

Expression of Herpes Simplex Virus Thymidine Kinase/Ganciclovir by RNA *Trans*-Splicing Induces Selective Killing of HIV-Producing Cells

Carin K. Ingemarsdotter,¹ Sushmita Poddar,² Sarah Mercier,¹ Volker Patzel,^{1,2} and Andrew M.L. Lever¹

¹Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK; ²Department of Microbiology & Immunology, Yong Loo Lin School of Medicine, National University of Singapore, 5 Science Drive 2, Singapore 117545, Singapore

Antiviral strategies targeting hijacked cellular processes are less easily evaded by the virus than viral targets. If selective for viral functions, they can have a high therapeutic index. We used RNA *trans*-splicing to deliver the herpes simplex virus thymidine kinase-ganciclovir (HSV-tk/GCV) cell suicide system into HIV-producing cells. Using an extensive *in silico* bioinformatics and RNA structural analysis approach, ten HIV RNA *trans*-splicing constructs were designed targeting eight different HIV splice donor or acceptor sites and were tested in cells expressing HIV. *Trans*-spliced mRNAs were identified in HIV-expressing cells using qRT-PCR with successful detection of fusion RNA transcripts between HIV RNA and the HSV-tk RNA transcripts from six of ten candidate RNA *trans*-splicing constructs. Conventional PCR and Sanger sequencing confirmed RNA *trans*-splicing junctions. Measuring cell viability in the presence or absence of GCV expression of HSV-tk by RNA *trans*-splicing led to selective killing of HIV-producing cells using either 3' exon replacement or 5' exon replacement in the presence of GCV. Five constructs targeting four HIV splice donor and acceptor sites, D4, A5, A7, and A8, involved in regulating the generation of multiple HIV RNA transcripts proved to be effective for *trans*-splicing mediated selective killing of HIV-infected cells, within which individual constructs targeting D4 and A8 were the most efficient.

INTRODUCTION

More than 39 million people worldwide are infected with HIV, according to the World Health Organization. Although highly active antiretroviral therapy (HAART) is effective at suppressing viral replication, treatment cannot eliminate the virus because it resides latently in a population of immune cells referred to as memory T cells (reviewed by Van Lint et al.¹). The existence of this reservoir, and the high mutation rate of HIV with the subsequent development of drug resistance,² means that alternative treatments are needed that not only suppress viral replication but can also eliminate infected cells. Conventional therapies do not selectively target infected cells, and adverse drug effects are common. In addition, no currently licensed antiviral drugs are able to eliminate cells in which the virus is latent. Latency in HIV does not equate to permanent transcrip-

tional silencing, and there is evidence that it is stochastic.³ Incomplete RNA transcripts may be found in cells that are not producing intact virus, and viral RNA-spliced products have been detected in resting CD4⁺ T cells isolated from HIV-infected patients on HAART.⁴ The viral RNA is thus an unexploited target to treat actively replicating virus and, at any one time, a significant proportion of latently infected cells. HIV uses alternative RNA splicing to process its genetic material into transcripts encoding different viral proteins (reviewed by Tazi et al.⁵). A recent study detected up to 109 different spliced HIV RNAs in HIV-infected T cells.⁶ The expression of these is tightly regulated, and when the HIV splicing process is modulated, viral replication and production is impaired.^{7–10} Given its importance for the viral life cycle, HIV splicing is an attractive target for the development of novel antiviral therapies.^{5,11,12} Recent work shows that the small-molecule inhibitors digoxin¹³ and 8-azaguanine¹⁴ inhibit HIV replication by altering HIV alternative splicing,^{13,14} and the compound ABX464 enhances HIV RNA splicing, thereby also compromising HIV replication.¹⁵

In addition to pharmacological inhibition of HIV, cell and gene therapies are attractive potential therapies against HIV that may lead to a more sustained control of viral rebound (reviewed by Hoxie et al.¹⁶ and Herrera-Carrillo et al.¹⁷). RNA *trans*-splicing is an elegant gene therapy method that has been used to correct acquired or inherited genetic disorders.¹⁸ It subverts the cellular RNA splicing machinery to exchange, through a *trans*-splicing reaction, a defective RNA transcript with a corrected mRNA molecule delivered *in trans*. In addition to correcting genetic mutations, RNA *trans*-splicing has been used to deliver cell death signals in order to specifically kill targeted cells (reviewed by Drude et al.¹⁹). RNA *trans*-splicing has been used to deliver the two-step *Herpes simplex* virus thymidine kinase/ganciclovir (HSV-tk/GCV) cell death system as a potential cancer therapy.^{20–24} HSV-tk acts by phosphorylating the pro-drug GCV, an analog of deoxyguanosine triphosphate, into an active compound, leading to

Received 19 September 2016; accepted 7 March 2017;
<http://dx.doi.org/10.1016/j.omtn.2017.03.004>.

Correspondence: Andrew M.L. Lever, Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, UK.
E-mail: amll1@medschl.cam.ac.uk

chain termination during DNA replication and cell death (reviewed by Duarte et al.²⁵). The HSV-tk/GCV system has been expressed in HIV-infected cells and proposed as a gene therapy against HIV.^{26–32} The extensive use of splicing by HIV to regulate its life cycle suggests that there are likely to be vulnerable splicing reactions susceptible to a *trans*-splicing approach. However, this has never previously been formally investigated using modern bioinformatic tools to identify those RNA species involved in individual splicing reactions that have structures favoring *trans*-splicing and that would be most effectively targeted.

We performed a comprehensive in silico analysis of HIV splicing and RNA structure to both find optimal targets and to design HIV binding domains in the *trans*-splicing constructs to, as near as possible, guarantee the specificity of the RNA *trans*-splicing reaction and facilitate the process. Using a combination of HIV splice site predictions, HIV RNA free energy calculations, and RNA secondary structure predictions together with an in-depth analysis of the literature, we were able to select a panel of ten 3' and 5' exon replacement constructs. These were then subject to testing for their ability to induce *trans*-splicing with transcripts from replicating virus. In those in which *trans*-splicing was confirmed at a molecular level, functional screening by MTT assay revealed four susceptible RNA target splice sites at which this approach led to specific killing of HIV-producing cells, with 3' exon replacement being overall most efficient.

RESULTS

In Silico Design of RNA *Trans*-Splicing Binding Domains and Selection of Target HIV Splice Sites

HIV splice sites were predicted using computer software as described in [Materials and Methods](#). The splice site prediction analysis of pNL4.3 gave similar results with both software packages used (data not shown). Predicted splice site sequences were compared with those previously described in the literature, and the most favorable were selected for targeting by *trans*-splicing ([Figure 1A](#)). Splice sites that had previously been reported in the literature matching our computer predictions, or that we identified as present within pNL4.3, were considered. In addition, two groups of two computer-predicted HIV splice sites each referred to as early acceptor (EA) and late donor (LD) sites were included, comprising the sites EA1 and EA2 (EA1/2) and LD1 and LD2 (LD1/2), respectively ([Figure 1A](#)).

RNA folding energies of the complete reverse complement HIV pNL4.3 genome in 50 nucleotide windows from 5' to 3' were calculated ([Figure 1B](#), left). The number of free nucleotides per folding window of selected binding domains subjected to further design is shown ([Figure 1B](#), right). Selected structures were then refolded in the context of the backbone 3' or 5' exon replacement cassette sequences to confirm lack of interference by flanking regions ([Figure 1C](#)), and final structures were generated by further sequence modification. From this detailed in silico analysis, three HIV splice donor sites, D4, D1a, and LD1/2, were targeted with five 3' exon replacement constructs, and five HIV splice acceptor site, EA1/2, A3, A5, A7, and A8, were targeted with 5' exon replacement ([Fig-](#)

[ure 1D](#)). Final HIV binding domain structures folded in the backbone cassettes, all displaying predominantly unstructured binding domain regions, are shown for 3' exon replacement in [Figure 2](#) and for 5' exon replacement in [Figure 3](#). For clarity, only the binding domain structures are shown. Target regions are shown in [Tables S1A](#) and [S1B](#), for 3' exon replacement and 5' exon replacement, respectively, and schematic diagrams of 3' exon replacement and 5' exon replacement are outlined in [Figures S1A](#) and [S1B](#) and of HSV-tk in [Figures S2A](#) and [S2B](#).

Sequence Conservation of HIV Binding Domain Target Regions

Given the high level of sequence variability of HIV, it was important to confirm that the sequences we used for the binding domains would target the maximum number of HIV strains. Sequence conservation of the HIV binding domain target regions was assessed using the AnalyzeAlign software.³³ For the 3' exon replacement construct BD2-D4, the maximum number of mutations of the HIV binding domain target region reached 32 in a maximum percentage of 21.6% from 3,944 aligned sequences, including all available HIV-1 sequences within the Los Alamos National Laboratory (LANL) filtered web alignment. With BD1-D4, this was reduced to 30 mutations in a maximum of 12.75% sequences from the consensus sequence. For the 5' exon replacement constructs, the number of mutations was small, ranging from 7, 8, 6, to 33, in a maximum of 14.79%, 19.72%, 27.55%, and 9.92% for BD-A7, BD-A5, BD-A3, and BD-A8 in 4,632, 2,712, 2,445, and 4,553 sequences analyzed, respectively ([Table S2](#)). Of the binding domain target sequences, BD-A8 was the least conserved, with 33 mutations but with a maximum of 9.92% variants from 4,553 sequences analyzed. Interestingly, when sequence variation was instead analyzed compared with the consensus sequence for the major subtypes, a decrease in the number of mutations of the maximum percentage of sequences was seen in half of the constructs, which could be diminished in all of the constructs to very low levels when sequence variation within subtype group B was analyzed. The number of mutations ranged from 3 in BD-A7 to 9 in BD-A8, with the exception of BD2-D4 showing the least conserved binding domain target sequence, with 27 mutations in a maximum of 20.16% of the sequences within subtype group B. Remarkably, when comparing the HIV target sequences for BD1-D4, the number of mutations could be reduced from 30 in 12.75% when comparing all HIV-1 sequences in the alignment to 7 mutations in 20.44% within subtype B sequences. Similarly, the number of mutations was reduced for BD-A8 target region from 33 in 9.92% of sequences when comparing all HIV-1 sequences to 9 mutations in 14.55% of sequences within subtype B, suggesting that the majority of our HIV binding domains are located in regions with high conservation within HIV-1 subtype B viruses.

We next aligned our HIV binding domain target sequences in pNL4.3 with the consensus sequences for each binding domain target region within subtype B using BLAST.³⁴ Needleman-Wunsch global alignments revealed sequence identities of 73% for BD1-D4, 55% for BD2-D4, 49% for BD-A3, 98% BD-A5, 96% for BD-A7, and 64% for BD-A8 compared with their respective subtype B consensus

