Review

U2, U6 and U5 snRNAs for Splicing and Gene Therapy

Olga V. Artemyeva-Isman *

Posted Date: 20 September 2026

doi: 10.20944/preprints202609.1663.v1

Keywords: U2 snRNA; U6 snRNA; pre-catalytic spliceosome; U5 snRNA; reverse splicing; intron lariat spliceosome; Duchenne muscular dystrophy; CRISPR/Cas; RNA interference; virus-like particles

 doi:10.20944/preprints202609.1663.v1

Review

U2, U6 and U5 snRNAs for Splicing and Gene Therapy

Therapeutic Applications of Human Splicing Ribozymes, Part 4 of 4

Olga V. Artemyeva-Isman

Branchpoint Therapeutics, London, UK; o.isman@branchpoint-therapeutics.com

Abstract

U2, U6 and U5 are the components of spliceosomal ribozymes that descend from mobile self-splicing introns. The fragmentation of ancestral introns in eukaryogenesis led to the evolution of the ribozyme assembly *in trans* using common snRNAs and protein-chaperones. This stepwise process provides opportunities to target distinct stages of ribozyme folding. Reviewing successive spliceosomal complexes, we will discuss how to adapt them for therapy, both splicing modulation and permanent gene therapy. Adapting U2 and U6 snRNAs can overcome the limitations of U1 and U7, currently used to target early spliceosome complexes. U2 can select alternative 3'ss, such as *VEGFA* switch between pro- and anti-angiogenic isoforms. U6 and U2 can be used together in the pre-catalytic complex, taking advantage of the U6/U2 Helix II that can be modified to prevent intermixing with WT U6 and U2 molecules. Extending 5'ss and BP helices with adapted U6 and U2 can create specific spliceosome species for individual introns. U5 snRNA, responsible for exon recognition, is the last to pair with pre-mRNA at pre-catalytic stage, but in reverse splicing, the exon recognition loop initiates intron insertion. Completed forward splicing produces an RNA-protein complex of an intron lariat paired with U6 and U2, the active ribozyme centre and U5 snRNA associated by protein interactions – the Intron Lariat Spliceosome. ILS is homologous to the mobile Group IIA intron particle capable of invading genomic loci by reverse splicing. Group II introns are used for microbial genome engineering. Can human ILS be adapted for specific genome insertions? Group IIA introns possess a second exon-recognition loop, which adds 6bp to the interaction with the 5'exon. Engineering an additional exon-recognition loop for spliceosomal U5 snRNA is the way to bring back reverse splicing. The field of genome engineering is dominated by CRISPR/*Cas* derivatives, so why do we need to develop reverse splicing? It is increasingly apparent that the evolutionary fate of RNPs defines their therapeutic utility: CRISPR/*Cas* are prokaryotic RNPs. Targeting human RNPs is the best solution, as the therapeutic success of RNAi using *RISC* complex indicates. The spliceosome is a human RNP worth exploring for splicing and gene therapy.

Keywords: U2 snRNA; U6 snRNA; pre-catalytic spliceosome; U5 snRNA; reverse splicing; intron lariat spliceosome; Duchenne muscular dystrophy; CRISPR/*Cas*; RNA interference; virus-like particles

1. Introduction

The last paper of the series provides a perspective on practical applications of the core spliceosomal snRNAs, U2, U6 and U5. These are the components of the splicing ribozyme that descend from ancestral mobile self-splicing introns (see *Part 1,* Artemyeva-Isman, 2026a). The fragmentation of these introns in eukaryogenesis led to the evolution of the ribozyme assembly *in trans* using common snRNAs. The current knowledge on spliceosome assembly (see *Part 2,* Artemyeva-Isman, 2026b) allows to target distinct stages of ribozyme folding. Reviewing successive spliceosomal complexes, we will focus on practical aspects of their adaptation for therapy.

 doi:10.20944/preprints202609.1663.v1

Halfway through, we will change from forward to reverse splicing, giving a new perspective on gene therapy. Splicing produces an RNP with an intron lariat and the active ribozyme centre – the Intron Lariat Spliceosome. ILS is homologous to the mobile Group IIA intron particle, and we will discuss how ILS can be adapted for genome insertions guided by U5 snRNA to specific targets.

Figures are numbered throughout this series of linked papers, *Part 1:* Figures 1–4; *Part 2:* Figures 5–7; *Part 3* (Artemyeva-Isman, 2026c): Figures 8–10, *Part 4* (this paper): Figures 11–14.

2. Early Splice Site Choice

Pre-clinical studies of **U1 snRNA** were at length reviewed in *Part 3*. Their accumulated experience provides a positive outlook on safety and AAV-mediated delivery (Figure 8), but despite the important initial success it also indicates a variation in efficacy and some activation of cryptic 5′ss. The limitations of using U1 alone can be remedied by core spliceosomal RNAs. Let us start with U2.

2.1. U2 snRNA

U2 is the first ribozyme component to pair with the pre-mRNA, forming the BP helix (see *Parts 1* and *2*). Splicing correction for branchpoint site (BP) mutations have not been attempted by the reciprocal U2 mutants since the first pioneering study of 1989 that proved the importance of Watson-Crick pairs between U2 and the BP site (Zhuang and Weiner, 1989). However, adapted U2 was again expressed in patient-derived immortalised cells (De Angelis et al., 2002), which seemed not to affect cell viability. Although the aim was DMD exon skipping, naturally, this was not achieved by U2 targeting the BP site.

Possible applications of U2 for splice site selection in the early spliceosome (complex **A**, Figure 5) are proposed in Figure 11A,B. U2 can be combined with U1 for targeting exon inclusion or U2 can be used to choose an alternative 3′ss site. The example of *VEGFA* pro-/anti-vascularisation splicing isoforms implies the choice between two 3′ss each combined with its branchpoint (see **3 Splicing switches as therapeutic targets in common diseases** in *Part 3*, Artemyeva-Isman, 2026c). U2 targeting BP sequence associated with the distal 3′ss should promote its usage and suppress vascularisation in solid tumours or in age-related macular degeneration.

More than one BP site used competitively and in a tissue-specific manner is possible for a single human 3′ss (Pineda and Bradley, 2018; Figure 11 shows multiple BP sites as dots). The best splicing booster must be selected from U2 versions made to match each of these competing sites or indeed any $\frac{U}{C}$N$\underline{A}$ that appears in the correct range from 3′ss. The idea is that Watson-Crick pairs of the BP Helix can be substituted by other Watson-Crick pairs, but $U_{BP\text{-}2}$ (or $C_{BP\text{-}2}$ as in Group II introns) is required for a non-Watson-Crick interaction with the BP $\underline{A}$ (Figure 2A,C in *Part 1*, Artemyeva-Isman, 2026a).

BP Helix stretch of 5bp upstream from the BP A is modelled as in *S. cerevisiae* where it is well conserved. In the human minor spliceosome and in Group II introns BP Helix is longer (Figure 2A,B). U2 $\Psi_{39}C^{m}{}_{40}\Psi_{41}$ nucleotides are available to pair with the intron, and human major BP Helix was recently modelled by CryoEM extending further to U2 Ψ_{44} (Cretu et al., 2021). However, other RNA interactions can limit the extent of the BP Helix with the adapted U2 (see next section and Figure 12).

 doi:10.20944/preprints202609.1663.v1

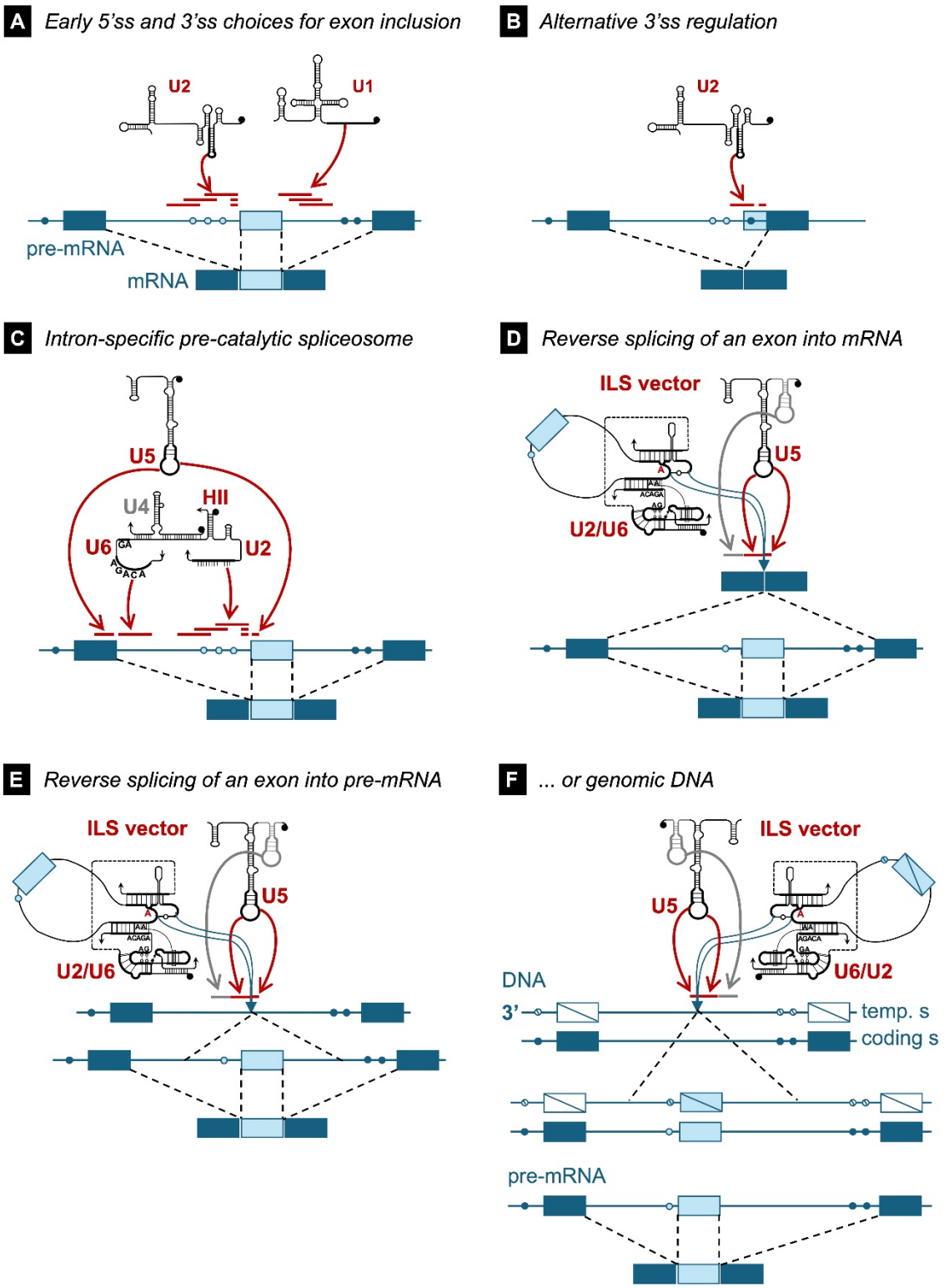

Figure 11. Practical applications of engineered snRNAs targeting early (**A-B**), pre-catalytic (**C**) and catalytic (**D-F**) spliceosomes (successive spliceosome complexes see Figure 5 in *Part 2*, Artemyeva-Isman, 2026b). **A.** In the early spliceosome, inclusion of an alternative or mutated exon can be boosted by adapted U1: array in red - compensatory or ExSpeU1s (Figure 10 in *Part 3*, Artemyeva-Isman, 2026c). Equally, affinity to U2 influences BP usage and exon inclusion: U2 array in red indicates alternative BPs targeted for optimisation. U1 and U2 can be used together to target complex **A**. **B.** Adapted U2 targeting a distal 3'ss with its branchpoint – an example of *VEGFA* anti-angiogenic isoform (see section **3 Splicing switches as therapeutic targets in common diseases** in *Part 3*, Artemyeva-Isman, 2026c). **C.** At the pre-catalytic stage splicing can be boosted by U6 and U2 engineered together, **HII,** U2/U6 Helix II can be altered to prevent intermixing with WT U2 and U6. Extension of

 doi:10.20944/preprints202609.1663.v1

complementarity of the 5′ss and BP Helices will create a spliceosome specific to a particular intron. U5 engineered to match the exons can be added, but intermixing with WT U5 cannot be prevented. **D.** Reverse splicing, guided by engineered U5 can be used to insert a deleted exon or exons. Additional U5 recognition loop (in grey), homologous to Group IIA IdI loop (Figure 2B in *Part 1*, Artemyeva-Isman, 2026a) will enhance efficiency and specificity. **ILS**, Intron Lariat Spliceosome vector carries the missing exon. Targeting mRNA implies that another round of splicing will take place to excise the integrated introns. **E.** Targeting pre-mRNA means that a target site is freely chosen within an intron with a deletion breakpoint. **F.** Reverse splicing into genomic DNA can target either the coding or the template strand. In the case of the coding strand the design of U5 and ILS vector is the same as for the pre-mRNA target (**E**). Illustrated here is the integration into the template strand.

