells and one well with uninfected cells were used per patient. Positive control serum was obtained from a patient with high levels of anti-SARS-CoV-2 IgG. After washing, cells were incubated for 1 h at room temperature with anti-human IgG FITC conjugate (Life Technologies, USA) diluted 1:40 in PBS containing 0.1% Evans blue and 5% FCS. Cells were washed three times with PBS and analyzed by microscopy. IFT results were independently evaluated by two staff members. Positive results indicated IFT titer ≥ 40.

Statistical analysis

GraphPad Prism version 8.0.2 was used for statistical analysis. Pearson correlation analysis was used to assess correlation between serological assays. Good correlation was assumed if $r \geq 0.5$ and moderate if $r \geq 0.3$ in combination with $p \leq 0.05$. One-way analysis of variance (ANOVA) was performed for comparison between groups after checking for normal distribution. The respective p values are given as $^{**}p \leq 0.01$ and $^{***}p \leq 0.001$.

Results

SARS-CoV-2 high-prevalence setting—social and working contacts of a German index patient

On 24 February 2020, a patient from the Heinsberg District, Germany was diagnosed SARS-CoV-2 positive by RT-PCR. By February 28, contact tracing from health authorities identified 37 secondary cases. In addition, this index patient was associated with a super spreading event held on 15 February 2020, and >1000 SARS-CoV-2 cases were linked to this event [8].

To assess different serological tests for detection of anti-SARS-CoV-2 antibodies in the context of a high-prevalence setting, blood samples of 42 social and working contacts of this index patient were collected on April 9, 2020 and subsequently analyzed. Importantly, since the index patient was hospitalized on February 24, contact to this patient must have occurred at least 6 weeks before sample collection. Despite reported symptom onset was around 10 days prior to admission to hospital, the patient continued to actively participate in social and business life.

The study population contained slightly more females than males (26/16 61.9%, 38.1%) and individuals were aged between 18 and 70 years (median 44). Although only eight of the 42 individuals were previously tested positive for SARS-CoV-2-RNA by RT-PCR, 26 described symptoms including fever (38.5%), cough (65.4%), fatigue (50%), shortness of breath, or difficulty of breathing (30.8%) while 16 reported no symptoms (Table S1).

Determining SARS-CoV-2 seroprevalence by an in-house SARS-CoV-2 immunofluorescence and neutralization test

First, an in-house neutralization test was performed to identify SARS-CoV-2 seropositive individuals in the described study population. The neutralization test, including the cut-off NT titer of ≥20, was previously validated with 30 SARS-CoV-2 negative sera (NT titer <10; PCR negative or sampled before December 2019) and 25 positive sera from RT-PCR-positive individuals (NT titer 20 to 10,240) and resulting NT titers reflected the current literature [9, 10].

Neutralizing antibodies (NT titers ≥20) were detected in 26 of the 42 serum samples (61.9%). Besides the RT-PCR-confirmed SARS-CoV-2 cases ($n = 8$), 13 out of the 19 symptomatic (68.4%) and 5 of the 15 asymptomatic (33.3%) individuals had neutralizing antibodies (Table S2). Neutralizing antibody levels in asymptomatic individuals were significantly lower compared to PCR-confirmed cases ($p \leq 0.01$, Fig. 1).

To support the NT-based finding of a high SARS-CoV-2 seroprevalence in our study population, an in-house immunofluorescence test (IFT) detecting anti-SARS-CoV-2 IgG was performed. Of the 26 sera positive in the neutralization test, 23 were also positive in the IFT (sensitivity 88.5%, 95% CI [0.710-0.960]). Additionally, negative IFT results were associated with low (≤40) NT titers. This overall supports the finding of a high seroprevalence in our study population as determined by NT.

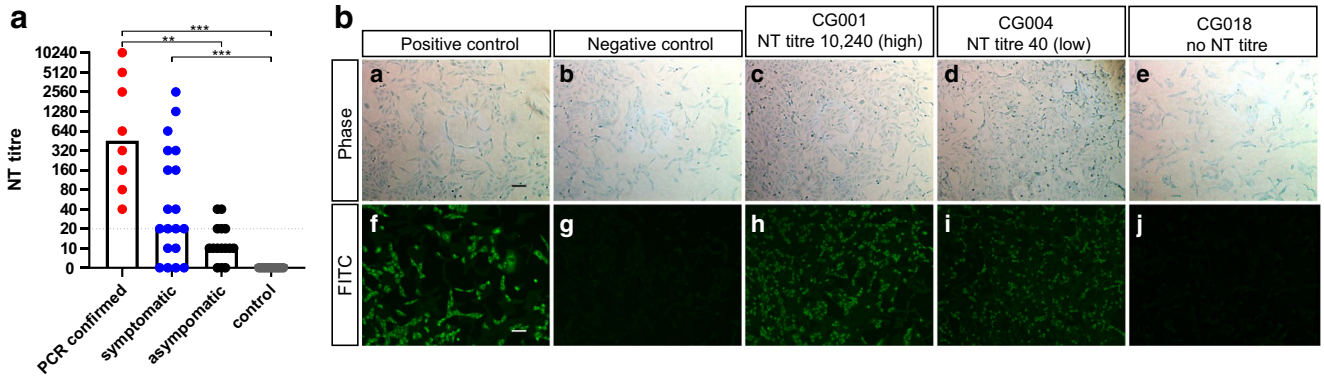


Fig. 1 SARS-CoV-2-neutralizing antibodies stratified according to status of study participants and exemplary immunofluorescence test results, Heinsberg District, Germany, April 2020 (n = 42). FITC: fluorescein isothiocyanate; NT: neutralization test. **a** Neutralization test results of 42 individuals grouped by their status in PCR confirmed (red), symptomatic (blue), and asymptomatic (black) and 11 control sera from healthy individuals sampled before December 2019. The reciprocal of the NT titer is depicted, and bars represent the respective median. The cut-off was defined as ≥20. One-way ANOVA was used to compare groups (**$p \leq 0.01$ and ***$p \leq 0.001$). **b** Exemplary anti-SARS-CoV-2 IgG immunofluorescence test results of 3 out of 42 tested individuals. Phase contrast (a–e) and FITC fluorescence detected at 488 nm (f–j). Serum of a severe hospitalized COVID-19 case served as a positive control (a+f) and (b+g) depict the result of a negative control serum. IFT results from a patient with a high NT titer (c+h; NT titer 10,240), a low NT titer (d+i NT titer 40), and no neutralization potential (e+j). Scale bar is 100 μm

Sensitivity of commercial high-throughput SARS-CoV-2 antibody assays

Based on in-house NT results, supported by IFT, anti-SARS-CoV-2 antibodies were found in 26 of the 42 sera. Since both methods are time-consuming and labor-intensive, suitability of antibody testing was analyzed with four different commercially available automated serological test systems targeting either the nucleocapsid protein (N) or the spike protein (S) of SARS-CoV-2 (Table 1, Figure S1/S2).

Our study included the (i) EUROIMMUN(EI)-anti-SARS-CoV-2 IgA and IgG ELISA test, which contains the S1 subunit of the spike protein (EI S1 IgG or EI S1 IgA); (ii) the LIAISON® SARS-CoV-2 S1/S2 IgG CLIA test, containing the S1 and S2 domain of the spike protein (DiaSorin S1/S2 IgG); (iii.)the SARS-CoV-2 IgG CMIA from Abbott detecting anti-nucleocapsid IgG antibodies (Abbott N IgG) and (iv) the Elecsys® anti-SARS-CoV-2 ECLIA test from Roche which uses biotinylated and ruthenylated nucleocapsid antigen for the determination of antibodies against SARS-CoV-2 (Roche N Ab). For comparison, test results from these commercially available assays were evaluated in relation to the previously described in-house NT.

Of the 26 sera that were tested positive by NT, 12 were also tested positive with the EI S1 IgG or IgA assay, while all 16 NT negative sera have been identified as negative. Of note, 10 of the 26 individuals were positive in the EI S1 IgG as well as the EI S1 IgA test (Table 1). Although the DiaSorin S1/S2 IgG test identified 16 of the 26 NT-positive individuals as positive, 5 of the 16 NT negative individuals were tested positive as well. The Abbott N IgG test detected 16 positive individuals while the Roche N Ab test determined 17 of the 26 NT-positive individuals as positive. In both tests, none of the NT negative sera was above the respective cut-off. Thus, the negative agreement between the NT and EI S1 IgG or IgA test, the Roche N Ab assay, and the Abbott N IgG test was 100%. However, the false-positive rate of the DiaSorin S1/S2 IgG assay was 31.3%.

Taking the performed in-house NT as standard, the EI S1 IgG or IgA test had the lowest sensitivity (46.2%, 95% CI [0.355-0.712]; IgA and/or IgG positive 53.8%). The sensitivity of the Abbott N IgG assay as well as the DiaSorin S1/S2 IgG test was 61.54% (95% CI [0.425-0.776]) in relation to NT results. Notably, the Roche N Ab assay had the highest sensitivity with 65.4% (95% CI [0.462-0.806]) (Table 2). Taken together, the N-restricted tests showed a better sensitivity compared to the S-restricted tests. Nevertheless, the use of the commercially available automated serological test systems described herein would result in the reporting of a lower seroprevalence compared to the in-house neutralization test.

Correlation of commercial SARS-CoV-2 antibody assay results with neutralization ability

To assess which SARS-CoV-2 antibody test platforms are more suitable for predicting neutralizing antibody levels, correlations of commercial SARS-CoV-2 antibody assay results with neutralization test results were determined. The neutralization titer, based on the in-house neutralization test, correlated strongly with all spike antigen-based antibody tests (EI S1 IgA $r = 0.7625$; EI S1 IgG $r = 0.6886$; DiaSorin S1/S2 IgG $r = 0.5641$) (Fig. 2). The weaker correlation of the commercial N-test systems (Abbott N IgG $r = 0.4579$ and Roche N Ab $r = 0.3523$) with the neutralizing antibody titers indicated that S-based systems are more likely to be predictive for functional antibodies.

Table 1 Patient characteristics and serological test results of all anti-SARS-CoV-2 assays performed in this study, Heinsberg District, Germany, April 2020 ($n = 42$)

Sample ID	Age in years	Gender	PCR	Status	Reciprocal NT titer	IgG IFT	EUROIMMUN		DiaSorin	Roche	Abbott
							ELISA IgA (OD ratio)*	ELISA IgG (OD ratio)*	CLIA IgG (AU/ml)*	ECLIA (COI)**	CMIA IgG (index S/C)**
CG001	50	m	pos	sym	10,240	++	>7	8,68	80.7	18.8	4.99
CG005	46	f	pos	sym	5120	+	4.36	>10	372	85.1	8.31
CG007	59	m	pos	sym	2560	++	>7	9,55	128	57.1	9.20
CG012	50	f	pos	as	40	+	0.24	0,62	12.2	0.1	0.06
CG015	54	m	pos	sym	320	+	2.11	0,56	16.2	29.2	4.34
CG031	29	m	pos	sym	80	+	3.41	4,84	68.4	55.7	4.72
CG042	24	f	pos	sym	640	+	4.55	6,63	82.8	23.4	3.75
CG043	43	m	pos	sym	160	+	0.4	2,15	63.4	68.1	6.76
CG002	45	f	neg	sym	10	−	0.51	0,16	4.84	<0.1	0.02
CG003	43	m	neg	sym	20	+	0.58	0,39	6.57	4.1	1.44
CG004	43	f	neg	sym	40	+	0.43	0,9	12.6	1	3.69
CG006	55	f	neg	sym	10	−	0.3	0,21	<3.8	<0.1	0.02
CG009	25	m	neg	sym	0	−	0.29	0,22	14.4	<0.1	0.03
CG011	18	f	n/a	sym	0	−	0.33	0,32	10.6	<0.1	0.04
CG013	43	f	n/a	sym	160	+	0.15	0,18	13.3	<0.1	0.02
CG014	55	f	neg	sym	2560	++	1.49	8,66	105	66.9	8.07
CG016	55	f	neg	sym	1280	+	1.19	1,81	30.7	96.1	9.44
CG017	59	m	neg	sym	160	+	0.54	0,97	10.2	22.2	4.60
CG020	38	f	neg	sym	320	+	1.11	7,84	123	75.9	9.03
CG021	39	m	n/a	sym	20	−	0.74	0,24	18.8	<0.1	0.02
CG022	41	m	neg	sym	640	+	0.68	4,47	60.6	91	8.74
CG026	22	f	neg	sym	0	−	0.2	0,29	16.5	<0.1	0.01
CG028	22	f	n/a	sym	0	−	0.18	0,27	11.2	<0.1	0.01
CG032	27	f	n/a	sym	40	+	1.57	3,97	79.5	10.3	2.17
CG033	29	f	n/a	sym	20	+	0.12	0,21	11.2	<0.1	0.12
CG040	23	f	n/a	sym	320	+	1.32	3,26	32.4	12	3.29
CG044	37	f	neg	sym	20	+	1.41	1,04	18	1.5	1.17
CG008	23	f	neg	as	10	−	0.3	0,21	8.15	<0.1	0.06
CG010	30	f	neg	as	10	−	0.38	0,18	6.5	0.1	0.02
CG018	46	m	n/a	as	0	−	0.27	0,15	13.5	<0.1	0.19
CG019	50	f	n/a	as	10	−	0.3	0,31	<3.8	<0.1	0.02
CG023	49	m	n/a	as	10	−	0.45	0,28	11.6	<0.1	0.01
CG024	46	f	n/a	as	10	−	0.19	0,17	<3.8	<0.1	0.03
CG025	70	f	neg	as	40	−	0.19	0,18	10.8	<0.1	0.02
CG027	47	f	neg	as	20	+	0.43	0,15	10.9	<0.1	0.01
CG029	69	m	n/a	as	40	−	0.21	0,17	22.3	<0.1	0.02
CG030	65	f	n/a	as	0	−	0.13	0,18	16.3	<0.1	0.02
CG034	55	f	neg	as	0	−	0.09	0,18	23.5	<0.1	0.01
CG035	59	m	neg	as	10	−	0.19	0,2	16.3	<0.1	0.02
CG036	31	m	n/a	as	10	−	0.44	0,23	20.4	<0.1	0.01
CG037	25	f	n/a	as	20	+	0.09	0,17	<3.8	<0.1	0.04
CG041	54	m	n/a	as	20	+	0.3	0.15	12.5	<0.1	0.01
Cohort summary $n = 42$											
			Positive, *n*		26	23	12	12	21	17	16
			Borderline, *n*		n/a	n/a	0	3	6	n/a	n/a
			Negative, *n*		16	19	30	27	15	25	26
			Seropositive		(61.9%)	(54.8%)	(28.6%)	(28.6%)	(50.0%)	(40.5%)	(38.1%)

Results are defined as positive according to the manufacturer's instructions: OD ratio ≥1.1; AU/ml ≥15; COI ≥1.0; index (S/C) ≥1.4

AU arbitrary units; *as* asymptomatic; *COI* cut-off index; *CMIA* chemiluminescent microparticle immunoassay; *CLIA* chemiluminescent immunoassay; *ECLIA* electrochemiluminescence immunoassay; *ELISA* enzyme-linked immunosorbent assay; *f* female; *ID* patient identification; *IFT* immunofluorescence test; *m* male; *n/a* not applicable; *NT* neutralization test; *OD* optical density; *S/C* sample/control; *sym* symptomatic

*Anti-spike

**Anti-nucleocapsid

Table 2 Performance characteristics of the EUROIMMUN, DiaSorin, Roche, and Abbott SARS-CoV-2 antibody platforms, Heinsberg District, Germany, April 2020 (n = 26)

		EUROIMMUN			DiaSorin	Roche	Abbott
		S1 IgA	S1 IgG	S1 IgA and/or IgG	S1/S2 IgG	N antibodies	N IgG
Overall NT positive ≥20	n/N	12/26	12/26	14/26	16/26	17/26	16/26
	Value	0.462	0.462	0.538	0.615	0.654	0.615
	(95% CI)	0.288-0.645	0.288-0.645	0.355-0.712	0.425-0.776	0.462-0.806	0.425-0.776

For sensitivity calculations of the commercial assays, only the NT-positive samples (≥20) were used

CI confidence interval; *N* nucleocapsid; *NT*: neutralization test; *S* spike

Discussion

We assessed and compared the sensitivity of four different available commercial antibody tests EUROIMMUN-anti-SARS-CoV-2 IgA and IgG ELISA, LIAISON® SARS-CoV-2 S1/S2 IgG (DiaSorin) CLIA, the SARS-CoV-2 IgG CMIA from Abbott, and the Elecsys® anti-SARS-CoV-2 ECLIA test from Roche as well as an in-house immunofluorescence and neutralization test, in a SARS-CoV-2 high-prevalence setting. For this, we collected serum samples of close contacts to the NRW index patients at least 6 weeks after possible contact occurred. With respect to serological assays, more than 1 month after a putative infection is sufficient to allow detection of SARS-CoV-2 antibody responses [11, 12]. Various reports suggest that virus-specific IgG levels in positive patients are most reliably detected between 17 days and 8 weeks post infection [12–14].

A peculiarity of this study is that the cohort included 42 individuals who had contact to the NRW index patient at the end of February 2020, a time of uncontained viral spread since health authorities had not yet taken containment measures. Although only 8 individuals were previously tested positive

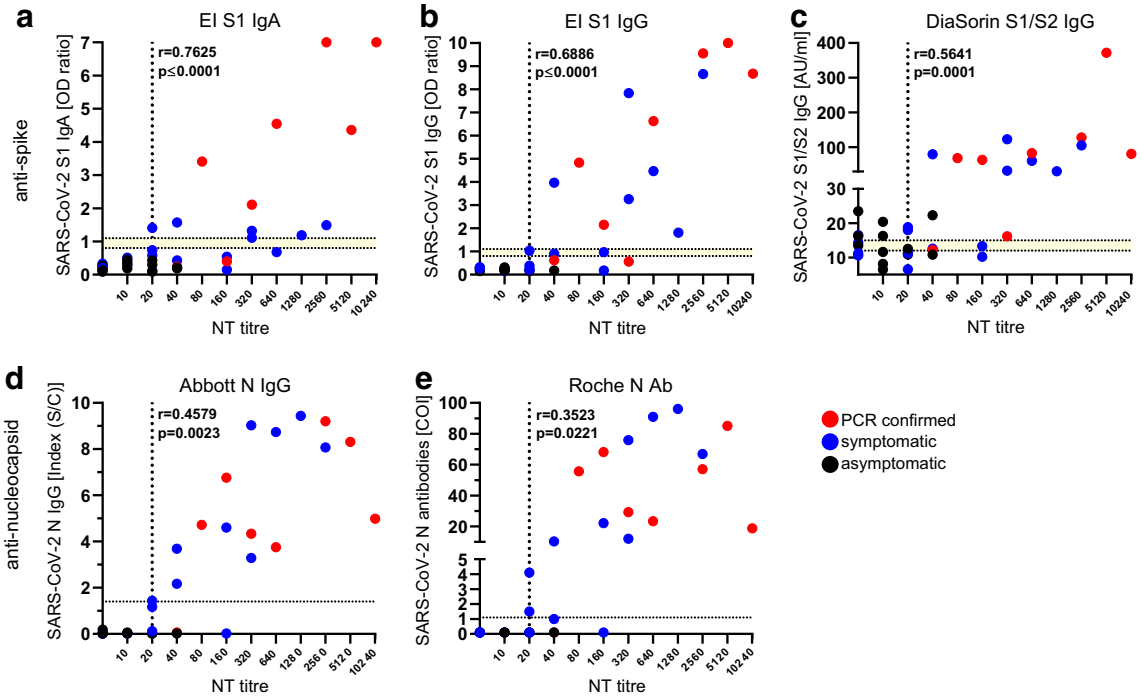


Fig. 2 Correlation between commercial SARS-CoV-2 antibody tests and the neutralization titer, Heinsberg District, Germany, April 2020 (n = 42). AU: arbitrary units; COI: cut-off index; EI: EUROIMMUN; N: nucleocapsid; NT: neutralization test; OD: optical density; r: correlation coefficient; S1: spike domain 1; S2: spike domain 2; S/C: sample/control; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2. The reciprocal of the NT titer is depicted. RT-PCR-confirmed SARS-CoV-2 infections are depicted in red and symptomatic individuals in blue. All asymptomatic individuals are displayed in black. **a** and **b** EUROIMMUN-anti-SARS-CoV-2 IgA and IgG ELISA (Euroimmun). **c** LIAISON® SARS-CoV-2 S1/S2 IgG (DiaSorin). **d** SARS-CoV-2 IgG CMIA (Abbott). **e** Elecsys® anti-SARS-CoV-2 ECLIA test (Roche). The dotted lines indicate the cut-off values recommended by the respective manufacturer to determine positive and negative test results. The borderline area if applicable is indicated in yellow and the vertical line represents the positive cut-off of an NT titer ≥20

for SARS-CoV-2 by PCR, we found that 26 of the 42 individuals had neutralizing antibodies (61.9%) in an in-house neutralizing test (NT). This high seroprevalence is consistent with data from a high school in France describing that 40.9% of pupils, teachers, and the school staff combined had SARS-CoV-2 antibodies [15].