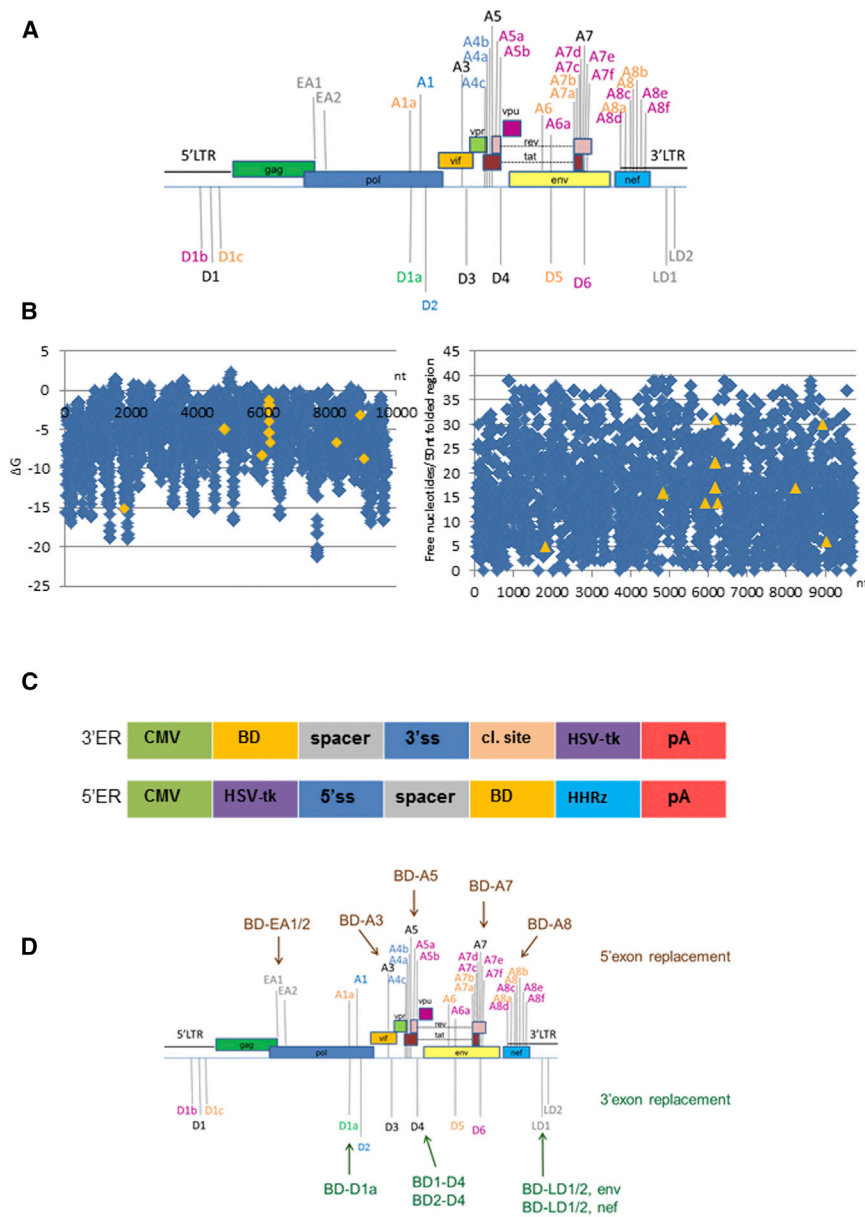


Figure 1. In Silico Analysis of HIV Splicing and Splice-Site Target Determination for RNA *Trans*-Splicing

(A) Summary of HIV splice sites considered for targeting by RNA *trans*-splicing. Conserved splice donor/acceptor sites predicted by computer splice site predictions of pNL4.3 are shown in black.⁸⁷ In green, previously reported cryptic splice donor site predicted by computer algorithms.⁴¹ In gray, late splice donor/early splice acceptor sites predicted in pNL4.3 by computer algorithms. In blue, conserved splice donor/acceptor sites not predicted in pNL4.3 by computer splice site predictions but previously published.⁸⁷ In orange, previously reported cryptic splice sites not predicted in pNL4.3 by computer algorithms (A1a, A6,⁵⁰⁻⁵² D5,^{50,51} A7a, A7b,⁸⁸ A8a, A8, A8b⁸⁹). In purple, recently described splice donor and splice acceptor sites not predicted by computer algorithms,⁶ A5a.^{6,89,90} (B) Selection of binding regions using the software Foldanalyze (HUSAR). MFE analysis of pNL4.3 throughout the complete reverse complement pNL4.3 genome per 50 nt folded region (left). Number of free nucleotides per 50 nt folded region of the complete reverse complement pNL4.3 genome (right). Target regions of binding domains selected for further design are shown in orange. (C) Maps of CMV promoter-driven RNA *trans*-splicing cassettes. 3' exon replacement (top) and 5' exon replacement (bottom) cassettes are shown. BD, binding domain; spacer, spacer sequence; 3'ss, 3' splice site domain; cl. site, P2A protein cleavage site⁹¹; HSV-tk, herpes simplex virus thymidine kinase gene; pA, polyA tail; HHRz, hammerhead ribozyme sequence. (D) Summary of selected HIV splice sites for targeting by 5' exon replacement (top) or 3' exon replacement (bottom). HIV splice site D4 was targeted with two different binding domain sequences (BD1-D4 and BD2-D4), and LD 1 and 2 (LD1 and LD2 respectively) were targeted with two BD-LD1/2 cassettes driven from either HIV Env or HIV Nef translational start sites.

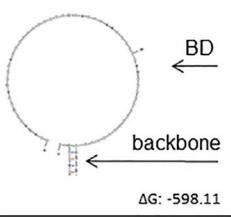
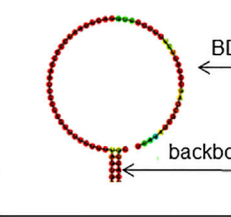
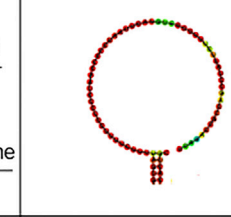
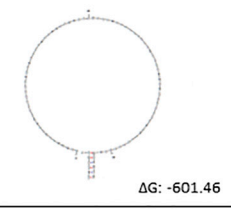
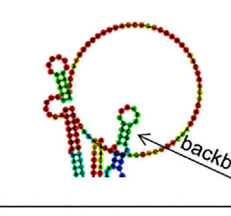
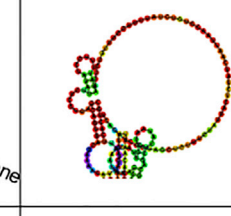
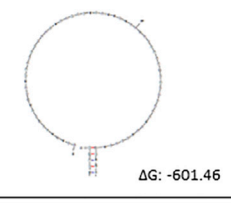
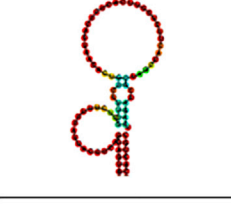
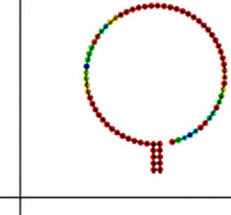
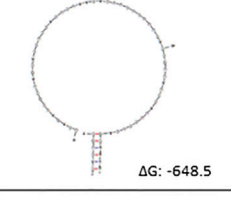
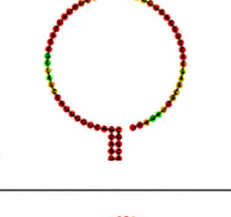
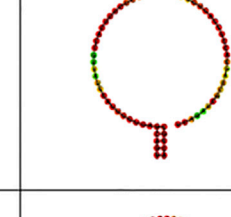
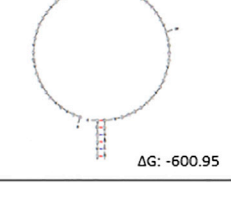
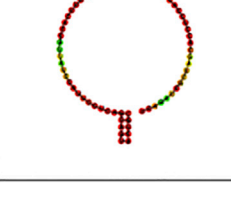
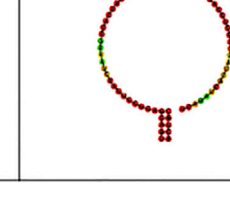
mRNAs; these were sought by qRT-PCR (Figure 4A). RNA *trans*-splicing amplicons were detected selectively in samples containing both HIV and RNA *trans*-splicing constructs BD1-D4 and BD2-D4, both in sequentially and co-transfected cells (Figure 4B). These could also be detected by conventional electrophoresis (Figure 4C); the PCR products were analyzed by Sanger sequencing, confirming junctional sequences (Figure 4D) between the HIV *tat* transcript and the *trans*-splicing construct.

We next sought chimeric mRNA upon 5' exon replacement. An analogous strategy to 3' exon replacement was used to detect RNA *trans*-splicing amplicons in the presence of pNL4.3 and 5' exon replacement constructs in either sequentially or co-transfected cells (Figure 5A). In four of five 5' exon replacement constructs, BD-A3, BD-A5, BD-A7, and BD-A8, chimeric RNA transcripts could be detected between

sequences, revealing that some of our binding domain target regions within pNL4.3 show some differences from the consensus sequence, although BD-A7 and BD-A5 show high levels of sequence identity.

Screening and Detection of RNA *Trans*-Splicing Junctions in HIV-Producing Cells

To test the activity of our 3' exon replacement RNA *trans*-splicing constructs, we performed an initial screen in 293T cells. Cells were either co-transfected with the proviral clone pNL4.3 and the 3' exon replacement constructs, or the latter was transfected 24 hr later. RNA *trans*-splicing should lead to the production of chimeric

	Mfold		RNA fold	
	mfe	mfe	mfe	centroid
BD-D1a				
BD1-D4				
BD2-D4				
BD-LD1/2, env				
BD-LD1/2, nef				

HSV-tk and HIV by qRT-PCR (Figure 5B). RNA *trans*-splicing amplicons could also be detected by conventional PCR and electrophoresis (Figure 5C), and Sanger sequencing confirmed junctional sequences (Figure 5D). Taken together, this demonstrates that 3' or 5' exon replacement between HIV and HSV-tk is feasible and that both generate the expected RNA *trans*-splicing products.

Expression of HSV-tk by RNA *Trans*-Splicing Leads to Selective Killing in HIV-Producing Cells

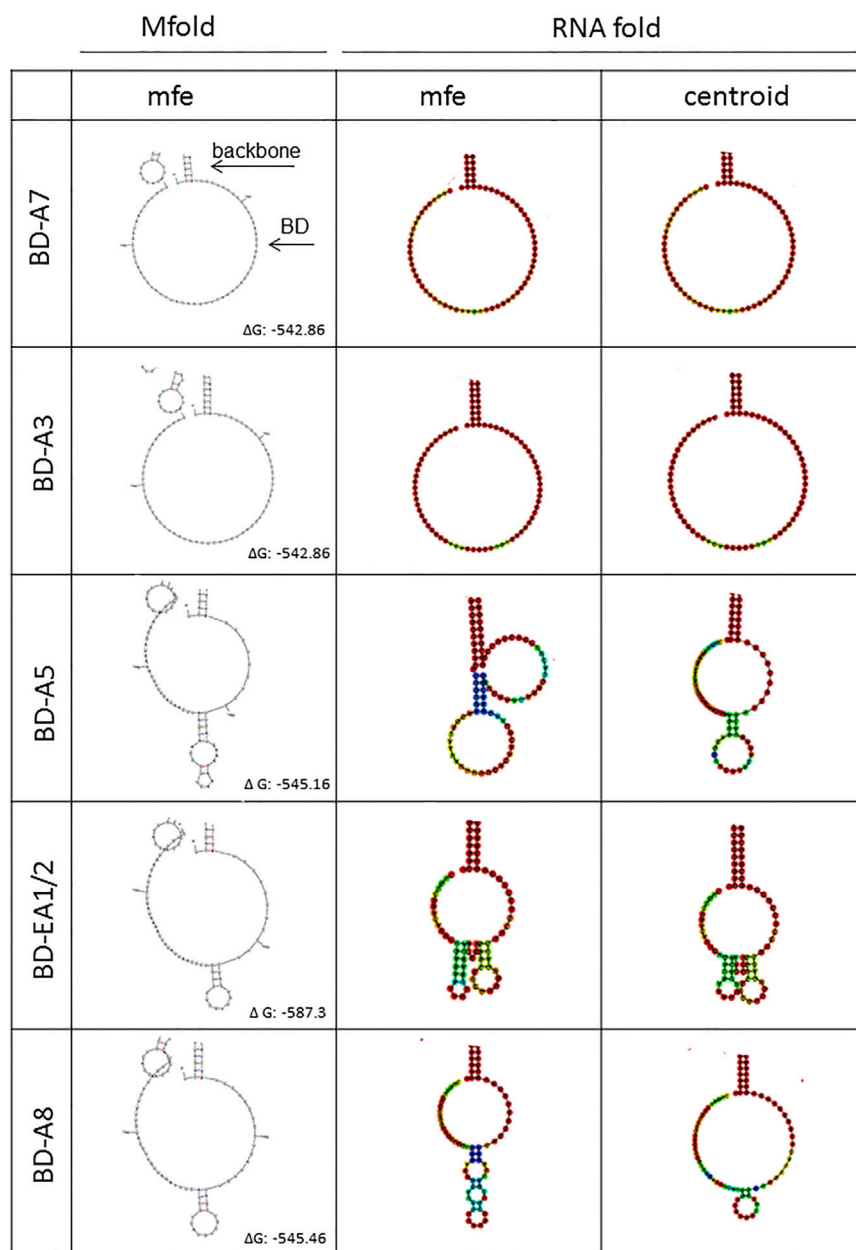
The expression of HSV-tk protein dependent on the presence of the target HIV RNA sequence should lead to selective killing in the pres-

Figure 2. RNA Secondary Structure Predictions of Designed 3' Exon Replacement HIV Binding Domains

RNA secondary structures of 3' exon replacement cassettes containing HIV binding domains were predicted using Mfold and RNAfold web servers. MFE RNA secondary structures of HIV binding domain regions predicted in Mfold are shown (left column). RNAfold secondary structure predictions showing MFE (middle column) and centroid folds are shown (right column). RNA secondary structures of HIV binding domains with only a small region of the backbone plasmid are shown for clarity. Predicted free energies for the complete fold (binding domain and backbone cassette) are shown as ΔG .

ence of ganciclovir (GCV). 293T cells were transfected with pNL4.3 and RNA *trans*-splicing constructs sequentially. Twenty-four hours later, cells and appropriate controls as described in the legend of Figure 6 were treated with one or two doses of 100 μ M GCV on 2 consecutive days 24 hr apart, and cell viability was measured by the MTT assay 3 days later (6 days post-transfection with pNL4.3). 3' exon replacement, *trans*-spliced HSV-tk, reduced cell viability in the presence of GCV. Compared with untreated controls (100% viability), the mean values of cell viability from representative experiments performed in multiple wells decreased to 26.2% for pNL4.3+BD1-D4 and 35.3% for pNL4.3+BD2-D4 (Figure 6A). Cell viability of cells transfected with pNL4.3+BD1-D4 compared with BD1-D4-transfected cells in the presence of GCV was also statistically significantly different ($p \leq 0.05$) in each of three independent experiments, with a reduction in cell viability in the presence of pNL4.3. Similarly, this was also observed with pNL4.3+BD2-D4 compared with BD2-D4-transfected cells in the presence of GCV ($p \leq 0.05$ in each of three independent experiments). A much smaller although still significant reduction in cell viability was seen in cells transfected with BD2-D4 alone compared with untransfected cells treated with GCV, from 92.8% to 64.9% ($p = 0.0035$).