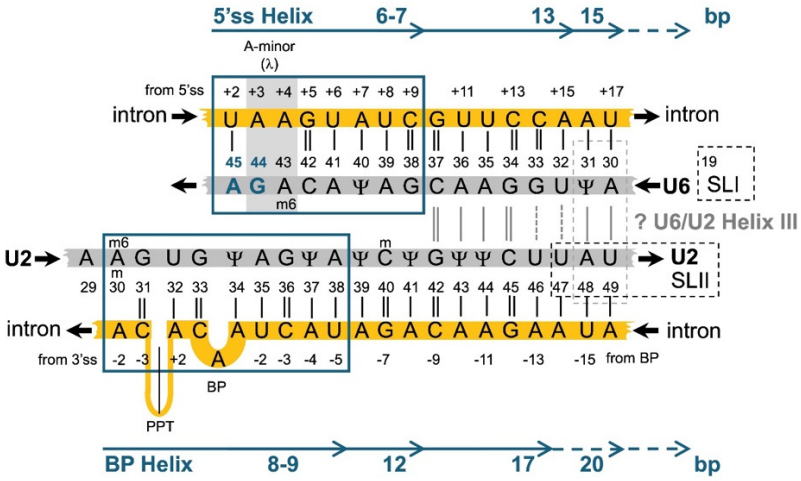

Figure 12. Extending 5′ss and BP helices with adapted U6 and U2 snRNAs can be limited by flanking RNA structures. U6 nucleotides in bold blue cannot be changed; Blue boxes – minimal helices. Grey filled box: A-minor link: conserved A_{+3} inserts into the minor groove of U6 SL of the catalytic site (Figure 2A,B in *Part 1*, Artemyeva-Isman, 2026a). A_{+4} can compensate for the absence of A_{+3}. If both adenines are absent another base in position +3 or +4 can be participating in a suboptimal interaction (see *Part 2*, Artemyeva-Isman, 2026b). Grey dotted box: positions potentially available for U6/U2 Helix III as shown by biochemical probing (Anokhina et al., 2013). U6/U2 Helix III is not featured in CryoEM studies (Bertram et al., 2017a; Cretu et al., 2021). Black dotted boxes: include end position of U6 SL I and start position of U2 SL II. SnRNA base modifications are introduced in sequence-specific manner, they will be lost if the sequence is changed in the vicinity. Assuming these modifications are needed to stabilise the helices with mismatched pairs, we can get away without them for perfect helices of exclusively Watson-Crick pairs. The loss of m6 modifications for U6 $A^{m6}{}_{43}$ and U2 $A^{m6,m}{}_{30}$ suggests that we can accommodate Watson-Crick pairs instead as in human minor introns (Figure 2A). U6 A_{43} can be changed for a base matching position +4 (only if A_{+3} is there for the optimal A-minor link) and U2 $A^{m6,m}{}_{30} \rightarrow U_{30}$ to pair the invariant A_{-2}.

The range for BP A distance from the 3′ss is very hard to define. In human introns poor conservation of BP sites is linked to great variability of the size and pyrimidine/purine composition of the spacer separating BP A from the 3′ss. This variability is at the basis of alternative 3′ss choices. Indeed, the initial selection combines BP, PPT and the 3′ss and is achieved by *U2AF35/65* and *SF1* protein components of the U2 snRNP (see *Part 2* - Artemyeva-Isman, 2026b). It is possible that U2 snRNA designed to form a perfect helix with the target BP site will be less reliant on protein support. Distant BPs possess spacers which harbour long PPTs (Gooding et al., 2006). When choosing $\frac{U}{C}$N$\underline{A}$ upstream of the 3′ss we cannot change the existing spacer composition – it will be impractical to create distant BPs. Therefore, U2 designs should favour minimal spacers or no spacer at all, if possible, choosing potential branchpoints between positions -6 (guided by U2 binding register) to -25 (the peak position for human introns) – see *Part 2*.

 doi:10.20944/preprints202609.1663.v1

U2 snRNA, when correctly bound to pre-mRNA has a split binding register, as it needs to skip BP A, which bulges out of the BP helix. *Part 2* reviews ribozyme folding with combined evidence (Sickmier et al., 2006; Chen et al., 2010; Biancon et al., 2022; Yoshida et al., 2020; Kent et al., 2003; Irimia and Roy, 2008; Corrionero et al., 2011) in favour of U2 snRNA bridging over the spacer to the 3'ss. We proposed that U2 G_{31} pairs with the conserved C_{-3} at the end of the intron. *U2AF65* acts as a cable clip, sharply bending the spacer and juxtaposing BP and the 3'ss. U2 design demands an experiment to verify the U2 G_{31}=C_{-3} pair. Double mutation analysis that transverses this pair can be a sufficient proof: U2 G_{31}→C_{31} and a splicing reporter with C_{-3}→G_{-3} mutation can be used in nuclear extracts or co-transfected in human cell lines. Another way will be to use the fact that the *U2AF35* S34F mutation in myelodysplasia specifically represses inclusion of exons if 3'ss lacks the conserved C_{-3} (Biancon et al., 2022, see *Part 2*). *U2AF35* WT normally supports binding to 3'ss with a variant -3 nucleotide: U_{-3} or a rare A_{-3}/G_{-3}. HEK293T cells that overexpress *U2AF35* S34F reproduce the aberrant splicing pattern (Biancon et al., 2022), and it will be interesting to see (RNAseq or long-read RNAseq) if U2 G_{31}→A_{31}/U_{31}/C_{31} can improve inclusion of exons preceded by the reciprocal U_{-3}/A_{-3}/G_{-3}.

3. Targeting Pre-Catalytic Complex

The pre-catalytic complex evolved as a splicing fidelity checkpoint that delays the formation of the catalytic metal binding site until U6 and U5 snRNA stably pair with pre-mRNA (see *Part 2* - Artemyeva-Isman, 2026b). Hwang and Cohen, 1996 demonstrated that 5'ss mutations can be supressed by compensatory changes in U6 snRNA complementary to intron positions +5 to +9. Three years previous Cortes et al., 1993 showed that U5 snRNA with changes in Loop1 promotes usage of cryptic splice sites. However, U5 base-pairing model then involved the 5' intron end, which unfortunately jumbled up experimental planning and conclusions. Since then, surprisingly little experimentation has been done with U6 and U5 snRNAs with a view to their therapeutic application.

3.1. U6 snRNA

Three attempts were made to adopt U6 alongside with U1 for the correction of disease-causing 5'ss mutations. These studies are reviewed for U1 snRNA in *Part 3* (Artemyeva-Isman, 2026c).

In the experiment by Carmel et al., 2004 partial correction of the splicing defect caused by *ELP1/IKBKAP* intron 20 +6T→C mutation was achieved by U1 snRNA. However, in a parallel experiment a reciprocal change in U6 snRNA 41A→G failed to improve inclusion of *ELP1* exon 20. Schmid et al., 2013 had more success when they used U1 and U6 together: the *BBS1* intron 5 +5G→U mutation was only partially corrected by U1 alone, however co-transfecting U1 and U6 markedly improved exon 5 inclusion. U6 42C→A was made complementary to the +5G→U mutation, but it worked even better with an additional change U6 41A→U producing a pair with the *BBS1* intron 5 unusual +6A. This indicates that extending U6 interaction into the intron is beneficial. Previously the 5'ss helix extended by mutating U6 position 38 to pair intron position +9 increased splicing efficacy (see *Part 2* - Artemyeva-Isman, 2026b; Crispino and Sharp,1995; Hwang and Cohen,1996). CryoEM models extend the 5'ss helix of human major introns until U6 A_{30}, which aligns with intron position +17 (Bertram et al., 2017b; Figure 6 in *Part 2*, Artemyeva-Isman, 2026b). Schmid et al., 2013 were changing U6 in the wrong direction, introducing complementarity with intron positions +4, +3 and +1, which makes no sense considering the RNA structure (Figure 2A,C - *Part 1*, Artemyeva-Isman, 2026a) and therefore these changes did not produce the desirable effect. Later Swirski et al., 2023 targeted the mouse *Opa1* intron 10 +5G→A mutation with compensatory U1, adding U6 double mutant 42C→U, 43A→U and U6 WT. They found that U6 addition significantly improved splicing correction, but overexpression of U6 WT was as good as U6 mutant or marginally better. Again, it is a pity that the U6 complementarity was extended to intron position +4, rather than further into the intron from position +5.

Future U6 design should take into account that the first RNA/RNA interaction to be formed in the pre-catalytic spliceosome is U6/U2 Helix II (in Pre-B complex Figures 6 and 5). U6 never acts on

 doi:10.20944/preprints202609.1663.v1

its own, it functions as a U6/U2 duplex; the role of U2 is obvious in anchoring U6 on the pre-mRNA. Minor spliceosome is distinct in this respect: U6atac and U12 do not have a helix homologous to U6/U2 Helix II (instead there is a U6atac 3'SL, Figure 2A), however U6atac forms a long well-conserved 5'ss Helix. Then again, U12 is initially associated with U11 via protein interactions in a di-snRNP selecting the 5'ss together with the 3'ss from the very beginning in the minor spliceosome assembly. Following major spliceosome assembly, U6 design should be considered in combination with U2 and U6/U2 Helix II can be used to the advantage, as it can be altered to prevent adapted U6 and U2 mixing with the endogenous counterparts.

Extending both 5'ss and BP helices is aimed at targeting unique introns with increased precision (Figure 11C). How far 5'ss and BP Helices can be extended depends on the predicted U6/U2 Helix III, which is not confirmed by biochemical methods apart from possibly U6 $A_{30}\Psi_{31}$ and U2 $A_{48}U_{49}$ (Anokhina et al., 2013) and seems absent in CryoEM models (Bertram et al., 2017b; Cretu et al., 2021). Despite its elusiveness, Helix III cannot be forgotten, unless designer U6 and U2 with 5'ss and BP helices excluding Helix III will prove more efficacious for splicing than U2 and U6, which can have Helix III interaction of some length. Accordingly, 5'ss Helix can include a minimum of 6-7 Watson-Crick pairs (Figure 12, Figure 2A) with the preceding 6nt of U6 potentially free for an extension of this interaction (Figure 12, Figure 6). BP Helix with a minimum of 8-9bp (Figure 12, Figure 2A) has the further 8nt of U2 possibly available for an extension (Figure 12, Figure 6). How far 5'ss and BP helices can be extended remains to be determined experimentally.

The length and GC content of the 5'ss and BP helices are important, as in primer and ASO design, but there is more to core spliceosomal snRNAs than Watson-Crick base-pairing. Next, tertiary links are formed: the non-Watson-Crick pair between the intron ends, the A-minor interaction of the 5'ss helix with the catalytic site and the subsequent BP A coordination (Figure 2A,C). We can consider that the ribozyme conducts recognition by tertiary structure, while ASOs operate at secondary structure level.

3.2. U5 snRNA

In the pre-catalytic complex U5 snRNA finishes splice site definition by binding the exons.

Targeting the pre-catalytic complex can include designing U5 snRNA alongside with the U6/U2 duplex, although U5 is not paired with U6 or U2 and the assembly of the designer U6/U2 duplex with the designer U5 is left to chance (Figure 11C).