The study by Streeck et al. [16], sampling a random cohort of 1007 people from the area where the German Heinsberg outbreak occurred, found an anti-SARS-CoV-2 seroprevalence in the range of 15%. With respect to this lower seroprevalence reported by Streeck et al., it is important to acknowledge the different sampling approaches. Nevertheless, our data suggest that in such a high-prevalence setting, a substantial number of convalescent COVID-19 cases may be missed with commercial serological assays. Although we found a high concordance of immunofluorescence test (IFT) positive with NT-positive individuals (23 of 26; 88.5% 95% CI [0.710-0.960]), the commercially available SARS-CoV-2 antibody assays from four companies evaluated in this study had led to fewer positive test results, suggesting a lower sensitivity compared to the NT or the IFT in this cohort. These results are in line with Kohmer et al. [17]. Furthermore, this adds to the difference observed between our study and the one of Streeck et al. as only ELISA IgG seropositive sera were analyzed in their NT [16].

Although none of the commercial assays detected more than 65.4% of SARS-CoV-2 NT-positive individuals, we see a slightly higher sensitivity of nucleocapsid assays compared to assays using spike, which is in line with previous findings [18]. However, since the median time of sera sampling after symptom onset was 43 days, this could not be attributed to an earlier anti-N response as described by Grzelak et al. [19]. Notably, both N-restricted assays gave no false-positive results in our small cohort even though a higher cross reactivity to human coronaviruses (HCoVs) has been proposed [5]. The sensitivity of the assays as reported by the manufacturers ranged between 93.8 and 100% ≥14 to >21 days post symptom onset. However, critical COVID-19 cases seem to mount a more robust antibody response than non-critical hospitalized patients [11]. Accordingly, all assays detected higher antibody levels in the 8 confirmed PCR-positive cases, a group that showed a more severe disease course than the other groups. It is important to note that the sensitivities calculated in the current study refer to a high-prevalence setting with mild and asymptomatic courses and only one non-critical hospitalized patient. In turn, the sensitivity might be insufficient for detection of all mild or asymptomatic cases as in this cohort. A study performed in South Korea found that serological testing of PCR confirmed but asymptomatic patients only identified 71% positive individuals, while neutralizing antibodies were detectable in all asymptomatic individuals [13].

In line with previous studies, ELISA and CLIA assays detecting anti-S or anti-N antibodies had a mild to strong correlation with neutralization titers [10, 20]. The EUROIMMUN-anti-SARS-CoV-2 IgA and IgG ELISA tests showed the strongest correlation with antibody function (IgA $r = 0.7625$, $p \leq 0.0001$; IgG $r = 0.6886$, $p \leq 0.0001$) followed by the LIAISON® SARS-CoV-2 S1/S2 IgG assay ($r = 0.5641$, $p = 0.0001$). In the current study, serological assays detecting spike antibodies showed better correlations, which might be due to the fact that the spike protein is the major target for neutralizing antibodies for related coronaviruses and proposedly as well for SARS-CoV-2 [21, 22]. Wu and colleagues as well report that the neutralizing antibody titers correlate with spike-binding antibodies which target the viral S1, RBD, and S2 regions [23].

Of note, the DiaSorin S1/S2 IgG assay rendered five false-positive results from NT assay-negative samples. This finding might suggest a cross reactivity to other endemic HCoVs, possibly because the spike S2 subunit is more conserved among HCoVs than the S1 domain, but this needs to be confirmed with further experiments [10, 20]. Since neutralizing antibody titers in SARS-COV-2-infected individuals varied widely, the EUROIMMUN-anti-SARS-CoV-2 IgA and IgG assay could be considered for pre-screenings to determine optimal donors for convalescent plasma or estimating the induction of virus-specific neutralizing antibodies after vaccination.

Calculation of sensitivity of all commercially available test systems was performed with the NT as reference for past SARS-CoV-2 infection. As 5.7% of hospitalized COVID-19 patients do not generate neutralizing antibodies neither at the time of discharge nor thereafter [23], we could not exclude the possibility that we potentially missed some SARS-CoV-2-infected individuals. Moreover, since we included volunteers from a high-risk area in this small sample study, the data might not be representative for a low-prevalence setting, which is the current situation in most areas of Europe.

In conclusion, the four commercially available high-throughput assays for the detection of SARS-CoV-2-specific antibodies differed in their sensitivity and their potential to predict the neutralization capacity of patient sera. The N-immunoassays tested here seemed to be more sensitive compared to S1 spike protein assays. However, sensitivity of the here described commercial SARS-CoV-2 antibody assays was insufficient for detection of all individuals that were shown to have neutralizing anti-SARS-CoV-2 antibodies. These results should be considered in future population-based seroprevalence studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10096-021-04169-7.

Acknowledgements We thank all participants who donated blood for this study.

Code availability Not applicable.

Author contribution Study design: FB, FK, HS, JT, and TS; conducting experiments and acquiring data: LM, PNO, RK; VDC, DFH, and TS; analyzing data: LM, PNO, AW, and TS; sample collection: TW, AM, VK, ID, and TS; development and implementation of serological testing: OA, SH, MA, and NL; TS wrote the manuscript with support from NL, AW, LM, PNO, and JT. All authors read and critically revised the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Jürgen Manchot Foundation and the Heinz-Ansmann Foundation for AIDS research.

Data availability The data supporting the findings of this study are available within the article and its supplementary material.

Declarations

Ethics approval and consent to participate The study was approved by the local ethics committee (study number: 5350). Written informed consent was given from each included individual.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273. https://doi.org/10.1038/s41586-020-2012-7
2. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, LLM P, Samborskiy DV, Sidorov IA, Sola I, Ziebuhr J, Coronaviridae Study Group of the International Committee on Taxonomy of V (2020) The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5(4):536–544. https://doi.org/10.1038/s41564-020-0695-z
3. COVID 19 Pandemic Emergency Information (2020) https://covid19.who.int/
4. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, Rathore U, Goldgof GM, Whitty C, Woo JM, Gallman AE, Miller TE, Levine AG, Nguyen DN, Bapat SP, Balcerek J, Bylsma SA, Lyons AM, Li S, Wong AW, Gillis-Buck EM, Steinhart ZB, Lee Y, Apathy R, Lipke MJ, Smith JA, Zheng T, Boothby IC, Isaza E, Chan J, Acenas DD 2nd, Lee J, Macrae TA, Kyaw TS, Wu D, Ng DL, Gu W, York VA, Eskandarian HA, Callaway PC, Warrier L, Moreno ME, Levan J, Torres L, Farrington LA, Loudermilk RP, Koshal K, Zorn KC, Garcia-Beltran WF, Yang D, Astudillo MG, Bernstein BE, Gelfand JA, Ryan ET, Charles RC, Iafrate AJ, Lennerz JK, Miller S, Chiu CY, Stramer SL, Wilson MR, Manglik A, Ye CJ, Krogan NJ, Anderson MS, Cyster JG, Ernst JD, AHB W, Lynch KL, Bern C, Hsu PD, Marson A (2020) Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. Nat Biotechnol. https://doi.org/10.1038/s41587-020-0659-0
5. Krammer F, Simon V (2020) Serology assays to manage COVID-19. Science eabc1227. https://doi.org/10.1126/science.abc1227
6. Walker A, Houwaart T, Wienemann T, Kohns Vasconcelos M, Strelow D, Senff T, Hülse L, Adams O, Andree M, Hauka S, Feldt T, Jensen B-E, Keitel V, Kindgen-Milles D, Timm J, Pfeffer K, Dilthey AT (2020) Genetic structure of SARS-CoV-2 in Western Germany reflects clonal superspreading and multiple independent introduction events. medRxiv: 2020.2004.2025.20079517. doi:https://doi.org/10.1101/2020.04.25.20079517
7. Gauger PC, Vincent AL (2014) Serum virus neutralization assay for detection and quantitation of serum-neutralizing antibodies to influenza A virus in swine. Methods Mol Biol 1161:313–324. https://doi.org/10.1007/978-1-4939-0758-8_26
8. RKI (2020) Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute 04/03/2020
9. Kohmer N, Westhaus S, Rühl C, Ciesek S, Rabenau HF Clinical performance of different SARS-CoV-2 IgG antibody tests. J Med Virol n/a (n/a). https://doi.org/10.1002/jmv.26145
10. Nisreen MAO, Marcel AM, Wentao L, Chunyan W, Corine HG, Victor MC, Mart ML, Reina SS, de Erwin B, Felicity DC, Yazdan Y, Le Quentin H, Diane D, Nadhira H-F, Chantal BEMR, Berend-Jan B, Christian D, Marion PGK, Bart LH (2020) Severe acute respiratory syndrome coronavirus 2−specific antibody responses in coronavirus disease 2019 patients. Emerg Infect Dis J 26(7). https://doi.org/10.3201/eid2607.200841
11. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, Wang X, Yuan J, Li T, Li J, Qian S, Hong C, Wang F, Liu Y, Wang Z, He Q, Li Z, He B, Zhang T, Fu Y, Ge S, Liu L, Zhang J, Xia N, Zhang Z (2020) Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis. https://doi.org/10.1093/cid/ciaa344
12. Long Q-X, Liu B-Z, Deng H-J, Wu G-C, Deng K, Chen Y-K, Liao P, Qiu J-F, Lin Y, Cai X-F, Wang D-Q, Hu Y, Ren J-H, Tang N, Xu Y-Y, Yu L-H, Mo Z, Gong F, Zhang X-L, Tian W-G, Hu L, Zhang X-X, Xiang J-L, Du H-X, Liu H-W, Lang C-H, Luo X-H, Wu S-B, Cui X-P, Zhou Z, Zhu M-M, Wang J, Xue C-J, Li X-F, Wang L, Li Z-J, Wang K, Niu C-C, Yang Q-J, Tang X-J, Zhang Y, Liu X-M, Li J-J, Zhang D-C, Zhang F, Liu P, Yuan J, Li Q, Hu J-L, Chen J, Huang A-L (2020) Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med. https://doi.org/10.1038/s41591-020-0897-1
13. Choe PG, Kang CK, Suh HJ, Jung J, Kang E, Lee SY, Song KH, Kim HB, Kim NJ, Park WB, Kim ES, Oh MD (2020) Antibody responses to SARS-CoV-2 at 8 weeks postinfection in asymptomatic patients. Emerg Infect Dis 26(10):2484–2487. https://doi.org/10.3201/eid2610.202211
14. Lou B, Li TD, Zheng SF, Su YY, Li ZY, Liu W, Yu F, Ge SX, Zou QD, Yuan Q, Lin S, Hong CM, Yao XY, Zhang XJ, Wu DH, Zhou GL, Hou WH, Li TT, Zhang YL, Zhang SY, Fan J, Zhang J, Xia

NS, Chen Y (2020) Serology characteristics of SARS-CoV-2 infection after exposure and post-symptom onset. Eur Respir J 56(2). https://doi.org/10.1183/13993003.00763-2020

15. Fontanet A, Tondeur L, Madec Y, Grant R, Besombes C, Jolly N, Fernandes Pellerin S, Ungeheuer M-N, Cailleau I, Kuhmel L, Temmam S, Huon C, Chen K-Y, Crescenzo B, Munier S, Demeret C, Grzelak L, Staropoli I, Bruel T, Gallian P, Cauchemez S, van der Werf S, Schwartz O, Eloit M, Hoen B (2020) Cluster of COVID-19 in northern France: a retrospective closed cohort study. medRxiv:2020.2004.2018.20071134. https://doi.org/10.1101/2020.04.18.20071134
16. Streeck H, Schulte B, Kuemmerer B, Richter E, Hoeller T, Fuhrmann C, Bartok E, Dolscheid R, Berger M, Wessendorf L, Eschbach-Bludau M, Kellings A, Schwaiger A, Coenen M, Hoffmann P, Noethen M, Eis-Huebinger A-M, Exner M, Schmithausen R, Schmid M, Hartmann G (2020) Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. medRxiv:2020.2005.2004.20090076. https://doi.org/10.1101/2020.05.04.20090076
17. Kohmer N, Westhaus S, Rühl C, Ciesek S, Rabenau HF (2020) Brief clinical evaluation of six high-throughput SARS-CoV-2 IgG antibody assays. J Clin Virol 129:104480. https://doi.org/10.1016/j.jcv.2020.104480
18. Burbelo PD, Riedo FX, Morishima C, Rawlings S, Smith D, Das S, Strich JR, Chertow DS, Davey RT Jr, Cohen JI (2020) Sensitivity in detection of antibodies to nucleocapsid and spike proteins of severe acute respiratory syndrome coronavirus 2 in patients with coronavirus disease 2019. J Infect Dis. https://doi.org/10.1093/infdis/jiaa273
19. Grzelak L, Temmam S, Planchais C, Demeret C, Huon C, Guivel F, Staropoli I, Chazal M, Dufloo J, Planas D, Buchrieser J, Rajah MM, Robinot R, Porrot F, Albert M, Chen K-Y, Crescenzo B, Donati F, Anna F, Souque P, Gransagne M, Bellalou J, Nowakowski M, Backovic M, Bouadma L, Le Fevre L, Le Hingrat Q, Descamps D, Pourbaix A, Yazdanpanah Y, Tondeur L, Besombes C, Ungeheuer M-N, Mellon G, Morel P, Rolland S, Rey F, Behillil S, Enouf V, Lemaitre A, Creach M-A, Petres S, Escriou N, Charneau P, Fontanet A, Hoen B, Bruel T, Eloit M, Mouquet H, Schwartz O, van der Werf S (2020) SARS-CoV-2 serological analysis of COVID-19 hospitalized patients, pauci-symptomatic individuals and blood donors. medRxiv:2020.2004.2021.20068858. https://doi.org/10.1101/2020.04.21.20068858
20. GeurtsvanKessel CH, Okba NMA, Igloi Z, Bogers S, Embregts CWE, Laksono BM, Leijten L, Rokx C, Rijnders B, Rahamat-Langendoen J, van den Akker JPC, van Kampen JJA, van der Eijk AA, van Binnendijk RS, Haagmans B, Koopmans M (2020) An evaluation of COVID-19 serological assays informs future diagnostics and exposure assessment. Nat Commun 11(1):3436. https://doi.org/10.1038/s41467-020-17317-y
21. Berry JD, Hay K, Rini JM, Yu M, Wang L, Plummer FA, Corbett CR, Andonov A (2010) Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. mAbs 2(1):53–66. https://doi.org/10.4161/mabs.2.1.10788
22. Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova V, McMahon M, Jiang K, Arunkumar GA, Jurczyszak D, Polanco J, Bermudez-Gonzalez M, Kleiner G, Aydillo T, Miorin L, Fierer DS, Lugo LA, Kojic EM, Stoever J, Liu STH, Cunningham-Rundles C, Felgner PL, Moran T, García-Sastre A, Caplivski D, Cheng AC, Kedzierska K, Vapalahti O, Hepojoki JM, Simon V, Krammer F (2020) A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat Med. https://doi.org/10.1038/s41591-020-0913-5
23. Wu F, Wang A, Liu M, Wang Q, Chen J, Xia S, Ling Y, Zhang Y, Xun J, Lu L, Jiang S, Lu H, Wen Y, Huang J (2020) Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. medRxiv: 2020.2003.2030.20047365.https://doi.org/10.1101/2020.03.30.20047365

3.3.3. Reconvalescent plasma/camostat mesylate in early SARS-CoV-2 q-PCR positive high-risk individuals (RES-Q-HR): a structured summary of a study protocol for a randomized controlled trial (Chapter 11)

The evaluation of strategies to treat COVID-19 and prevent further progression of the disease in the early stages of infection is an important pillar in the pandemic management. This is important for patients with pre-existing risk factors such as old age and conditions including diabetes and obesity, chronic obstructive pulmonary disease, and others, which increase the risk of moderate to severe COVID-19. In this 4-arm, multicenter, randomized, partly double-blind, controlled study the safety and efficacy of the convalescent plasma and camostat mesylate administered early after diagnosis of SARS-CoV-2 infection is evaluated. Convalescent plasma screened for high SARS-CoV-2 specific antibody titers and especially high titers of neutralizing antibodies as determined by the in-house developed neutralization assay using replication-competent SARS-CoV-2 represents an important antiviral strategy. Camostat mesylate as an interventional drug acts as an inhibitor of the host cell serine protease TMPRSS2 and prevents the virus from entering the cell. The working hypothesis to be tested in the RES-Q-HR study is that the early administration of convalescent plasma or camostat mesylate reduces the likelihood of disease progression into severe stages in high-risk patients.

The following study protocol is published in Trials. 2021 May 17;22(1):343. doi: 10.1186/s13063-021-05181-0. by

Keitel V, Jensen B, Feldt T, Fischer JC, Bode JG, Matuschek C, Bölke E, Budach W, Plettenberg C, Scheckenbach K, Kindgen-Milles D, Timm J, **Müller L**, Kolbe H, Stöhr A, Calles C, Hippe A, Verde P, Spinner CD, Schneider J, Wolf T, Kern WV, Nattermann J, Zoufaly A, Ohmann C, Luedde T; RES-Q-HR Trial Team.

Contributions:
VK, TF, BJ, JF, and EB had the idea for the study; VK, TF, and BJ generated the initial concept for the study. All authors made a substantial contribution to the design and the concept of the study. All authors approved this summary.
Individual contribution: 5% L.M. established and performed the neutralization test for convalescent plasma donors.

Keitel et al. Trials (2021) 22:343
https://doi.org/10.1186/s13063-021-05181-0

Trials

LETTER Open Access

Reconvalescent plasma/camostat mesylate in early SARS-CoV-2 Q-PCR positive high-risk individuals (RES-Q-HR): a structured summary of a study protocol for a randomized controlled trial

Verena Keitel[1*], Björn Jensen[1], Torsten Feldt[1], Johannes C. Fischer[2], Johannes G. Bode[1], Christiane Matuschek[3], Edwin Bölke[3], Wilfried Budach[3], Christian Plettenberg[4], Kathrin Scheckenbach[4], Detlef Kindgen-Milles[5], Jörg Timm[6], Lisa Müller[6], Henrike Kolbe[7], Andreas Stöhr[7], Christian Calles[7], Andreas Hippe[7], Pablo Verde[7], Christoph D. Spinner[8], Jochen Schneider[8], Timo Wolf[9], Winfried V. Kern[10], Jacob Nattermann[11], Alexander Zoufaly[12], Christian Ohmann[7], Tom Luedde[1] RES-Q-HR Trial Team

Abstract

Objectives: Currently, there are no approved treatments for early disease stages of COVID-19 and few strategies to prevent disease progression after infection with SARS-CoV-2. The objective of this study is to evaluate the safety and efficacy of convalescent plasma (CP) or camostat mesylate administered within 72 h of diagnosis of SARS-CoV-2 infection in adult individuals with pre-existing risk factors at higher risk of getting seriously ill with COVID-19. Camostat mesylate acts as an inhibitor of the host cell serine protease TMPRSS2 and prevents the virus from entering the cell. CP represents another antiviral strategy in terms of passive immunization. The working hypothesis to be tested in the RES-Q-HR study is that the early use of CP or camostat mesylate reduces the likelihood of disease progression to (modified) WHO stages 4b-8 in SARS-CoV-2-positive adult patients at high risk of moderate or severe COVID-19 progression.

Trial design: This study is a 4-arm (parallel group), multicenter, randomized (2:2:1:1 ratio), partly double-blind, controlled trial to evaluate the safety and efficacy of convalescent plasma (CP) or camostat mesylate with control or placebo in adult patients diagnosed with SARS-CoV-2 infection and high risk for progression to moderate/severe COVID-19. Superiority of the intervention arms will be tested.

(Continued on next page)

* Correspondence: Verena.Keitel@med.uni-duesseldorf.de
[1]Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany
Full list of author information is available at the end of the article

(Continued from previous page)

Participants: The trial is conducted at 10–15 tertiary care centers in Germany. Individuals aged 18 years or above with ability to provide written informed consent with SARS-CoV-2 infection, confirmed by PCR within 3 days or less before enrolment and the presence of at least one SARS-CoV-2 symptom (such as fever, cough, shortness of breath, sore throat, headache, fatigue, smell/and or taste disorder, diarrhea, abdominal symptoms, exanthema) and symptom duration of not more than 3 days.
Further inclusion criteria comprise:
Presence of at least one of the following criteria indicating increased risk for severe COVID-19:

- Age > 75 years
- Chronic obstructive pulmonary disease (COPD) and/or pulmonary fibrosis
- BMI > 40 kg/m^2
- Age > 65 years with at least one other risk factor (BMI > 35 kg/m^2, coronary artery disease (CAD), chronic kidney disease (CKD) with GFR < 60 ml/min but ≥ 30 ml/min, diabetes mellitus, active tumor disease)
- BMI > 35 kg/m^2 with at least one other risk factor (CAD, CKD with GFR < 60 ml/min but ≥ 30 ml/min, diabetes mellitus, active tumor disease)

Exclusion criteria:

1. Age < 18 years
2. Unable to give informed consent
3. Pregnant women or breastfeeding mothers
4. Previous transfusion reaction or other contraindication to a plasma transfusion
5. Known hypersensitivity to camostat mesylate and/or severe pancreatitis
6. Volume stress due to CP administration would be intolerable
7. Known IgA deficiency
8. Life expectancy < 6 months
9. Duration SARS-CoV-2 typical symptoms > 3 days
10. SARS-CoV-2 PCR detection older than 3 days
11. SARS-CoV-2 associated clinical condition ≥ WHO stage 3 (patients hospitalized for other reasons than COVID-19 may be included if they fulfill all inclusion and none of the exclusion criteria)
12. Previously or currently hospitalized due to SARS-CoV-2
13. Previous antiviral therapy for SARS-CoV-2
14. ALT or AST > 5 x ULN at screening
15. Liver cirrhosis > Child A (patients with Child B/C cirrhosis are excluded from the trial)
16. Chronic kidney disease with GFR < 30 ml/min
17. Concurrent or planned anticancer treatment during trial period
18. Accommodation in an institution due to legal orders (§40(4) AMG).
19. Any psycho-social condition hampering compliance with the study protocol.
20. Evidence of current drug or alcohol abuse
21. Use of other investigational treatment within 5 half-lives of enrolment is prohibited
22. Previous use of convalescent plasma for COVID-19
23. Concomitant proven influenza A infection
24. Patients with organ or bone marrow transplant in the three months prior to screening visit

(Continued on next page)

(Continued from previous page)

Intervention and comparator: Participants will be randomized to the following 4 groups:

1) Convalescent plasma (CP), 2 units at screening/baseline visit (day 0) or day 1; CP is defined by the presence of neutralizing anti-SARS-CoV-2 antibodies with titers ≥ 1:160; individuals with body weight ≥ 150 kg will receive a third unit of plasma on day 3
2) Camostat mesylate (200 mg per capsule, one capsule taken each in the morning, afternoon and evening on days 1–7)
3) Standard of care (SOC, control for CP)
4) Placebo (identical in appearance to camostat mesylate capsules, one capsule taken each morning, afternoon and evening on days 1–7; for camostat mesylate control group)

Participants will be monitored after screening/baseline on day 3, day 5, day 8, and day 14. On day 28 and day 56, telephone visits and on day 90, another outpatient visit are scheduled.
Adverse events and serious adverse events will be monitored and reported until the end of the study. An independent data safety monitoring committee will review trial progression and safety.