The effect of HSV-tk/GCV on cell viability was also investigated using 5' exon replacement. 5' exon replacement reduced cell viability in the presence of GCV to 34.61% for pNL4.3+BD-A7, 34.93% for pNL4.3+BD-A5, 51.63% for pNL4.3+BD-A8, and 73.51% for pNL4.3+BD-A3, compared with control cells (Figure 6B). When cell viability was assessed as above in cells transfected with pNL4.3+RNA *trans*-splicing construct, compared with single-transfected cells



solely expressing the RNA *trans*-splicing construct, both in the presence of GCV, a reduction in cell viability for pNL4.3+BD-A3, pNL4.3+BD-A7, pNL4.3+BD-A5, and pNL4.3+BD-A8 was observed, of which pNL4.3+BD-A8 reached statistical significance in each of three independent experiments ($p \leq 0.05$). BD-A7 and BD-A5 were effective in reducing cell viability in the presence of pNL4.3, and BD-A3 had a modest effect, although for these constructs this did not reach statistical significance in all three independent experiments. The 3'exon replacement construct BD1-D4 was the most efficient of our panel of RNA *trans*-splicing constructs. To confirm that RNA *trans*-splicing activity is not restricted to the

spliced and unspliced transcripts by binding to the Rev responsive element (RRE) within the *tat/rev* intron, leading to further production of viral proteins in the cytoplasm (reviewed by Tazi et al.⁵).

The central importance of splicing to HIV has led to exploration of possible ways to interfere with the process as a way to inhibit viral replication and control viral growth (reviewed by Tazi et al.⁵). Previous approaches have been largely empirical and have not used a detailed and extensive bioinformatic analysis to optimize the targeting. We have now refined this approach taking advantage of

Figure 3. RNA Secondary Structure Predictions of Designed 5' Exon Replacement HIV Binding Domains after HHRz RNA Self-Cleavage to Release the Binding Domain

RNA secondary structures of 5' exon replacement cassettes containing HIV binding domains were predicted using Mfold and RNAfold web servers. MFE RNA secondary structures of HIV binding domain regions predicted in Mfold are shown (left column). RNAfold secondary structure predictions showing MFE (middle column) and centroid folds (right column). RNA secondary structures of HIV binding domains with only a small region of the backbone plasmid are shown for clarity. See Figure S1 B for a schematic diagram of HHRz self-cleavage.

HIV-1 proviral clone pNL4.3, we tested BD1-D4 against the HIV-1 proviral clone HXB2 (SVC21). A significant reduction in cell viability was observed in the presence of the RNA *trans*-splicing construct BD1-D4 in HXB2-producing cells compared with cells transfected with BD1-D4 alone in the presence of GCV ($p < 0.001$), demonstrating that induction of cell death by RNA *trans*-splicing is not restricted to the proviral clone pNL4.3 (Figure S3).

DISCUSSION

HIV RNA splicing is central to the HIV viral life cycle. Three classes of mRNA transcripts are generated in infected cells: unspliced, partially spliced, and multiply spliced mRNAs. HIV uses alternative splice donor and acceptor sites in different combinations to generate a large repertoire of mRNA transcripts in a highly regulated fashion. The multiply spliced transcripts are produced early during infection to generate the *tat*, *rev*, and *nef* transcripts encoding the regulatory proteins Tat, Rev, and Nef. Tat drives transcription by binding to TAR in the 5'LTR, and Rev mediates nuclear export of the partially

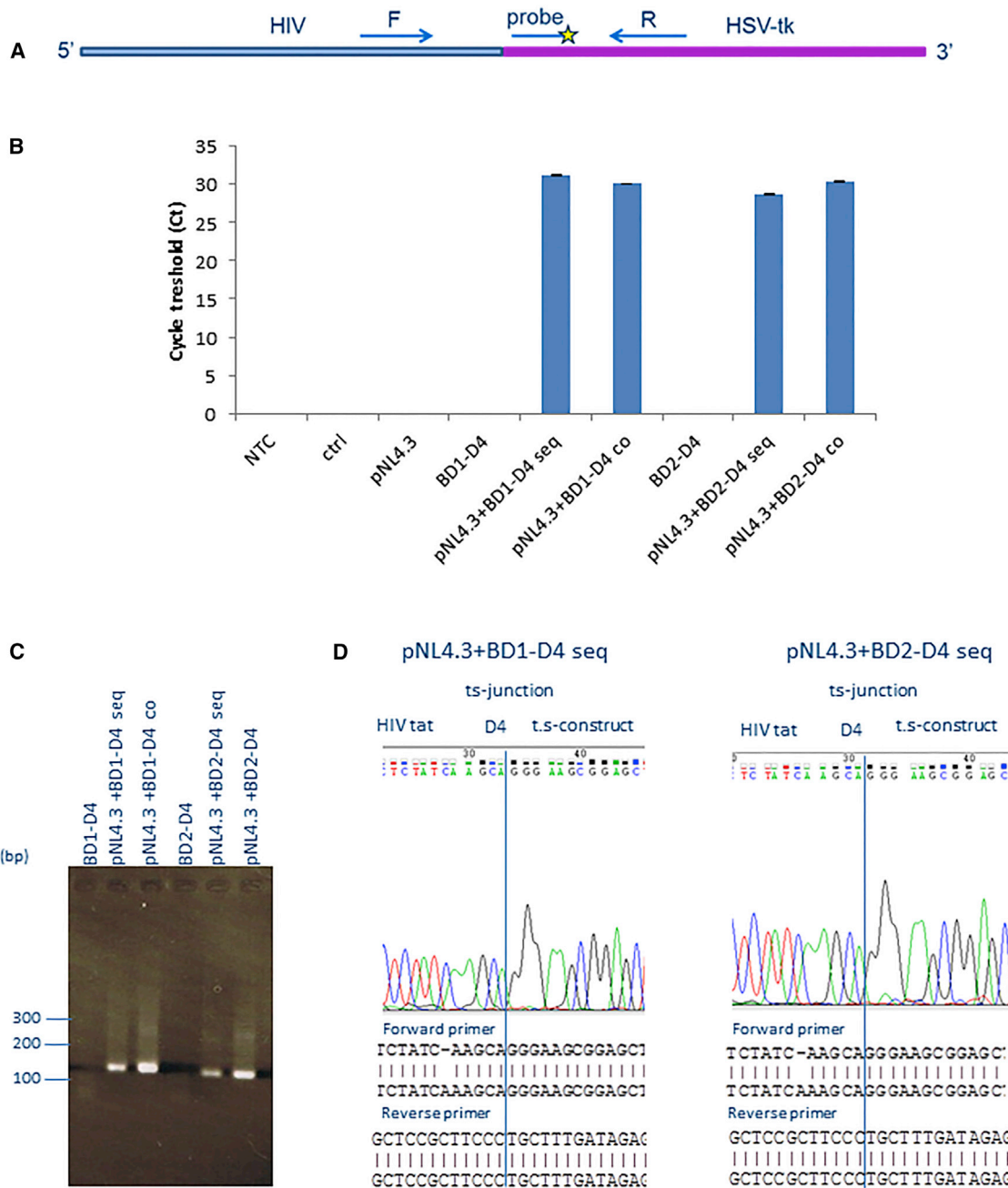
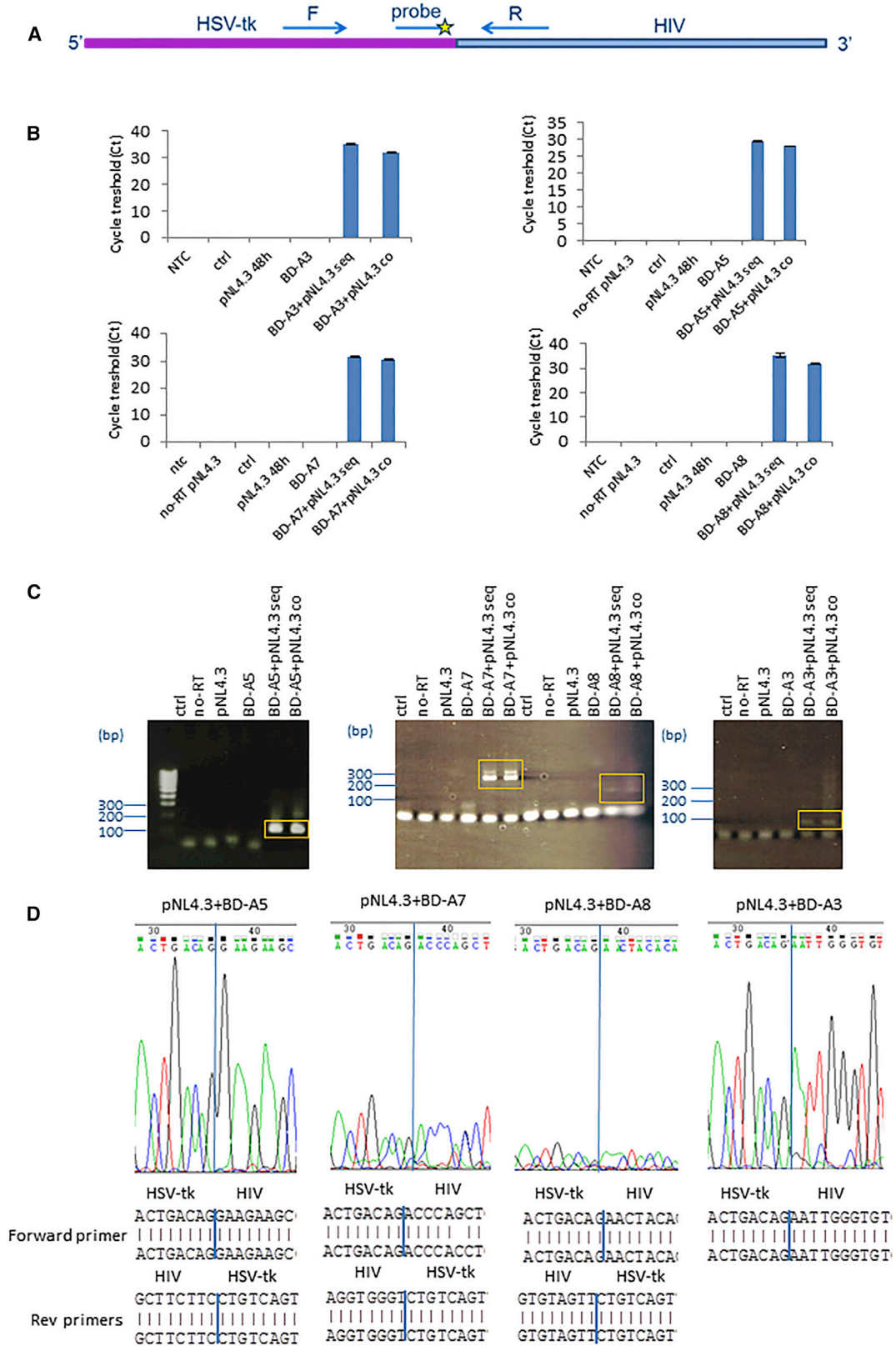