U5 binding register and the discrepancy between CryoEM and statistical dependencies at human splice sites (Artemyeva-Isman and Porter, 2021) was discussed at length in *Part 2* (Artemyeva-Isman, 2026b – section **3.3 Exon recognition by U5 snRNA Loop1**). Experimental verification of the U5 binding register is necessary and we need to choose good models and approaches for this task. Recent experiments with transient and stable transfections of U5 variants with mutated Loop1 in two species of yeast (Ishigami et al., 2021; Negi et al., 2025) involved RT-PCR for a few selected mRNAs and the results do not support one or the other U5 binding register conclusively (see *Part 2*). In addition, as exon ends are not conserved in yeast, U5 must be more important for splice site recognition in human cells, where exon ends have a clear consensus and therefore human cells will be a better experimental model. The recent discoveries (Jackson et al., 2025; Nava et al., 2025) that linked U5 Loop1 motif mutations in *RNU5B* and *RNU5A* genes to neurodevelopmental diseases brought fresh interest in U5 interactions with exons. Different U5 Loop1 mutations have distinct effects on splicing in patients' blood samples, affecting either 5'ss or 3'ss or both. A possible experimental approach to explore this further is to assess, using RNASeq or long RNASeq, the effects of substitutions in the U5 Loop1 motif in embryonic mammalian neurons and iPSC-derived human neurons. Another experimental approach to fix the U5 binding register is to take advantage of the fact that the *U2AF35* Q157R mutation in myelodysplasia supresses the usage of the 3'ss if the first exon base is not a G conserved in 50% of human exons. According to our binding register the exon-start G pairs with U5 38C. HEK293 cells expressing *U2AF35* Q157R that reproduce the aberrant splicing pattern (Biancon et al., 2022) can be transfected with U5 38C→U/A/G and analysed by

 doi:10.20944/preprints202609.1663.v1

RNASeq for changes in splicing. However, base-pairing of the first base of the exon is not defined by the CryoEM model, but it can be inferred as U5 C_{39}=G_{+1}, presuming that it is next to the pair of the last base of the 5'exon – U5 U_{40}•G_{-1}. Accordingly, this can be tested with U5 39C→U/A/G. General approaches, RNASeq or long run RNASeq, are preferable to inspecting individual transcripts, as we can expect large variability depending on the sequence environment and the expression of RBPs in model cells. Finally, directed evolution of U5 Loop1 sequence can be tried for a pre-mRNA template with the conserved guanines at the end of the 5' exon and at the start of the 3' exon changed to cytosines.

4. Adapting Intron Lariat Complex for Reverse Splicing

The ILS complex, the Intron Lariat Spliceosome (Figure 5 in *Part 2*, Artemyeva-Isman, 2026b), is released when pre-mRNA splicing is completed. ILS is structurally homologous to the Group II intron particle – an RNP that can insert intron RNA into a specific target site. Importantly, Group II introns reverse splice into RNA and DNA with comparable efficiency, indicating that 2'O of the target sugar-phosphate backbone does not affect intron integration (Griffin et al., 1995). Group II intron-based 'targetrons' have been used for decades in academic and industrial labs for genetic engineering in bacteria (Karberg et al., 2001; Mohr et al., 2013; Sang et al., 2024). Despite the efforts to transfer Group II introns as gene engineering tools to human cells, their catalytic activity is suppressed by lower free Mg^{2+} concentrations, which is 1-4mM in bacterial cells compared to 0.2-1.0mM in eukaryotic cells and even lower in the nucleus, where Mg^{2+} is chelated to chromosomal DNA. Accordingly, $MgCl_2$ microinjection is required for Group II intron integration in *Xenopus laevis* oocytes or *Drosophila melanogaster* and zebrafish embryos (Lambowitz and Belfort, 2015). It is known that the intron encoded protein *IEP* supporting Group II intron activity *in vitro* can compensate for low Mg^{2+} levels. Both *IEP's* role as a chaperone and high Mg^{2+} assist correct ribozyme folding. Spliceosomal introns that operate at lower Mg^{2+} evolved to rely on chaperone support of the *IEP*-derived *Prp8* and the *NTC* complex, which binds the catalytic core prior to the catalysis (Figure 5).

ILS, like Group II introns, is capable of reverse splicing. As reviewed in *Parts 1* and *2* (Artemyeva-Isman, 2026a,b), both catalytic steps of splicing are reversible in *S. cerevisiae* nuclear extracts (Tseng and Cheng, 2008). Genomic integration of a reporter intron by reverse splicing in *S. cerevisiae* was also demonstrated – albeit as very rare events at a frequency in the order of 10^{-11} (Lee and Stevens, 2016). However, in *S. cerevisiae* introns were purged out by purifying selection, because of a large effective population size and a very efficient homologous recombination: HR is the basic mechanism for intron elimination via the cDNA template (see *Part 1*). The same forces are likely to interfere with any significant intron integration by ILS complexes.

Admittedly, reverse splicing has never been explored in human cells, where the background conditions are radically different from *S. cerevisiae*: the genome is intron-rich, splicing is abundant and critical for genome regulation. Human introns were preserved by evolution as HR rates are pitifully low. In fact, low HR in human cells and the ensuing fiasco of gene targeting (see for example Isman et al., 2008) brought about the rise of the CRISPR-*Cas* technology. It much exceeds gene targeting in efficiency and rather than relying on endogenous HR in human cells, imports prokaryotic functions. Present-day undergraduates are taught that CRISPR-*Cas* is synonymous to gene engineering, which leaves little chance for the development of reverse splicing for therapeutic gene repair. Therefore, it is necessary to say a few words here about CRISPR-*Cas* limitations and the reverse splicing potential.

4.1. CRISPR-Cas Inadequacy for Therapy

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), coupled with their *Cas CRISPR-associated* endonucleases, evolved in bacteria and archaea to destroy foreign genomes. The connection with antiviral defence was originally proved by the agricultural and yoghurt industry research aimed at *Streptococcus thermophilus* culture optimisation (Bolotin et al., 2005; Barrangou et al., 2007). These 'adaptive immunity' systems acquire and store copies of foreign DNA in CRISPR

 doi:10.20944/preprints202609.1663.v1

arrays (adaptation), that is transcribed and processed into crRNAs (expression) and used to guide *Cas* nucleases to cleave memorised targets (interference). However, CRISPR-*Cas* is just one component in the ensemble of prokaryotic defences – there are more than a 100 distinct strategic systems now discovered, and the mechanisms of action for many of them are still unknown (Olijslager et al., 2024; Millman et al., 2022; Doron et al., 2018). These defence systems coordinate their pathways to either inactivate the invader or commit suicide, that is eliminate the infected cell before the phage particles emerge and spread the damage in a vast bacterial population. The most prevalent defences targeting foreign DNA are restriction-modification systems (Anton et al., 2025) and CRISPR-*Cas*, along with less prevalent prokaryotic *Argonaut* systems (p*Ago*s). Suicidal pathways are otherwise known as abortive infection (*Abi*) including classical toxin-antitoxin (T-A) *Abi* (Lopatina et al., 2020), CBASS cyclic oligoadenylate-based antiphage signalling system, retrons and many more others (Olijslager et al., 2024; Millman et al., 2020). The genes for various defence systems are clustered together in 'defence islands' on the prokaryotic chromosome and encoded by MGEs, which facilitates continuous horizontal transfer (Makarova et al., 2011; Doron et al., 2018). It is important to understand that prokaryotic streamlined genomes cannot carry all possible defence repertoires owing to their burden on fitness. In fact, *"the 'effective' immune system is not the one encoded by the genome of a single microorganism but rather by its pan-genome, comprising the sum of all immune systems available for a microorganism to horizontally acquire and use"* (pan-immune system model Bernheim and Sorek, 2020). Although the defence systems are traditionally divided into the two categories that either destroy the foreign DNA or trigger cell death, the CRISPR-*Cas* systems evolved by combination of these two functions. Responsible for specific recognition and activation control of nucleases, the adaptation module comprising CRISPRs and *Cas1,2,(4)* evolved from caspasons and their associated *Cas1*-like integrase (Koonin and Makarova, 2019). CRISPRs are linked with a large variety of effectors that include nucleases of diverse origins (Makarova et al., 2025). Class 1 effectors are multiprotein modules descended from cyclase-polymerase CBASS-like systems that in the simplest form use cyclic oligoadenylate 2',3'-c-diAMP to trigger cell death, often by indiscriminate RNase with *HEPN*[1] domain. Class 2 effectors, which are single multidomain proteins (nucleases used in gene engineering), are also related to cell destruction pathways. *Cas9* and *Cas12* evolved from RNA-guided ω-endonucleases[2] *IscB* and *TnpBs* of the IS200/IS605 transposons. These nucleases are likely to have a regulatory function, and one of the proposed roles is that they are toxins, to which transposon is the antitoxin. *'Such toxins would kill cells lacking the inserted IS200/605 locus, thus making the host addicted to the IS200/605 transposon'* (Altae-Tran et al., 2021, extended discussion). The third Class 2 effector *Cas13* contains two *HEPN* domains. *Cas13* evolved from T-A *AbiF/D* systems (Zilberzwige-Tal et al., 2025; Yoon et al., 2024). *AbiF* is a non-specific toxin RNA nuclease, that can be blocked by a ncRNA antitoxin. Hardly surprising is that modern CRISPR-*Cas* systems switch between destruction of foreign DNA/RNA and self-destruction functions. Early on, it was apparent that *Cas13* activation by the specific RNA target and cleavage of this target leads to non-specific, otherwise termed 'collateral' or '*trans*', destruction of host cell RNAs leading to metabolic arrest, dormancy and cell death (East-Seletsky et al., 2016). Then Sundaresan et al., 2017 indicated that the other two class 2 effectors *Cas9* and *Cas12a*, when purified, indiscriminately nick dsDNA and degrade ssDNA in the absence of guide RNA and in the presence of Mn^{2+} and ATP. Both HNH domain, which usually cuts the target strand complementary to gRNA and RuvC domain, responsible for cutting the opposite strand, are involved in the non-specific activity. By 2023 Dmytrenko et al. and Bravo et al. established that Class 2 DNA-targeting nuclease *Cas12a2*, after specific activation mediated by guide RNA, cleaves its DNA target and subsequently drives host cell self-destruction by non-specific cleavage of ssRNA, ssDNA and dsDNA. Finally, Chen et al., 2025a discovered that *Cas9* is not an exception, as it also proceeds to

[1] *HEPN* - Higher Eukaryotes and Prokaryotes Nucleotide-binding. *HEPN* RNases are found in all domains of Life. Catalytic centre requires *cis*- or *trans*- dimerization. Functions are diverse: apart from antiviral defence in prokaryotes, in eukaryotes *HEPN* RNases are involved in ribosome biogenesis, RNA processing and decay, and are implicated in neurological diseases and cancer (Pillon et al., 2021).

[2] w-nucleases complexed with wRNA, OMEGA - obligate mobile element-guided activity

 doi:10.20944/preprints202609.1663.v1

cleave non-specific ssDNA and ssRNA targets after activation by specific DNA target cleavage. Although these reports are worrying for therapeutic applications, they led to new developments in diagnostics: a tiny amount of a specific target recognised by CRISPR-*Cas9*, *12* or *13* can be detected due to subsequent non-specific activity on reporter substrates (Chen et al., 2025b; Gootenberg et al., 2017). The fact that specific activation achieves only a very limited control over the lethal activity of *Cas* effectors might explain why RNA guided endonucleases are not easy to find in eukaryotes. Indeed, it was not until 2013, that Bao & Jurka discovered homologues of bacterial *TnpB_IS605* ω-endonuclease (the ancestor of *Cas12* as later designated in Altae-Tran et al., 2021, see above), which they termed *Fanzors* in giant eukaryotic dsDNA viruses and transposons of some diverse eukaryotes: monocellular algae, *Amoebozoa* inclusive of slime moulds, stramenopiles, fungi inclusive of *S. cerevisiae*, and oddly, in two species of metazoans, *Hydra magnipapillata*, and a fly, *Mayetiola destructor* - a pest of cereal crops. A recent study (Saito et al., 2023) used structural mining of an AlphaFold database for *Fanzor* orthologues and found a total of 649 proteins in eukaryotes and 80 in eukaryotic viruses. In metazoans, apart from some arthropods, *Fanzors* were discovered in some molluscs, for example, the clam *Mercenaria mercenaria*. Eukaryotes possess two separate clades of *Fanzors*, designated *Fz1* and *Fz2*, which originate from different *TnpB* ancestors. In addition, there are some single *Fz* that seem to have originated independently from other *TnpBs*. This is very much like the multiple origins of prokaryotic *Cas12s*: at least 50 independent events were observed linked to distinct *TnpBs* (Altae-Tran et al., 2023). The likely evolutionary scenario involves horizontal transfer of *TnpBs* from prokaryotes to eukaryotic hosts or bacterivores, and further horizontal transfers of *Fz* among eukaryotes, facilitated by viruses. Saito et al., 2023 show that, like *TnpBs*, *Fanzors* are nucleases associated with ωRNA and capable of specific dsDNA cleavage *in vitro* without *'any robust cleavage activity on collateral dsDNA, ssDNA, dsRNA or ssRNA substrate'*. Moreover, the authors optimised *Fz1* from the fungus *Spizellomyces punctatus* to target 12 human genomic loci in HEK293T cells, achieving 8-18% indel activity. Nevertheless, the unknown role of *Fanzors* in their native hosts clearly have not evolved into an indispensable function for eukaryotes. Their prevalence amongst eukaryotes is limited and possibly indicates that chance transfers from prokaryotes prompted an evolution of suppression mechanisms to prevent toxicity to eukaryotic hosts.