Main outcomes: The primary endpoint of the study is the cumulative number of individuals who progress to or beyond category 4b on the modified WHO COVID-19 ordinal scale (defined as hospitalization with COVID-19 pneumonia and additional oxygen demand via nasal cannula or mask) within 28 days after randomization.

Randomization: Participants will be randomized using the Alea-Tool (aleaclinical.com) in a 2:2:1:1 ratio to the treatment arms (1) CP, (2) camostat mesylate, (3) standard of care (SoC), and (4) placebo matching camostat mesylate. Randomization will be stratified by study center.

Blinding (masking): The camostat mesylate treatment arm and the respective placebo will be blinded for participants, caregivers, and those assessing outcomes.
The treatment arms convalescent plasma and standard of care will not be blinded and thus are open-labeled, unblinded.

Numbers to be randomized (sample size): Overall, $n = 994$ participants will be randomized to the following groups: $n = 331$ to convalescent plasma (CP), $n = 331$ to camostat mesylate, $n = 166$ to standard of care (SoC), and $n = 166$ to placebo matching camostat mesylate.

Trial status: The RES-Q-HR protocol (V04F) was approved on the 18 December 2020 by the local ethics committee and by the regulatory institutions PEI/BfARM on the 2 December 2020. The trial was opened for recruitment on 26 December 2020; the first patient was enrolled on 7 January 2021 and randomized on 8 January 2021.
Recruitment shall be completed by June 2021. The current protocol version RES-Q HR V05F is from 4 January 2021, which was approved on the 18 January 2021.

Trial registration: EudraCT Number 2020-004695-18. Registered on September 29, 2020.
ClinicalTrial.gov NCT04681430. Registered on December 23, 2020, prior to the start of the enrollment (which was opened on December 26, 2020).

Full protocol: The full protocol (V05F) is attached as an additional file, accessible from the Trials website (Additional file 1). In the interest in expediting dissemination of this material, the familiar formatting has been eliminated; this letter serves as a summary of the key elements of the full protocol.
The study protocol has been reported in accordance with the Standard Protocol Items: Recommendations for Clinical Interventional Trials (SPIRIT) guidelines (Additional file 2).

Keywords: COVID-19, Randomized controlled trial, Protocol, Convalescent plasma, Camostat mesylate, Antiviral therapy, Early phase of SARS-CoV-2 infection

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13063-021-05181-0.

Additional file 1. Full study protocol.

Additional file 2. SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*.

Acknowledgements
We would like to thank Sabine Tebbe, Martha Holtfreter, Bettina See, Lisa Knopp, and Stefan Boxnick for coordination of the study site in Duesseldorf, design and revision of the eCRF. We thank Hogrefe for contributing the SF-12 questionnaire free of charge for this trial.

RES-Q-HR Trial Team
Simon Labuhn[1], Noemi Freise[1], Alexander Killer[1], Caroline Klindt[1], Carola Dröge[1], Anselm Kunstein[1], David Schoeler[1], Sandra Jost[1], Erik Lehnert[2], Stefanie Ackerstaff[2], Timo Brandenburger[5], Christina Westhoff[11], Christine Fritsch[11], Stephanie Laer[12], Andrea Icks[13]

[1] Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany.
[2] Institute for Transplant Diagnostics and Cell Therapeutics, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany.
[5] Department of Anesthesiology, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany.
[11] University Pharmacy, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, D-40225 Duesseldorf, Germany.
[12] Clinical Pharmacy and Pharmacotherapy, Mathematics and Natural Science Faculty, Heinrich Heine University, Duesseldorf, Germany.
[13] Institute for Health Services Research and Health Economics, Heinrich Heine University, Duesseldorf, Germany.
Email Address Trial Team:
Simon Labuhn simon.labuhn@med.uni-duesseldorf.de
Noemi Freise noemi.freise@med.uni-duesseldorf.de
Alexander Killer alexander.killer@med.uni-duesseldorf.de
Caroline Klindt caroline.klindt@med.uni-duesseldorf.de
Carola Dröge carola.droege@med.uni-duesseldorf.de
Anselm Kunstein anselm.kunstein@med.uni-duesseldorf.de
David Schoeler david.schoeler@med.uni-duesseldorf.de
Sandra Jost sandra.jost@med.uni-duesseldorf.de
Erik Lehnert erik.lehnert@med.uni-duesseldorf.de
Stefanie Ackerstaff stefanie.ackerstaff@med.uni-duesseldorf.de
Timo Brandenburger timo.brandenburger@med.uni-duesseldorf.de
Christina Westhoff christina.westhoff@med.uni-duesseldorf.de
Christine Fritsch christiane.fritsch@med.uni-duesseldorf.de
Stephanie Läer stephanie.laeer@uni-duesseldorf.de
Andrea Icks andrea.icks@uni-duesseldorf.de

Authors' contributions

VK, TF, BJ, JF, and EB had the idea for the study; VK, TF, and BJ generated the initial concept for the study. All authors made a substantial contribution to the design and the concept of the study. All authors approved this summary.

Funding

The study is funded by the Bundesministerium für Gesundheit (BMG, German Ministry of Health). The funder did not contribute to the study design and will not contribute to data analysis and interpretation. The funder will be informed of the trial results prior to publication but has no right of veto of any publication. AI is also supported by the MethodCOV project, which is part of the "National Network University Medicine (NUM)" funded by the Federal Ministry of Education and Research (BMBF) (grant number 01KX2021). The National Network University Medicine is coordinated at the Charité and supervised by the German Aerospace Center (DLR Project Management Agency).

Availability of data and materials

All investigators will have access to the final trial dataset. Data will be available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Medical Faculty at Heinrich Heine University on 12/18/2020 (MC-LKP-1186).
Informed written consent will be obtained from all participants prior to enrolment into the study. All participants will be aged 18 years or above.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany. [2]Institute for Transplant Diagnostics and Cell Therapeutics, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany. [3]Department of Radiation Oncology, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany. [4]Department of Ear, Nose and Throat disease, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany. [5]Department of Anesthesiology, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany. [6]Institute for Virology, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University Duesseldorf, Universitaetsstr. 1, D-40225 Duesseldorf, Germany. [7]Coordination Center for Clinical Studies, Medical Faculty, Heinrich Heine University Duesseldorf, D-40225 Duesseldorf, Germany. [8]Technical University of Munich, School of Medicine, University Hospital Rechts der Isar, Ismaninger Str. 22, D-81675 Munich, Germany. [9]Department of Internal Medicine, Infectious Diseases, University Hospital Frankfurt, Goethe University, Theodor-Stern-Kai 7, D-60590 Frankfurt am Main, Germany. [10]Department of Internal Medicine II, Infectious Diseases, University Hospital Freiburg, Hugstetter Str. 55, D-79106 Freiburg, Germany. [11]Department of Internal Medicine I, University Hospital Bonn, Sigmund-Freud-Str. 25, D-53127 Bonn, Germany. [12]Department of Medicine IV, Kaiser Franz Josef Hospital, Kundratstrasse 3, 1100 Vienna, Austria.

Received: 5 March 2021 Accepted: 10 March 2021
Published online: 17 May 2021

Publisher's Note

3.3.4. Convalescent plasma achieves SARS-CoV-2 viral clearance in a patient with persistently high viral replication over 8 weeks due to severe combined immunodeficiency (SCID) and graft failure (Chapter 12)

This case report describes the disease course and cure of SARS-CoV-2 infection in a patient with severe combined immunodeficiency (SCID) and graft failure. The clinical characteristic of SCID is the absence of a humoral immune response to infections, an important feature to achieve viral clearance. The patient persistently showed high SARS-CoV-2 RNA concentrations in respiratory samples and shedding of infectious viral particles over a prolonged period. After developing signs of pneumonia, the patient was first treated with Remdesivir, a drug that stalls SARS-CoV-2 polymerase activity. However, sustained viral clearance was only achieved after transfusion of convalescent plasma with high antibody titers. This observation underscores the necessity of the humoral immune response for SARS-CoV-2 clearance.

The following article is published in Front. Immunol. 2021 12:645989. doi: 10.3389/fimmu.2021.645989. by

Keitel V*, Bode JG*, Feldt T, Walker A, **Müller L**, Kunstein A, Klindt C, Killer A, Senff T, Timm J, Ostermann P, Damagnez M, Lübke N, Adams O, Schaal H, Antoch G, Neubert J, Albrecht P, Meuth S, Elben S, Mohring A, Fischer JC, Bölke E, Hoenig M, Schulz AS, Luedde T, Jensen B*.
* contributed equally

Contributions:
VK, JB, TF, and BJ initiated this work, supervised the study, and drafted the manuscript. VK, JB, TF, BJ, AKu, CK, AKi, TL, AM, AS, MH, PA, GA, JN, SM, and SE took care of the patient, analyzed the clinical data and phenotype, determined diagnostic procedures and treatment plan, and interpreted treatment responses. AW, LM, TS, JT, PO, MD, NL, OA, and HS developed virological test strategies (ELISAs, testing for neutralizing SARS-CoV-2 antibodies, viral sequencing), and performed and interpreted virological data. All authors critically revised the manuscript. All authors contributed to the article and approved the submitted version.
Individual contribution: 15 %L.M. established and performed experiments on the infectivity of the patients' swabs. L.M. established and performed neutralization tests.

frontiers in Immunology

CASE REPORT
published: 03 May 2021
doi: 10.3389/fimmu.2021.645989

Case Report: Convalescent Plasma Achieves SARS-CoV-2 Viral Clearance in a Patient With Persistently High Viral Replication Over 8 Weeks Due to Severe Combined Immunodeficiency (SCID) and Graft Failure

Verena Keitel[1†], Johannes Georg Bode[1*†], Torsten Feldt[1], Andreas Walker[2], Lisa Müller[2], Anselm Kunstein[1], Caroline Klindt[1], Alexander Killer[1], Tina Senff[2], Jörg Timm[2], Philipp Ostermann[2], Maximilian Damagnez[2], Nadine Lübke[2], Ortwin Adams[2], Heiner Schaal[2], Gerald Antoch[3], Jennifer Neubert[4], Philipp Albrecht[5], Sven Meuth[5], Saskia Elben[5], Annemarie Mohring[6], Johannes C. Fischer[7], Edwin Bölke[8], Manfred Hoenig[9], Ansgar S. Schulz[9], Tom Luedde[1] and Björn Jensen[1*†]*

[1] Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [2] Institute of Virology, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [3] Department of Diagnostic and Interventional Radiology, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [4] Department of Pediatric Oncology, Hematology and Clinical Immunology, University Children's Hospital, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [5] Department of Neurology, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [6] Department of Hematology, Oncology and Clinical Immunology, Medical Faculty, Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, [7] Institute for Transplant Diagnostics and Cell Therapeutics, University Hospital Duesseldorf, Medical Faculty, Heinrich-Heine-University, Duesseldorf, Germany, [8] Department of Radiation Oncology, University Hospital Duesseldorf, Medical Faculty Heinrich-Heine-University, Duesseldorf, Germany, [9] Pediatric Stem Cell Transplantation Unit, University Hospital Ulm, Ulm, Germany

We describe the unique disease course and cure of SARS-CoV-2 infection in a patient with SCID and graft failure. In absence of a humoral immune response, viral clearance was only achieved after transfusion of convalescent plasma. This observation underscores the necessity of the humoral immune response for SARS-CoV-2 clearance.

Keywords: SARS-CoV-2, severe combined immunodeficiency, humoral immune response, convalescent plasma, remdesivir

OPEN ACCESS

Edited by:
Antonio Condino-Neto, University of São Paulo, Brazil

Reviewed by:
Mike Joyner, Mayo Clinic, United States
Yun Ling, Fudan University, China

***Correspondence:**
Verena Keitel
verena.keitel@med.uni-duesseldorf.de
Björn Jensen
Bjoern-ErikOle.Jensen@med.uni-duesseldorf.de
Johannes Bode
johannes.bode@med.uni-duesseldorf.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Primary Immunodeficiencies, a section of the journal Frontiers in Immunology

Received: *03 February 2021*
Accepted: *22 March 2021*
Published: *03 May 2021*

Citation:
Keitel V, Bode JG, Feldt T, Walker A, Müller L, Kunstein A, Klindt C, Killer A, Senff T, Timm J, Ostermann P, Damagnez M, Lübke N, Adams O, Schaal H, Antoch G, Neubert J, Albrecht P, Meuth S, Elben S, Mohring A, Fischer JC, Bölke E, Hoenig M, Schulz AS, Luedde T and Jensen B (2021) Case Report: Convalescent Plasma Achieves SARS-CoV-2 Viral Clearance in a Patient With Persistently High Viral Replication Over 8 Weeks Due to Severe Combined Immunodeficiency (SCID) and Graft Failure. Front. Immunol. 12:645989. doi: 10.3389/fimmu.2021.645989

INTRODUCTION

We describe a 25-year-old female patient with severe combined immunodeficiency (SCID) due to a RAG1 variant (1, 2) with persistently high SARS-CoV-2-RNA concentrations in respiratory samples over 60 days. Immunocompromised patients have not only an increased risk of acquiring severe Corona virus disease 2019 (COVID-19) (3, 4) but may fail to achieve viral clearance with prolonged shedding of viable virus (5, 6).

Our patient was first treated with remdesivir and subsequently received convalescent plasma (CP), which achieved sustained viral clearance.

CASE DESCRIPTION AND DIAGNOSTIC ASSESSMENT

The patient was diagnosed with $T^-/B^-/NK^+$ SCID and received unconditioned haploidentical hematopoietic stem cell transplantation (HSCT) from her father at 4 months of age (7). Due to incomplete immune reconstitution with poor T cell- and no B cell-engraftment she received a stem cell boost without preconditioning at 4 years of age, repetitive donor lymphocyte infusions (5 times, last infusion 11/2019) and regular immunoglobulin substitution therapy.

She suffered from recurrent bronchopulmonary infections and chronic obstructive pulmonary disease. Due to progressive graft failure she was scheduled for another HSCT.

After a close friend tested positive for SARS-CoV-2, testing was performed while she was asymptomatic and results were positive for SARS-CoV-2 on 30th of April 2020 (day 0). Since patients with SCID are prone to severe systemic viral infections (e.g. cytomegalovirus, adenovirus, parainfluenza virus) (8–10) she was admitted for clinical observation.

Upon admission, her physical examination, vital signs, chest radiography and a CT scan were unremarkable (**Figure 1**). The patient experienced a mild headache for one day but no other COVID-19 associated symptoms. The initial SARS-CoV-2-RNA concentration in the nasopharyngeal swab was 4.89 x 10^8 copies/ml. SARS-CoV-2 could not be PCR-amplified from the patient's EDTA blood, bone marrow, urine and stool samples. Over the course of 30 days, the patient did not develop any overt symptoms despite persistent high-level viral replication.

On initial admission (day 0) the patient had a reduced neutrophil count (nadir of 115/µl on day 4), lymphopenia (389/µl) with reduced T-cells 250/µl ($CD4^+CD45RA^+$T-cells 6.4/µl; $CD4^+CD45RO^+$T-cells 63/µl; $CD8^+CD45RA^+$T-cells 29/µl; $CD8^+CD45RO^+$T-cells 68/µl). NK-cells ($CD3^-CD56^+$) were reduced to 1.3% (4.8/µl). Monocytes were 285/µl and B-cells were absent, which was in line with undetectable IgA and IgM levels (IgG was substituted). Neutrophils were reduced shortly after infection and recovered preceding development of pneumonia (**Table 1**). The patient received prophylactic antibiotic and antifungal treatment.

On d33 of follow-up the patient presented without overt symptoms, but oxygen saturation was 93% and a CT-scan showed signs of COVID-19 pneumonia (**Figure 1**). SARS-CoV-2-RNA was 1.95 x 10^7 and 4.07 × 10^6 copies/ml in nasopharyngeal and bronchial fluid samples, respectively. Thus, COVID-19 pneumonia was diagnosed and the patient received remdesivir (200 mg i.v. on d33, 100 mg/d i.v. d34-42) over 10 days (11). Remdesivir treatment reduced viral concentrations from 1.95 x 10^7 copies/ml to 5.35 x 10^4 copies/ml (**Figure 2**). Whole genome sequencing of SARS-CoV-2 showed no remdesivir resistance development. Clinical

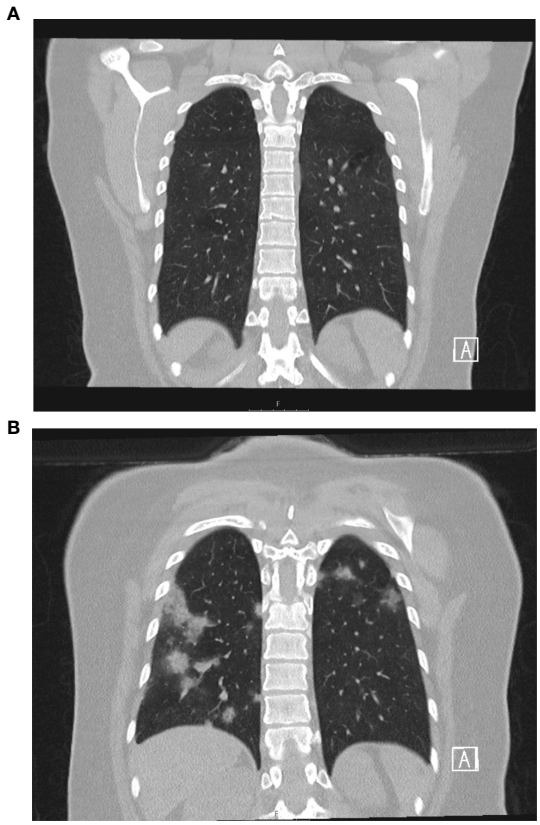


FIGURE 1 | Chest CT scans on day 3 after admission **(A)** without signs of COVID-19 and day 34 **(B)** showing COVID-19 pneumonia.

symptoms of pneumonia improved, however, virus concentrations increased again to levels of 1.48 x 10^8 copies/ml on d54. To achieve viral clearance, the patient received two units of convalescent plasma (CP, 250 ml each) from donor-1 on day 55 (12). This contained spike-specific IgA- and IgG-antibodies (OD-ratios were 1.94 and 3.26, respectively) and had a neutralizing antibody titer (NT-titer) of 1:80. On d57 a third unit of donor-1 CP was administered. Viral concentration dropped from 3.8 x 10^7 copies/ml (d55) to 6.75 x 10^4 copies/ml (d59, 2.75-log reduction). Infusion of three additional units of CP from a different donor (donor-2; d60, d62, d64; IgA/IgG OD-ratio: 8.58/6.44; NT-titer: 1:80) resulted in undetectable viral concentration on NP swabs and increased anti-SARS-CoV-2 antibodies in the patient's serum above the detection limit (IgA/IgG OD-ratio: 2.78/2.96) (**Figure 1**). The patient's symptoms cleared completely and SARS-CoV-2 RNA remained negative even after anti-SARS-CoV-2 antibodies decreased below the detection limit on day 111. The patient received the planned second HSCT on day 138 following conditioning with treosulfan (42g/m^2). Despite this immunosuppressive and -modulatory procedure, the SARS-CoV-2-RNA was not detected by PCR on NP swabs or in the patient's blood (last test from day 158).

TABLE 1 | Laboratory and virological findings; n.d., not detected; NPS, nasopharyngeal swab; CRP, C-reactive protein; PCT, procalcitonin; WBC, white blood cell count (absolute numbers) and differentiation by FACS.

	01/2019	d1	d4	d14	d21	d33	d43	d46	d54/55	d64	d75	d82	d109
Viral load NPS $*10^6$	not appl.	490	116	227	202	19	0.5	0.1	148	n.d.	n.d.	n.d.	n.d.
CRP (mg/dl) <0.5		0.8	0.6	3.4	0.3	4.4	0.3	0.2	0.3	<0.1	<0.1	0.1	0.5
PCT (ng/ml) <0.05		0.07	0.07	0.03	0.03	0.07		0.1	0.08	0.06	0.05	0.04	
IL-6 (pg/ml)		3.9				24.6	9.5		5.3				
Ferritin (µg/ml)		33	77	69	29	90	87	47	32	42	26	24	19
WBC $*10^4$/µl	5.5	0.8	0.6	1.0	2.6	3.5	3.1	4.9	3.5	4.9	3.4	4.0	4.5
Neutrophils (n/µl)		190	125	1238	1134	2479	1135	1928	1322	2628	1623	2329	3045
CD20+ B-cells (n/µl)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
CD3+ T-cells (n/µl)	574	250	373	428	522	375	435	617	816	676	711	709	1151
CD3+/CD4+ (n/µl)	125		71	72	92	56	88	96	114	97	92	118	152
CD3+/CD8+ (n/µl)	224		108	86	187	157	215	327	426	358	343	339	530
CD3-/CD56+/ CD16+ (n/µl)	79		4.8	7.3	7.3	5.2	6.1	16.2	19.3	16.7	6.21	12.0	21.5

Yellow indicates values before SARS-CoV-2 infection. Grey indicates remdesivir application (d33-d43), green indicates application of 6 units of convalescent plasma (CP) from 2 different donors (d55-d64).