Figure 4. Confirmation of RNA Trans-Splicing between HIV and 3' Exon Replacement Constructs

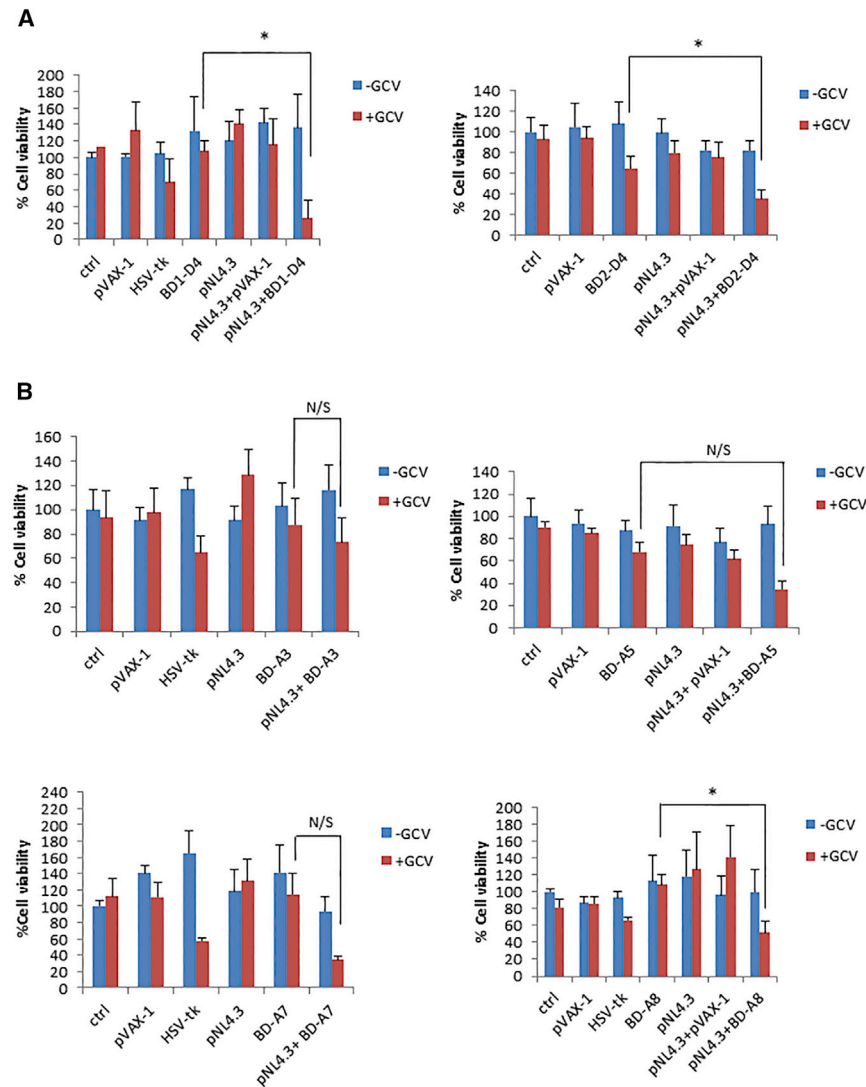
(A) qRT-PCR primer and probe design to detect RNA *trans*-splicing amplicons. The forward primer is located in the HIV region and the reverse primer in the HSV-tk region. The PCR amplicon was detected with a probe in the HSV-tk region. (B) qRT-PCR data in cells transfected with the HIV proviral clone pNL4.3 followed by transfection with 3' exon replacement construct BD1-D4 or BD2-D4. Cells were analyzed 48 hr post-transfection with the RNA *trans*-splicing constructs in sequentially transfected cells (seq) or co-transfected cells (co). (C) Confirmation of RNA *trans*-splicing by conventional PCR in sequentially (seq) or co-transfected cells (co). (D) Confirmation of RNA *trans*-splicing products by Sanger sequencing between HIV *tat* and HSV-tk constructs. BD1-D4 (left) and BD2-D4 (right). The error bars represent the standard deviation (STDEV) of average Ct values.

sophisticated predictive and RNA structural software and performing an extensive in silico analysis of HIV splice sites with the aim of HIV dependent expression of the HSV-tk/GCV cell suicide system by

RNA *trans*-splicing. The specificity of the RNA *trans*-splicing reaction is guaranteed by having complementary sequences binding to the target pre-mRNA. Binding domains are generally based on target



(legend on next page)



sequence (reviewed by Walsh³⁵), but optimization of binding domains to include stem-loop structures has been tested,³⁶ and fluorescence-based screens of binding domain libraries have been developed.³⁷⁻³⁹ In addition to the complementary nucleotide sequence, we considered RNA structure as an important element for efficient RNA *trans*-splicing. It was previously demonstrated by Patzel and Sczakiel⁴⁰ that the secondary structure of artificial antisense

sites A3, A5, A7, and A8, all of which were confirmed by PCR and sequencing.

A comprehensive analysis of HIV splicing patterns by Ocwieja et al.⁶ revealed more than 100 spliced mRNA species. Of the donor splice sites identified (Figure 1A; Table S1), D4 that we targeted is one of the major conserved HIV splice sites used in vivo to generate the

Figure 5. Confirmation of RNA *Trans*-Splicing between HSV-tk and 5' Exon Replacement Constructs

(A) qRT-PCR primer and probe design to detect RNA *trans*-splicing amplicons. The forward primer is located in the HSV-tk region and the reverse primer in the HIV region. The PCR amplicon was detected with a probe in the HIV region. (B) qRT-PCR data from cells transfected with the HIV proviral clone pNL4.3 followed by transfection with 5' exon replacement constructs sequentially (seq) or after co-transfection of the two constructs (co). Top: data for BD-A3 and BD-A5 (left to right). Bottom: data for BD-A7 and BD-A8 (left to right). (C) Detection of HIV RNA *trans*-splicing products by conventional PCR. RNA *trans*-splicing was analyzed by with primers specific for BD-A5 (left), BD-A7 and BD-A8 (middle), and BD-A3 (right). PCR products highlighted in orange were PCR-purified and further analyzed by Sanger sequencing, as in (D). (D) Confirmation of RNA *trans*-splicing junctions by Sanger sequencing in sequentially transfected cells. Chromatographs showing sequence reads are included (top). PCR products as shown in (C) were analyzed by sequencing with a 5' ER forward primer and HIV-specific reverse primers. Sequencing data are shown from left to right for BD-A5, BD-A7, BD-A8, and BD-A3. RNA *trans*-splicing junctions are shown by blue lines on sequence chromatographs and sequence reads.

Figure 6. Cell Viability Screen of Exon Replacement Constructs in HIV-Producing Cells

293T cells were transfected with pNL4.3 and sequentially transfected with RNA *trans*-splicing constructs the following day. Cells were treated with one or two doses of GCV at a concentration of 100 μ M, and cell viability was measured by the MTT assay 6 days after the initial transfection. Ctrl refers to untransfected 293T cells, pVAX-1 (empty vector backbone) was used as a negative control, and HSV-tk (full length HSV-tk) was included as a positive control. (A) 3' exon replacement in HIV-producing 293T cells in the presence or absence of GCV. (B) Cell viability screen of 5' exon replacement constructs in HIV-producing 293T cells in the presence or absence of GCV. BD-A3 (top left), BD-A5 (top right), BD-A7 (bottom left), and BD-A8 (bottom right). The data in (A) and (B) represent the average mean of representative experiments performed in multiple wells (three to eight wells). The error bars represents the SD of the average percentage cell viability. p values (two-tailed) are shown for "t test: two sample assuming unequal variances"; * $p \leq 0.05$, where statistical significance was observed in each of three independent experiments.

RNA is important for efficient annealing to target RNA and that target binding represents the rate-limiting step of antisense RNA-triggered inhibition of gene expression. A fast annealing rate to the target region was shown to correlate with terminal unpaired nucleotides.⁴⁰ Taking all of these factors into consideration, HIV binding domains were designed on the basis of stepwise minimum free energy (MFE) calculations. Using our approach, we confirmed successful RNA *trans*-splicing in six of ten designed constructs demonstrating use of the HIV donor site D4 and acceptor

vpr, *tat*, *rev*, *nef*, and two novel transcripts. It is most commonly used in combination with the A7 splice acceptor site, followed by A8c, a recently identified site.⁶

We also confirmed RNA *trans*-splicing at HIV splice acceptor sites. The HIV splice acceptor site A3 is used to generate the *tat* transcripts and is most common used in combination with the D1 major splice donor, followed by D2 and then D3.⁶ Similarly, splicing at the HIV acceptor site A5 that generates the *env/vpu*, *rev*, and *nef* transcripts is most frequently used in combination with splice donor site D1, followed by D3 and D2. HIV splice acceptor site A7 generates the *vpr*, *tat*, *rev*, *nef*, and recently identified novel transcripts.⁶ In addition, A7 mainly splices onto donor site D4 followed by D1. HIV acceptor site A8 is a cryptic splice site splicing mainly onto D4, but low levels of splicing onto D1 and D1b has been shown.⁶ Although we were able to confirm RNA *trans*-splicing to the splice acceptor sites A5, A7, A3, and A8, we cannot exclude that low levels of RNA *trans*-splicing events are occurring at proximal acceptor sites. An additional band was observed when RNA *trans*-splicing amplicons were analyzed by conventional PCR with BD-A7 in the presence of HIV (Figure 5C). The A7 cluster of splice sites, A7c, A7d, A7, A7e, and A7f, are located within close proximity and with the reverse primer used in the PCR, whose target sequence is located downstream of A7f, it would be possible to amplify RNA *trans*-splicing products arising from splicing events occurring at proximal splice sites upstream of A7, which would give rise to a PCR amplicon of increased size. This could be observed, but the main component in the sequencing analysis revealed splicing from A7. Similarly, with BD-A5 and BD-A8, we cannot exclude low levels of RNA *trans*-splicing in the presence of HIV from the proximal A5a and A5b site in addition to A5 and from the A8 cluster of splice sites A8a, A8b, A8e, and A8f, respectively. Again, the main components in the sequencing analysis revealed splicing from A5 and A8. Taken together, this suggests that RNA *trans*-splicing events at the A5, A7, and A8 clusters of splice sites are dominant despite proximal splice sites being present.

We were not able to confirm RNA *trans*-splicing to the D1a splice donor site, which is an infrequently used cryptic site, but is suggested to have a role in RNA stability.⁴¹ Nor could we confirm use of EA sites designated EA1/2 or LD sites LD1/2 identified by our *in silico* predictions. However, these are putative sites that may not be favorable or may perhaps be silenced by splicing silencing regulatory domains. It is intriguing that these sites could not be activated even when RNA *trans*-splicing constructs provide splicing domains delivered in *trans*, possibly suggesting that they are tightly suppressed. No HIV splicing silencing domains have so far been identified in the vicinity of these predicted EAs and LD sites, suggesting that there may be additional HIV splice site silencing regulatory domains in the HIV genome.

Of the six constructs confirmed to *trans*-splice onto HIV, five led to cell killing in 293T cells in the presence of HIV. The 3' exon replacement RNA *trans*-splicing candidates (BD1-D4 and BD2-D4) targeting HIV splice sites D4 are good candidates for further study. Of

the 5' exon replacement constructs, BD-A7, BD-A5, and BD-A8 reduced cell viability, of which BD-A8 was the most effective. Splicing to the D4 site by 3' exon replacement with BD1-D4 gave the most pronounced reduction in cell viability. HIV splice donor site D4 has a high usage, second only to the major HIV splice donor site D1, onto which all RNA transcripts splice.⁶ Recently, Sherrill-Mix et al.⁴² observed chimeric RNA transcripts between HIV RNA and host-cell RNA in infected cells spliced from the D4 site onto cellular acceptor sites, and RNA *trans*-splicing from D4 to upstream acceptors has recently been observed, suggesting *trans*-splicing between different HIV RNA transcripts,⁴³ which is in agreement with our finding that splicing in *trans* from D4 is feasible. In cells transfected with BD2-D4 alone, we observed a small reduction in cell viability, possibly due to non-specific binding of the binding domain to cellular pre-mRNA targets. The 3' exon replacement constructs BD1-D4 and BD2-D4 have overlapping binding domains, but the binding domain nucleotide sequence of BD1-D4 binds 53 nucleotides upstream of BD2-D4 and is 17 nt longer, suggesting that increasing the length of the binding domain may enhance the specificity of the RNA *trans*-splicing reaction, which is in agreement with others.⁴⁴ Indeed, of the two binding domains targeting D4, BD1-D4 was the most efficient construct in reducing cell viability specifically and, notably, the only binding domain with a length greater than 50 nucleotides. Importantly, the efficiencies of these two binding domains in reducing viability are directly comparable, as the expression of HSV-tk is driven from the same Tat translational start codon. The association of the longest binding domain with the greatest effect on GCV-mediated, HIV-dependent cell killing may be significant.