An interesting parallel between CRISPR-*Cas* and *Argonauts* was discussed by Chen et al., 2023 with respect to both these systems switching from specific invader inactivation to induction of self-destruction. Remarkably, the very different evolutionary trajectories of these two systems in prokaryotes and eukaryotes have a direct impact on therapeutics (Figure 13). CRISPR-*Cas* are part of the 'immune' system in ~40% of all examined bacteria and archaea species (the prevalence varies depending on a habitat – Olijslager et al., 2024), their eukaryotic cousins *Fanzors*, with an unknown biological role are only sighted in a few hundred genomes, predominantly of monocellular organisms. In contrast, *Argonauts* are generally less prevalent in prokaryotes than CRISPR-*Cas*: *Ago*s are found in ~10% of bacterial and ~30% of archaeal genomes. The roles of *pAgo*s are diverse, they can either specifically cleave foreign DNA or RNA or trigger abortive infection (Chen et al., 2023). However, *Ago*s were first discovered in eukaryotes, where apart from antiviral defence they assumed a new critical regulatory role: gene silencing by RNA interference. Humans possess an *AGO* superfamily (reviewed in Pantazopoulou et al., 2020), *AGO2* is the main slicer-nuclease of the RNA-Induced Silencing Complex *RISC* for the host gene expression control, and *AGO4* is involved in antiviral defence in mammalian immune cells by targeting viral RNA and boosting *IFN* (Adiliaghdam et al., 2020). Therapeutic applications of RNAi are now a reality, because the mechanism is native to human cells. On the contrary, using CRISPR-*Cas* or ωRNA-*Fz* for therapy implies importing alien agents, which human cells cannot control. Major efforts and resources have been mounted to study and prevent *Abi*-style non-specific DNA cleavage activity by *Cas* after specific activation (Montagud-Martínez et al., 2026), converting *Cas* nucleases to nickases to avoid particularly deleterious DSBs (Cong et al., 2013) and increase specificity by using 2x guide RNA (Ran et al., 2013), engineering specific base-editor and prime-editor enzymes by fusing to mutant *Cas* (Gaudelli et al., 2017; Chen and Liu, 2023), inventing delivery methods for *Cas* mRNA or pre-

 doi:10.20944/preprints202609.1663.v1

assembled RNP to prevent dangerous long-term expression of CRISPR-*Cas* (see next section), and finally, attenuating human immune response to bacterial *Cas* proteins (Raghavan et al., 2025). Nevertheless, are there alternatives to CRISPR-*Cas* for gene engineering based on human cell biology? Reverse splicing can be one such approach worth exploring (Figure 13).

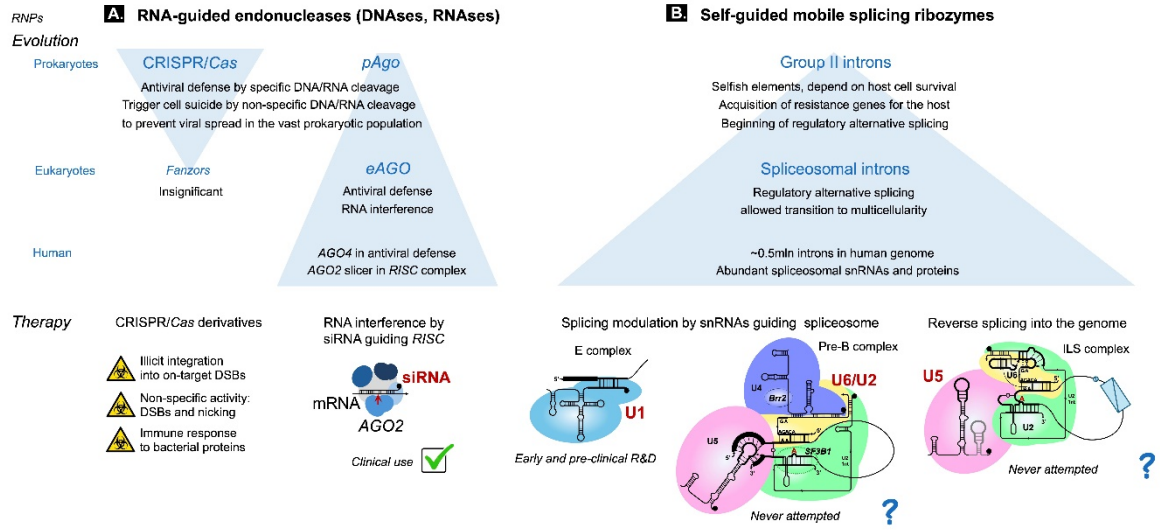

Figure 13. Evolutionary fate and therapeutic utility of RNA-protein complexes. RNPs take time to evolve, and we share many RNPs for basic cellular functions with bacteria; ribosome is one of them. However, not all prokaryotic RNPs were adopted by eukaryotes, while on the contrary other RNPs acquired additional complexity and prominent functions in complex organisms. Hence, evolutionary trajectories of RNPs have a direct impact on their utility for gene therapy. **A.** In recent decades RNA-guided endonucleases of bacterial 'immune' system became central in molecular biology. Bacterial CRISPR/*Cas* made gene targeting efficient in mammalian cells for functional genomics studies. However, using CRISPRs for therapy is problematic: human cells cannot control imported alien functions and bacterial endonucleases, or their modified derivatives are foreign for the human immune system. The only CRISPR/*Cas* analogues in eukaryotes, *Fanzors* are confined to some monocellular species and appear in rare molluscs and arthropods, likely acquired by horizontal transfer from bacterial endosymbionts or food. In contrast, argonauts (*Ago*), another RNA-guided enzyme of prokaryotic 'immune' system targeting viral RNA, became more prominent in eukaryotes and humans possess an *AGO* superfamily, with *AGO2* dedicated to transcription control by RNA interference. Using RNAi for therapy is now a reality. **B.** Human genes are replete with introns which are, in fact, functional splicing ribozymes. Abundantly expressed snRNAs U2, U5 and U6 assemble with every intron into an active ribozyme at the core of a large RNP, the spliceosome. U1 snRNP, which is not part of the ribozyme, evolved for selection of alternative splice sites. U1, U2, U6 and U5 snRNAs together guide the modular assembly of the ribozyme, which makes them the best candidates for splicing modulation. Moreover, splicing can be adapted for genome alterations. Human intron ribozymes are the decedents of prokaryotic mobile ribozymes (termed Group II introns) associated with a single protein *IEP*, the precursor of the largest spliceosomal protein *Prp8*. Group II intron RNPs are capable to invade new genomic loci by reverse splicing and are for decades used for genome engineering in microbial industry. While in prokaryotes Group II introns are rare, because populations are large and any superfluous additions are purged out of the genome, in eukaryogenesis a Group IIA intron escaped selective pressure and spread by reverse splicing. However, now Group II introns are inactive in eukaryotic cells because of low Mg^{2+}, which is not a problem for the spliceosome. We know that spliceosome can still catalyse reverse splicing. U5 snRNA is homologous to Group IIA intron subdomain responsible for target recognition that triggers reverse splicing. However, the recognition regions of U5 got curtailed to suppress intron mobility. Directed evolution can be tried to bring back reverse splicing catalysed by the human spliceosome. **DSB**, DNA double strand break; ***RISC***, RNA-induced silencing complex; **ILS**, Intron Lariat Spliceosome.

 doi:10.20944/preprints202609.1663.v1

4.2. Transient Delivery of RNA for Genome Engineering: eVLP

Enveloped virus particles eVLP are empty retrovirus capsids enveloped by the cellular membrane. Producer cells engineered to express viral proteins generate VLPs that do not carry virus genetic material and can be loaded instead with the therapeutic RNA or RNP. These vehicles are ideal for nuclear delivery: the inner capsid derived from HIV1 is a regular conical structure with a base diameter of 60nm that evolved to pass through 64nm wide nuclear pores (Zila et al., 2021 – EM images). The recent development of eVLPs is prompted by the idea that transient delivery of CRISPR-*Cas*-derived genome editing agents can reduce associated genotoxic events (reviewed in Lyu and Lu, 2022 and Lyu et al., 2020). Yadav et al., 2022 designed all-in-one eVLPs that carry sgRNA cassette with the lentivirus Ψ packaging signal instead of the RNA genome attached to nucleocapsid *NC*, and *Cas9* mRNA fixed by bacteriophage Mu-derived com aptamer to the *Com* protein fused with *NC*. Hamilton et al., 2024 developed enveloped delivery vehicles that carry a pre-assembled RNP - *Cas9* in complex with sgRNA. *Cas9* was fused to the 3′ end of HIV-1 *Gag* capsid gene via a linker that is cleaved by HIV-1 protease during capsid maturation. An alternative to fusion strategy is the use of bacteriophage aptamers attached to the RNA component of the pre-assembled RNP, which equally ensures the RNP inclusion into the enveloped vehicles (reviewed in Lyu and Lu, 2022). Lentivirus-like-particles are naturally enveloped by cellular membrane at the initial budding stage. This envelope is commonly pseudotyped by engineering producer cells to express VSVGmut (vesicular stomatitis virus glycoprotein mutant), which cannot bind its natural target *LDLR* (the low density lipoprotein receptor on the surface of many cell types) but retains its endosomal escape function. Specific cell targeting of the envelope can be achieved by co-expressing a single chain antibody variable fragment (scFv) or fusing VSVGmut with scFv. Hamilton et al., 2024 co-expressed VSVGmut and scFv targeting T-cells, while Gao and Bergman, 2023 used *Her2*-specific scFv fused with VSVGmut targeting breast cancer cells.

A breakthrough discovery by Segel et al., 2021 shows that eVLPs can be produced using human endogenous retroelements, rather than HIV-1 commonly employed for this purpose (see also comment Riecken et al., 2021). Indeed, the mammalian endogenous Paternally Expressed Gene 10 ***PEG10*** (7q21) protein (imprinted gene involved in placenta formation) is a homologue of retroviral *Gag*, expressed by an LTR retrotransposon. *PEG10* binds many cellular RNAs in trophoblast stem cells and is known to promote vesicle budding like HIV-1 Capsid *CA* protein - *Gag* p24 (Abed et al., 2019). *PEG10* also binds its own RNA at 5′ and 3′ UTR regions, forming a capsid around it, which is then enveloped producing a VLP and can be transferred to other cells. This RNA can be substituted for a therapeutic RNA, and the envelope can be pseudotyped with fusogens to induce transduction. Segel et al., 2021 transferred *Cas9* mRNA and sgRNA to HEK293 cells using *PEG10* eVLPs pseudotyped with VSVG.