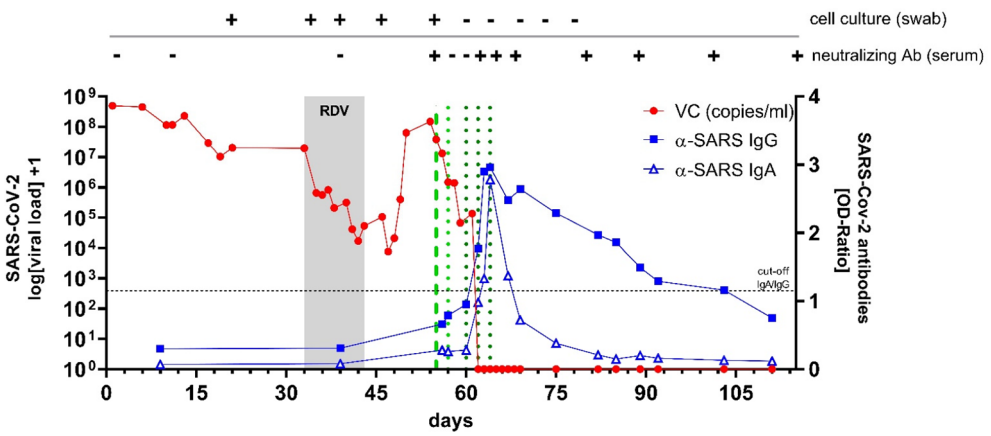


FIGURE 2 | SARS-CoV-2 viral concentration (VC) in nasopharyngeal swabs (in red) and spike specific SARS-CoV-2 antibody titers over time (in blue). Cell culture was used to determine viral replication as well as presence of neutralizing antibodies (Ab). Application of remdesivir (RDV) over 10 days is depicted in grey. Application of convalescent plasma from 2 different donors is depicted as dotted green lines (light green = 3 units from donor-1; dark green = 3 units from donor-2).

DISCUSSION

This unique case illustrates the course of COVID-19 in a situation where the functionality of innate and especially adaptive humoral and cellular immunity is severely limited. Development of COVID-19 pneumonia was significantly delayed despite high viral concentrations and only developed after partial recovery of the cellular immune response. As expected, viral clearance is not achieved with severely impaired T-cell and absent B-cell mediated responses (13, 14). This case and the detection of viral replication in cell culture beyond d50 highlights the need for prolonged quarantine measures and monitoring in patients with immune defects (6).

While remdesivir treatment reduced virus concentrations by 2.6-log, however, after stopping of the drug virus concentrations quickly recovered. CP administration from two different donors achieved sustained viral clearance even after anti-SARS-CoV-2 antibodies dropped below the detection limit, which is in line with reports from patients with primary and secondary immunodeficiency as well as with hematological malignancies (15–17). This therapeutic effect was retained even during a second HSCT on day 138. This case report underscores the importance of the humoral immune response, substituted here by CP transfusions, to successfully clear SARS-CoV-2 infection.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. Consensus Sequences are available on GISAID: EPI_ISL_572330, EPI_ISL_572331, EPI_ISL_572333, EPI_ISL_573152, EPI_ISL_574259, EPI_ISL_572397. See also **Supplementary Material**.

ETHICS STATEMENT

The examinations were carried out in accordance with the Declaration of Helsinki and the patient gave written informed consent for use of CP as well as for publication of the pseudonymized results and patient history.

AUTHOR CONTRIBUTIONS

VK, JB, TF, and BJ initiated this work, supervised the study, and drafted the manuscript. VK, JB, TF, BJ, AKu, CK, AKi, TL, AM, AS, MH, PA, GA, JN, SM, and SE took care of the patient, analyzed the clinical data and phenotype, determined diagnostic procedures and treatment plan, and interpreted treatment responses. AW, LM, TS, JT, PO, MD, NL, OA, and HS developed virological test strategies (ELISAs, testing for neutralizing SARS-CoV-2 antibodies, viral sequencing), and performed and interpreted virological data. All authors critically revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

JT received acknowledges funding through BMBF B-FAST.

ACKNOWLEDGMENTS

Expert technical assistance by Lisa Knopp is thankfully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.645989/full#supplementary-material

REFERENCES

1. Notarangelo LD, Kim MS, Walter JE, Lee YN. Human RAG Mutations: Biochemistry and Clinical Implications. *Nat Rev Immunol* (2016) 16(4):234–46. doi: 10.1038/nri.2016.28
2. Fischer A, Notarangelo LD, Neven B, Cavazzana M, Puck JM. Severe Combined Immunodeficiencies and Related Disorders. *Nat Rev Dis Primers* (2015) 1:15061. doi: 10.1038/nrdp.2015.61
3. Goodall JW, Reed TAN, Ardissino M, Bassett P, Whittington AM, Cohen DL, et al. Risk Factors for Severe Disease in Patients Admitted With COVID-19 to a Hospital in London, England: A Retrospective Cohort Study. *Epidemiol Infect* (2020) 148:e251. doi: 10.1017/S0950268820002472
4. Kim L, Garg S, O'Halloran A, Whitaker M, Pham H, Anderson EJ, et al. Risk Factors for Intensive Care Unit Admission and In-hospital Mortality Among Hospitalized Adults Identified Through the U.S. Coronavirus Disease 2019 (Covid-19)-Associated Hospitalization Surveillance Network (Covid-Net). *Clin Infect Dis* (2020). doi: 10.1093/cid/ciaa1012
5. Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, et al. Case Study: Prolonged Infectious SARS-Cov-2 Shedding From an Asymptomatic Immunocompromised Individual With Cancer. *Cell* (2020) 183:1901–12. doi: 10.1016/j.cell.2020.10.049
6. Aydillo T, Gonzalez-Reiche AS, Aslam S, van de Guchte A, Khan Z, Obla A, et al. Shedding of Viable Sars-CoV-2 After Immunosuppressive Therapy for Cancer. *N Engl J Med* (2020) 383:2586–8. doi: 10.1056/NEJMc2031670
7. Pai SY, Logan BR, Griffith LM, Buckley RH, Parrott RE, Dvorak CC, et al. Transplantation Outcomes for Severe Combined Immunodeficiency, 2000-2009. *N Engl J Med* (2014) 371(5):434–46. doi: 10.1056/NEJMoa1401177
8. Ruffner MA, Sullivan KE, Henrickson SE. Recurrent and Sustained Viral Infections in Primary Immunodeficiencies. *Front Immunol* (2017) 8:665. doi: 10.3389/fimmu.2017.00665
9. Dropulic LK, Cohen JI. Severe Viral Infections and Primary Immunodeficiencies. *Clin Infect Dis* (2011) 53(9):897–909. doi: 10.1093/cid/cir610
10. Al-Herz W, Essa S. Spectrum of Viral Infections Among Primary Immunodeficient Children: Report From a National Registry. *Front Immunol* (2019) 10:1231. doi: 10.3389/fimmu.2019.01231
11. Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate Use of Remdesivir for Patients With Severe Covid-19. *N Engl J Med* (2020) 382(24):2327–36. doi: 10.1056/NEJMc2015312
12. Bloch EM, Shoham S, Casadevall A, Sachais BS, Shaz B, Winters JL, et al. Deployment of Convalescent Plasma for the Prevention and Treatment of COVID-19. *J Clin Invest* (2020) 130(6):2757–65. doi: 10.1172/JCI138745
13. Vardhana SA, Wolchok JD. The Many Faces of the anti-COVID Immune Response. *J Exp Med* (2020) 217(6):e20200678. doi: 10.1084/jem.20200678
14. Liu STH, Lin HM, Baine I, Wajnberg A, Gumprecht JP, Rahman F, et al. Convalescent Plasma Treatment of Severe COVID-19: A Propensity Score-Matched Control Study. *Nat Med* (2020) 26:1708–13. doi: 10.1038/s41591-020-1088-9
15. Senefeld JW, Klassen SA, Ford SK, Wiggins CC, Bostrom BC, Thompson MA, et al. Therapeutic Use of Convalescent Plasma in COVID-19 Patients With Immunodeficiency. *medRxiv* (2020) 2020.11.08.20224790. doi: 10.1101/2020.11.08.20224790
16. Thompson MA, Henderson JP, Shah PK, Rubinstein SM, Joyner MJ, Choueiri TK, et al. Convalescent Plasma and Improved Survival in Patients With Hematologic Malignancies and COVID-19. *medRxiv* (2021) 2021.02.05.21250953. doi: 10.1101/2021.02.05.21250953
17. Hueso T, Pouderoux C, Péré H, Beaumont A-L, Raillon L-A, Ader F, et al. Convalescent Plasma Therapy for B-cell–depleted Patients With Protracted COVID-19. *Blood* (2020) 136(20):2290–5. doi: 10.1182/blood.2020008423

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.3.5. Age-dependent immune response to the Biontech/Pfizer BNT162b2 COVID-19 vaccination (Chapter 13)

Despite therapeutic interventions, prophylactic vaccinations are a fundamental tool to overcome the pandemic. This has led to the rapid development and testing of vaccine candidates and in late 2020, first vaccines were approved and vaccinations were rolled out in a prioritization procedure. Elderly adults were prioritized in vaccination campaigns since old age is considered a risk factor. However, data on the immune responses elicited in this group was underrepresented in approval studies. In this cohort study with two age groups, young vaccinees below 60 and elderly vaccinees over 80, antibody responses to the first and second dose of the BNT162b2 COVID-19 vaccination were compared. While the majority of participants produced antibody titers, titers were significantly lower in elderly participants. Furthermore, one third of the elderly had no detectable neutralizing antibodies after the second vaccination in contrast to the younger group. Therefore, close monitoring of this group is required and suggests potential earlier revaccination or an increased vaccine dose to ensure stronger and long-lasting immunity and protection against SARS-CoV-2 infection.

This article is published in Clin Infect Dis. 2021 Apr 27:ciab381. doi: 10.1093/cid/ciab381. by

Müller L*, Andrée M*, Moskorz W, Drexler I, Walotka L, Grothmann R, Ptok J, Hillebrandt J, Ritchie A, Rabl D, Ostermann PN, Robitzsch R, Hauka S, Walker A, Menne C, Grutza R, Timm J, Adams O#, Schaal H#.

* contributed equally, # contributed equally

Contributions:

Conceptualization: HS, OA, MA, LM, Formal analysis: OA, WM, Investigation: LM, MA, WM, ID, LW, RG, JP, JH, AR, DR, OA, HS, Writing – original draft preparation: HS, OA, MA, LM, Writing – review and editing: LM, MA, WM, ID, LW, RG, JP, JH, AR, DR, PNO, RR, SH, AW, CM, RG, JT, OA, HS, Supervision: HS, OA, MA, LM.

Individual contribution: 25% L.M. wrote the ethics proposal and designed the study plan. L.M. generated and curated data and performed experiments for all figures. L.M. wrote the first manuscript draft.

The article is inserted as the submitted manuscript for copyright reasons.

Age-dependent immune response to the Biontech/Pfizer BNT162b2 COVID-19 vaccination

Lisa Müller#+[1], Marcel Andrée#[1], Wiebke Moskorz[1], Ingo Drexler[1], Lara Walotka[1], Ramona Grothmann[1], Johannes Ptok[1], Jonas Hillebrandt[1,2], Anastasia Ritchie[1], Denise Rabl[1], Philipp Niklas Ostermann[1], Rebekka Robitzsch, Sandra Hauka[1], Andreas Walker[1], Christopher Menne[1], Ralf Grutza[1], Jörg Timm[1], Ortwin Adams*[1] and Heiner Schaal*[1]

# contributed equally

* contributed equally

+ Corresponding author

[1]Institute of Virology, Medical Faculty, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[2] Department of Nephrology, Medical Faculty, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

40-word summary:

This study compared antibody responses in two age groups (<60/ >80 years) after first and second BNT162b2 COVID-19 vaccination. While the majority in both groups developed SARS-CoV-2 spike-specific antibodies, IgG and neutralization titers were significantly lower in the elderly group.

Abstract

Background:

The SARS-CoV-2 pandemic has led to the development of various vaccines. Real-life data on immune responses elicited in the most vulnerable group of vaccinees over 80 years old is still underrepresented despite the prioritization of the elderly in vaccination campaigns.

Methods:

We conducted a cohort study with two age groups, young vaccinees below the age of 60 and elderly vaccinees over the age of 80, to compare their antibody responses to the first and second dose of the BNT162b2 COVID-19 vaccination.

Results:

While the majority of participants in both groups produced specific IgG antibody titers against SARS-CoV-2 spike protein, titers were significantly lower in elderly participants. Although the increment of antibody levels after the second immunization was higher in elderly participants, the absolute mean titer of this group remained lower than the <60 group. After the second vaccination, 31.3 % of the elderly had no detectable neutralizing antibodies in contrast to the younger group, in which only 2.2% had no detectable neutralizing antibodies.

Conclusion:

Our data showed differences between the antibody responses raised after the first and second BNT162b2 vaccination, in particular lower frequencies of neutralizing antibodies in the elderly group. This suggests that this population needs to be closely monitored and may require earlier revaccination or/and an increased vaccine dose to ensure stronger long lasting immunity and protection against infection.

Introduction

In December 2019, authorities in China's Wuhan province reported a lung disease of unknown cause. Back in January 2020, the sequence of a novel coronavirus was published and identified as the causative agent of this disease [1]. In March of the same year, the World Health Organization (WHO) declared the spread of this virus a public health emergency of international concern. With limited drug treatment options available, research on prophylactic immunization, especially for high-risk groups, became a priority [2].

Hence, rapid vaccine development became a global effort, which led to the emergency approval of 13 COVID-vaccines as of now [3-6], with many others in different advanced stages of development. The types of vaccines that are currently in use or under investigation in various clinical stages include non-replicating viral vector vaccines, formulations based on replicating viral vectors or virus like particles as well as inactivated vaccines and vaccines based on protein subunits (reviewed in [7]). A novel development in vaccine formulation that also received emergency approval, are mRNA based vaccines. These are also main vaccine types currently used in the western world, in particular the two mRNA technology vaccines Comirnaty (BNT162b2) by Biontech/Pfizer and mRNA-1273 by Moderna.

The Biontech/Pfizer and Moderna vaccines were not only the first approved COVID-vaccines, they are also the first approved drugs to employ the novel mRNA technology. While mRNA has long been discussed as a potent alternative to conventional vaccine formulations [8], the hurdle of low RNA stability and inefficient delivery had to be overcome to make full use of this technology. In recent years, the use of modified nucleosides, in particular modified uridine, the removal of double-stranded RNA by HPLC, codon optimization and the delivery via lipid-nanoparticles were developed. These advances helped to decrease innate sensing of the synthetic mRNA and thus, paved the way to efficient use of RNA vaccines. Currently approved vaccines also employ these methods [9, 10]. However, this new class of vaccines also carries certain disadvantages just as other drug formulations. This includes the stability of mRNA during transport and storage as well as still limited cellular uptake compared to other systems such viral vectors [11].

Early studies on mRNA vaccines from Biontech/Pfizer [5] and Moderna [12] showed high efficacy and safety of the formulations. With mass vaccinations being carried out using these vaccines, more promising reports on the effectiveness of the vaccines after completing the full vaccination schedule (prime and boost dose) were published [13-15]. The current vaccination strategy for the Biontech/Pfizer Comirnaty (BNT162b2) is a two-step "prime and boost" procedure in which the first vaccination is followed by a second vaccination with the same dose at least 21 days later [5]. Studies suggest that effectiveness of the vaccine is lower in individuals who received only the first dose compared to individuals who received the full vaccination regimen [16, 17].

In Germany and many other countries worldwide, COVID-vaccinations at the beginning of 2021 were offered in a prioritization procedure. First, individuals who are at particularly high risk for severe courses of COVID-19 disease or who are professionally in close contact with such vulnerable people were vaccinated. These two prioritized groups included senior residents of nursing homes aged ≥ 80 years, and their caregivers typically aged ≤ 65 years. This is of particular importance since SARS-CoV-2 and its associated disease COVID-19, can result in a remarkable variable severity of clinical symptoms, from asymptomatic infection to severe COVID-19 with lung manifestation and acute respiratory distress syndrome in up to 14% of patients [18]. Here, the elderly population is primarily at risk for severe disease, as adults over 65 years of age accounted for approximately 80% of hospitalizations [19, 20]. Additionally, prolonged disease, delayed viral clearance, and a higher fatality rate is also reported to be age-related [21].

Although vaccination is key to prevent infections, vaccine responses are often found to be lower in elderly adults. In numerous studies, the markedly reduced vaccine success in older adults has been attributed to adaptive immunosenescence. Reduced vaccination success in elderly adults is especially known for hepatitis B, pneumococcal, and influenza vaccinations [22, 23]. Although hallmarks of immunosenescence depend on multifaceted factors and vary greatly between individuals, they are considered to be related to i) the decreased ability to respond to new antigens associated with a reduced peripheral plasmablast response; (ii) decreased capacity of memory T cells and (iii) a low level of persistent chronic inflammation. This leads to declining immune efficiency and fidelity, resulting in

increased susceptibility to infectious diseases and decreased response to vaccinations [23-26].

With the experience from previous vaccinations, the question arose whether there are also differences in the immune response between younger and older people after immunization against SARS-CoV-2. We therefore started a daily practice study in a nursing home immediately after the start of the official vaccination campaign in Germany at the end of December 2020. In order to accommodate two distinctly different populations in this study, we compared the induction of immune responses between young and older vaccinees (< 60 years and > 80 years, respectively) who received their first and second vaccination on the same day. For this purpose, IgG titers against SARS-CoV-2 spike S1 and neutralization titers were determined after both the first and the second vaccination since antibody titers and in particular, neutralization titers, together with T-cell responses are the main arms of the adaptive immune response and hence, levels of protection are suggested to be potentially estimated based on neutralizing antibody titers [27]. Finally, the self-reported side effects corresponding to the sum of symptoms after vaccination were examined for a potential correlation between the severity of the symptoms and antibody response.

Methods

Study population

The ethics committee of the Medical Faculty at the Heinrich-Heine University Düsseldorf, Germany (study no. 2021-1287), approved the study. Participants were volunteers from the SBK nursing home in Cologne, Germany. Characteristics of the study population are summarized in Table 1. Informed consent was obtained from all volunteers (N = 179) before sampling.

Medical questionnaires

In order to assess the subjective perception of post-vaccination reactions, medical questionnaires including the following categories were scored according to the sum of reported reactions: i) elevated temperature and fever, ii) chills, iii) pain at the injection

site, iv) head/limb pain, v) fatigue/tiredness, vi) nausea/dizziness, vii) other complaints (unscored).

Sample processing:

All blood samples were collected on January 15th, 2021 (first collection, 17—19 days after first immunization) and February 5th, 2021 (second collection, 17 days after second immunization) and stored at 4 °C. Samples were subjected to the respective assays within 72h after each collection. For cross validation, a subset of samples from the first blood collection were run during analysis of samples from the second blood collection. Positive and negative samples, which were previously tested, were included in all assay.

Commercially available Anti-SARS-CoV-2 tests systems

Samples were tested for Anti-SARS-CoV-2 antibodies using two commercially available test systems: Euroimmun Anti-SARS-CoV-2-QuantiVac-ELISA measuring IgG levels against SARS-CoV-2 spike S1 subunit and Abbott Architect SARS-CoV-2 IgG recognizing SARS-CoV-2 nucleocapsid (N) antibodies.

Euroimmun ELISA was performed on the Euroimmun Analyzer I-2P according to the manufacturer's instructions. The assay encompasses a 6-point calibration curve and issues the IgG antibody concentration as standardized units (BAU/ml = Binding Antibody Units). Results < 25.6 BAU/ml were considered as negative, ≥ 25.6 BAU/ml ≤ 35.2 BAU/ml as indeterminate, and > 35.2 BAU/ml as positive. The lower detection limit for undiluted samples was < 3.2 BAU/ml, the upper detection limit was > 384 BAU/ml. For samples over the detection limit, 1:10 or 1:100 dilutions were performed in IgG sample buffer according to the manufacturer's instruction. The SARS-CoV-2 IgG chemiluminescent microparticle immunoassay (CMIA) from Abbott was performed on an ARCHITECT i2000 SR after the second blood collection. The relation of chemiluminescent RLU and the calibrator is given as the calculated index (S/C). An index (S/C) <1.4 as was considered negative, ≥1.4 was considered positive.

In-house SARS-CoV-2 neutralization test

A serial dilution endpoint neutralization test [28] with the infectious SARS-CoV-2 isolate (EPI_ISL_425126) was performed in a BSL-3 facility to determine the SARS-CoV-2 neutralization capacity of the serum samples after the first and second vaccination. Serial dilutions of heat-inactivated (56°C, 30 minutes) serum samples were pre-incubated in cell-free plates with 100 TCID50 units of SARS-CoV-2 for 1 hour at 37° C. After pre-incubation, 100µl of cell suspension containing 7×10^4/ml Vero cells (ATTC-CCL-81) were added. Plates were incubated at 37°C, 5% CO2 for 4 days before microscopic inspection for virus-induced cytopathic effect (CPE). The neutralization titer was determined as the highest serum dilution without CPE. Tests were performed as independent duplicates for each sample. Positive, negative, virus only, and cell growth controls were run during each assay.