Both donor site D4 and acceptor site A3 are involved in generating the different *tat* RNA transcripts. D4 is used to generate the completely spliced *tat* RNA transcripts and A3 for both the completely spliced and partially spliced *tat* transcripts.⁶ Within the pNL4.3 genome, these sites are located 267 nt apart, and there are multiple splice site regulatory domains located between them.⁴⁵ Although both sites are involved in generating Tat transcripts and are close together, we were not able to detect a significant reduction in cell viability with BD-A3 targeting the A3 site. During the preparation of this manuscript, Emery et al.⁴³ reported that splice acceptor A3 is used to low frequency in pNL4.3. In addition, the HIV binding domains BD-A3 and BD2-D4 bind 946 nt upstream of splice site A3 and 149 nt downstream of D4 respectively, so it is possible that the distance between the binding domain target sequence and the splice site targeted is important. This may be an important consideration for *trans*-splicing binding domain design in general, as previously suggested.⁴⁶ Intriguingly, BD-A3 targeting splice acceptor site A3 and BD-D1a targeting the cryptic splice donor site D1a have overlapping binding domain sequences, but RNA *trans*-splicing junctions could not be detected from D1a, confirming that the binding domain does not confer RNA *trans*-splicing per se.

In vivo, the conserved HIV splice acceptor sites included in our screen are used with the following frequency: A7 > A5 > A3 > A8 in the HIV-1 isolate 89.6.⁶ No clear correlation was seen between splice

site usage and reduction in cell viability, although they are all conserved and widely used sites, except A8, a cryptic site.⁶

On the basis of previous findings from antisense RNA design,⁴⁰ we selected and designed HIV binding domain sequences to have a high degree of free nucleotides and unstructured regions by in silico RNA secondary structure predictions, aiming for fast annealing rates. We also compared our binding domain candidates with previously published SHAPE reactivity data of the complete HIV genome,⁴⁷ which revealed that our binding domains are targeted to regions containing highly reactive and thus predicted unpaired nucleotides, in agreement with our in silico predictions. The 5' exon replacement binding domains BD-A5, BD-EA1/2, and BD-A8 contain small structural domains with hairpin loops, although the majority of nucleotides are free and unstructured. These hairpins were difficult to resolve; nevertheless, BD-A5 and BD-A8 had *trans*-splicing activity, suggesting that they had little effect on *trans*-splicing efficiency.

We investigated HIV binding domain target sequence variation and compared the number of mutations in the maximum percentage of sequences between sequences from all HIV-1 groups, the major subtypes, or within subtype B. The majority of the HIV binding domain targets are located in regions with a low maximum percentage of sequence variants, ranging from 9.92%–27.55% when all groups were included in the analysis. Sequence variation could be reduced when comparing HIV binding domain target sequences in the major subgroups and even further within subtype B. Interestingly, a recent study highlighted the nucleotide genomic diversity across the HIV genome at the large-scale population level,⁴⁸ and Li et al.⁴⁸ demonstrated that the nucleotide genomic diversity between HIV-1 groups was 37.5%, 14.7% between HIV-1 subtypes, and 8.2% within the different HIV-1 subtypes, which supports our findings that fewer variants were present in the maximum percentage of sequences within subtype B. Global sequence alignments of our binding domain target sequences in pNL4.3 with their respective subtype B consensus sequences revealed that in some cases, overall sequence identity was moderate between pNL4.3 and the subtype B consensus sequence, ranging from 49%–73%. Although it would be ambitious to claim that a single binding domain sequence would be universally applicable for *trans*-splicing, our data suggest that targeting individual subtypes such as subtype B with a single construct is highly feasible on the basis of consensus sequences. Developing HIV-1 therapeutics on the basis of consensus sequences has previously been suggested for HIV-1 vaccine development⁴⁹ and would be an attractive and pragmatic approach for the development of next-generation HIV-1 RNA *trans*-splicing binding domains.

Despite the similarity in BD1-D4 HIV-1 binding domain target sequences between pNL4.3 and HXB2, these are distinct lab strains and show differences in splice site usage, HXB2 for example having an additional splice acceptor site A6,^{50–52} and the splice donor site D5,^{50,51} which was not predicted in our in silico analysis in pNL4.3. Notwithstanding these differences, we were able to confirm a signif-

icant reduction in cell viability upon RNA *trans*-splicing with BD1-D4 in the HXB2 clone SVC21, indicating that it is feasible to target HIV-1 strains with some variations in splice site usage and that this did not affect RNA *trans*-splicing at the conserved HIV splice donor site D4 site.

The expression of HSV-tk lead to cell death in HIV-producing cells in the presence of GCV. Our 3' exon replacement constructs lack the first translational initiation codon of HSV-tk and are designed so that an upstream in-frame HIV-1 translational initiation codon initiates HSV-tk expression. Intriguingly, HSV-tk has an unusual mechanism for translational initiation with additional translational initiation sites located downstream of the first ATG start codon, generating additional HSV-tk polypeptides,^{53–55} with an observed increase in usage when the first ATG is mutated.^{56,57} We cannot exclude that some of the HSV-tk-induced cell death observed in the absence of HIV-1 stems from HSV-tk peptide expression driven from downstream translational initiation codons or other off-target effects (reviewed by Berger et al.⁵⁸). Splicing in *cis* and protein expression from putative translational initiation codons within RNA *trans*-splicing molecules has previously been reported.⁵⁹ Future work will aim at investigating the relative contribution of different HSV-tk peptides in HSV-tk/GCV-mediated cell death and potential off-target effects.

HSV-tk/GCV has been widely used in gene therapy to induce cell killing of malignant cells in a number of different cancers, in vitro and in vivo, with promising pre-clinical data (reviewed by Karjoo et al.⁶⁰). HSV-tk-activated GCV has been shown to induce cell death through cellular DNA damage and apoptosis,^{61,62} but non-apoptotic death has also been observed.^{63–66} HIV infection leads to T cell depletion through apoptosis in permissive cells (reviewed by Février et al.⁶⁷). However, the majority of HIV-driven CD4⁺ T cell depletion in lymphoid tissue occurs through pyroptosis of nonpermissive cells,⁶⁸ with cell-to-cell viral spread being a crucial factor.⁶⁹ By eliminating the population of HIV-producing cells through HSV-tk/GCV-induced cell death, it is reasonable to hypothesize that T cell depletion by pyroptosis may be reduced. Further research is needed to explore this, and to elucidate the precise mechanism of HSV-tk/GCV-induced cell death in the context of HIV infection.

HSV-tk/GCV-mediated gene therapy has reached phase I/II clinical trials and been shown to be safe, although with limited clinical efficacy. The vector system most widely used for HSV-tk/GCV cancer gene therapy is based on adenoviral vectors (reviewed by Karjoo et al.⁶⁰), but because of the lack of Coxsackie and adenovirus receptor (CAR) on leukocytes,^{70–72} and the recent observation that Ad5-specific T cells are more susceptible to HIV,⁷³ a conventional adenoviral-based vector may not easily be translatable to T cell-targeted HIV RNA gene therapy. Lentiviral vectors may be better candidates. Lentiviral-mediated gene transfer of RNA therapeutics against HIV has been studied both in pre-clinical studies in vivo and in clinical trials (reviewed by Hoxie et al.¹⁶). In addition, lentiviral-mediated gene delivery of the HSV-tk/GCV suicide system has been used in

HIV-infected cells³² and to control graft-versus-host disease in a phase I/II clinical trial in patients with leukemia.⁷⁴ Recently, it has been shown that HIV can infect cells of the hematopoietic lineage *in vivo*,⁷⁵ and novel lentiviral vectors targeting CD34⁺ hematopoietic progenitor cells have been described.⁷⁶ Bone marrow gene therapy is an attractive technology for HIV antiviral therapies (reviewed by Herrera-Carrillo et al.¹⁷), and lentiviral-mediated gene transfer of bone-marrow cells has been demonstrated to be feasible in patients,^{77–79} which could be a potential delivery method for RNA *trans*-splicing technology in HIV-infected individuals. Recently, Zhou et al.⁸⁰ described the generation of an engineered lentiviral vector that can selectively transduce memory CD4⁺ T cells. Thus, in the future, it may be feasible to deliver RNA *trans*-splicing gene therapy systemically. Encouragingly, GlaxoSmithKline (GSK) recently received positive recommendation for market authorization of a gene therapy from the Committee for Medicinal Products for Human Use (CHMP), followed by marketing authorization for use within the European Union by the European Commission. The gene therapy constitutes autologous CD34⁺ cells transduced with a retroviral vector to express adenosine deaminase (ADA), and is targeted to treat severe combined immunodeficiency due to adenosine deaminase deficiency (ADA-SCID).⁸¹ Our future work will focus on targeting latent transcripts which may contain the major splice donor site D1 and generating RNA *trans*-splicing cassettes in suitable vectors for delivery into CD4⁺ T cells for use in *in vitro* and *in vivo* studies.

To conclude, expressing HSV-tk/GCV by RNA *trans*-splicing is an efficient means to selectively induce cell death in HIV-producing cells. RNA *trans*-splicing by either 3' exon replacement targeting splice donor site D4 or 5' exon replacement targeting splice acceptor site A8 were the most efficient HIV splice sites to target. A detailed bioinformatic approach can facilitate design of effective *trans*-splicing constructs.

MATERIALS AND METHODS

To develop RNA *trans*-splicing vectors targeting HIV RNA splice sites, we used the HIV proviral clone pNL4.3 as a template sequence. HIV splice sites of this clone nt 1–9,709 (GenBank: AF324493.2) were predicted using the Splice Site Prediction by Neural Network Server within the Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/seq_tools/splice.html),⁸² with minimum scores for 5' and 3' splice sites at 0.4. The probability for cryptic splice site activation was predicted for pNL4-3 with the CrypSkip software within the Bioinformatics HUSAR server, German Cancer Research Centre (https://genome.inet.dkfz-heidelberg.de/husar/hs_home.html). HIV binding domains were designed on the basis of stepwise MFE calculations of the pNL4-3 genome. The RNA folding energies for the reverse complement of pNL4-3 nt 1–9,709 was predicted with the Foldanalyze software within the Bioinformatics HUSAR server, with a window size of 50 and a step size of 1. Potential regions of HIV binding domains with high free energy and a large number of unpaired bases in the vicinity of predicted and selected HIV splice sites were subjected to MFE RNA secondary structure predictions using Mfold

(<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>),⁸³ with an upper boundary on the number of computing foldings set to 1, and the percentage suboptimality number to 5. The MFE fold and partition function were predicted and calculated with the RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). Structures displaying both a large number of unpaired nucleotides and similar predicted structures with both software packages were selected for further design. Selected structures were then refolded in the backbone 3' or 5' exon replacement cassettes to exclude long-distance effects of folding of the vector backbone, again using the web-servers Mfold and RNAfold and subjected to further design as described below. One or 2 mismatch nucleotides were introduced into binding domain sequences at every 20–25 nucleotides to prevent effects triggered by long double-stranded RNA (reviewed by Chalupnikova et al.⁸⁴). When structured helical domains were present in the RNA secondary structures of binding domains, C-to-U or A-to-G base exchanges that trigger GU or UG wobble base pairs with the target were introduced to resolve duplexes and promote an unstructured conformation.

Sequence Conservation Analysis

Sequence conservation of HIV binding domain target regions were determined with the AnalyzeAlign software³³ using the filtered web LANL database alignment in HIV-1, for either all subtypes or major subtypes. The nucleotide range number refers to residues of HXB2 (GenBank: K03455). A 95% cutoff for calculating the frequency by position was used. For finding variants, the consensus of the alignment was used as master sequence. Nucleotide alignments of pNL4.3 target sequences with subtype B consensus sequence were performed with BLAST Needleman-Wunsch global alignments.³⁴ Gaps within consensus sequences were replaced with N for any nucleotide prior to the alignments.