These methods developed for transient expression of CRISPR-*Cas* systems, in particular RNA transfer to the nucleus, will be equally adaptable for reverse splicing (see next section). Furthermore, pseudotyping for muscle can be achieved using *Mymk* and *Mymg* fusogens specific to activated muscle precursor cells (Hindi et al., 2024, Figure 14). In humans, skeletal muscle damage associated with DMD is greater than in mouse models; active muscle regeneration is a lot more widespread and it is reasonable to expect that RNA transfer to muscle using regeneration-specific fusogens will be greater in humans than observed in mice.

 doi:10.20944/preprints202609.1663.v1

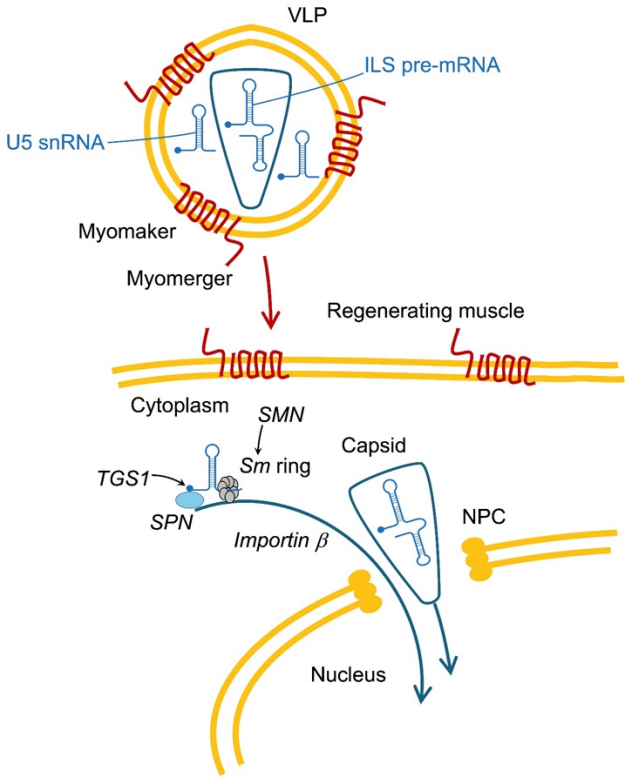

Figure 14. Virus-like particles (VLP) delivery of RNA to regenerating muscle. VLPs are empty capsids derived from lentivirus (Hamilton et al., 2024) or human domesticated retrovirus element *PEG10* (Segel et al., 2021) which can package RNA expressed in producer cells. Activated muscle precursor cells and damaged fibres switch on specific fusogens Myomaker *Mymk* and Myomerger *Mymg*, that mediate muscle regeneration by cell fusion. VLPs are enveloped with cell membrane and if produced in cells that express *Mymk* and *Mymg* will specifically deliver their cargo to the cytoplasm of regenerating muscle (Hindi et al., 2024). The inner capsid is shaped as a cone with a base diameter just fitting the nuclear pore (HIV1 capsid base Ø 60nm, pore Ø 64nm, Zila et al., 2021). RNA can be engineered to be packaged inside the capsid. It is essential to deliver pre-mRNA to the nucleus to produce ILS by splicing. snRNA can be delivered to the cytoplasm and will be picked up by the processing pathway and subsequently imported to the nucleus. Alternatively, if snRNA is included in the capsids, it will be exported to the cytoplasm and re-imported back. **ILS** – Intron Lariat Spliceosome complex; ***SMN*** – survival motor neuron complex adorns snRNAs with Sm proteins; ***TGS1*** – Trimethylguanosine Synthase converts the m7G cap (m7G-5′-ppp-5′) acquired in the nucleus when transcribed by Pol II into 2,2,7m3G or TMG cap (2,2,7m3G-5′-ppp-5′); ***SPN*** -Snurportin, a protein adaptor that binds the TMG cap and interacts with importin β; **NPC** -nuclear pore complex.

4.3. Human ILS for Gene Repair: Concept and New Research Avenues

The eukaryotic intron progenitor evolved for specific genomic insertions. At the same time, host survival was paramount to the persistence of the intron. The primary catalytic activity is that of the self-splicing ribozyme; no endonucleases causing deleterious DSBs are involved. The spliceosome emerged largely by neutral evolution, which provided a pool for the selection of adaptative advantages, such as the alternative splicing mechanism that combines precision with variation, necessary for cell differentiation in multicellular organisms. Indeed, splicing is the more prevalent the more complex is the organism - perhaps the peak of splicing complexity can be attributed to the human brain. The human spliceosome is by far 'fitter' than that of the well-studied monocellular model *S. cerevisiae*, as a phenomenal number of introns is spliced out in a precisely regulated manner in human cells. Admittedly, suppressed intron mobility is also a feature of the modern spliceosome, as eukaryotic U5 snRNA has only one recognition loop and not two like mobile bacterial Group IIA introns (Figure 2 A,B in *Part 1* - Artemyeva-Isman, 2026a). In fact, the structure of the ribozyme and the mechanism of catalysis are unchanged, but reverse splicing needs to be initiated by U5. Increasing

 doi:10.20944/preprints202609.1663.v1

U5 Loop1 affinity to the target and engineering additional loop to extend the target recognition is the path to explore reverse splicing. We can expect high precision from a system that evolved to avoid the damage to the host cell and became integral to eukaryotic gene regulation (see *Part 1*).

To explain deletion repair, it will be easier to refer to the example of Duchenne muscular dystrophy (DMD), a sporadic X-linked progressive muscle wasting disease affecting 1:3,500 newborn boys. 60-80% of cases are caused by deletions of variable size that frequently originate within this chromosomal region of genomic instability, a common fragile site in human populations worldwide. The aim is to insert a missing exon, or multiple exons, between the exons that flank the deletion. DNA breaks that cause deletions usually occur in introns for a simple reason that the coding sequence of the dystrophin gene amounts for only ~0.6% of its DNA, as the 79 exons are dotted in the giant 2Mb locus. There are two deletion/duplication 'hotspots': breaks occur most frequently in introns 42-52 and intron 7. Dystrophin gene deletions of several exons and introns can be large enough to be visible on chromosome spreads. The proposed repair will not restore the in-between introns when multiple exons are deleted but will restore the coding sequence of the gene by inserting these exons fused together. Although individual deletions have unique breakpoints within the introns, a common ILS vector can repair deletions of the same exon(s).

The intron lariat spliceosome (ILS) vector can be produced by splicing in the target cells themselves if the ILS minigene and the engineered U5 gene are delivered together by AAV. Single AAV packaging capacity is by far sufficient even if we add the genes for the engineered U6 and U2. AAVs pseudotyped with muscle-specific peptides in the exposed variable regions of the capsid proteins (Figure 8 in *Part 3* - Artemyeva-Isman, 2026c; Tabebordbar et al., 2021; Weinmann et al., 2020; El Andari et al., 2022) are optimal for delivery to muscle and ensure reduced liver toxicity. Using a muscle-specific promoter for the ILS minigene can further warrant efficient expression in muscle and abolish the ectopic expression (Darbey et al., 2024). However, reflecting on the delivery options, if we are targeting the DMD transcript, a long-term expression from an AAV is desired; but if targeting the DMD genomic locus, only a short-term expression is the goal (see explained below and Figure 11D-F). Controlled inducible expression of an ILS minigene is possible, as reported for tetracycline transactivator, when expression of a transgene is induced by tetracyclin and abolished in less than 3 days in mice after the antibiotic is discontinued (Chtarto et al., 2003). A better alternative for the transient ILS expression is to avoid delivering the ILS minigene and the U5 gene altogether and instead deliver ILS pre-mRNA and U5 snRNA, taking advantage of the recent developments for *Cas* mRNA and gRNA delivery, such as lentiviral-like particles (Hamilton et al., 2024) and especially the mammalian *PEG10* retrovirus-like VLPs (Segel et al., 2021), both of which deliver RNA in a capsid to the nucleus. Their outer membrane envelope can be pseudotyped by *Mymk* and *Myng* to target specifically regenerating muscle (Hindi et al., 2023, Figure 14, see previous section). Delivery of an ILS RNP itself might be impractical, as it is much larger than a *Cas*-gRNA RNP and might be tricky to package into VLPs.

Engineering U5 with an additional recognition loop is our principal task, which can be tackled by modifying *In vitro* Selection, a technique that shares its principles with Systematic Evolution of Ligands by Exponential Enrichment (SELEX) – the former is applied to the selection of ribozymes, the latter is widely used to produce specific RNA or DNA aptamers (Higgs and Muller, 2025). Indeed, re-creating ancient ribozyme functions in a laboratory is an advancing field of study inspired by the hypothesis of the ribozyme and peptide world as a precondition for the development of life on Earth (Attwater et al., 2025; Li et al., 2022; Scheitl et al., 2020). *In vitro* ribozyme selection techniques work very well, but the assembly of U5 with the other RNA and protein components of the spliceosome requires transfections of human cells, although reverse splicing reactions can be set up in nuclear extracts and cloning can be avoided, as opposed to the laborious *in vivo* selection procedures.

ILS can be directed to three different targets: mRNA, pre-mRNA and genomic DNA (Figure 11D-F); each of these pathways need to be explored. Inserting into mRNA (Figure 11D) is straightforward to explain: the intron lariat containing a missing exon is spliced in between the exons, flanking the deletion. In this case the engineered U5 snRNA targets the junction of these flanking exons. We assume that another round of splicing will occur to remove the introns around the exon

 doi:10.20944/preprints202609.1663.v1

we are inserting. If targeting the pre-mRNA (Figure 11E), we are sure to have the introns spliced out after the lariat is spliced in, and we have a free choice of a target site inside the intron with the deletion. Lariat insertion results in an intron within an intron and we need to consider splice site competition and prevent the exclusion of the exon we aim to insert. For example, lariat 5′ and 3′ss can be suboptimal, but this will downregulate the ILS production. Alternatively, we can produce an ILS with U6 and U2 engineered to pair intron sequences of the ILS distinct from WT splice sites. A permanent solution would be to insert the deleted exon(s) into the genomic locus. In this case we are targeting either the coding or the template DNA strands. ILS vector that targets pre-mRNA also targets the coding strand. On the contrary, lariat insertion into the template strand (Figure 11F) does not create an intron within an intron, avoiding splice site competition. There can be a difference in accessibility between the coding and the template strands, and it may vary depending on a locus.

Importantly, RNA reverse splicing is initiated and completed by the ribozyme itself. However, there remain theoretical questions: how the second DNA strand will be synthesized? How the inserted RNA can be changed for DNA? And do we require cellular replication for this? When attacking a genomic site, Group II introns use endonuclease *EN* activity of their *IEP* for the second DNA strand cleavage opposite to the insertion site. This primes second strand DNA synthesis by *RT* domain of the *IEP* on the RNA template of the insertion. Cellular repair functions then change the RNA insert to DNA. However, this is by no means the only integration pathway: it depends both on Group II intron and the host cell. Indeed, when inserting into the intronless allele of its home *LtrB* gene (‘retrohoming’) in its host bacterium *Lactococcus lactis, Ll.LtrB* Group IIA intron inserts into the double stranded DNA target following the pathway described above. By contrast, *Ll.LtrB* intron prefers to insert into ssDNA target when attacking new genomic loci (retrotransposition) in its native host, but, curiously, when switched to *E. coli* dsDNA is still preferred for retrotransposition, indicting the influence of the host cell factors. Another well-studied example is the Group IIB *RmInt1* in bacterium *Sinorhizobium meliloti*, a nitrogen-fixing symbiont of the forage crop lucerne (alfalfa) that is *EN*-deficient and actively mobile. When retrohoming, *RmInt1* inserts into single stranded DNA template for the lagging strand at replication forks and employs nascent lagging strand primers (Okazaki fragments) for the second strand DNA synthesis. Curiously, the replication-dependent mobility of *EN*-deficient mutants of *Ll.LtrB* intron has an opposite strand bias: the intron reverse splices into dsDNA to appear in the leading strand template for the approaching replication fork and copied by the leading strand synthesis (Lambowitz and Belfort, 2015; Beauregard et al., 2008; Martínez-Abarca et al., 2004; Zhong and Lambowitz, 2003). We conclude that *IEP EN* function is dispensable and can be compensated by priming at the replication fork. What about *RT* function? Group II introns in yeast mitochondria manage to sustain their mobility at 40% if *IEP RT* domain is inactivated by mutations (Moran et al., 1995; Eskes et al., 1997, 2000). This shows that a eukaryotic host cell can supply factors to complete integration without the intron-associated *RT* function. The mechanism of this process can be only inferred and was suggested to rely on double-strand break repair DSBR (Eskes et al., 1997; Dai and Zimmerly, 2002). Indeed, this mechanism was earlier attributed to Group I intron mobility, where the intron-encoded meganuclease specifically recognises the dsDNA target and introduces a DSB. Repair of this DSB leads to conversion of the intronless allele into the allele with the intron, accompanied with co-conversion of 5′ and 3′ exon markers (Belfort et al., 1995; Hedberg and Johansen, 2013). However, co-conversion was observed in the majority, but not in all integration events for Group II intron *RT*-deficient mutant in yeast mitochondria (Eskes et al., 1997, 2000). Moreover, as opposed to DSB-generating Group I associated meganucleases, Group II intron *IEP EN* only nicks the strand opposite to the intron insertion site. There is also a possibility, that an *RT* function to assist intron integration can be supplied by the host cell[3]. Indeed, Group II