Statistical analysis

The data were analyzed using SPSS Statistics 25 (IBM©) and GraphPad Prism 9.0.00 (GraphPad Software, San Diego, CA, USA). Categorical data were studied using Fisher's exact test or Pearson's chi-square test, depending on the sample size. Quantitative data were analyzed by the non-parametric Mann-Whitney U test for two groups of paired and unpaired samples. Simple linear regression was performed using GraphPad Prism version 9.0.0 (the coefficient of determination R^2 and p-values are given in the figures).

Results

Participant characteristics

In total, blood samples from 176 volunteers, young and elderly vaccinees (<60 / >80 years of age) were analysed for vaccine-induced SARS-CoV-2 spike specific IgG titers and SARS-CoV-2 neutralizing antibodies after a prime and boost vaccination campaign using BNT162b2 (Comirnaty Biontech/Pfizer) to screen for age-related differences in their immune response. Therefore, samples were collected at two time

points, 17—19 days after the first vaccination and 17 days after the second vaccination. To be able to distinguish the immune response of the vaccinees from those who had already undergone a previous SARS-CoV-2 infection we also determined infection-induced SARS-CoV-2 nucleocapsid specific antibodies using the SARS-CoV-2 IgG chemiluminescent microparticle immunoassay (CMIA). Three vaccinees were tested positive and therefore were excluded from the dataset. While group sizes were comparable (93 participants <60 years of age versus 83 participants >80 years of age), there was an overrepresentation of female participants compared with males (124 female to 52 male) (Table 1).

Table 1: *Characteristics of the study population.*

Characteristics	**< 60 years of age (younger vaccinees)**	**> 80 years of age (elderly vaccinees)**	**Total**
Total N (%)	91 (53%)	85 (47%)	176 (100%)
Gender			
Male N (%)	29 (32%)	23 (27%)	52 (30%)
Female N (%)	62 (68%)	62 (73%)	124 (70%)
Mean years (min - max)	42.2 (19.5 - 59.5)	87.9 (80.1 - 100.5)	

Vaccination-induced SARS-CoV-2 spike specific IgG levels differ between young and elderly vaccinees after the first and second vaccination

The first sample collection was carried out 17—19 days after the volunteers received their first vaccination in late December 2020. At this time point, quantitative SARS-CoV-2 spike S1 specific IgG levels between the two groups differed significantly ($p < 0.0001$). For the younger group of vaccinees, IgG titers ranged between 0—3840.0 BAU/ml with a mean of 313.3 BAU/ml after the first vaccination. Only 4.4 % of the participants had titers below the cut-off, and 2.3% were indeterminate (Figure 1A). The mean titer for the group > 80 years of age was 41.2 BAU/ml with titers ranging from 0—484.7 BAU/ml. In this group, 65.9% showed titers below the cut-off (>35.6), and 9.4% were indeterminate.

The second sample collection was carried out 17 days after the volunteers received their second vaccination, at a time point when full protection is suggested (>7 days according to [5]). Nevertheless, there was still a significant difference in IgG levels between the two groups ($p < 0.0001$). The mean titer of the younger group increased more than 10-fold (3702.0 BAU/ml) and ranged from 81.6—32000.0 with no participant testing below cut-off (Figure 1B). While the mean titer for elderly vaccinees increased to 1332.0 BAU/ml (0—16891.0 BAU/ml), 10.6% of the participants in this group still had titers below the cut-off.

The comparison of SARS-CoV-2 spike specific IgG titers showed an extremely significant ($p<0.0001$) difference between the two age groups, after both the first and second vaccination, suggesting an attenuated antibody response in the group of elderly vaccinees > 80 years of age. While the gap in mean values narrowed after the second vaccination, which in particular underlines once again the necessity of a second vaccination, several elderly participants remained below the detection limit of the anti-SARS-CoV-2 assay. A general age-dependent negative correlation in SARS-CoV-2 spike specific IgG after both vaccinations is noticed throughout the entire cohort (Figure 1D/1E).

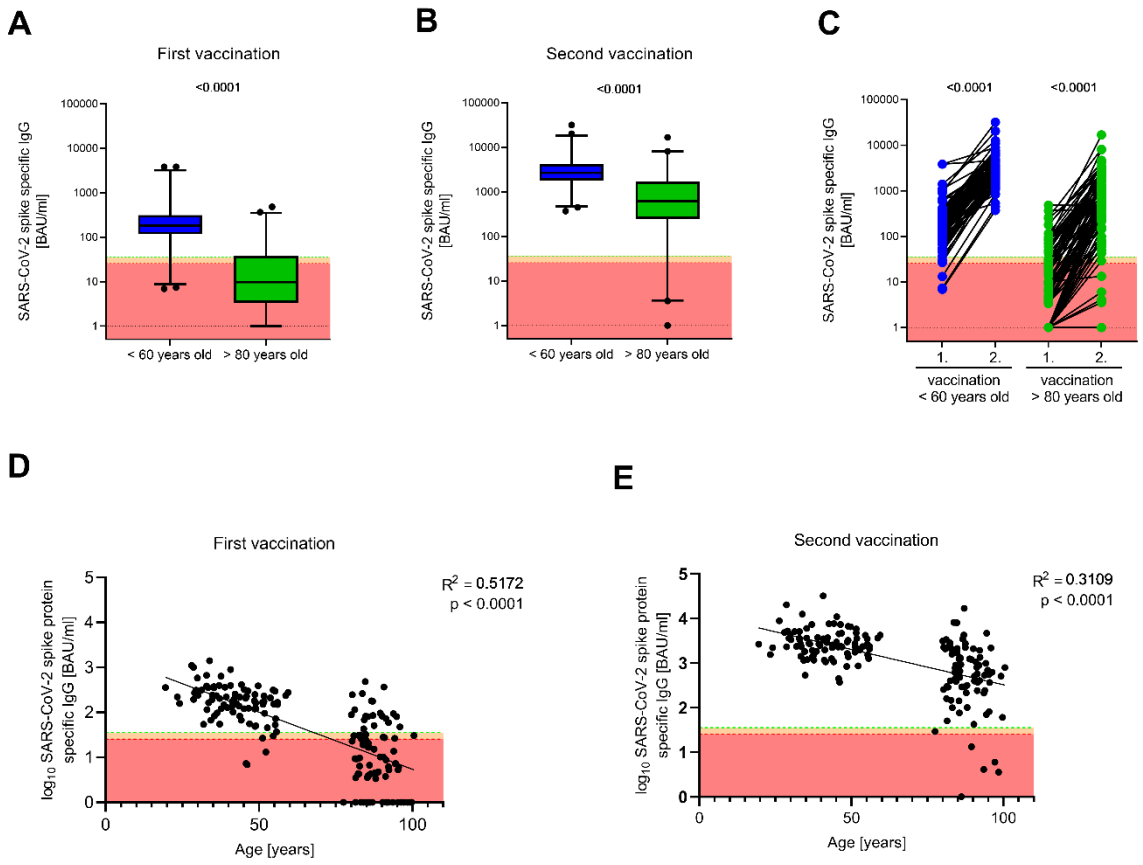


Figure 1

SARS-CoV-2 spike protein specific antibody titers were determined using Euroimmun Anti-SARS-CoV-2-QuantiVac-ELISA. Antibody titers below the detection limit were set to 1.0. ***A*** *and* ***B*** *Antibody titers 17—19 day after first (A) and second (B) vaccination are shown. Boxes span the interquartile range; the line within each box denotes the median and whiskers indicate the 2.5 and 97.5 percentile values.* ***C*** *The pairwise comparison of IgG antibody titers within the two analysed age groups are shown.* ***D*** *and* ***E*** *Linear correlations between participant's age and SARS-CoV-2 specific antibody titer after first vaccination (D) and second vaccination (E). Results < 25.6 BAU/ml as negative (red area), ≥ 25.6 BAU/ml ≤ 35.6 BAU/ml as indeterminate (orange), and > 35.6 BAU/ml were considered positive. For comparison of two groups either two-tailed parametric unpaired t-tests or paired t-test were performed. Correlation was analysed by simple linear regression. P-values < 0.05 were considered statistically significant. P-Values are depicted in the figures.*

Elderly vaccinees showed reduced SARS-CoV-2 neutralizing capacity compared to younger vaccinees

We next determined the neutralization capacity in our cohort after the first and second dose of vaccination. At 17—19 days after the first vaccination, the majority of participants, regardless of their age, failed to display neutralizing antibody titers. In the group of younger vaccines, 16.1 % displayed neutralizing antibodies with titers ranging

between 1:10 to 1:2560. In the group of elderly vaccinees, only 1.2 % had developed neutralizing antibodies after the first vaccination (Figure 2A).

After the second dose, a neutralization titer was attained by 97.8% of the younger vaccinees. In the elderly group, 68.7% showed titers ranging from 1:10 to 1:320. Remarkably, in 31.3% of the elderly vaccinees neutralizing antibodies were not detectable after the second vaccination, and thus, were potentially without seroprotection (Figure 2B).

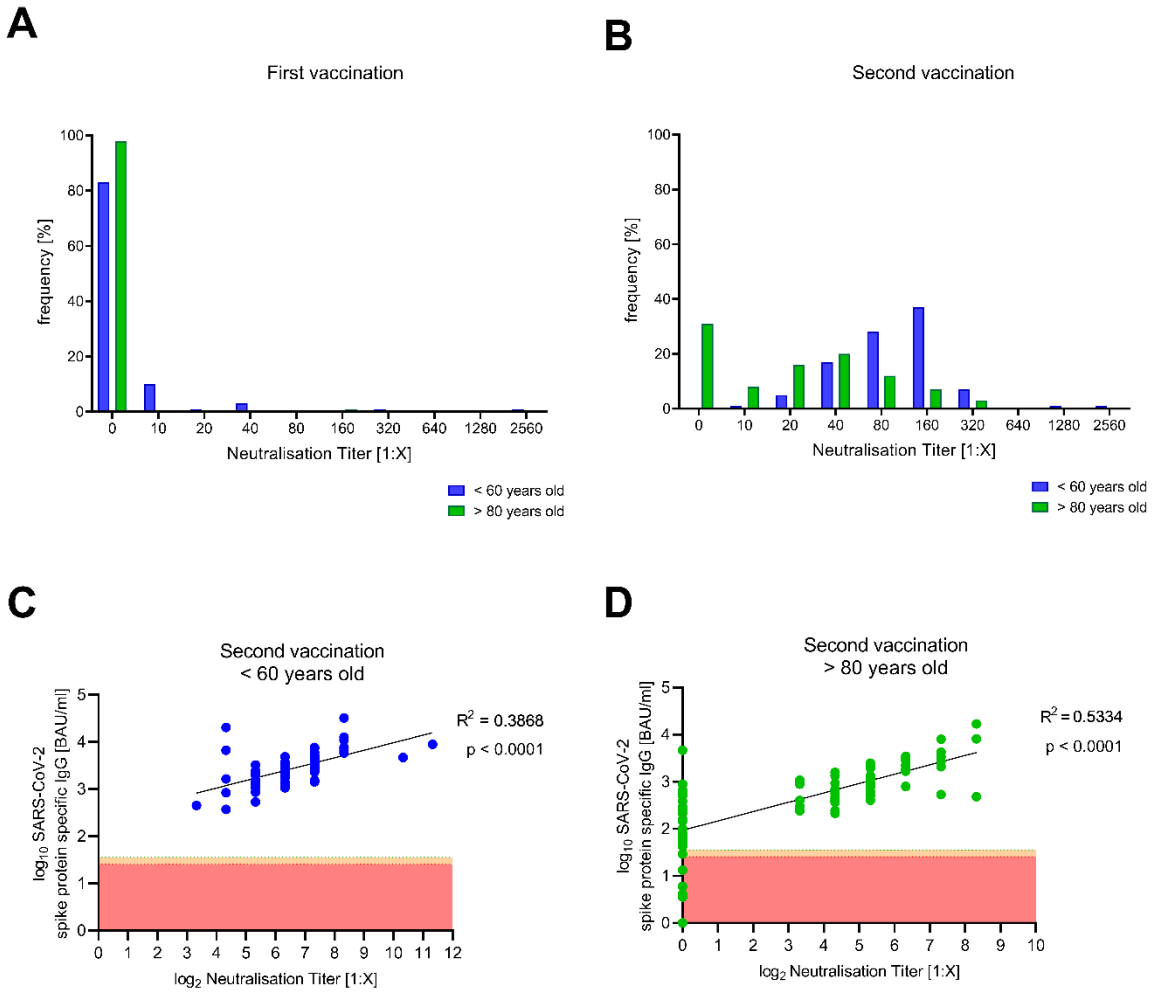


Figure 2

Neutralization antibody titers were determined as described in the methods section. The frequencies of individuals with a certain neutralizing antibody titer after the first vaccination (A) and the second vaccination (B) are shown.

The severity of post-vaccination reactions does not correlate with antibody response

To assess differences in post-vaccination reactions between the age groups and to evaluate a potential correlation with antibody titers, medical questionnaires were completed at the two collection time points.

After the first vaccination, half of the younger cohort (51.6%) reported no reactions to the vaccination, the remaining vaccinees recorded reactions with a score ranging between 1 and 4 of combined reactions. In turn, 93.9 % of elderly vaccinees reported no post-vaccination reactions; the remaining 6.1% reported either one or two of the scored reactions (Figure 3A).

After the second dose, only 25.8% of the younger vaccinees had no post-vaccination reactions. While 38.7% of this group reported only one of the scored post-vaccination reactions, 35.5% reported a combination of reactions scoring between 2 and 6. Among the elderly, 83.1% reported no reaction, and the remaining 16.9% of this group reported combined reactions up to a score of 3 (Figure 3B). However, there was no general correlation between vaccination-induced SARS-CoV-2 spike specific IgG or neutralizing antibody production and the presence or absence of individual post-vaccination reaction reports.

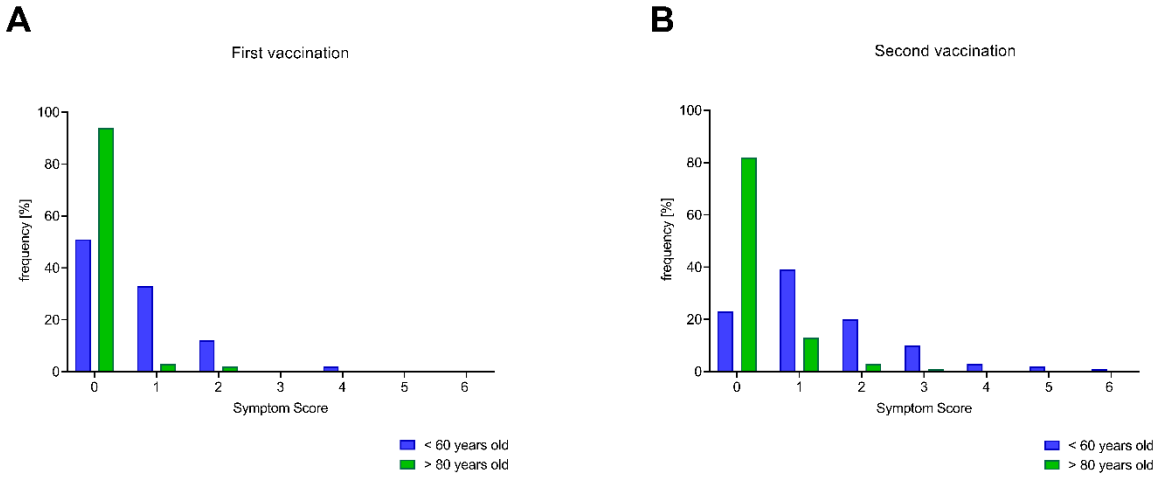


Figure 3

Post-vaccination reaction scores after first (A) and second (B) vaccination were determined as the sum of cumulative reactions using to the predefined categories (see method section).

Discussion

The SARS-CoV-2 pandemic has led to the development of various vaccines and vaccine strategies, which have been made available to the public by either emergency use designation or conditional marketing authorization. Inevitably, data on populations that are difficult to enroll, including immunocompromised or cohorts <16 years or >80 years who might show reduced vaccine reactiveness, are limited. The main goal of this real-life study was to investigate the immunogenicity of the current vaccination strategy in the most vulnerable group of vaccinees (>80 years old) compared to those younger than 60 years who received the Biontech/Pfizer BNT162b2 COVID-19 vaccination. We compared the induction of immune responses in these two age groups after the first and second vaccination by measuring vaccine-induced SARS-CoV-2 spike specific IgG and SARS-CoV-2 neutralizing antibodies. While the majority of both young and elderly vaccinees raised IgG responses after their second vaccination, the induction of ELISA-IgG and in particular neutralizing antibody levels were significantly lower in the elderly vaccinees.

The main differences between the two groups are likely a consequence of immunosenescence, which describes the reduced adaptive immune responses in the elderly [29]. It is well described that elderly individuals not only have higher rates of morbidity due to infection but also respond less to vaccination [30-32], mainly due to a decline in cellular as well as humoral immunity. For vaccinations including the influenza vaccine, this limitation is bypassed by increasing the vaccine doses [33].

The notion that humoral vaccination responses are impaired with increasing age is well depicted in our cohort, as the mean titer of SARS-CoV-2 spike specific IgG remained 2.8-fold lower after the second vaccination for the elderly group of vaccinees compared to the younger cohort (Figure 1B). Additionally, a general intra- and inter-group trend in negative correlation between age and IgG titer is visible after both vaccinations (Figure 1C/1D). More importantly, a similar age-dependent trend can be seen for SARS-CoV-2 specific neutralizing antibody titers: While neutralization antibody titers were attained by 97.8% of the younger vaccinees, 31.3% of the elderly remained without neutralization antibody titers after the second vaccination (Figure 2B).

The lack of neutralizing antibody responses in about one-third of the elderly group raises the questions whether the effectiveness of vaccine-induced immune protection may be transferred to this population without explicit testing. In a large cohort study using the Biontech/Pfizer vaccine BNT162b2 and the related BNT162b1 vaccine candidate, the humoral responses in two adult age groups (18-55 and 65 to 85 years) were compared after the second vaccination. They reported that immunogenicity as measured by antibody responses including neutralization titers was lower in the elderly cohort and also discussed immunosenescence as potential cause [34]. The role of neutralizing antibodies is in particular crucial since neutralizing antibody levels correlate with protection against many viruses including SARS-CoV-2 in humans [35, 36] and recent data suggest that high neutralizing titers are particularly important for protection against novel circulating SARS-CoV-2 variants conferring immune escape [37-39].

Currently, different vaccination schedules for the same vaccines have been adopted in several countries. These include a delay of the second vaccination, as implemented by the UK or Israel, to allow for the initial primary vaccination to a larger proportion of the population, a strategy that is controversially discussed [40, 41]. The observation that single-dose vaccinees broadly lacked neutralizing antibody responses in our cohort raises the question, whether these individuals might still acquire infections and may transmit the disease while remaining asymptomatic. This assumption is supported by recent results of a large Israeli study which reports a 46% effectiveness in preventing a documented infection 14 to 20 days after the first dose, the BNT162b2 vaccine [13]. Smaller studies report similar results with incomplete Biontech/Pfizer vaccinations [16, 17]. However, other large scale population studies on the experience with COVID vaccinations report that even after the first mRNA SARS-CoV-2 vaccination, a significant decrease in hospitalizations and severe disease is seen in the overall vaccinated population, but also the >80 year old group [38]. These reports emphasize that not only direct protection of vulnerable groups but also indirect protection by generating a community immunity can contribute to the decrease of severe COVID cases, hospitalizations and death, which ultimately eases the economic burden of the pandemic. However, it is not yet clear how long this protective effect of mRNA vaccination lasts, hence monitoring effectiveness after the vaccine deployment is inevitable [42].

Our data presented here suggests that it might be necessary to have strategies at hand to overcome possible age-related limitations for COVID-19 vaccination. Moderna has recently demonstrated an increased immune response determined by higher binding and neutralizing antibody titers by increasing the dose of the second vaccination from 25 µl to 100 µl [14]. Strategies to enhance immunogenicity such as the use of adjuvants, application of increased amounts or multiple doses of the same vaccine, or the combination of different vaccines for a heterologous prime/boost should be rapidly tested and implemented in COVID-19 vaccination protocols where necessary.

This study provides insight into age-dependent limitations of immune responses elicited after the first and second dose of the BNT162b2 vaccine. By comparing similar-sized cohorts of vaccinees aged < 60 years and > 80 years, we found that more than 30% of elderly vaccinees did not attain neutralizing antibody responses after their second vaccination. Despite the fact that the elderly age group is most vulnerable, this population was underrepresented in previous studies. Nevertheless, promising studies show that even after the first vaccination with the mRNA vaccines, at least severe courses of COVID-19 are attenuated.

Notes

Financial support. This work was supported by Stiftung für Altersforschung, Düsseldorf for [to H.S., L.W.], Jürgen Manchot foundation [to H.S., I.D. L.M., L.W., A.R., J.T., R.G., P.N.O.]. This work was supported by research funding from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant GK1949/1 and project number 452147069 [to I.D.]. This work was supported by the Forschungskommission of the Medical Faculty, Heinrich-Heine-Universität Düsseldorf [to H.S., J.P., J.H.].