Cells and Chemicals

293T cells were maintained at 37°C and 5% CO₂ in DMEM (GIBCO/Life Technologies) supplemented with 10% fetal bovine serum (GIBCO, Life Technologies) and penicillin 100 U/mL and streptomycin 100 µg/mL final concentration (GIBCO, Life Technologies).

Thiazolyl Blue Tetrazolium Salt (3-[4,5-Dimethyl-2-thiazolyl-2,5-diphenyl-2H-tetrazolium bromide] [MTT]) (M5655, Sigma) was resuspended in PBS, filtered through a 0.22 µm filter, and stored in aliquots of 5 mg/mL at –20°C. GCV (G2536, Sigma) was resuspended in 0.1 M HCl to a concentration of 39.6 mM, and aliquots were stored at –20°C.

Plasmids, Cloning, and Transfections

3' and 5' exon replacement cassettes were subcloned into the pVAX-1 backbone (Invitrogen) using *Spe*I and *Bbs*I. HIV binding domains were synthesized by gene synthesis (Life Technologies, Invitrogen) and subcloned into 3' exon replacement cassettes in the pVAX-1 backbone (Invitrogen) using *Nhe*I and *Mlu*I (Thermo Scientific) restriction enzyme digestion and subcloning. The *Mlu*I site in the pVAX-1 backbone at position 30 was removed by ethidium

bromide-mediated partial digestion prior to inserting the HIV binding domains into the 3' exon replacement cassettes. HIV binding domains were inserted into the 5' exon replacement cassettes in the pVAX-1 backbone using KpnI and BbvCI (New England Biolabs) restriction enzyme digestion and subcloning.

The infectious HIV clone pNL4-3 was obtained from Dr. Malcom Martin through the AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health (NIH).⁸⁵

Plasmid DNA was transfected into 293T cells with the Trans-IT-LT1 Transfection reagent (Mirus) in DMEM (GIBCO/Life Technologies) free from fetal bovine serum and antibiotics.

PCR and Sequencing

Ten nanograms of template cDNA was amplified by two-step PCR with 1.25 U of GoTaq DNA polymerase (Promega) in a PCR containing 5× GoTaq Reaction Buffer (Promega), 2.5 mM PCR Nucleotide mix (Promega), and 500 nM of forward and reverse primers (see Table S3 for primer sequences). Two-step PCR cycling conditions to detect 3' exon replacement RNA *trans*-splicing amplicons were as follows: 95°C for 2 min, 30 cycles of 95°C, 30 sec, 55°C, 30 sec, 72°C, 30 sec (step 1), followed by another 30 cycles of the step 1 cycling conditions above (step 2). A final extension step was performed at 72°C for 5 min, followed by hold at 4°C. PCR products were visualized on 2% agarose gels, and PCR fragments were purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions (Qiagen). The presence of RNA *trans*-splicing junctions in purified PCR products were confirmed by Sanger sequencing with the same primer pairs used as in the two-step PCR. RNA *trans*-splicing amplicons after 5' exon replacement were detected by two-step PCR as above but with 25 cycles of steps 1 and 2.

qRT-PCR

Total cellular RNA was extracted with the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Extracted RNA was treated with DNase I (New England Biolabs) for 20 min at 37°C. The DNase was inactivated by the addition of EDTA to a final concentration of 5 mM and incubated for 10 min at 75°C. One microgram of DNase-treated RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. cDNA to an equivalent of 10 ng of DNase-treated RNA was used as a template in qPCR. The qPCRs were as follows: 2× TaqMan Fast Advanced Master Mix (Applied Biosystems), 500 nM forward and reverse primers, and 150 nM probe (see Table S3 for primer and probe sequences). The PCR products were amplified at 50°C for 2 min, 95°C for 20 s, and 50 cycles of 95°C for 3 s and 60°C for 30 s on a 7500 Fast Real Time PCR System (Applied Biosystems). A Ct value of 40 was used as a cutoff for background and mis-priming, and values above Ct 40 were omitted during the analysis.

Analysis of Cell Viability by the MTT Assay

293T cells (2×10^4) were plated per well in 96-well plates on day 1 and transfected with 100 ng pNL4.3 and RNA *trans*-splicing construct on days 2 and 3, respectively. 293T cells were treated with one or two doses of 100 μM GCV during the 2 following days. Cell viability was assayed on day 8 by the MTT assay on the basis of a previously published method.⁸⁶ MTT was added to each well to a final concentration of 0.5 mg/ml and incubated at 37°C, 5% CO₂ for 2 hr. Following this, the MTT-containing medium was aspirated, and the formazan crystals were resuspended in acidified (0.04 N HCl) isopropanol/6% Triton X-100 and incubated for 15 min at room temperature to inactivate HIV viral particles. Absorbance was read at 540 and 690 nm on a Multiskan Ascent absorbance plate reader, and the 690 nm reading was subtracted from the 540 nm reading for each well. In Figure S3, absorbance was measured at 595 and 655 nm on a iMark Microplate Reader, BioRad. The 655 nm reading was subtracted from the 595 nm reading for each well.

Statistical Analysis

Statistical analysis was performed with Microsoft Excel 2010 software using the "t test: two sample assuming unequal variances" function, with p values shown for two-tailed analysis.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.omtn.2017.03.004>.

AUTHOR CONTRIBUTIONS

Conceptualization, C.K.I., S.P., V.P., A.M.L.; Methodology, C.K.I., S.P., V.P., A.M.L.; Investigation, C.K.I., S.P., S.M., V.P.; Resources, V.P.; Writing-Original Draft, C.K.I., A.M.L.; Writing-Review & Editing, C.K.I., S.P., S.M., V.P., A.M.L.; Funding Acquisition, V.P., A.M.L.; Supervision, V.P., A.M.L.

CONFLICTS OF INTEREST

S.P. and V.P. declare competing financial interests. A patent application covering the design of *trans*-splicing RNA is pending.

ACKNOWLEDGMENTS

We are grateful for the help and advice from all members of Andrew Lever's laboratory, and we thank Julia Kenyon, Christopher Smith, John Sinclair, Mark Wills, Lars Dölken, and members of the infectious disease unit for helpful discussions (University of Cambridge). We thank Jane Greatorex (PHE), Axel Fun, Hoi-Ping Mok, and Timothy Fitzmaurice for technical advice and Sophia Ho for technical assistance (University of Cambridge). The reagent pNL4-3 was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Malcom Martin.⁸⁵ This work was supported by the National University of Singapore (NUS-Cambridge Start-Up grant number R-182-000-163-646), Medical Research Council grant (G0800142), Medical Research Council Confidence in Concept award, the Biomedical Research Centre, and

a Higher Education Funding Council for England award via the Stevenage Bioscience Catalyst.

REFERENCES

- Van Lint, C., Bouchat, S., and Marcello, A. (2013). HIV-1 transcription and latency: an update. *Retrovirology* 10, 67.
- Wainberg, M.A., Zaharatos, G.J., and Brenner, B.G. (2011). Development of antiretroviral drug resistance. *N. Engl. J. Med.* 365, 637–646.
- Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155, 540–551.
- Lassen, K.G., Bailey, J.R., and Siliciano, R.F. (2004). Analysis of human immunodeficiency virus type 1 transcriptional elongation in resting CD4+ T cells in vivo. *J. Virol.* 78, 9105–9114.
- Tazi, J., Bakkour, N., Marchand, V., Ayadi, L., Aboufirassi, A., and Branlant, C. (2010). Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. *FEBS J.* 277, 867–876.
- Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K., et al. (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* 40, 10345–10355.
- Madsen, J.M., and Stoltzfus, C.M. (2005). An exonic splicing silencer downstream of the 3' splice site A2 is required for efficient human immunodeficiency virus type 1 replication. *J. Virol.* 79, 10478–10486.
- Madsen, J.M., and Stoltzfus, C.M. (2006). A suboptimal 5' splice site downstream of HIV-1 splice site A1 is required for unspliced viral mRNA accumulation and efficient virus replication. *Retrovirology* 3, 10.
- Zheng, Y.H., Yu, H.F., and Peterlin, B.M. (2003). Human p32 protein relieves a post-transcriptional block to HIV replication in murine cells. *Nat. Cell Biol.* 5, 611–618.
- Mandal, D., Feng, Z., and Stoltzfus, C.M. (2008). Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5' splice site. *J. Virol.* 82, 1600–1604.
- Bakkour, N., Lin, Y.L., Maire, S., Ayadi, L., Mahuteau-Betzer, F., Nguyen, C.H., Mettling, C., Portales, P., Grierson, D., Chabot, B., et al. (2007). Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. *PLoS Pathog.* 3, 1530–1539.
- Soret, J., Bakkour, N., Maire, S., Durand, S., Zekri, L., Gabut, M., Fic, W., Divita, G., Rivalle, C., Dauzonne, D., et al. (2005). Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors. *Proc. Natl. Acad. Sci. USA* 102, 8764–8769.
- Wong, R.W., Balachandran, A., Ostrowski, M.A., and Cochrane, A. (2013). Digoxin suppresses HIV-1 replication by altering viral RNA processing. *PLoS Pathog.* 9, e1003241.
- Wong, R.W., Balachandran, A., Haaland, M., Stoilov, P., and Cochrane, A. (2013). Characterization of novel inhibitors of HIV-1 replication that function via alteration of viral RNA processing and rev function. *Nucleic Acids Res.* 41, 9471–9483.
- Campos, N., Myburgh, R., Garcel, A., Vautrin, A., Lapasset, L., Nadal, E.S., Mahuteau-Betzer, F., Najman, R., Fornarelli, P., Tantale, K., et al. (2015). Long lasting control of viral rebound with a new drug ABX464 targeting Rev - mediated viral RNA biogenesis. *Retrovirology* 12, 30.
- Hoxie, J.A., and June, C.H. (2012). Novel cell and gene therapies for HIV. *Cold Spring Harb. Perspect. Med.* 2, 2.
- Herrera-Carrillo, E., and Berkhout, B. (2015). Bone marrow gene therapy for HIV/AIDS. *Viruses* 7, 3910–3936.
- Wally, V., Murauer, E.M., and Bauer, J.W. (2012). Spliceosome-mediated trans-splicing: the therapeutic cut and paste. *J. Invest. Dermatol.* 132, 1959–1966.
- Drude, I., Dombos, V., Vauléon, S., and Müller, S. (2007). Drugs made of RNA: development and application of engineered RNAs for gene therapy. *Mini Rev. Med. Chem.* 7, 912–931.
- Song, M.S., and Lee, S.W. (2006). Cancer-selective induction of cytotoxicity by tissue-specific expression of targeted trans-splicing ribozyme. *FEBS Lett.* 580, 5033–5043.
- Jung, H.S., and Lee, S.W. (2006). Ribozyme-mediated selective killing of cancer cells expressing carcinoembryonic antigen RNA by targeted trans-splicing. *Biochem. Biophys. Res. Commun.* 349, 556–563.
- Song, M.S., Jeong, J.S., Ban, G., Lee, J.H., Won, Y.S., Cho, K.S., Kim, I.H., and Lee, S.W. (2009). Validation of tissue-specific promoter-driven tumor-targeting trans-splicing ribozyme system as a multifunctional cancer gene therapy device in vivo. *Cancer Gene Ther.* 16, 113–125.
- Won, Y.S., and Lee, S.W. (2007). Targeted retardation of hepatocarcinoma cells by specific replacement of alpha-fetoprotein RNA. *J. Biotechnol.* 129, 614–619.
- Won, Y.S., and Lee, S.W. (2012). Selective regression of cancer cells expressing a splicing variant of AIMP2 through targeted RNA replacement by trans-splicing ribozyme. *J. Biotechnol.* 158, 44–49.
- Duarte, S., Carle, G., Faneca, H., de Lima, M.C., and Pierrefite-Carle, V. (2012). Suicide gene therapy in cancer: where do we stand now? *Cancer Lett.* 324, 160–170.
- Brady, H.J., Miles, C.G., Pennington, D.J., and Dzierzak, E.A. (1994). Specific ablation of human immunodeficiency virus Tat-expressing cells by conditionally toxic retroviruses. *Proc. Natl. Acad. Sci. USA* 91, 365–369.
- Caruso, M., and Klatzmann, D. (1992). Selective killing of CD4+ cells harboring a human immunodeficiency virus-inducible suicide gene prevents viral spread in an infected cell population. *Proc. Natl. Acad. Sci. USA* 89, 182–186.
- Caruso, M., Salomon, B., Zhang, S., Brisson, E., Clavel, F., Lowy, I., and Klatzmann, D. (1995). Expression of a Tat-inducible herpes simplex virus-thymidine kinase gene protects acyclovir-treated CD4 cells from HIV-1 spread by conditional suicide and inhibition of reverse transcription. *Virology* 206, 495–503.
- Marcello, A., and Giaretta, I. (1998). Inducible expression of herpes simplex virus thymidine kinase from a bicistronic HIV1 vector. *Res. Virol.* 149, 419–431.
- Smith, S.M., Markham, R.B., and Jeang, K.T. (1996). Conditional reduction of human immunodeficiency virus type 1 replication by a gain-of-herpes simplex virus 1 thymidine kinase function. *Proc. Natl. Acad. Sci. USA* 93, 7955–7960.
- Venkatesh, L.K., Arens, M.Q., Subramanian, T., and Chinnadurai, G. (1990). Selective induction of toxicity to human cells expressing human immunodeficiency virus type 1 Tat by a conditionally cytotoxic adenovirus vector. *Proc. Natl. Acad. Sci. USA* 87, 8746–8750.
- Garg, H., and Joshi, A. (2016). Conditional cytotoxic anti-HIV gene therapy for selectable cell modification. *Hum. Gene Ther.* 27, 400–415.
- Los Alamos National Laboratory. HIV sequence database: AnalyzeAlign. https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html
- National Center for Biotechnology Information. BLAST®: global alignment: Needleman-Wunsch global align nucleotide sequences. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink.
- Walsh, C.E. (2004). New paradigms for gene transfer: RNA trans-splicing and small interfering RNA as therapeutic strategies. *Semin. Hematol.* 41, 297–302.
- Mansfield, S.G., Kole, J., Puttaraju, M., Yang, C.C., Garcia-Blanco, M.A., Cohn, J.A., and Mitchell, L.G. (2000). Repair of CFTR mRNA by spliceosome-mediated RNA trans-splicing. *Gene Ther.* 7, 1885–1895.
- Gruber, C., Koller, U., Murauer, E.M., Hainzl, S., Hüttner, C., Kocher, T., South, A.P., Hintner, H., and Bauer, J.W. (2013). The design and optimization of RNA trans-splicing molecules for skin cancer therapy. *Mol. Oncol.* 7, 1056–1068.
- Bauer, J.W., Murauer, E.M., Wally, V., and Koller, U. (2013). RNA trans-splicing for genodermatoses. *Methods Mol. Biol.* 961, 441–455.
- Murauer, E.M., Koller, U., Hainzl, S., Wally, V., and Bauer, J.W. (2013). A reporter-based screen to identify potent 3' trans-splicing molecules for endogenous RNA repair. *Hum. Gene Ther. Methods* 24, 19–27.
- Patzel, V., and Sczakiel, G. (1998). Theoretical design of antisense RNA structures substantially improves annealing kinetics and efficacy in human cells. *Nat. Biotechnol.* 16, 64–68.