[3] Ty are active retrotransposons in *S. cerevisiae,* that constitute 3.5% of its nuclear genome. Recently it became apparent that Ty behaviour is influenced by mitochondria. Curiously, Ty mobility in yeast hybrids depends on the provenance of the inherited mitochondria, suggesting the participation of hitherto unknown mitochondria-encoded factors (Smukowski et al., 2021). On the other hand, mature Ty virus-like particles (VLPs) accumulate with protein aggregates around mitochondria (Schneider et al., 2024).

 doi:10.20944/preprints202609.1663.v1

introns always rely on host functions to complete the integration. There is a marked dependence on the type of host: *EN*-deficient introns are mobile in bacteria, but not in yeast mitochondria, while it is the opposite for *RT*-deficient introns (Dai & Zimmerly, 2002). Could it be because eukaryotic genomes are replete with retroelements that can supply the *RT* function? Fine tuning of the intron to the host was mentioned before: mobility of *Ll.LtrB* intron with fully competent *IEP* changes the preference between ssDNA and dsDNA insertion sites if transferred from the native *L. lactis* to *E. coli* (Beauregard et al., 2008). What we know for sure is that direct reverse splicing into DNA by Group II introns is their main mechanism when invading new loci (Dickson et al., 2001). It is impossible to predict what a human cell will do if an ectopic reverse splicing takes place. Assuming human *Prp8 EN* and *RT* domains are both deficient in their original activity and considering that Group II introns compensate for both *EN* and *RT* deficiency borrowing functions from the host cells to complete genomic integration after the reverse splicing, we need to see what happens in human cells if U5 snRNA can be engineered to bring back the prolific reverse splicing activity. There are DNA polymerases that have *RT* activity, in particular *Pol η*, which is involved in DNA repair and emerged as a major reverse transcriptase in human cells (Su et al., 2019). Finally, it is uncertain if cell proliferation is absolutely required, as ssDNA can be targeted in transcribed genes. A nick of the second DNA strand can be provided by a topoisomerase that relives supercoiling tension and integration completed by transcription-associated DNA repair. In fact, both *TOP1* and *TOP2B* activities are necessary for transcription of long genes, and while the former acts as nickase, the latter generates DSBs (reviewed in Pommier et al., 2016, 2022; King et al., 2013– topoisomerase in transcription of long genes linked to autism). What is certain is that exploring reverse splicing pathways into RNA and DNA targets in human cells is not only of theoretical interest, but of practical importance. Returning to the example of DMD deletion repair we reflect on the importance of muscle tissue composition in the choice of target cells. Mature muscle fibres are multinucleated quiescent syncytia, that actively transcribe dystrophin gene. Disease-associated muscle damage stimulates proliferation of muscle precursor cells, that do not transcribe dystrophin until a later differentiation stage, just before they fuse with the muscle fibres in need of repair. Fusogens *Mymk* and *Mymg* (Hindi et al., 2024; see above, Figure 14) ensure the delivery to activated muscle progenitor cells and the damaged myofibres, which enables targeting both replication in proliferating progenitors and transcription in myofibres.

5. Discussion

Stepwise modular assembly of spliceosomal ribozymes allows for targeting early splice site choices, pre-catalytic complexes or catalytic spliceosomes. Early complexes involve U1 and U2 snRNAs. Initially, U1 snRNA selects the 5′ss. The bulk of the existing experimental data was produced for U1 snRNA (reviewed in *Part 3* - Artemyeva-Isman, 2026c). U1 appears to be safe and can be readily delivered by rAAVs in mice. U2 snRNA is involved in branchpoint and 3′ss selection. Crucially, U2 snRNP interacts with U1 snRNP to match 5′ and 3′ss at an early stage (complex A – Figure 5, see *Part 2* - Artemyeva-Isman, 2026b). Yet U2 was not tried for therapy since a pioneering mutation study in 1989. In the pre-catalytic spliceosome the 5′ end of U2 pairs with the 3′ end of U6, forming U2/U6 Helix II. U6 displaces U1 at the 5′ss and U5 binds the exons. The pre-catalytic complex is the checkpoint for splicing fidelity: the formation of the catalytic metal binding site is delayed by U4 snRNA, while the recognition is completed by U2/U6 and U5 snRNAs. 5′ and 3′ss vaguely selected by the early complexes are double checked before splicing commitment. This enhanced precision makes the pre-catalytic complex an attractive target. It is important to understand that target recognition depends not only on the extent of complementarity but is conducted by specific 3D folding: 5′ss and BP helices are integral to the ribozyme, as opposed to U1/5′ss Helix. The limitations of U1 technology – variable efficacy and cryptic 5′ss activation – can be addressed by adapting U2 and U6 together to make perfect 5′ss and BP Helices and further altering U2/U6 Helix II, creating a spliceosome species specific for a target intron. U6 snRNA was used in three studies in combination with U1, but the U6 binding register was not consulted – the present review is intended to assist U6 and U2 design (Figure 12, and see *Parts 1* and *2* - Artemyeva-Isman, 2026a,b, Figures 2A and 6). U5 completes the

 doi:10.20944/preprints202609.1663.v1

target recognition in the pre-catalytic spliceosome; it can be adapted to match the target exon junction, but it cannot be attached to the adapted U6/U2 duplex, as U5 is joined with U6/U2 by protein interactions. However, U5 is central if we are interested in reverse splicing by catalytic spliceosomes (see below).

The real purpose of this series of papers is to point out that U5 snRNA has a potential for gene therapy that rivals CRISPR-*Cas*. Eukaryotic spliceosome conserves reversible catalysis of the ancestral mobile ribozyme, which was capable of specific integration into genomic DNA by reverse splicing. The mechanism of intron mobility is well studied in related modern bacterial Group IIA introns. Reverse splicing is triggered by the recognition loops of the Group IIA intron ribozyme binding a target sequence, similar to the junction of its exons. We have previously suggested that this similarity is defined by base-pair geometry: mismatches that can assume the shape of Watson-Crick pairs by tautomerisation of one of the bases are preferred to the mismatches that perturb the helix architecture (see *Parts 1* and *2* - Artemyeva-Isman, 2026a,b). We also hypothesised that U5 Loop1 inherited this mechanism for the recognition of exons. Here I propose that engineering U5 with strong affinity of Loop 1 to the target and an additional loop to extend this interaction, as in ancestral mobile intron, is the way to bring back reverse splicing. Pursuing reverse splicing as a research aim makes sense if we consider that the evolutionary fate of cellular functions defines their utility for therapy. The prokaryotic CRISPR-*Cas* complex never developed into a major eukaryotic agent, contrary to prokaryotic Argonauts *pAgo*, which became more prominent in eukaryotes. The human *AGO2* slicer in the *RISC* complex makes RNAi a valid therapeutic approach, as opposed to imported bacterial CRISPR-*Cas* derivatives that will require adjustments against collateral DNA/RNA cleavage activities and the immune response to bacterial proteins. Introns assembled with snRNAs are active ribozymes abundant in higher eukaryotes, contrary to their suppressed prokaryotic Group II intron ancestors. Yet for some decades Group II introns have been industrial gene engineering tools in bacteria. The human intron lariat spliceosome guided by adapted U5 must be developed to repair deletions in the dystrophin gene as a better therapy for Duchenne muscular dystrophy.

Acknowledgements: I could not have prepared this series of papers without the continual encouragement from Dr Rimma Belotserkovskaya (ICR, Cambridge). I thank Alexandra Isman for the careful editing and corrections of all the four manuscripts.

Conflicts of Interest: The author declares no conflicts of interest.

References

Abed, M., Verschueren, E., Budayeva, H., Liu, P., Kirkpatrick, D. S., Reja, R., Kummerfeld, S. K., Webster, J. D., Gierke, S., Reichelt, M., Anderson, K. R., Newman, R. J., Roose-Girma, M., Modrusan, Z., Pektas, H., Maltepe, E., Newton, K., & Dixit, V. M. (2019). The Gag protein PEG10 binds to RNA and regulates trophoblast stem cell lineage specification. *PloS one*, *14*(4), e0214110. https://doi.org/10.1371/journal.pone.0214110

Adiliaghdam, F., Basavappa, M., Saunders, T. L., Harjanto, D., Prior, J. T., Cronkite, D. A., Papavasiliou, N., & Jeffrey, K. L. (2020). A Requirement for Argonaute 4 in Mammalian Antiviral Defense. *Cell reports*, *30*(6), 1690–1701.e4. https://doi.org/10.1016/j.celrep.2020.01.021

Altae-Tran, H., Kannan, S., Demircioglu, F. E., Oshiro, R., Nety, S. P., McKay, L. J., Dlakić, M., Inskeep, W. P., Makarova, K. S., Macrae, R. K., Koonin, E. V., & Zhang, F. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science (New York, N.Y.)*, *374*(6563), 57–65. https://doi.org/10.1126/science.abj6856

Altae-Tran, H., Kannan, S., Suberski, A. J., Mears, K. S., Demircioglu, F. E., Moeller, L., Kocalar, S., Oshiro, R., Makarova, K. S., Macrae, R. K., Koonin, E. V., & Zhang, F. (2023). Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. *Science (New York, N.Y.)*, *382*(6673), eadi1910. https://doi.org/10.1126/science.adi1910

Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., & Lührmann, R. (2013). RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *The EMBO journal*, *32*(21), 2804–2818. https://doi.org/10.1038/emboj.2013.198

 doi:10.20944/preprints202609.1663.v1

Anton, B. P., Blumenthal, R., Eaglesham, J. B., Mruk, I., Roberts, R. J., Xu, S.-y, Weigele, P. R., & Raleigh, E. A. (2025). Biology of host-dependent restriction-modification in prokaryotes. *EcoSal Plus*, *13*(1), eesp00142022. https://doi.org/10.1128/ecosalplus.esp-0014-2022

Artemyeva-Isman, O. V. (2026a). The origins of spliceosomal ribozymes. *Preprints.org*

Artemyeva-Isman, O. V. (2026b). Spliceosomal ribozyme assembly. *Preprints.org*

Artemyeva-Isman, O. V. (2026c). U1 (and U7) snRNA for splicing therapy. *Preprints.org*

Artemyeva-Isman, O. V., & Porter, A. C. G. (2021). U5 snRNA Interactions With Exons Ensure Splicing Precision. *Frontiers in genetics*, *12*, 676971. https://doi.org/10.3389/fgene.2021.676971

Attwater, J., Augustin, T. L., Curran, J. F., Kwok, S. L. Y., Ohlendorf, L., Gianni, E., & Holliger, P. (2025). Trinucleotide substrates under pH-freeze-thaw cycles enable open-ended exponential RNA replication by a polymerase ribozyme. *Nature chemistry*, *17*(7), 1129–1137. https://doi.org/10.1038/s41557-025-01830-y

Bao, W., & Jurka, J. (2013). Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mobile DNA*, *4*(1), 12. https://doi.org/10.1186/1759-8753-4-12

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Beauregard, A., Curcio, M. J., & Belfort, M. (2008). The take and give between retrotransposable elements and their hosts. *Annual review of genetics*, *42*, 587–617. https://doi.org/10.1146/annurev.genet.42.110807.091549

Belfort, M., Reaban, M. E., Coetzee, T., & Dalgaard, J. Z. (1995). Prokaryotic introns and inteins: a panoply of form and function. *Journal of bacteriology*, *177*(14), 3897–3903. https://doi.org/10.1128/jb.177.14.3897-3903.1995

Bernheim, A., & Sorek, R. (2020). The pan-immune system of bacteria: antiviral defence as a community resource. *Nature reviews. Microbiology*, *18*(2), 113–119. https://doi.org/10.1038/s41579-019-0278-2

Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., Urlaub, H., Kastner, B., Lührmann, R., & Stark, H. (2017b). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell*, *170*(4), 701–713.e11. https://doi.org/10.1016/j.cell.2017.07.011

Biancon, G., Joshi, P., Zimmer, J. T., Hunck, T., Gao, Y., Lessard, M. D., Courchaine, E., Barentine, A. E. S., Machyna, M., Botti, V., Qin, A., Gbyli, R., Patel, A., Song, Y., Kiefer, L., Viero, G., Neuenkirchen, N., Lin, H., Bewersdorf, J., Simon, M. D., … Halene, S. (2022). Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies. *Molecular cell*, *82*(6), 1107–1122.e7. https://doi.org/10.1016/j.molcel.2022.02.025