Acknowledgments: The authors thank Dr. Anna Seelentag (Sozial Betriebe Köln, SBK) for excellent organization and support, Gabriele Patzke (SBK) for the opportunity to conduct this study, Dr. Irene Spiertz-Schmidt, Christine Mavinga and the caregivers at SBK for their support at the blood collection days. We would like to thank all volunteers at the SBK nursing home. We thank all members of the Virology department.

Potential conflicts of interest: The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

Author contributions: Conceptualization: HS, OA, MA, LM, Formal analysis: OA, WM, Investigation: LM, MA, WM, ID, LW, RG, JP, JH, AR, DR, OA, HS, Writing – original draft preparation: HS, OA, MA, LM, Writing – review and editing: LM, MA, WM, ID, LW, RG, JP, JH, AR, DR, PNO, RR, SH, AW, CM, RG, JT, OA, HS, Supervision: HS, OA, MA, LM

References

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med **2020**; 382(8): 727-33.
2. Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. Immunity **2020**; 52(4): 583-9.
3. Jackson LA, Anderson EJ, Rouphael NG, et al. An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. N Engl J Med **2020**; 383(20): 1920-31.
4. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. Virus research **2020**; 288: 198114.
5. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med **2020**; 383(27): 2603-15.
6. Voysey M, Clemens SAC, Madhi SA, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet **2021**; 397(10269): 99-111.
7. Rawat K, Kumari P, Saha L. COVID-19 vaccine: A recent update in pipeline vaccines, their design and development strategies. Eur J Pharmacol **2021**; 892: 173751.
8. Schlake T, Thess A, Fotin-Mleczek M, Kallen KJ. Developing mRNA-vaccine technologies. RNA Biol **2012**; 9(11): 1319-30.
9. Fuller DH, Berglund P. Amplifying RNA Vaccine Development. N Engl J Med **2020**; 382(25): 2469-71.
10. Pardi N, Hogan MJ, Weissman D. Recent advances in mRNA vaccine technology. Curr Opin Immunol **2020**; 65: 14-20.
11. Wang J, Peng Y, Xu H, Cui Z, Williams RO, 3rd. The COVID-19 Vaccine Race: Challenges and Opportunities in Vaccine Formulation. AAPS PharmSciTech **2020**; 21(6): 225.
12. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med **2021**; 384(5): 403-16.
13. Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. N Engl J Med **2021**.
14. Anderson EJ, Rouphael NG, Widge AT, et al. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. N Engl J Med **2020**; 383(25): 2427-38.
15. Amit S, Regev-Yochay G, Afek A, Kreiss Y, Leshem E. Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients. Lancet **2021**.
16. Thompson MG, Burgess JL, Naleway AL, et al. Interim Estimates of Vaccine Effectiveness of BNT162b2 and mRNA-1273 COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Health Care Personnel, First Responders, and Other Essential and Frontline Workers - Eight U.S. Locations, December 2020-March 2021. MMWR Morb Mortal Wkly Rep **2021**; 70(13): 495-500.
17. Britton A, Jacobs Slifka KM, Edens C, et al. Effectiveness of the Pfizer-BioNTech COVID-19 Vaccine Among Residents of Two Skilled Nursing Facilities Experiencing COVID-19 Outbreaks - Connecticut, December 2020-February 2021. MMWR Morb Mortal Wkly Rep **2021**; 70(11): 396-401.

18. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA **2020**; 323(13): 1239-42.
19. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? Aging (Albany NY) **2020**; 12(10): 9959-81.
20. Liu Y, Mao B, Liang S, et al. Association between age and clinical characteristics and outcomes of COVID-19. Eur Respir J **2020**; 55(5).
21. Leung C. Clinical features of deaths in the novel coronavirus epidemic in China. Rev Med Virol **2020**; 30(3): e2103.
22. Schenkelberg T. Vaccine-induced protection in aging adults and pandemic response. Biochemical and biophysical research communications **2021**; 538: 218-20.
23. Crooke SN, Ovsyannikova IG, Poland GA, Kennedy RB. Immunosenescence and human vaccine immune responses. Immun Ageing **2019**; 16: 25.
24. Pawelec G. Hallmarks of human "immunosenescence": adaptation or dysregulation? Immun Ageing **2012**; 9(1): 15.
25. Aiello A, Farzaneh F, Candore G, et al. Immunosenescence and Its Hallmarks: How to Oppose Aging Strategically? A Review of Potential Options for Therapeutic Intervention. Front Immunol **2019**; 10: 2247.
26. Sasaki S, Sullivan M, Narvaez CF, et al. Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies. The Journal of clinical investigation **2011**; 121(8): 3109-19.
27. Lau EHY, Tsang OTY, Hui DSC, et al. Neutralizing antibody titres in SARS-CoV-2 infections. Nat Commun **2021**; 12(1): 63.
28. Muller L, Ostermann PN, Walker A, et al. Sensitivity of anti-SARS-CoV-2 serological assays in a high-prevalence setting. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology **2021**.
29. Blomberg BB, Frasca D. Quantity, not quality, of antibody response decreased in the elderly. The Journal of clinical investigation **2011**; 121(8): 2981-3.
30. Gustafson CE, Kim C, Weyand CM, Goronzy JJ. Influence of immune aging on vaccine responses. J Allergy Clin Immunol **2020**; 145(5): 1309-21.
31. Pera A, Campos C, Lopez N, et al. Immunosenescence: Implications for response to infection and vaccination in older people. Maturitas **2015**; 82(1): 50-5.
32. Frasca D, Blomberg BB. Aging induces B cell defects and decreased antibody responses to influenza infection and vaccination. Immun Ageing **2020**; 17(1): 37.
33. DiazGranados CA, Dunning AJ, Kimmel M, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. N Engl J Med **2014**; 371(7): 635-45.
34. Walsh EE, Frenck RW, Jr., Falsey AR, et al. Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. N Engl J Med **2020**; 383(25): 2439-50.
35. Plotkin SA. Correlates of protection induced by vaccination. Clin Vaccine Immunol **2010**; 17(7): 1055-65.
36. Iyer AS, Jones FK, Nodoushani A, et al. Dynamics and significance of the antibody response to SARS-CoV-2 infection. medRxiv **2020**.
37. Garcia-Beltran WF, Lam EC, Denis KS, et al. Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. medRxiv **2021**.
38. Jangra S, Ye C, Rathnasinghe R, et al. The E484K mutation in the SARS-CoV-2 spike protein reduces but does not abolish neutralizing activity of human convalescent and post-vaccination sera. medRxiv **2021**.
39. Kuzmina A, Khalaila Y, Voloshin O, et al. SARS CoV-2 escape variants exhibit differential infectivity and neutralization sensitivity to convalescent or post-vaccination sera. Preprint **2021**.
40. Robertson JFR, Sewell HF, Stewart M. Delayed second dose of the BNT162b2 vaccine: innovation or misguided conjecture? Lancet **2021**.
41. Bieniasz P. The case against delaying SARS-CoV-2 mRNA vaccine boosting doses. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2021**.
42. Forman R, Shah S, Jeurissen P, Jit M, Mossialos E. COVID-19 vaccine challenges: What have we learned so far and what remains to be done? Health Policy **2021**.

4. Summary and Conclusion

RNA is not only a mediator between the information stored in the DNA and its translation into proteins it additionally carries regulatory information as a central molecule of various biological processes. Therefore, the precise regulation of its processing is inevitable to maintain an error-free function. One central processing step, especially for eukaryotic RNA, is pre-mRNA splicing which describes the excision of introns and the ligation of exons. To carry out this orchestrated process, different layers of regulation are needed. This includes the precise recognition of the splice sites that constitute exon/intron borders by the spliceosome.

As reviewed in **chapter 1**, the exclusive recognition of splice sites without recognition of additional supportive splicing regulatory elements is most often insufficient. SREs and their respective protein binding partners highly contribute to distinguishing true from false splice sites, i.e. sequences that resemble splice sites but that are not used as such. Therefore, changes in SRE capacity and thus in the binding of the two main regulatory protein families, SR proteins and hnRNPs, can often lead to missplicing. Giving way for the use of *de novo* splice sites, activation of cryptic splice sites or the decreased use of physiological splice sites can have detrimental effects on the individual affected. To aid the evaluation of nucleotide exchanges in splice sites or their surroundings, various computational tools were developed, however, their underlying approaches highly differ. While robust testing of the predictions made with these tools is inevitable, they can ultimately assist not only in the evaluation of mutations but also in the prediction of functional targets for therapeutic approaches.

In **chapter 2**, the prediction and robust modeling of the splicing outcome by one particular tool, the HEXplorer tool, was tested as it was used to artificially design splicing regulatory elements with a prescribed function that were combined in different 5’ splice site settings. The *in silico* designed sequences did mimic the effect of their natural counterparts and were even shown to pose as binding sites for specific splicing regulatory proteins. As shown here and as discussed in chapter 1, splice site recognition highly relies on SREs apart from the sequence complementarity of the splice sites. Therefore, in this experimental approach, a novel RNA-seq-based 5'ss utilization landscape was developed. This combines the intrinsic strength of splice sites

with the calculated support of surrounding SREs, which provides a unified functional splice site strength. With this unique combination of information, it might be possible to more accurately predict the consequences of human pathogenic mutations in diagnostics.

In a similar approach as presented above, **chapter 3** provides insight into the regulation of 3' splice sites. While 5' splice sites follow a strict position-dependent regulation from surrounding SREs, their effect on 3'ss and in particular, their role in a setting with two competing 3'ss is yet to be fully elucidated. Using several experimental approaches, it was possible to show that SREs act in a similar position-dependent way on splice acceptor sites and thus, on general exon recognition and the crosstalk between 3'ss and 5'ss. This regulatory feature is an important concept in the evaluation of human pathogenic mutations as it provides another regulatory mechanism.

While SREs binding SR or hnRNP proteins are described as main regulators of splice site recognition for both, 5'ss and 3'ss (chapter 1-3), several classes of intronic non-coding RNAs such as microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs) have also been found to influence splice site recognition and usage. In **chapter 4**, *in silico* and *in vitro* experiments were used to identify a new class of intronic splicing regulatory sequences. Nuclear intronic mitochondrial-derived tRNAs homologs (nimtRNAs) located within human introns were found to have similar *cis*-acting functions as bona fide SREs. They were used in different splicing reporters to elucidate their function and showed to increase splice site recognition and use following the position dependency concept of SR and hnRNP proteins. This new class of splicing regulatory sequences adds another layer of regulation to the intricate process of splice site selection.

The tight splicing regulation by direct splicing signals and position dependent SREs, with intronic nimtRNAs adding to the repertoire of splicing regulators, is demonstrated here (chapter 1-4). However, an additional determinator of exon inclusion is given by exon and intron size since this contributes to the process of exon or intron definition which is highly dependent on exon length (168,169). Yet to be fully elucidated, however, is the process of the removal of several kilobase long introns. Long intron removal often requires intermediate steps with unstable RNA. This makes the elucidation and full tracking of the process by standard RNA seq, that mainly captures

mature RNA, more difficult (170). Two main features that enhance the removal of long introns are recursive splicing as well as intra-splicing. The former rather eliminates smaller parts of introns than the full long intron at once, often by the use of “zero nucleotide exons”. These exons consist of only the AG-GT dinucleotides required for basal spliceosome recognition, so that the inclusion of these exon leave no trace in the final transcript. Recent studies in *Drosophila melanogaster* suggest that recursive splicing adds to splicing fidelity since it had been shown that long introns that contain recursive splice sites are recognized more accurately than those without (171). Intra-splicing, however, is basically a form of nested splicing, where so-called intra-introns are removed consecutively to bring the long intron splice sites in closer proximity for a final splicing reaction (172), this regulative process, however, is not yet fully elucidated and has only been shown in few genes (170).

Furthermore, despite their direct role in splicing regulation, SREs and in particular their protein binding partners play an extended role in RNA processing as they can highly contribute to mRNA export or the inhibition thereof. The export of mRNA from the nucleus to the cytoplasm is another highly regulated process in gene expression that requires several protein-protein or protein-RNA interaction. After capping, splicing and the deposition of exon-junction complexes (EJC), the transcription-export (TREX) complex is fitted to the mRNA to be exported. Then, one of the main mediators of mRNA export, the nuclear export factor 1 (NXF1) binds and facilitates export by interaction with nuclear pores (173,174). Recently, NXF1 has been shown to interact with SR-proteins as adaptors. In particular, NXF1 iCLIP data revealed that the protein has a binding preference to sequence that overlap with SRSF3 binding sites. Furthermore, an interaction with SRSF7 was shown (175). On the other hand, hnRNP proteins have been shown to be associated with the retention of transcripts in the nucleus. In fact, hnNRPC, which is almost exclusively found in nuclear fractions, also contains a nuclear localization signal (NLS) and thus, its binding contributes to nuclear retention of transcripts which also plays a role in viral gene expression (176,177).

Respectively, unspliced viral RNAs including the HIV-1 9 kb message are exported via the CRM1 pathway. Here, a direct interaction of CRM1 with SR of hnRNP proteins as it was shown for NXF1, however, is not elucidated so far (178).

Despite the high level of splicing regulation, missplicing occurs due to various underlying mechanisms and highly affects the splicing outcome. In **chapter 5**, misregulated splicing of the receptor for hyaluronan mediated motility (RHAMM) is described. A shift of splice isoforms from full-length transcripts to a truncated version is described to be associated with human disease, multiple myeloma. To elucidate potential underlying causes, single nucleotide variants were bioinformatically evaluated by the HEXplorer tool and their potential effects on the splicing outcome is described in this report. Furthermore, it is speculated that the identification of nucleotide exchanges that contribute to a different splicing outcome might be used for patient screening and could eventually be exploited as experimental new drug targets.
Apart from eukaryotic gene expression, RNA plays a central role as a storage form of genetic information for retroviruses and RNA viruses. HIV-1 not only carries an RNA genome, but it also fully exploits the host cellular splicing machinery to replicate. In **chapter 6**, the regulation of viral splice sites by a dense regulatory network is described. In particular, the focus is laid on exon 2/2b that is interspersed with splicing regulatory elements and contains two splice donors, SD2 and SD2b. Depending on the use of one of these two donors, more or less Vif is produced, an antagonist of the host restriction factor APOBEC, and its amount determines the efficiency of virus replication. By bioinformatic prediction and subsequent experimental analysis, it was possible to identify a new splicing regulatory element in this region termed ESS2b. Furthermore, protein binding partners were identified and the functional relevance of the SREs was tested by mutational analysis and masking of the sequences by locked nucleic acids (chapter 8).
The importance of the integrity and the functional relevance of SREs in the exon 2/2b region is further elucidated in **chapter 7**, where drug resistance-associated mutations located here were analyzed for their potential at the same time also to alter the splicing outcome. In particular, two missense mutations with different calculated potential to affect splice donor choice were screened, where R263K had a severe effect on the splicing pattern. This mutation is of particular interest since it had previously been shown that it is associated with drug resistance in therapy naïve patients. To further elucidate its mode of action a silent version of the mutation with a comparable calculated potential was bioinformatically designed. R263R showed a similar effect on

the splicing outcome as its missense counterpart, underlining the importance of taking silent nucleotide exchanges into account when performing resistance profiling.

Emerging drug resistance, especially in therapy-naïve patients, underscores the importance of continuous research into new potential targets for antiviral therapies. Since HIV-1 highly relies on a balanced splicing outcome for successful replication (chapter 6 and 7) it might be possible to target this aspect to disturb viral gene expression. The use of locked nucleic acids (LNAs), antisense oligonucleotides with an additional methylene bridge, that masks viral splicing regulatory elements, is presented in **chapter 8**. Here, LNAs are targeted against the known, conserved splicing regulatory elements G_{I3}-2 and ESE_{tat}. Specifically, this region influences the formation of *tat* mRNA, whose translation product is essential for the conversion of the viral transcription complex into a processive transcription complex for efficient HIV-1 replication. Delivery was performed via transfection, however, to make the approach feasible as a potential antiviral strategy, the LNAs were also administered gymnotically, hence without a transfection reagent. Both approaches showed efficient masking of the sequences that resulted in the abrogation of viral replication in HIV-1 infected cells *in vitro*.

Both, viral and eukaryotic gene expression and subsequently, homeostasis, highly relies on functional splicing regulation and thus, changes in splicing regulatory elements either by intrinsic factors such as mutations or external factors such as drug interventions can have detrimental effects as shown in chapters 5 – 8. For mutational analysis, a highly desirable tool is the prediction of the splicing outcome solely based on the nucleotide sequences. Here, recent developments include the use of deep learning approaches (79). This becomes a particular attractive tool due to the rise of diagnostic sequencing, especially in cancer genetics, which in turn often results in the detection of so-called variants of uncertain significance (VUS) that need to be evaluated for their clinical relevance (179). The hereditary breast and ovarian cancer-associated genes BRCA1 and BRCA2 are two well-known examples for diagnostic sequencing and thus, the detection of VUS. Both genes are prone to spliceopathies and therefore, pathologic splicing alterations that can have direct consequence for the clinical management of patients. Recently, guided by bioinformatic prediction, two novel pathogenic variants causing missplicing in BRCA1 have been identified (180).

In December 2019, reports about a novel disease emerged in the city of Wuhan, China. Shortly after, a novel coronavirus termed SARS-CoV-2 was identified as the causative agent. The situation rapidly developed into a pandemic that is still ongoing in 2021. The situation led to a shift in research around the world to quickly understand the viruses' features and to identify possible targets for therapy.

While both, HIV-1 and SARS-CoV-2 use RNA as their form of genome storage, the replication of the viruses differs. While HIV-1 hijacks the hosts' cellular splicing machinery for extensive alternative splicing in the nucleus, SARS-CoV-2 fully exploits the hosts' transcription machinery in the cytoplasm. However, both share the common challenge that they have to make full use of their limited-sized genome to successfully replicate. SARS-CoV-2 therefore makes use of discontinuous transcription instead of splicing, allowing both viruses, albeit in different ways, to expose the 5' CAP distal ORFs. Also, a shared strategy is the use of a programmed –1 ribosomal shift via a slippery sequence that allows the expression of downstream ORF and even in a different reading frame from a single translational start codon. Despite their differences in gene expression, the life cycles and replication strategies of both viruses overlap at various points, e.g. the glycoprotein mediated entry, so that it was possible to quickly transfer experimental approaches between these two viruses.

Since early in the SARS-CoV-2 pandemic, reports about infected individuals with neurological symptoms emerged, the molecular basis of this feature was to be further studied. In **chapter 9**, 3D human brain organoids were employed to elucidate the preferred host cell population within the organoids and to study the effect of infection in these cells, adding to the understanding of clinical reports as it was possible to show that the virus targets primarily neurons and that infections lead to an altered distribution of Tau, a protein that modulates the stability of axonal microtubules.

Despite the understanding of the molecular features of the infection, epidemiological monitoring and the assessment of immunological features is an important pillar of pandemic management. A cohort study performed early in the pandemic (**chapter 10**) when there were yet no safety measures in place, showed the high seroprevalence in a patient cohort that had contact with a so-called super spreader. Furthermore, this study was used to compare serological testing by either commercial or in-house developed assays to assess their respective reliability, since these tools are highly

used to assess the immune response of large patient cohorts. Here, especially assays that measure antibody levels against SARS-CoV-2 spike protein generated more reliable results than those measuring nucleocapsid antibodies. However, a classic virological method still considered to be the gold standard in antibody response assessment, the full virus endpoint dilution neutralization test, outperformed the commercial assays.

Reliable assessment of antibody titers becomes especially crucial when screening for potential plasma donors for convalescent plasma therapy for severely ill patients. The safety and efficacy of convalescent plasma therapy compared to the administration of camostat mesylate, a protease inhibitor, is assessed in a multicentric, randomized, controlled study as described in **chapter 11**. To select suitable donors, commercial antibody tests, as well as the in-house developed neutralization test, are employed.

The importance and efficacy of convalescent plasma in individual therapy are demonstrated in **chapter 12**, where a SARS-CoV-2 infected patient suffering severe combined immunodeficiency (SCID) was unable to clear the infection for a prolonged period of time. After developing more severe symptoms, the patient was first treated with Remdesivir, however, full viral clearance was only achieved after administration of convalescent plasma from a donor with high neutralizing antibody titers.

Apart from research into treatment options for infected patients, an early focus was laid on the development of prophylactic vaccinations. In December 2020, one year after the first report of the new disease, the first vaccines based on a novel mRNA technology received emergency approval, and vaccination campaigns were rolled out worldwide. Here, in particular, high-risk individuals were prioritized which included elderly individuals in care homes as well as their caregivers. While it is known that elderly individuals generally have a lower immune response to vaccination due to the phenomenon of immunosenescence, it was of particular importance to study their immune reaction to the COVID vaccination in comparison to a younger cohort since a novel class of vaccines, based on mRNA, was used. In **chapter 13**, the results of this large cohort study are presented and show that the majority of vaccinees developed an immune response to the vaccination, at least after the second dose of the prime and boost vaccination schedule, however, IgG and in particular, the levels of neutralizing antibodies were significantly lower in the elderly compared to the younger

cohort. With the pandemic still ongoing, this underlines the necessity of close monitoring and might hint towards an adjustment of the vaccination strategy in this group, whether by an increased dosage or an earlier re-vaccination.