41. Lützelberger, M., Reinert, L.S., Das, A.T., Berkhout, B., and Kjemis, J. (2006). A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. *J. Biol. Chem.* *281*, 18644–18651.
42. Sherrill-Mix, S., Ocwieja, K.E., and Bushman, F.D. (2015). Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. *Retrovirology* *12*, 79.
43. Emery, A., Zhou, S., Pollom, E., and Swanstrom, R. (2017). Characterizing HIV-1 splicing using next generation sequencing. *J Virol.*, Published online January 11, 2017. <http://dx.doi.org/10.1128/JVI.02515-16>.
44. Puttaraju, M., DiPasquale, J., Baker, C.C., Mitchell, L.G., and Garcia-Blanco, M.A. (2001). Messenger RNA repair and restoration of protein function by spliceosome-mediated RNA trans-splicing. *Mol. Ther.* *4*, 105–114.
45. Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O., and Schaal, H. (2014). Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic Acids Res.* *42*, 10681–10697.
46. Mansfield, S.G., Clark, R.H., Puttaraju, M., Kole, J., Cohn, J.A., Mitchell, L.G., and Garcia-Blanco, M.A. (2003). 5' exon replacement and repair by spliceosome-mediated RNA trans-splicing. *RNA* *9*, 1290–1297.
47. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* *460*, 711–716.
48. Li, G., Piampongsant, S., Faria, N.R., Voet, A., Pineda-Peña, A.C., Khouri, R., Lemey, P., Vandamme, A.M., and Theys, K. (2015). An integrated map of HIV genome-wide variation from a population perspective. *Retrovirology* *12*, 18.
49. Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T., and Korber, B. (2002). Diversity considerations in HIV-1 vaccine selection. *Science* *296*, 2354–2360.
50. Benko, D.M., Schwartz, S., Pavlakis, G.N., and Felber, B.K. (1990). A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. *J. Virol.* *64*, 2505–2518.
51. Salfeld, J., Göttlinger, H.G., Sia, R.A., Park, R.E., Sodroski, J.G., and Haseltine, W.A. (1990). A tripartite HIV-1 tat-env-rev fusion protein. *EMBO J.* *9*, 965–970.
52. Göttlinger, H.G., Dorfman, T., Cohen, E.A., and Haseltine, W.A. (1992). The role of the tvn protein and tvn RNA splicing signals in replication of HIV-1 IIIB isolates. *Virology* *189*, 618–628.
53. Marsden, H.S., Haarr, L., and Preston, C.M. (1983). Processing of herpes simplex virus proteins and evidence that translation of thymidine kinase mRNA is initiated at three separate AUG codons. *J. Virol.* *46*, 434–445.
54. Preston, C.M., and McGeoch, D.J. (1981). Identification and mapping of two polypeptides encoded within the herpes simplex virus type 1 thymidine kinase gene sequences. *J. Virol.* *38*, 593–605.
55. Smiley, J.R., Wagner, M.J., Summers, W.P., and Summers, W.C. (1980). Genetic and physical evidence for the polarity of transcription of the thymidine kinase gene of herpes simplex virus. *Virology* *102*, 83–93.
56. Ellison, A.R., and Bishop, J.O. (1996). Initiation of herpes simplex virus thymidine kinase polypeptides. *Nucleic Acids Res.* *24*, 2073–2079.
57. Haarr, L., Marsden, H.S., Preston, C.M., Smiley, J.R., Summers, W.C., and Summers, W.P. (1985). Utilization of internal AUG codons for initiation of protein synthesis directed by mRNAs from normal and mutant genes encoding herpes simplex virus-specified thymidine kinase. *J. Virol.* *56*, 512–519.
58. Berger, A., Maire, S., Gaillard, M.C., Sahel, J.A., Hantraye, P., and Bemelmans, A.P. (2016). mRNA trans-splicing in gene therapy for genetic diseases. *Wiley Interdiscip. Rev. RNA* *7*, 487–498.
59. Monjaret, F., Bourg, N., Suel, L., Roudaut, C., Le Roy, F., Richard, I., and Chartron, K. (2014). Cis-splicing and translation of the pre-trans-splicing molecule combine with efficiency in spliceosome-mediated RNA trans-splicing. *Mol. Ther.* *22*, 1176–1187.
60. Karjoo, Z., Chen, X., and Hatefi, A. (2016). Progress and problems with the use of suicide genes for targeted cancer therapy. *Adv. Drug Deliv. Rev.* *99* (Pt A), 113–128.
61. Ladd, B., O'Konek, J.J., Ostruszka, L.J., and Shewach, D.S. (2011). Unrepairable DNA double-strand breaks initiate cytotoxicity with HSV-TK/ganciclovir. *Cancer Gene Ther.* *18*, 751–759.
62. Beltinger, C., Fulda, S., Kammertoens, T., Meyer, E., Uckert, W., and Debatin, K.M. (1999). Herpes simplex virus thymidine kinase/ganciclovir-induced apoptosis involves ligand-independent death receptor aggregation and activation of caspases. *Proc. Natl. Acad. Sci. USA* *96*, 8699–8704.
63. Halloran, P.J., and Fenton, R.G. (1998). Irreversible G2-M arrest and cytoskeletal reorganization induced by cytotoxic nucleoside analogues. *Cancer Res.* *58*, 3855–3865.
64. Melcher, A., Todryk, S., Hardwick, N., Ford, M., Jacobson, M., and Vile, R.G. (1998). Tumor immunogenicity is determined by the mechanism of cell death via induction of heat shock protein expression. *Nat. Med.* *4*, 581–587.
65. Kaneko, Y., and Tsukamoto, A. (1995). Gene therapy of hepatoma: bystander effects and non-apoptotic cell death induced by thymidine kinase and ganciclovir. *Cancer Lett.* *96*, 105–110.
66. Vile, R.G., Castleden, S., Marshall, J., Camplejohn, R., Upton, C., and Chong, H. (1997). Generation of an anti-tumour immune response in a non-immunogenic tumour: HSVtk killing in vivo stimulates a mononuclear cell infiltrate and a Th1-like profile of intratumoural cytokine expression. *Int. J. Cancer* *71*, 267–274.
67. Février, M., Dorgham, K., and Rebollo, A. (2011). CD4+ T cell depletion in human immunodeficiency virus (HIV) infection: role of apoptosis. *Viruses* *3*, 586–612.
68. Doitsh, G., Galloway, N.L., Geng, X., Yang, Z., Monroe, K.M., Zepeda, O., Hunt, P.W., Hatano, H., Sowinski, S., Muñoz-Arias, I., and Greene, W.C. (2014). Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* *505*, 509–514.
69. Galloway, N.L., Doitsh, G., Monroe, K.M., Yang, Z., Muñoz-Arias, I., Levy, D.N., and Greene, W.C. (2015). Cell-to-cell transmission of HIV-1 is required to trigger pyroptotic death of lymphoid-tissue-derived CD4 T cells. *Cell Rep.* *12*, 1555–1563.
70. Huang, S., Endo, R.I., and Nemerow, G.R. (1995). Upregulation of integrins alpha v beta 3 and alpha v beta 5 on human monocytes and T lymphocytes facilitates adenovirus-mediated gene delivery. *J. Virol.* *69*, 2257–2263.
71. Leon, R.P., Hedlund, T., Meech, S.J., Li, S., Schaack, J., Hunger, S.P., Duke, R.C., and DeGregori, J. (1998). Adenoviral-mediated gene transfer in lymphocytes. *Proc. Natl. Acad. Sci. USA* *95*, 13159–13164.
72. Wan, Y.Y., Leon, R.P., Marks, R., Cham, C.M., Schaack, J., Gajewski, T.F., and DeGregori, J. (2000). Transgenic expression of the Coxsackie/adenovirus receptor enables adenoviral-mediated gene delivery in naive T cells. *Proc. Natl. Acad. Sci. USA* *97*, 13784–13789.
73. Hu, H., Eller, M.A., Zafar, S., Zhou, Y., Gu, M., Wei, Z., Currier, J.R., Marovich, M.A., Kibuuka, H.N., Bailer, R.T., et al. (2014). Preferential infection of human Ad5-specific CD4 T cells by HIV in Ad5 naturally exposed and recombinant Ad5-HIV vaccinated individuals. *Proc. Natl. Acad. Sci. USA* *111*, 13439–13444.
74. Ciceri, F., Bonini, C., Stanghellini, M.T., Bondanza, A., Traversari, C., Salomoni, M., Turchetto, L., Colombi, S., Bernardi, M., Peccatori, J., et al. (2009). Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol.* *10*, 489–500.
75. Nixon, C.C., Vatakis, D.N., Reichelderfer, S.N., Dixit, D., Kim, S.G., Uittenbogaart, C.H., and Zack, J.A. (2013). HIV-1 infection of hematopoietic progenitor cells in vivo in humanized mice. *Blood* *122*, 2195–2204.
76. Frecha, C., Costa, C., Nègre, D., Amirache, F., Trono, D., Rio, P., Bueren, J., Cosset, F.L., and Verhoeyen, E. (2012). A novel lentiviral vector targets gene transfer into human hematopoietic stem cells in marrow from patients with bone marrow failure syndrome and in vivo in humanized mice. *Blood* *119*, 1139–1150.
77. Biffi, A., Montini, E., Lorioli, L., Cesani, M., Fumagalli, F., Plati, T., Baldoli, C., Martino, S., Calabria, A., Canale, S., et al. (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* *341*, 1233158.
78. Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* *341*, 1233151.
79. Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., et al. (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* *326*, 818–823.