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, *151*(Pt 8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Bravo, J. P. K., Hallmark, T., Naegle, B., Beisel, C. L., Jackson, R. N., & Taylor, D. W. (2023). RNA targeting unleashes indiscriminate nuclease activity of CRISPR-Cas12a2. *Nature*, *613*(7944), 582–587. https://doi.org/10.1038/s41586-022-05560-w

Carmel, I., Tal, S., Vig, I., & Ast, G. (2004). Comparative analysis detects dependencies among the 5' splice-site positions. *RNA (New York, N.Y.)*, *10*(5), 828–840. https://doi.org/10.1261/rna.5196404

Chen, C., Zhao, X., Kierzek, R., & Yu, Y. T. (2010). A flexible RNA backbone within the polypyrimidine tract is required for U2AF65 binding and pre-mRNA splicing in vivo. *Molecular and cellular biology*, *30*(17), 4108–4119. https://doi.org/10.1128/MCB.00531-10

Chen, J., Chen, Y., Huang, L., Lin, X., Chen, H., Xiang, W., & Liu, L. (2025a). Trans-nuclease activity of Cas9 activated by DNA or RNA target binding. *Nature biotechnology*, *43*(4), 558–568. https://doi.org/10.1038/s41587-024-02255-7

Chen, P. J., & Liu, D. R. (2023). Prime editing for precise and highly versatile genome manipulation. *Nature reviews. Genetics*, *24*(3), 161–177. https://doi.org/10.1038/s41576-022-00541-1

Chen, Y., Chen, J., & Liu, L. (2025b). The discovery of Cas9's trans-cleavage activity: Unlocking new molecular diagnostic tools. *Clinical and translational medicine*, *15*(6), e70384. https://doi.org/10.1002/ctm2.70384

Chen, Y., Zeng, Z., She, Q., & Han, W. (2023). The abortive infection functions of CRISPR-Cas and Argonaute. *Trends in microbiology*, *31*(4), 405–418. https://doi.org/10.1016/j.tim.2022.11.005

Chtarto, A., Bender, H. U., Hanemann, C. O., Kemp, T., Lehtonen, E., Levivier, M., Brotchi, J., Velu, T., & Tenenbaum, L. (2003). Tetracycline-inducible transgene expression mediated by a single AAV vector. *Gene therapy*, *10*(1), 84–94. https://doi.org/10.1038/sj.gt.3301838

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Corrionero, A., Raker, V. A., Izquierdo, J. M., & Valcárcel, J. (2011). Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease. *RNA (New York, N.Y.)*, *17*(3), 401–411. https://doi.org/10.1261/rna.2444811

Cortes, J. J., Sontheimer, E. J., Seiwert, S. D., & Steitz, J. A. (1993). Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5' splice sites in vivo. *The EMBO journal*, *12*(13), 5181–5189. https://doi.org/10.1002/j.1460-2075.1993.tb06213.x

Cretu, C., Gee, P., Liu, X., Agrawal, A., Nguyen, T. V., Ghosh, A. K., Cook, A., Jurica, M., Larsen, N. A., & Pena, V. (2021). Structural basis of intron selection by U2 snRNP in the presence of covalent inhibitors. *Nature communications*, *12*(1), 4491. https://doi.org/10.1038/s41467-021-24741-1

Crispino, J. D., & Sharp, P. A. (1995). A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes & development*, *9*(18), 2314–2323. https://doi.org/10.1101/gad.9.18.2314

Dai, L., & Zimmerly, S. (2002). Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic acids research*, *30*(5), 1091–1102. https://doi.org/10.1093/nar/30.5.1091

Darbey, A., Jin, W., Greensmith L., Sleigh, J. N., Counsell, J., Fratta, P. (2024). Refining gene delivery to skeletal muscle with a dual-strategy approach of muscle-tropic AAV capsids and muscle-specific promoters. *bioRxiv* 2024.08.02.605568; doi: https://doi.org/10.1101/2024.08.02.605568

De Angelis, F. G., Sthandier, O., Berarducci, B., Toso, S., Galluzzi, G., Ricci, E., Cossu, G., & Bozzoni, I. (2002). Chimeric snRNA molecules carrying antisense sequences against the splice junctions of exon 51 of the dystrophin pre-mRNA induce exon skipping and restoration of a dystrophin synthesis in Delta 48-50 DMD cells. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9456–9461. https://doi.org/10.1073/pnas.142302299

Dickson, L., Huang, H. R., Liu, L., Matsuura, M., Lambowitz, A. M., & Perlman, P. S. (2001). Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13207–13212. https://doi.org/10.1073/pnas.231494498

Dmytrenko, O., Neumann, G. C., Hallmark, T., Keiser, D. J., Crowley, V. M., Vialetto, E., Mougiakos, I., Wandera, K. G., Domgaard, H., Weber, J., Gaudin, T., Metcalf, J., Gray, B. N., Begemann, M. B., Jackson, R. N., & Beisel, C. L. (2023). Cas12a2 elicits abortive infection through RNA-triggered destruction of dsDNA. *Nature*, *613*(7944), 588–594. https://doi.org/10.1038/s41586-022-05559-3

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (New York, N.Y.)*, *359*(6379), eaar4120. https://doi.org/10.1126/science.aar4120

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270–273. https://doi.org/10.1038/nature19802

El Andari, J., Renaud-Gabardos, E., Tulalamba, W., Weinmann, J., Mangin, L., Pham, Q. H., Hille, S., Bennett, A., Attebi, E., Bourges, E., Leborgne, C., Guerchet, N., Fakhiri, J., Krämer, C., Wiedtke, E., McKenna, R., Guianvarc'h, L., Toueille, M., Ronzitti, G., Hebben, M., … Grimm, D. (2022). Semirational bioengineering of AAV vectors with increased potency and specificity for systemic gene therapy of muscle disorders. *Science advances*, *8*(38), eabn4704. https://doi.org/10.1126/sciadv.abn4704

Eskes, R., Liu, L., Ma, H., Chao, M. Y., Dickson, L., Lambowitz, A. M., & Perlman, P. S. (2000). Multiple homing pathways used by yeast mitochondrial group II introns. *Molecular and cellular biology*, *20*(22), 8432–8446. https://doi.org/10.1128/MCB.20.22.8432-8446.2000

Eskes, R., Yang, J., Lambowitz, A. M., & Perlman, P. S. (1997). Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell*, *88*(6), 865–874. https://doi.org/10.1016/s0092-8674(00)81932-7

Gao, Y., & Bergman, I. (2023). Vesicular Stomatitis Virus (VSV) G Glycoprotein Can Be Modified to Create a Her2/Neu-Targeted VSV That Eliminates Large Implanted Mammary Tumors. *Journal of virology*, *97*(6), e0037223. https://doi.org/10.1128/jvi.00372-23

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

Gooding, C., Clark, F., Wollerton, M. C., Grellscheid, S. N., Groom, H., & Smith, C. W. (2006). A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome biology*, *7*(1), R1. https://doi.org/10.1186/gb-2006-7-1-r1

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A., Myhrvold, C., Bhattacharyya, R. P., Livny, J., Regev, A., Koonin, E. V., Hung, D. T., Sabeti, P. C., Collins, J. J., & Zhang, F. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science (New York, N.Y.)*, *356*(6336), 438–442. https://doi.org/10.1126/science.aam9321

Griffin, E. A., Jr, Qin, Z., Michels, W. J., Jr, & Pyle, A. M. (1995). Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chemistry & biology*, *2*(11), 761–770. https://doi.org/10.1016/1074-5521(95)90104-3

Hamilton, J. R., Chen, E., Perez, B. S., Sandoval Espinoza, C. R., Kang, M. H., Trinidad, M., Ngo, W., & Doudna, J. A. (2024). In vivo human T cell engineering with enveloped delivery vehicles. *Nature biotechnology*, *42*(11), 1684–1692. https://doi.org/10.1038/s41587-023-02085-z

Hedberg, A., & Johansen, S. D. (2013). Nuclear group I introns in self-splicing and beyond. *Mobile DNA*, *4*(1), 17. https://doi.org/10.1186/1759-8753-4-17

Higgs, P. G., & Muller, U. F. (2025). Principles of *in vitro* selection of ribozymes from random sequence libraries. *Journal of the Royal Society, Interface*, *22*(225), 20240878. https://doi.org/10.1098/rsif.2024.0878

Hindi, S. M., Petrany, M. J., Greenfeld, E., Focke, L. C., Cramer, A. A. W., Whitt, M. A., Khairallah, R. J., Ward, C. W., Chamberlain, J. S., Prasad, V., Podbilewicz, B., & Millay, D. P. (2023). Enveloped viruses pseudotyped with mammalian myogenic cell fusogens target skeletal muscle for gene delivery. *Cell*, *186*(16), 3520. https://doi.org/10.1016/j.cell.2023.06.025

Hwang, D. Y., & Cohen, J. B. (1996). U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells. *Genes & development*, *10*(3), 338–350. https://doi.org/10.1101/gad.10.3.338

Irimia, M., & Roy, S. W. (2008). Evolutionary convergence on highly-conserved 3' intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *PLoS genetics*, *4*(8), e1000148. https://doi.org/10.1371/journal.pgen.1000148

Ishigami, Y., Ohira, T., Isokawa, Y., Suzuki, Y., & Suzuki, T. (2021). A single m^{6}A modification in U6 snRNA diversifies exon sequence at the 5' splice site. *Nature communications*, *12*(1), 3244. https://doi.org/10.1038/s41467-021-23457-6

Isman, O., Roberts, M. L., Morgan, J. E., Graham, I. R., Goldring, K., Lawrence-Watt, D. J., Lu, Q. L., Dunckley, M. G., Porter, A. C., Partridge, T. A., & Dickson, G. (2008). Adenovirus-based targeting in myoblasts is hampered by nonhomologous vector integration. *Human gene therapy*, *19*(10), 1000–1008. https://doi.org/10.1089/hum.2008.063

Jackson, A., Thaker, N., Blakes, A., Rice, G., Griffiths-Jones, S., Balasubramanian, M., Campbell, J., Shannon, N., Choi, J., Hong, J., Hunt, D., de Burca, A., Kim, S. Y., Kim, T., Lee, S., Redman, M., Rius, R., Simons, C., Tan, T. Y., Ellingford, J., … Banka, S. (2025). Analysis of R-loop forming regions identifies RNU2-2 and RNU5B-1 as neurodevelopmental disorder genes. *Nature genetics*, *57*(6), 1362–1366. https://doi.org/10.1038/s41588-025-02209-y

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., & Lambowitz, A. M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nature biotechnology*, *19*(12), 1162–1167. https://doi.org/10.1038/nbt1201-1162

 doi:10.20944/preprints202609.1663.v1

Kent, O. A., Reayi, A., Foong, L., Chilibeck, K. A., & MacMillan, A. M. (2003). Structuring of the 3' splice site by U2AF65. *The Journal of biological chemistry*, *278*(50), 50572–50577. https://doi.org/10.1074/jbc.M307976200

King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H. S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T., Chamberlain, S. J., Philpot, B. D., & Zylka, M. J. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature*, *501*(7465), 58–62. https://doi.org/10.1038/nature12504

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *374*(1772), 20180087. https://doi.org/10.1098/rstb.2018.0087

Lambowitz, A. M., & Belfort, M. (2015). Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiology spectrum*, *3*(1), MDNA3–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0050-2014

Lee, S., & Stevens, S. W. (2016). Spliceosomal intronogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6514–6519. https://doi.org/10.1073/pnas.1605113113

Li, P., Holliger, P., & Tagami, S. (2022). Hydrophobic-cationic peptides modulate RNA polymerase ribozyme activity by accretion. *Nature communications*, *13*(1), 3050. https://doi.org/10.1038/s41467-022-30590-3

Lopatina, A., Tal, N., & Sorek, R. (2020). Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annual review of virology*, *7*(1), 371–384. https://doi.org/10.1146/annurev-virology-011620-040628

Lyu, P., & Lu, B. (2022). New Advances in Using Virus-like Particles and Related Technologies for Eukaryotic Genome Editing Delivery. *International journal of molecular sciences*, *23*(15), 8750. https://doi.org/10.3390/ijms23158750