As shown here, anti-SARS-CoV-2 neutralizing antibodies that are generated after infection and vaccination play a pivotal role in disease management. Similarly, the generation of broadly neutralizing antibodies (bNAbs) against the HIV-1 Env trimer is one goal of HIV-1 vaccine and treatment development (181). This is recently revisited due to improved single cell antibody cloning methods (182) after the first generation of bNAbs that was reported in the 90s' had low clinical potency (183). Features of bNAbs are described in several studies to include the neutralization of viral particles, the clearing of infected cells and inhibition of transmission of HIV-1. In the last ten years, several bNAbs have entered trials. One formulation, VRC01, was tested for efficacy but was unfortunately found to not prevent infection in this cohort study, but as proof-of principle it showed that this bNab is potent against sensitive strains that are included in the neutralizing range (184).

While SARS-CoV-2 does not rely on the hosts' cellular splicing machinery as HIV-1 does, splicing was very recently suggested to play a role in the COVID vaccine related side effects, in particular, the rare occurrence of vaccine-induced immune thrombotic thrombocytopenia (VITT) that can cause cerebral venous sinus thrombosis (CVST) which was reported after vector based immunizations (185). While all processing steps of SARS-CoV-2 RNA occur outside of the nucleus, the route of adenoviral DNA that is used for COVID vector vaccines encompasses the nucleus where it is transcribed, capped and spliced by the hosts' cellular machinery. Due to the size of the Spike ORF, alternative splicing events are likely to occur. The results of the study show that due to splicing, a membrane anchor is lost which was shown to result in secreted, soluble Spike protein variants that might play a role in VITT and CVST (186). However, the report is not yet peer reviewed, but provides an interesting hypothesis which underlines the evolutionary difference between genomic replication of RNA and DNA viruses.

This work shows the versatile role of RNA as a central molecule in biological processes. It emphasizes the requirement of a favorable sequence neighborhood for subsequent splice site recognition and underlines the susceptibility of splice site usage

towards changes in splicing regulatory elements for both, eukaryotic and viral transcripts. Recent experimental and bioinformatic approaches are increasingly suggesting that the codon degeneracy of the amino acid sequence is probably a consequence of the messenger ribonucleoprotein "mRNP" code, the plethora of protein occupying the RNA during processing steps, underlying the genetic code. This code is necessary to create the prerequisite for RNA-binding proteins, which are required for mRNA processing, all while maintaining the protein-coding information. For viruses replicating in the nucleus and employing the hosts' splicing machinery, such as HIV-1, splicing regulatory proteins near splice sites influence their utilization. For viruses replicating in the cytoplasm, such as SARS-CoV-2, replication of the full-length genome and sub genomic RNA require the host' protein synthesis machinery that allows the expression of viral proteins that then interact with crucial host cellular factors. Although both, HIV-1 and SARS-CoV-2 carry an RNA genome that is used for storage and replication of their genetic information, their life cycles differ, but several connections to study the pathogenesis and immunological features of SARS-CoV-2 can be draw from HIV-1 research.

5. Curriculum Vitae

Personal information:

Lisa Müller, M.Sc.

Born on September 9th 1991 in Idar-Oberstein, Germany

Education:

Since 05/2017	PhD Thesis, Heinrich-Heine-University Düsseldorf, Germany Supervision: Prof. Dr. Heiner Schaal
01/2016 – 05/2017	M.Sc. Biomedical Sciences, Edinburgh Napier University, UK 'Modelling splicing outcome dependent on splicing regulatory element activity and splice donor strength' Supervision: Dr. Peter Barlow
10/2012 – 12/2015	B.Sc. Biology, Heinrich-Heine-University Düsseldorf, Germany 'Activation of cryptic splice sites: HEXplorer-based analysis to predict the splicing outcome of pathogenic splice mutations' Supervision: Prof. Dr. Heiner Schaal

Publications:

Smola, A, Samazadeh, S, Müller, L, Adams, O, Homey, B, Albrecht, P Meller S. Omalizumab prevents anaphylactoid reactions to mRNA COVID-19 vaccine. Journal of the European Academy of Dermatology and Venereology. July 2021. doi: 10.1111/jdv.17549. Epub ahead of print.

Kolb T*, Fischer S*, Müller L*, Lübke N, Hillebrandt J, Andrée M, Schmitz M, Schmidt C, Kücükköylü S, Koster L, Kittel M, Weiland L, Dreyling K, Hetzel G, Adams O, Schaal H, Ivens K, Rump L C, Timm J#, Stegbauer J#. Impaired immune response to SARS-

CoV-2 vaccination in dialysis patients and in kidney transplant recipients. Kidney360 July 2021, doi: 10.34067/KID.0003512021. Epub ahead of print.

Rothenburg J, Rink-Baron S, Mueller L, Ostermann PN, Fischer J, Stegbauer J, Moldenhauer A. COVID-19 antibody donation using immunoadsorption: Report of two cases. Transfus Apher Sci. 2021 Jun 15:103193. doi: 10.1016/j.transci.2021.103193.

Ptok J, Müller L, Ostermann PN, Ritchie A, Dilthey AT, Theiss S, Schaal H. Modifying splice site usage with ModCon: Maintaining the genetic code while changing the underlying mRNP code. Comput Struct Biotechnol J. 2021 May 21;19:3069-3076. doi: 10.1016/j.csbj.2021.05.033.

Keitel V*, Bode JG*, Feldt T, Walker A, Müller L, Kunstein A, Klindt C, Killer A, Senff T, Timm J, Ostermann P, Damagnez M, Lübke N, Adams O, Schaal H, Antoch G, Neubert J, Albrecht P, Meuth S, Elben S, Mohring A, Fischer JC, Bölke E, Hoenig M, Schulz AS, Luedde T, Jensen B*. Case Report: Convalescent Plasma Achieves SARS-CoV-2 Viral Clearance in a Patient With Persistently High Viral Replication Over 8 Weeks Due to Severe Combined Immunodeficiency (SCID) and Graft Failure. Front Immunol. 2021 May 3;12:645989. doi: 10.3389/fimmu.2021.645989.

Keitel V, Jensen B, Feldt T, Fischer JC, Bode JG, Matuschek C, Bölke E, Budach W, Plettenberg C, Scheckenbach K, Kindgen-Milles D, Timm J, Müller L, Kolbe H, Stöhr A, Calles C, Hippe A, Verde P, Spinner CD, Schneider J, Wolf T, Kern WV, Nattermann J, Zoufaly A, Ohmann C, Luedde T; RES-Q-HR Trial Team. Reconvalescent plasma/camostat mesylate in early SARS-CoV-2 Q-PCR positive high-risk individuals (RES-Q-HR): a structured summary of a study protocol for a randomized controlled trial. Trials. 2021 May 17;22(1):343. doi: 10.1186/s13063-021-05181-0.

Werner J*, Kronberg RM*, Stachura P, Ostermann PN, Müller L, Schaal H, Bhatia S, Kather JN, Borkhardt A, Pandyra AA, Lang KS, Lang PA. Deep Transfer Learning Approach for Automatic Recognition of Drug Toxicity and Inhibition of SARS-CoV-2. Viruses. 2021 Apr 2;13(4):610. doi: 10.3390/v13040610.

Müller L*, Andrée M*, Moskorz W, Drexler I, Walotka L, Grothmann R, Ptok J, Hillebrandt J, Ritchie A, Rabl D, Ostermann PN, Robitzsch R, Hauka S, Walker A, Menne C, Grutza R, Timm J, Adams O#, Schaal H#. Age-dependent immune response to the Biontech/Pfizer BNT162b2 COVID-19 vaccination. Clin Infect Dis. 2021 Apr 27:ciab381. doi: 10.1093/cid/ciab381. Epub ahead of print.

Müller L*, Ostermann PN*, Walker A, Wienemann T, Mertens A, Adams O, Andree M, Hauka S, Lübke N, Keitel V, Drexler I, Di Cristanziano V, Hermsen DF, Kaiser R, Boege F, Klein F, Schaal H, Timm J, Senff T. Sensitivity of anti-SARS-CoV-2 serological

assays in a high-prevalence setting. Eur J Clin Microbiol Infect Dis. 2021 May;40(5):1063-1071. doi: 10.1007/s10096-021-04169-7. Epub 2021 Feb 3.

Hoser SM, Hoffmann A, Meindl A, Gamper M, Fallmann J, Bernhart SH, Müller L, Ploner M, Misslinger M, Kremser L, Lindner H, Geley S, Schaal H, Stadler PF, Huettenhofer A. Intronic tRNAs of mitochondrial origin regulate constitutive and alternative splicing. Genome Biol. 2020 Dec 8;21(1):299. doi: 10.1186/s13059-020-02199-6.

Ramani A*, Müller L*, Ostermann PN*, Gabriel E, Abida-Islam P, Müller-Schiffmann A, Mariappan A, Goureau O, Gruell H, Walker A, Andrée M, Hauka S, Houwaart T, Dilthey A, Wohlgemuth K, Omran H, Klein F, Wieczorek D, Adams O, Timm J, Korth C, Schaal H, Gopalakrishnan J. SARS-CoV-2 targets neurons of 3D human brain organoids. EMBO J. 2020 Oct 15;39(20):e106230. doi: 10.15252/embj.2020106230. Epub 2020 Sep 23.

Erkelenz S, Poschmann G, Ptok J, Müller L, Schaal H. Profiling of cis- and trans-acting factors supporting noncanonical splice site activation. RNA Biol. 2021 Jan;18(1):118-130. doi: 10.1080/15476286.2020.1798111. Epub 2020 Aug 5.

Ptok J*, Müller L*, Theiss S, Schaal H. Context matters: Regulation of splice donor usage. Biochim Biophys Acta Gene Regul Mech. 2019 Nov-Dec;1862(11-12):194391. doi: 10.1016/j.bbagrm.2019.06.002. Epub 2019 Jun 13.

Hillebrand F, Ostermann PN, Müller L, Degrandi D, Erkelenz S, Widera M, Pfeffer K, Schaal H. Gymnotic Delivery of LNA Mixmers Targeting Viral SREs Induces HIV-1 mRNA Degradation. Int J Mol Sci. 2019 Mar 3;20(5):1088. doi: 10.3390/ijms20051088.

Erkelenz S, Theiss S, Kaisers W, Ptok J, Walotka L, Müller L, Hillebrand F, Brillen AL, Sladek M, Schaal H. Ranking noncanonical 5' splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. Genome Res. 2018 Dec;28(12):1826-1840. doi: 10.1101/gr.235861.118. Epub 2018 Oct 24.

Brillen AL, Walotka L, Hillebrand F, Müller L, Widera M, Theiss S, Schaal H. Analysis of Competing HIV-1 Splice Donor Sites Uncovers a Tight Cluster of Splicing Regulatory Elements within Exon 2/2b. J Virol. 2017 Jun 26;91(14):e00389-17. doi: 10.1128/JVI.00389-17.

Brillen AL, Schöneweis K, Walotka L, Hartmann L, Müller L, Ptok J, Kaisers W, Poschmann G, Stühler K, Buratti E, Theiss S, Schaal H. Succession of splicing regulatory elements determines cryptic 5´ss functionality. Nucleic Acids Res. 2017 Apr 20;45(7):4202-4216. doi: 10.1093/nar/gkw1317.

6. Acknowledgements

Die vorliegende Arbeit wurde im Zeitraum Mai 2017 bis Juli 2021 am Institut für Virologie des Uniklinikums Düsseldorf durchgeführt.

Zu danken gilt es an dieser Stelle so vielen Menschen, dass es unmöglich wäre alle namentlich zu nennen, trotzdem versuche ich mich hier an einer Auswahl:

Zu aller erst möchte ich mich herzlichst bei Herrn Prof. Dr. Heiner Schaal bedanken. Nicht nur für die Vergabe des interessanten Themas, die steten Hilfestellungen, die wissenschaftlichen Freiheiten und die fachliche Unterstützung, sondern auch für die exzellente Betreuung über diese Arbeit heraus. Ihm ist es gelungen eine Atmosphäre in der Arbeitsgruppe zu schaffen, die neben der fachlichen Weiterentwicklung auf verschiedenste Weisen auch den Freiraum bietet, persönlich zu wachsen. Dafür gilt ihm mein größter Dank.

Bei Herrn Prof. Dr. Michael Feldbrügge möchte ich mich für die freundliche Übernahme des Mentorings und sein Interesse an dieser Arbeit bedanken.

Bei Prof. Dr. Jörg Timm möchte ich mich für die Möglichkeit bedanken, diese Arbeit am Institut für Virologie anzufertigen.

Bei der Jürgen Manchot Stiftung und der MOI Graduiertenschule möchte ich mich für die finanzielle Unterstützung bedanken.

Allen Kollaborationspartnern im In- und Ausland möchte ich für die gute Zusammenarbeit danken.

Allen meinen (ehemaligen) Kolleginnen und Kollegen am Institut für Virologie und besonders unserer Arbeitsgruppe möchte ich für die tolle Zeit im und außerhalb des Labors danken. Besonderer Dank hier gilt Lara Walotka, Anna-Lena Brillen, Nora Diehl, Philipp Ostermann, Johannes Ptok, Björn Wefers, Yvonne Dickschen, Jennifer Antemann, Niklas Kiel, Ju-Young Kim sowie Ramona Grothmann und Wiebke

Moskorz. Auch Stephan Theiss möchte ich herzlich für seine Unterstützung danken. Ebenfalls gilt ein besonderer Dank Marcel Andrée und Sandra Hauka.

Auch möchte ich meinen Freunden und meiner Familie danken, die mich auf diesem Weg begleitet haben und deshalb auch öfter auf meine Anwesenheit verzichten mussten.

Frank möchte ich von ganzem Herzen für seine unendliche Unterstützung, seine Geduld und sein Verständnis danken.

Diese Arbeit ist meinen Eltern gewidmet und ihnen gebührt mein Dank für ihre uneingeschränkte Unterstützung und ihre unermüdliche Geduld mit mir. Gerne möchte ich diese Arbeit auch meinen Großeltern, vor allem meinen Großmüttern widmen.

7. Declaration

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Lisa Müller

Düsseldorf, den

8. References

1. Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799-816.
2. Nikolov, D.B. and Burley, S.K. (1997) RNA polymerase II transcription initiation: a structural view. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 15-22.
3. Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna*, **14**, 802-813.
4. Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 3171-3175.
5. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell*, **19**, 981-992.
6. Gilbert, W. (1978) Why genes in pieces? *Nature*, **271**, 501.
7. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470-476.
8. Tress, M.L., Abascal, F. and Valencia, A. (2017) Most Alternative Isoforms Are Not Functionally Important. *Trends Biochem Sci*, **42**, 408-410.
9. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457-463.
10. Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. (2013) Function of alternative splicing. *Gene*, **514**, 1-30.
11. Singh, R. (2002) RNA-protein interactions that regulate pre-mRNA splicing. *Gene Expr*, **10**, 79-92.
12. Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell*, **12**, 5-14.
13. Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, **3**.
14. Matera, A.G. and Wang, Z. (2014) A day in the life of the spliceosome. *Nature reviews. Molecular cell biology*, **15**, 108-121.
15. Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701-718.
16. Patel, S.B. and Bellini, M. (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic acids research*, **36**, 6482-6493.
17. Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T. and Luhrmann, R. (2004) The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *Rna*, **10**, 929-941.
18. Matera, A.G., Terns, R.M. and Terns, M.P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature reviews. Molecular cell biology*, **8**, 209-220.
19. Dergai, O., Cousin, P., Gouge, J., Satia, K., Praz, V., Kuhlman, T., Lhote, P., Vannini, A. and Hernandez, N. (2018) Mechanism of selective recruitment of

RNA polymerases II and III to snRNA gene promoters. *Genes & development*, **32**, 711-722.
20. Chen, J. and Wagner, E.J. (2010) snRNA 3' end formation: the dawn of the Integrator complex. *Biochemical Society transactions*, **38**, 1082-1087.
21. Albrecht, T.R. and Wagner, E.J. (2012) snRNA 3' end formation requires heterodimeric association of integrator subunits. *Molecular and cellular biology*, **32**, 1112-1123.
22. Becker, D., Hirsch, A.G., Bender, L., Lingner, T., Salinas, G. and Krebber, H. (2019) Nuclear Pre-snRNA Export Is an Essential Quality Assurance Mechanism for Functional Spliceosomes. *Cell reports*, **27**, 3199-3214 e3193.
23. Masuyama, K., Taniguchi, I., Kataoka, N. and Ohno, M. (2004) RNA length defines RNA export pathway. *Genes & development*, **18**, 2074-2085.
24. Singh, R.N. and Singh, N.N. (2018) Mechanism of Splicing Regulation of Spinal Muscular Atrophy Genes. *Adv Neurobiol*, **20**, 31-61.
25. Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J. and Bordonne, R. (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep*, **4**, 616-622.
26. Lott, K. and Cingolani, G. (2011) The importin beta binding domain as a master regulator of nucleocytoplasmic transport. *Biochimica et biophysica acta*, **1813**, 1578-1592.
27. Spector, D.L. and Lamond, A.I. (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol*, **3**.
28. Wieben, E.D., Madore, S.J. and Pederson, T. (1983) Protein binding sites are conserved in U1 small nuclear RNA from insects and mammals. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 1217-1220.
29. Kattah, N.H., Kattah, M.G. and Utz, P.J. (2010) The U1-snRNP complex: structural properties relating to autoimmune pathogenesis in rheumatic diseases. *Immunol Rev*, **233**, 126-145.
30. Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, **419**, 86-90.
31. Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K., Tange, T.O., Lund, M.K., Kjems, J., Scheid, A. and Schaal, H. (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.
32. Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., Meyer, I., Gierling, W., Theiss, S., Preuss, T. *et al.* (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic acids research*, **31**, 6963-6975.
33. Hartmann, L.T., Stephan; Niederacher, Dieter; Schaal, Heiner. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Frontiers in bioscience : a journal and virtual library*, **13**, 3252-3272.
34. Burset, M., Seledtsov, I.A. and Solovyev, V.V. (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic acids research*, **28**, 4364-4375.
35. Erkelenz, S., Theiss, S., Kaisers, W., Ptok, J., Walotka, L., Muller, L., Hillebrand, F., Brillen, A.L., Sladek, M. and Schaal, H. (2018) Ranking noncanonical 5'

splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. *Genome research*, **28**, 1826-1840.
36. Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, **3**, 285-298.
37. Smith, C.W., Chu, T.T. and Nadal-Ginard, B. (1993) Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Molecular and cellular biology*, **13**, 4939-4952.
38. Sibley, C.R., Blazquez, L. and Ule, J. (2016) Lessons from non-canonical splicing. *Nat Rev Genet*, **17**, 407-421.
39. Parada, G.E., Munita, R., Cerda, C.A. and Gysling, K. (2014) A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic acids research*, **42**, 10564-10578.
40. Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R. and Sachidanandam, R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic acids research*, **34**, 3955-3967.
41. Roscigno, R.F., Weiner, M. and Garcia-Blanco, M.A. (1993) A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *The Journal of biological chemistry*, **268**, 11222-11229.
42. Valcarcel, J., Gaur, R.K., Singh, R. and Green, M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*, **273**, 1706-1709.
43. Voith von Voithenberg, L., Sanchez-Rico, C., Kang, H.S., Madl, T., Zanier, K., Barth, A., Warner, L.R., Sattler, M. and Lamb, D.C. (2016) Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, E7169-E7175.
44. Green, M.R. (1986) Pre-mRNA splicing. *Annu Rev Genet*, **20**, 671-708.
45. Corvelo, A., Hallegger, M., Smith, C.W. and Eyras, E. (2010) Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol*, **6**, e1001016.
46. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Molecular cell*, **11**, 965-976.
47. Sun, C. (2020) The SF3b complex: splicing and beyond. *Cell Mol Life Sci*, **77**, 3583-3595.
48. Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37-47.
49. Ptok, J., Muller, L., Theiss, S. and Schaal, H. (2019) Context matters: Regulation of splice donor usage. *Biochim Biophys Acta Gene Regul Mech*, **1862**, 194391.
50. Manley, J.L. and Krainer, A.R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & development*, **24**, 1073-1074.
51. Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical journal*, **417**, 15-27.

52. Caceres, J.F., Screaton, G.R. and Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development*, **12**, 55-66.
53. Aznarez, I., Nomakuchi, T.T., Tetenbaum-Novatt, J., Rahman, M.A., Fregoso, O., Rees, H. and Krainer, A.R. (2018) Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. *Cell reports*, **23**, 2186-2198.
54. Müller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K. and Neugebauer, K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & development*, **30**, 553-566.
55. Boucher, L., Ouzounis, C.A., Enright, A.J. and Blencowe, B.J. (2001) A genome-wide survey of RS domain proteins. *Rna*, **7**, 1693-1701.
56. Pinol-Roma, S. (1997) HnRNP proteins and the nuclear export of mRNA. *Seminars in cell & developmental biology*, **8**, 57-63.
57. Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nature reviews. Molecular cell biology*, **3**, 195-205.
58. Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y. and Darnell, R.B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, **25**, 359-371.
59. Warzecha, C.C., Jiang, P., Amirikian, K., Dittmar, K.A., Lu, H., Shen, S., Guo, W., Xing, Y. and Carstens, R.P. (2010) An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *The EMBO journal*, **29**, 3286-3300.
60. Rahman, L., Bliskovski, V., Reinhold, W. and Zajac-Kaye, M. (2002) Alternative splicing of brain-specific PTB defines a tissue-specific isoform pattern that predicts distinct functional roles. *Genomics*, **80**, 245-249.
61. Han, J. and Cooper, T.A. (2005) Identification of CELF splicing activation and repression domains in vivo. *Nucleic acids research*, **33**, 2769-2780.
62. Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schoneweis, K., Hertel, K.J. and Schaal, H. (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *Rna*, **19**, 96-102.
63. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature*, **583**, 711-719.
64. Badr, E. and Heath, L.S. (2014) Identifying splicing regulatory elements with de Bruijn graphs. *J Comput Biol*, **21**, 880-897.
65. Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A. *et al.* (2018) Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Molecular cell*, **70**, 854-867 e859.
66. Michaud, S. and Reed, R. (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & development*, **5**, 2534-2546.

67. Kosowski, T.R., Keys, H.R., Quan, T.K. and Ruby, S.W. (2009) DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle. *Rna*, **15**, 1345-1362.
68. Berglund, J.A., Rosbash, M. and Schultz, S.C. (2001) Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *Rna*, **7**, 682-691.
69. Zhan, X., Yan, C., Zhang, X., Lei, J. and Shi, Y. (2018) Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res*, **28**, 1129-1140.
70. Chanarat, S. and Strasser, K. (2013) Splicing and beyond: the many faces of the Prp19 complex. *Biochimica et biophysica acta*, **1833**, 2126-2134.
71. Liu, Y.C. and Cheng, S.C. (2015) Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. *J Biomed Sci*, **22**, 54.
72. Wahl, M.C. and Luhrmann, R. (2015) SnapShot: Spliceosome Dynamics I. *Cell*, **161**, 1474-e1471.
73. Le Hir, H. and Andersen, G.R. (2008) Structural insights into the exon junction complex. *Curr Opin Struct Biol*, **18**, 112-119.
74. Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, **72**, 291-336.
75. Matlin, A.J., Clark, F. and Smith, C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nature reviews. Molecular cell biology*, **6**, 386-398.
76. Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J. and Chasin, L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome research*, **21**, 1360-1374.
77. Grodecka, L., Buratti, E. and Freiberger, T. (2017) Mutations of Pre-mRNA Splicing Regulatory Elements: Are Predictions Moving Forward to Clinical Diagnostics? *International journal of molecular sciences*, **18**.
78. Cheng, J., Nguyen, T.Y.D., Cygan, K.J., Celik, M.H., Fairbrother, W.G., Avsec, Z. and Gagneur, J. (2019) MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome biology*, **20**, 48.
79. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B. *et al.* (2019) Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, **176**, 535-548 e524.
80. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, **11**, 377-394.
81. de Miguel, F.J., Pajares, M.J., Martinez-Terroba, E., Ajona, D., Morales, X., Sharma, R.D., Pardo, F.J., Rouzaut, A., Rubio, A., Montuenga, L.M. *et al.* (2016) A large-scale analysis of alternative splicing reveals a key role of QKI in lung cancer. *Mol Oncol*, **10**, 1437-1449.
82. Singh, R.K. and Cooper, T.A. (2012) Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med*, **18**, 472-482.
83. Wang, J., Zhang, J., Li, K., Zhao, W. and Cui, Q. (2012) SpliceDisease database: linking RNA splicing and disease. *Nucleic acids research*, **40**, D1055-1059.
84. Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nat Rev Genet*, **17**, 19-32.

85. Abramowicz, A. and Gos, M. (2019) Correction to: Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of applied genetics*, **60**, 231.
86. Pawlicka, K., Kalathiya, U. and Alfaro, J. (2020) Nonsense-Mediated mRNA Decay: Pathologies and the Potential for Novel Therapeutics. *Cancers (Basel)*, **12**.
87. Tazi, J., Bakkour, N. and Stamm, S. (2009) Alternative splicing and disease. *Biochimica et biophysica acta*, **1792**, 14-26.
88. Ruzickova, S. and Stanek, D. (2017) Mutations in spliceosomal proteins and retina degeneration. *RNA Biol*, **14**, 544-552.
89. Taylor, J. and Lee, S.C. (2019) Mutations in spliceosome genes and therapeutic opportunities in myeloid malignancies. *Genes Chromosomes Cancer*, **58**, 889-902.
90. Baltimore, D. (1971) Viral genetic systems. *Trans N Y Acad Sci*, **33**, 327-332.
91. Poltronieri, P., Sun, B. and Mallardo, M. (2015) RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. *Curr Genomics*, **16**, 327-335.
92. Choi, K.H. (2012) Viral polymerases. *Advances in experimental medicine and biology*, **726**, 267-304.
93. Hiscox, J.A. (2007) RNA viruses: hijacking the dynamic nucleolus. *Nat Rev Microbiol*, **5**, 119-127.
94. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C. *et al.* (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
95. Popovic, M., Sarngadharan, M.G., Read, E. and Gallo, R.C. (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, **224**, 497-500.
96. Castro-Nallar, E., Perez-Losada, M., Burton, G.F. and Crandall, K.A. (2012) The evolution of HIV: inferences using phylogenetics. *Mol Phylogenet Evol*, **62**, 777-792.
97. Archer, J., Pinney, J.W., Fan, J., Simon-Loriere, E., Arts, E.J., Negroni, M. and Robertson, D.L. (2008) Identifying the important HIV-1 recombination breakpoints. *PLoS Comput Biol*, **4**, e1000178.
98. German Advisory Committee Blood, S.A.o.P.T.b.B. (2016) Human Immunodeficiency Virus (HIV). *Transfus Med Hemother*, **43**, 203-222.
99. Bell, N.M. and Lever, A.M. (2013) HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol*, **21**, 136-144.
100. Pancera, M., Majeed, S., Ban, Y.E., Chen, L., Huang, C.C., Kong, L., Kwon, Y.D., Stuckey, J., Zhou, T., Robinson, J.E. *et al.* (2010) Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 1166-1171.
101. Engelman, A. and Cherepanov, P. (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol*, **10**, 279-290.
102. Adams, O., Schaal, H. and Scheid, A. (2000) Natural variation in the amino acid sequence around the HIV type 1 glycoprotein 160 cleavage site and its effect on cleavability, subunit association, and membrane fusion. *AIDS Res Hum Retroviruses*, **16**, 1235-1245.

103. Taylor, B.S., Sobieszczyk, M.E., McCutchan, F.E. and Hammer, S.M. (2008) The challenge of HIV-1 subtype diversity. *N Engl J Med*, **358**, 1590-1602.
104. Arif, M.S., Hunter, J., Leda, A.R., Zukurov, J.P.L., Samer, S., Camargo, M., Galinskas, J., Kallas, E.G., Komninakis, S.V., Janini, L.M. *et al.* (2017) Pace of Coreceptor Tropism Switch in HIV-1-Infected Individuals after Recent Infection. *Journal of virology*, **91**.
105. Zila, V., Margiotta, E., Turonova, B., Muller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Borner, K., Rada, J., Muller, B. *et al.* (2021) Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*, **184**, 1032-1046 e1018.
106. Ambrose, Z. and Aiken, C. (2014) HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology*, **454-455**, 371-379.
107. Metifiot, M., Maddali, K., Naumova, A., Zhang, X., Marchand, C. and Pommier, Y. (2010) Biochemical and pharmacological analyses of HIV-1 integrase flexible loop mutants resistant to raltegravir. *Biochemistry*, **49**, 3715-3722.
108. Chiu, T.K. and Davies, D.R. (2004) Structure and function of HIV-1 integrase. *Curr Top Med Chem*, **4**, 965-977.
109. Brady, J. and Kashanchi, F. (2005) Tat gets the "green" light on transcription initiation. *Retrovirology*, **2**, 69.
110. Pasternak, A.O., Lukashov, V.V. and Berkhout, B. (2013) Cell-associated HIV RNA: a dynamic biomarker of viral persistence. *Retrovirology*, **10**, 41.
111. Carrera, C., Pinilla, M., Perez-Alvarez, L. and Thomson, M.M. (2010) Identification of unusual and novel HIV type 1 spliced transcripts generated in vivo. *AIDS Res Hum Retroviruses*, **26**, 815-820.
112. Karn, J. and Stoltzfus, C.M. (2012) Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*, **2**, a006916.
113. Purcell, D.F. and Martin, M.A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of virology*, **67**, 6365-6378.
114. Emery, A. and Swanstrom, R. (2021) HIV-1: To Splice or Not to Splice, That Is the Question. *Viruses*, **13**.
115. Anderson, J.L., Johnson, A.T., Howard, J.L. and Purcell, D.F. (2007) Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. *Journal of virology*, **81**, 4664-4676.
116. Krummheuer, J., Johnson, A.T., Hauber, I., Kammler, S., Anderson, J.L., Hauber, J., Purcell, D.F. and Schaal, H. (2007) A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG. *Virology*, **363**, 261-271.
117. Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K. *et al.* (2012) Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic acids research*, **40**, 10345-10355.
118. Suhasini, M. and Reddy, T.R. (2009) Cellular proteins and HIV-1 Rev function. *Curr HIV Res*, **7**, 91-100.

119. O'Reilly, M.M., McNally, M.T. and Beemon, K.L. (1995) Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology*, **213**, 373-385.
120. Kammler, S., Otte, M., Hauber, I., Kjems, J., Hauber, J. and Schaal, H. (2006) The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*, **3**, 89.
121. Sertznig, H., Hillebrand, F., Erkelenz, S., Schaal, H. and Widera, M. (2018) Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology*, **516**, 176-188.
122. Dlamini, Z. and Hull, R. (2017) Can the HIV-1 splicing machinery be targeted for drug discovery? *HIV AIDS (Auckl)*, **9**, 63-75.
123. Mitsuya, H., Weinhold, K.J., Furman, P.A., St Clair, M.H., Lehrman, S.N., Gallo, R.C., Bolognesi, D., Barry, D.W. and Broder, S. (1985) 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 7096-7100.
124. Montaner, J.S., Lima, V.D., Harrigan, P.R., Lourenco, L., Yip, B., Nosyk, B., Wood, E., Kerr, T., Shannon, K., Moore, D. *et al.* (2014) Expansion of HAART coverage is associated with sustained decreases in HIV/AIDS morbidity, mortality and HIV transmission: the "HIV Treatment as Prevention" experience in a Canadian setting. *PloS one*, **9**, e87872.
125. Quinn, T.C. (2008) HIV epidemiology and the effects of antiviral therapy on long-term consequences. *Aids*, **22 Suppl 3**, S7-12.
126. Panos, G., Samonis, G., Alexiou, V.G., Kavarnou, G.A., Charatsis, G. and Falagas, M.E. (2008) Mortality and morbidity of HIV infected patients receiving HAART: a cohort study. *Curr HIV Res*, **6**, 257-260.
127. Koenig, S.P., Kuritzkes, D.R., Hirsch, M.S., Leandre, F., Mukherjee, J.S., Farmer, P.E. and del Rio, C. (2006) Monitoring HIV treatment in developing countries. *Bmj*, **332**, 602-604.
128. Nachega, J.B., Marconi, V.C., van Zyl, G.U., Gardner, E.M., Preiser, W., Hong, S.Y., Mills, E.J. and Gross, R. (2011) HIV treatment adherence, drug resistance, virologic failure: evolving concepts. *Infect Disord Drug Targets*, **11**, 167-174.
129. Maskew, M., Bor, J., MacLeod, W., Carmona, S., Sherman, G.G. and Fox, M.P. (2019) Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. *Lancet HIV*, **6**, e760-e768.
130. Hu, W.S. and Hughes, S.H. (2012) HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*, **2**.
131. Lubke, N., Di Cristanziano, V., Sierra, S., Knops, E., Schulter, E., Jensen, B., Oette, M., Lengauer, T. and Kaiser, R. (2015) Proviral DNA as a Target for HIV-1 Resistance Analysis. *Intervirology*, **58**, 184-189.
132. Nanfack, A.J., Redd, A.D., Bimela, J.S., Ncham, G., Achem, E., Banin, A.N., Kirkpatrick, A.R., Porcella, S.F., Agyingi, L.A., Meli, J. *et al.* (2017) Multimethod Longitudinal HIV Drug Resistance Analysis in Antiretroviral-Therapy-Naive Patients. *J Clin Microbiol*, **55**, 2785-2800.
133. Arts, E.J. and Hazuda, D.J. (2012) HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med*, **2**, a007161.

134. Bhatti, A.B., Usman, M. and Kandi, V. (2016) Current Scenario of HIV/AIDS, Treatment Options, and Major Challenges with Compliance to Antiretroviral Therapy. *Cureus*, **8**, e515.
135. Archin, N.M., Kirchherr, J.L., Sung, J.A., Clutton, G., Sholtis, K., Xu, Y., Allard, B., Stuelke, E., Kashuba, A.D., Kuruc, J.D. *et al.* (2017) Interval dosing with the HDAC inhibitor vorinostat effectively reverses HIV latency. *The Journal of clinical investigation*, **127**, 3126-3135.
136. Asamitsu, K., Fujinaga, K. and Okamoto, T. (2018) HIV Tat/P-TEFb Interaction: A Potential Target for Novel Anti-HIV Therapies. *Molecules*, **23**.
137. Karpinski, J., Hauber, I., Chemnitz, J., Schafer, C., Paszkowski-Rogacz, M., Chakraborty, D., Beschorner, N., Hofmann-Sieber, H., Lange, U.C., Grundhoff, A. *et al.* (2016) Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat Biotechnol*, **34**, 401-409.
138. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R. *et al.* (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*, **382**, 727-733.
139. Hatmal, M.M., Alshaer, W., Al-Hatamleh, M.A.I., Hatmal, M., Smadi, O., Taha, M.O., Oweida, A.J., Boer, J.C., Mohamud, R. and Plebanski, M. (2020) Comprehensive Structural and Molecular Comparison of Spike Proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and Their Interactions with ACE2. *Cells*, **9**.
140. Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M. and Hassan, M.I. (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis*, **1866**, 165878.
141. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z. *et al.* (2020) Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, **183**, 730-738 e713.
142. Letko, M., Marzi, A. and Munster, V. (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol*, **5**, 562-569.
143. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L. *et al.* (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**, 215-220.
144. Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G. and Decroly, E. (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research*, **176**, 104742.
145. Chan, J.F., Kok, K.H., Zhu, Z., Chu, H., To, K.K., Yuan, S. and Yuen, K.Y. (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*, **9**, 221-236.
146. Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N. and Chang, H. (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell*, **181**, 914-921 e910.
147. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L. *et al.* (2020) A SARS-CoV-

2 protein interaction map reveals targets for drug repurposing. *Nature*, **583**, 459-468.
148. Padmanabhan, P., Desikan, R. and Dixit, N.M. (2020) Targeting TMPRSS2 and Cathepsin B/L together may be synergistic against SARS-CoV-2 infection. *PLoS Comput Biol*, **16**, e1008461.
149. Choppin, P.W. and Scheid, A. (1980) The role of viral glycoproteins in adsorption, penetration, and pathogenicity of viruses. *Rev Infect Dis*, **2**, 40-61.
150. Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J. *et al.* (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*, **27**, 325-328.
151. Kelly, J.A., Olson, A.N., Neupane, K., Munshi, S., San Emeterio, J., Pollack, L., Woodside, M.T. and Dinman, J.D. (2020) Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *The Journal of biological chemistry*, **295**, 10741-10748.
152. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. and Thiel, V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*, **19**, 155-170.
153. Lebeau, G., Vagner, D., Frumence, E., Ah-Pine, F., Guillot, X., Nobecourt, E., Raffray, L. and Gasque, P. (2020) Deciphering SARS-CoV-2 Virologic and Immunologic Features. *International journal of molecular sciences*, **21**.
154. Harrison, A.G., Lin, T. and Wang, P. (2020) Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends Immunol*, **41**, 1100-1115.
155. Chu, C.M., Cheng, V.C., Hung, I.F., Wong, M.M., Chan, K.H., Chan, K.S., Kao, R.Y., Poon, L.L., Wong, C.L., Guan, Y. *et al.* (2004) Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax*, **59**, 252-256.
156. Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M. *et al.* (2020) A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N Engl J Med*, **382**, 1787-1799.
157. Self, W.H., Semler, M.W., Leither, L.M., Casey, J.D., Angus, D.C., Brower, R.G., Chang, S.Y., Collins, S.P., Eppensteiner, J.C., Filbin, M.R. *et al.* (2020) Effect of Hydroxychloroquine on Clinical Status at 14 Days in Hospitalized Patients With COVID-19: A Randomized Clinical Trial. *JAMA*, **324**, 2165-2176.
158. Hoffmann, M., Hofmann-Winkler, H., Smith, J.C., Kruger, N., Arora, P., Sorensen, L.K., Sogaard, O.S., Hasselstrom, J.B., Winkler, M., Hempel, T. *et al.* (2021) Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *EBioMedicine*, **65**, 103255.
159. Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S. *et al.* (2020) Remdesivir for the Treatment of Covid-19 - Final Report. *N Engl J Med*, **383**, 1813-1826.
160. Group, R.C., Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A. *et al.* (2021) Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*, **384**, 693-704.

161. Katz, L.M. (2021) (A Little) Clarity on Convalescent Plasma for Covid-19. *N Engl J Med*, **384**, 666-668.
162. Taylor, P.C., Adams, A.C., Hufford, M.M., de la Torre, I., Winthrop, K. and Gottlieb, R.L. (2021) Neutralizing monoclonal antibodies for treatment of COVID-19. *Nat Rev Immunol*, **21**, 382-393.
163. Deb, P., Molla, M.M.A. and Saif-Ur-Rahman, K.M. (2021) An update to monoclonal antibody as therapeutic option against COVID-19. *Biosaf Health*, **3**, 87-91.
164. Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J. *et al.* (2020) An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl J Med*, **383**, 1920-1931.
165. Kaur, S.P. and Gupta, V. (2020) COVID-19 Vaccine: A comprehensive status report. *Virus research*, **288**, 198114.
166. Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Perez Marc, G., Moreira, E.D., Zerbini, C. *et al.* (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med*, **383**, 2603-2615.
167. Voysey, M., Clemens, S.A.C., Madhi, S.A., Weckx, L.Y., Folegatti, P.M., Aley, P.K., Angus, B., Baillie, V.L., Barnabas, S.L., Bhorat, Q.E. *et al.* (2021) Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*, **397**, 99-111.
168. Arias, M.A., Lubkin, A. and Chasin, L.A. (2015) Splicing of designer exons informs a biophysical model for exon definition. *Rna*, **21**, 213-229.
169. De Conti, L., Baralle, M. and Buratti, E. (2013) Exon and intron definition in pre-mRNA splicing. *Wiley interdisciplinary reviews. RNA*, **4**, 49-60.
170. Conboy, J.G. (2021) Unannotated splicing regulatory elements in deep intron space. *Wiley interdisciplinary reviews. RNA*, e1656.
171. Pai, A.A., Paggi, J.M., Yan, P., Adelman, K. and Burge, C.B. (2018) Numerous recursive sites contribute to accuracy of splicing in long introns in flies. *PLoS genetics*, **14**, e1007588.
172. Ott, S., Tamada, Y., Bannai, H., Nakai, K. and Miyano, S. (2003) Intrasplicing--analysis of long intron sequences. *Pac Symp Biocomput*, 339-350.
173. Carmody, S.R. and Wente, S.R. (2009) mRNA nuclear export at a glance. *J Cell Sci*, **122**, 1933-1937.
174. Aibara, S., Katahira, J., Valkov, E. and Stewart, M. (2015) The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA. *Nucleic acids research*, **43**, 1883-1893.
175. Muller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K. and Neugebauer, K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes & development*, **30**, 553-566.
176. Geuens, T., Bouhy, D. and Timmerman, V. (2016) The hnRNP family: insights into their role in health and disease. *Hum Genet*, **135**, 851-867.
177. Ostermann, P.N., Ritchie, A., Ptok, J. and Schaal, H. (2021) Let It Go: HIV-1 cis-Acting Repressive Sequences. *Journal of virology*, **95**, e0034221.

178. Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. and Luhrmann, R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, **82**, 475-483.
179. Gelli, E., Colombo, M., Pinto, A.M., De Vecchi, G., Foglia, C., Amitrano, S., Morbidoni, V., Imperatore, V., Manoukian, S., Baldassarri, M. *et al.* (2019) Usefulness and Limitations of Comprehensive Characterization of mRNA Splicing Profiles in the Definition of the Clinical Relevance of BRCA1/2 Variants of Uncertain Significance. *Cancers (Basel)*, **11**.
180. Wangensteen, T., Felde, C.N., Ahmed, D., Maehle, L. and Ariansen, S.L. (2019) Diagnostic mRNA splicing assay for variants in BRCA1 and BRCA2 identified two novel pathogenic splicing aberrations. *Hered Cancer Clin Pract*, **17**, 14.
181. Scheid, J.F., Mouquet, H., Feldhahn, N., Seaman, M.S., Velinzon, K., Pietzsch, J., Ott, R.G., Anthony, R.M., Zebroski, H., Hurley, A. *et al.* (2009) Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature*, **458**, 636-640.
182. Caskey, M., Klein, F. and Nussenzweig, M.C. (2019) Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. *Nature medicine*, **25**, 547-553.
183. West, A.P., Jr., Scharf, L., Horwitz, J., Klein, F., Nussenzweig, M.C. and Bjorkman, P.J. (2013) Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10598-10603.
184. Corey, L., Gilbert, P.B., Juraska, M., Montefiori, D.C., Morris, L., Karuna, S.T., Edupuganti, S., Mgodi, N.M., deCamp, A.C., Rudnicki, E. *et al.* (2021) Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. *N Engl J Med*, **384**, 1003-1014.
185. Cines, D.B. and Bussel, J.B. (2021) SARS-CoV-2 Vaccine-Induced Immune Thrombotic Thrombocytopenia. *N Engl J Med*, **384**, 2254-2256.
186. Kowarz, E.K., L.; Reis, J.; Bracharz, S.; Kochanek, S.; Marschalek, R . (2021) “Vaccine-Induced Covid-19 Mimicry” Syndrome:Splice reactions within the SARS-CoV-2 Spike open reading frame result in Spike protein variants that may cause thromboembolic events in patients immunized with vector-based vaccines. *Preprint*.