Lyu, P., Wang, L., & Lu, B. (2020). Virus-Like Particle Mediated CRISPR/Cas9 Delivery for Efficient and Safe Genome Editing. *Life (Basel, Switzerland)*, *10*(12), 366. https://doi.org/10.3390/life10120366

Makarova, K. S., Shmakov, S. A., Wolf, Y. I., Mutz, P., Altae-Tran, H., Beisel, C. L., Brouns, S. J. J., Charpentier, E., Cheng, D., Doudna, J., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Pausch, P., Pinilla-Redondo, R., Shah, S. A., Siksnys, V., Terns, M. P., Tordoff, J., … Koonin, E. V. (2025). An updated evolutionary classification of CRISPR-Cas systems including rare variants. *Nature microbiology*, *10*(12), 3346–3361. https://doi.org/10.1038/s41564-025-02180-8

Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology*, *193*(21), 6039–6056. https://doi.org/10.1128/JB.05535-11

Martínez-Abarca, F., Barrientos-Durán, A., Fernández-López, M., & Toro, N. (2004). The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic acids research*, *32*(9), 2880–2888. https://doi.org/10.1093/nar/gkh616

Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y., & Sorek, R. (2020). Bacterial Retrons Function In Anti-Phage Defense. *Cell*, *183*(6), 1551–1561.e12. https://doi.org/10.1016/j.cell.2020.09.065

Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hör, J., Garb, J., Bechon, N., Brandis, A., Lopatina, A., Ofir, G., Hochhauser, D., Stokar-Avihail, A., Tal, N., Sharir, S., Voichek, M., Erez, Z., Ferrer, J. L. M., Dar, D., Kacen, A., … Sorek, R. (2022). An expanded arsenal of immune systems that protect bacteria from phages. *Cell host & microbe*, *30*(11), 1556–1569.e5. https://doi.org/10.1016/j.chom.2022.09.017

Mohr, G., Hong, W., Zhang, J., Cui, G. Z., Yang, Y., Cui, Q., Liu, Y. J., & Lambowitz, A. M. (2013). A targetron system for gene targeting in thermophiles and its application in Clostridium thermocellum. *PloS one*, *8*(7), e69032. https://doi.org/10.1371/journal.pone.0069032

Montagud-Martínez, R., Ruiz, R., Baldanta, S., Delicado-Mateo, R., & Rodrigo, G. (2026). CRISPR-Cas9 trans-cleavage is hindered by a flanked R-loop, an elongated spacer, and an inactive HNH domain. *Nature communications*, 10.1038/s41467-026-68789-3. Advance online publication. https://doi.org/10.1038/s41467-026-68789-3

Moran, J. V., Zimmerly, S., Eskes, R., Kennell, J. C., Lambowitz, A. M., Butow, R. A., & Perlman, P. S. (1995). Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Molecular and cellular biology*, *15*(5), 2828–2838. https://doi.org/10.1128/MCB.15.5.2828

Nava, C., Cogne, B., Santini, A., Leitão, E., Lecoquierre, F., Chen, Y., Stenton, S. L., Besnard, T., Heide, S., Baer, S., Jakhar, A., Neuser, S., Keren, B., Faudet, A., Forlani, S., Faoucher, M., Uguen, K., Platzer, K., Afenjar, A., Alessandri, J. L., … Depienne, C. (2025). Dominant variants in major spliceosome U4 and U5 small nuclear

RNA genes cause neurodevelopmental disorders through splicing disruption. *Nature genetics*, *57*(6), 1374–1388. https://doi.org/10.1038/s41588-025-02184-4

Negi, M. S., Krishnan, V. P., Saraf, N., & Vijayraghavan, U. (2025). Prp16 enables efficient splicing of introns with diverse exonic consensus elements in the short-intron rich *Cryptococcus neoformans* transcriptome. *RNA biology*, *22*(1), 1–14. https://doi.org/10.1080/15476286.2025.2477844

Olijslager, L. H., Weijers, D., & Swarts, D. C. (2024). Distribution of specific prokaryotic immune systems correlates with host optimal growth temperature. *NAR genomics and bioinformatics*, *6*(3), lqae105. https://doi.org/10.1093/nargab/lqae105

Pantazopoulou, V. I., Georgiou, S., Kakoulidis, P., Giannakopoulou, S. N., Tseleni, S., Stravopodis, D. J., & Anastasiadou, E. (2020). From the Argonauts Mythological Sailors to the Argonautes RNA-Silencing Navigators: Their Emerging Roles in Human-Cell Pathologies. *International journal of molecular sciences*, *21*(11), 4007. https://doi.org/10.3390/ijms21114007

Pillon, M. C., Gordon, J., Frazier, M. N., & Stanley, R. E. (2021). HEPN RNases - an emerging class of functionally distinct RNA processing and degradation enzymes. *Critical reviews in biochemistry and molecular biology*, *56*(1), 88–108. https://doi.org/10.1080/10409238.2020.1856769

Pineda, J. M. B., & Bradley, R. K. (2018). Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes & development*, *32*(7-8), 577–591. https://doi.org/10.1101/gad.312058.118

Pommier, Y., Nussenzweig, A., Takeda, S., & Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. *Nature reviews. Molecular cell biology*, *23*(6), 407–427. https://doi.org/10.1038/s41580-022-00452-3

Pommier, Y., Sun, Y., Huang, S. N., & Nitiss, J. L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature reviews. Molecular cell biology*, *17*(11), 703–721. https://doi.org/10.1038/nrm.2016.111

Raghavan, R., Friedrich, M. J., King, I., Chau-Duy-Tam Vo, S., Strebinger, D., Lash, B., Kilian, M., Platten, M., Macrae, R. K., Song, Y., Nivon, L., & Zhang, F. (2025). Rational engineering of minimally immunogenic nucleases for gene therapy. *Nature communications*, *16*(1), 105. https://doi.org/10.1038/s41467-024-55522-1

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Riecken, K., Głów, D., & Fehse, B. (2021). How to package and SEND mRNA: a novel "humanized" vector system based on endogenous retroviruses. *Signal transduction and targeted therapy*, *6*(1), 384. https://doi.org/10.1038/s41392-021-00803-0

Saito, M., Xu, P., Faure, G., Maguire, S., Kannan, S., Altae-Tran, H., Vo, S., Desimone, A., Macrae, R. K., & Zhang, F. (2023). Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature*, *620*(7974), 660–668. https://doi.org/10.1038/s41586-023-06356-2

Sang, Z., Li, X., Yan, H., Wang, W., & Wen, Y. (2024). Development of a group II intron-based genetic manipulation tool for Streptomyces. *Microbial biotechnology*, *17*(5), e14472. https://doi.org/10.1111/1751-7915.14472

Scheitl, C. P. M., Ghaem Maghami, M., Lenz, A. K., & Höbartner, C. (2020). Site-specific RNA methylation by a methyltransferase ribozyme. *Nature*, *587*(7835), 663–667. https://doi.org/10.1038/s41586-020-2854-z

Schmid, F., Hiller, T., Korner, G., Glaus, E., Berger, W., & Neidhardt, J. (2013). A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. *Human gene therapy*, *24*(1), 97–104. https://doi.org/10.1089/hum.2012.110

Schneider, K. L., Hao, X., Keuenhof, K. S., Berglund, L. L., Fischbach, A., Ahmadpour, D., Chawla, S., Gómez, P., Höög, J. L., Widlund, P. O., & Nyström, T. (2024). Elimination of virus-like particles reduces protein aggregation and extends replicative lifespan in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(14), e2313538121. https://doi.org/10.1073/pnas.2313538121

Segel, M., Lash, B., Song, J., Ladha, A., Liu, C. C., Jin, X., Mekhedov, S. L., Macrae, R. K., Koonin, E. V., & Zhang, F. (2021). Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science (New York, N.Y.)*, *373*(6557), 882–889. https://doi.org/10.1126/science.abg6155

Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., & Kielkopf, C. L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Molecular cell*, *23*(1), 49–59. https://doi.org/10.1016/j.molcel.2006.05.025

Smukowski Heil, C., Patterson, K., Hickey, A. S., Alcantara, E., & Dunham, M. J. (2021). Transposable Element Mobilization in Interspecific Yeast Hybrids. *Genome biology and evolution*, *13*(3), evab033. https://doi.org/10.1093/gbe/evab033

Su, Y., Ghodke, P. P., Egli, M., Li, L., Wang, Y., & Guengerich, F. P. (2019). Human DNA polymerase η has reverse transcriptase activity in cellular environments. *The Journal of biological chemistry*, *294*(15), 6073–6081. https://doi.org/10.1074/jbc.RA119.007925

Sundaresan, R., Parameshwaran, H. P., Yogesha, S. D., Keilbarth, M. W., & Rajan, R. (2017). RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell reports*, *21*(13), 3728–3739. https://doi.org/10.1016/j.celrep.2017.11.100

Swirski, S., May, O., Ahlers, M., Wissinger, B., Greschner, M., Jüschke, C., & Neidhardt, J. (2023). In Vivo Efficacy and Safety Evaluations of Therapeutic Splicing Correction Using U1 snRNA in the Mouse Retina. *Cells*, *12*(6), 955. https://doi.org/10.3390/cells12060955

Tabebordbar, M., Lagerborg, K. A., Stanton, A., King, E. M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J. J., Messemer, K. A., Troiano, E. C., Moghadaszadeh, B., Peacker, B. L., Leacock, K. A., Horwitz, N., Beggs, A. H., Wagers, A. J., & Sabeti, P. C. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell*, *184*(19), 4919–4938.e22. https://doi.org/10.1016/j.cell.2021.08.028

Tseng, C. K., & Cheng, S. C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science (New York, N.Y.)*, *320*(5884), 1782–1784. https://doi.org/10.1126/science.1158993

Weinmann, J., Weis, S., Sippel, J., Tulalamba, W., Remes, A., El Andari, J., Herrmann, A. K., Pham, Q. H., Borowski, C., Hille, S., Schönberger, T., Frey, N., Lenter, M., VandenDriessche, T., Müller, O. J., Chuah, M. K., Lamla, T., & Grimm, D. (2020). Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants. *Nature communications*, *11*(1), 5432. https://doi.org/10.1038/s41467-020-19230-w

Yadav, M., Atala, A., & Lu, B. (2022). Developing all-in-one virus-like particles for Cas9 mRNA/single guide RNA co-delivery and aptamer-containing lentiviral vectors for improved gene expression. *International journal of biological macromolecules*, *209*(Pt A), 1260–1270. https://doi.org/10.1016/j.ijbiomac.2022.04.114

Yoon, P. H., Zhang, Z., Loi, K. J., Adler, B. A., Lahiri, A., Vohra, K., Shi, H., Rabelo, D. B., Trinidad, M., Boger, R. S., Al-Shimary, M. J., & Doudna, J. A. (2024). Structure-guided discovery of ancestral CRISPR-Cas13 ribonucleases. *Science (New York, N.Y.)*, *385*(6708), 538–543. https://doi.org/10.1126/science.adq0553

Yoshida, H., Park, S. Y., Sakashita, G., Nariai, Y., Kuwasako, K., Muto, Y., Urano, T., & Obayashi, E. (2020). Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1. *Nature communications*, *11*(1), 4744. https://doi.org/10.1038/s41467-020-18559-6

Zhong, J., & Lambowitz, A. M. (2003). Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *The EMBO journal*, *22*(17), 4555–4565. https://doi.org/10.1093/emboj/cdg433

Zhuang, Y., & Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes & development*, *3*(10), 1545–1552. https://doi.org/10.1101/gad.3.10.1545

Zila, V., Margiotta, E., Turoňová, B., Müller, T. G., Zimmerli, C. E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., Lusic, M., Kräusslich, H. G., & Beck, M. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*, *184*(4), 1032–1046.e18. https://doi.org/10.1016/j.cell.2021.01.025

Zilberzwige-Tal, S., Altae-Tran, H., Kannan, S., Wilkinson, M. E., Vo, S. C., Strebinger, D., Edmonds, K. K., Yao, C. J., Mears, K. S., Shmakov, S. A., Makarova, K. S., Macrae, R. K., Koonin, E. V., & Zhang, F. (2025). Reprogrammable RNA-targeting CRISPR systems evolved from RNA toxin-antitoxins. *Cell*, *188*(7), 1925–1940.e20. https://doi.org/10.1016/j.cell.2025.01.034