80. Zhou, Q., Uhlig, K.M., Muth, A., Kimpel, J., Lévy, C., Münch, R.C., Seifried, J., Pfeiffer, A., Trkola, A., Coulibaly, C., et al. (2015). Exclusive transduction of human CD4+ T cells upon systemic delivery of CD4-targeted lentiviral vectors. *J. Immunol.* *195*, 2493–2501.
81. European Medicines Agency. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/003854/human_med_001985.jsp&mid=WC0b01ac058001d124.
82. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* *4*, 311–323.
83. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* *31*, 3406–3415.
84. Chalupnikova, K., Nejepinska, J., and Svoboda, P. (2013). Production and application of long dsRNA in mammalian cells. *Methods Mol. Biol.* *942*, 291–314.
85. Adachi, A., Gendelman, H.E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M.A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* *59*, 284–291.
86. Pannecouque, C., Daelemans, D., and De Clercq, E. (2008). Tetrazolium-based colorimetric assay for the detection of HIV replication inhibitors: revisited 20 years later. *Nat. Protoc.* *3*, 427–434.
87. Purcell, D.F., and Martin, M.A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* *67*, 6365–6378.
88. Schwartz, S., Felber, B.K., Benko, D.M., Fenyö, E.M., and Pavlakis, G.N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J. Virol.* *64*, 2519–2529.
89. Carrera, C., Pinilla, M., Pérez-Alvarez, L., and Thomson, M.M. (2010). Identification of unusual and novel HIV type 1 spliced transcripts generated in vivo. *AIDS Res. Hum. Retroviruses* *26*, 815–820.
90. Furtado, M.R., Balachandran, R., Gupta, P., and Wolinsky, S.M. (1991). Analysis of alternatively spliced human immunodeficiency virus type-1 mRNA species, one of which encodes a novel tat-env fusion protein. *Virology* *185*, 258–270.
91. Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., and Choi, S.Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE* *6*, e18556.

OMTN, Volume 7

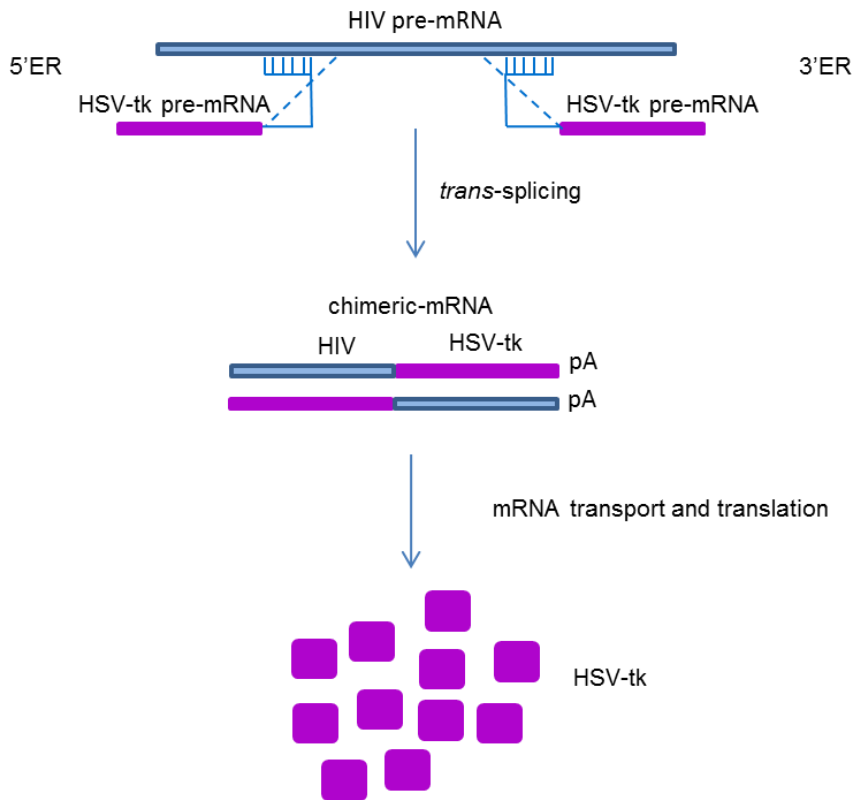
Supplemental Information

Expression of Herpes Simplex Virus Thymidine Kinase/Ganciclovir by RNA *Trans*-Splicing Induces Selective Killing of HIV-Producing Cells

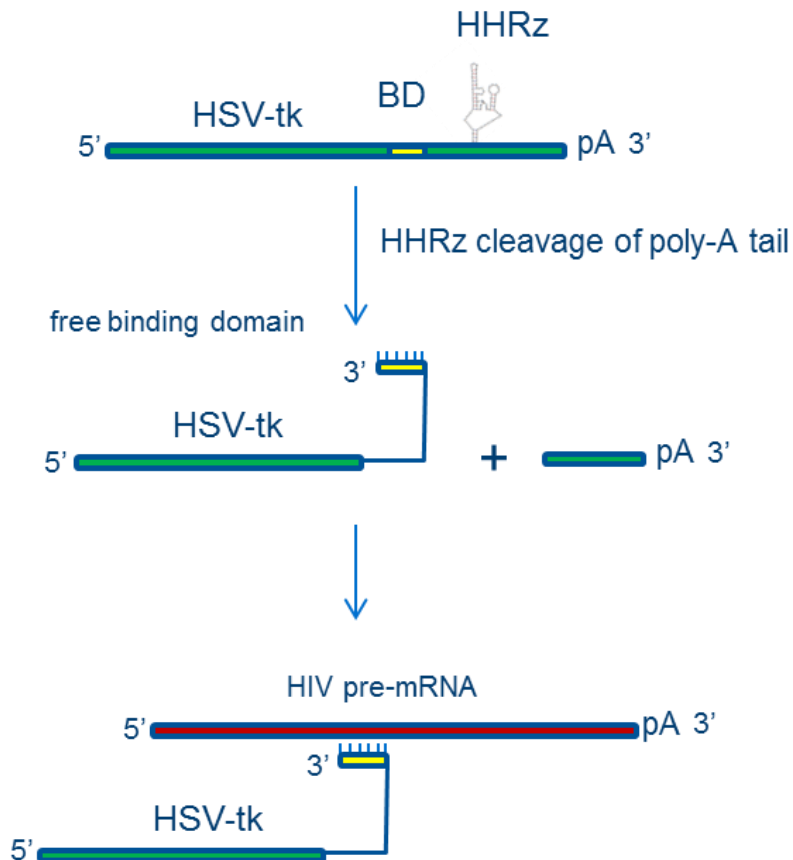
Carin K. Ingemarsdotter, Sushmita Poddar, Sarah Mercier, Volker Patzel, and Andrew M.L. Lever

Supplementary figure 1.

a



b



Supplementary figure 1. Schematic diagram of 3'exon replacement and 5'exon replacement.

(a) RNA *trans*-splicing between HSV-tk pre-mRNA (purple) by 5' exon replacement (5'ER) or 3' exon replacement (3'ER) to HIV pre-mRNA (blue). RNA *trans*-splicing lead to the generation of a chimeric RNA transcript that will be transported to the cytoplasm where it will be translated and processed into HSV-tk protein.

(b) Schematic diagram of HHRz self-cleavage. The HHRz¹ sequence is located downstream of HSV-tk and the binding domain (yellow). HHRz RNA self-cleavage will lead to the removal of the polyA tail and exposure of the binding domain. The HSV-tk *trans*-splicing molecule (green) will anneal to the HIV pre-mRNA (red) via the binding domain prior to the RNA *trans*-splicing reaction as shown in Supplementary figure 1a.

1. Saksmerprome, V, Roychowdhury-Saha, M, Jayasena, S, Khvorova, A, Burke, DH (2004). Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures. *RNA* **10**: 1916-1924.

Supplementary figure 2.

a

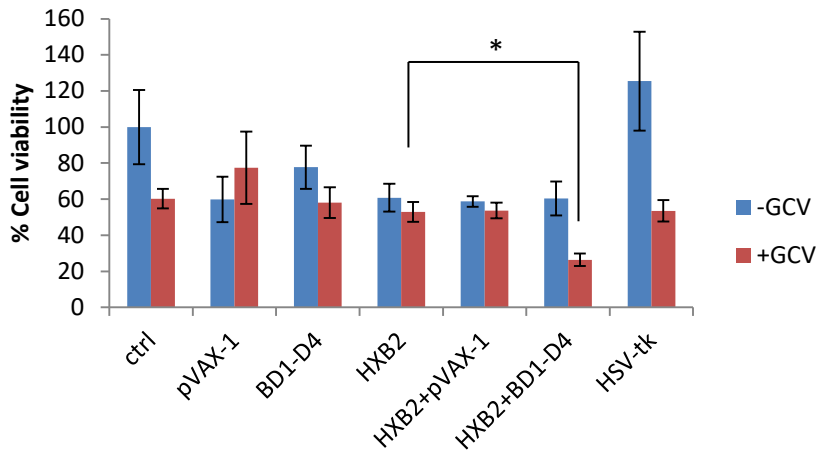


b



Supplementary figure 2. Diagrams of HSV-tk. (a) Diagram of HSV-tk for 3' exon replacement. The first ATG of HSV-tk is deleted and HSV-tk starts at amino acid 2. An introduced mini-intron is shown in grey. (b) Diagram of HSV-tk for 5' exon replacement. The first ATG is present at amino acid position 1. The mini intron is outlined in grey as above. The numbering refers to amino acid numbering in HSV-tk.

Supplementary figure 3.



Supplementary figure 3. Cell viability after RNA *trans*-splicing using HXB2 as HIV-1 target strain.

Cell viability was assessed by MTT assay. 293T cells were sequentially transfected with HXB2 (SVC21) and RNA *trans*-splicing construct (BD1-D4) twenty-four hours apart, treated with 100 μ M GCV on day 1 and 2 after transfection with BD1-D4 (n=8 wells per experimental condition). Cell viability was assessed on day 8, five days post-transfection with the RNA *trans*-splicing construct. Error bars represent the standard deviation (STDEV) of the average % cell viability. P values (two-tailed) are shown for t-test: two sample assuming unequal variances; *p \leq 0.05)

Supplementary Table 1.

a

Binding domain	Splice site target	HIV transcript target	HIV-1 pNL4.3 BD target seq.	HIV splice site location (end of exon)	Nts between binding domain and splice site
BD-D1a	D1a	Cryptic splice site	4781-4830	4721	59
BD1-D4	D4	Vpr, Tat, Rev, Nef, Novel	6141-6204	6044	96
BD2-D4	D4	Vpr, Tat, Rev, Nef, Novel	6194-6240	6044	149
BD-LD1/2,env BD-LD1/2,nef	LD1/LD2	Computer algorithm predicted	9302-9339	LD1-9242, LD2-9287	59 (LD1) 14 (LD2)

b

Binding domain	Splice site target	HIV transcript target	HIV-1 pNL4.3 BD target seq.	HIV splice site location (end of intron)	Nts between binding domain and splice site
BD-EA1/2	EA1/EA2	Computer algorithm predicted	1773-1820	EA1-2139, EA2-2258	EA1-318, EA2-437
BD-A3	A3	Tat	4780-4829	5776	946
BD-A5	A5	Env/Vpu Rev Nef	5899-5947	5975	27
BD-A7	A7	Vpr, Tat, Rev, Nef, Novel	8188-8235	8368	132
BD-A8	A8	Novel Cryptic	8896-8943	9161	217

Supplementary Table 1. Table of RNA *trans*-splicing constructs with target information for 3'exon replacement in a, and 5'exon replacement in b.

Supplementary Table 2.

Number of HIV-1 sequences (filtered web)	Binding domain	HIV-1 target coordinates in HXB2	Number of mutations in max % of sequences	Max % of sequences
	3944 BD1-D4	Vpu 84-147	30	12.75
	3944 BD2-D4	Vpu 137-183	32	21.6
	4632 BD-A7	Env 1974-2021	7	14.79
	2445 BD-A3	Pol 551-600	6	27.55
	2712 BD-A5	Tat 70-118	8	19.72
	4553 BD-A8	Nef 110-157	33	9.92
HIV-1 sequences (major subtypes A, B, C, D, F, G, CRF01, CRF02)	Binding domain	HIV-1 target coordinates in HXB2	Number of mutations in max % of sequences	Max % of sequences
	2960 BD1-D4	Vpu 84-147	23	10.13
	2960 BD2-D4	Vpu 137-183	30	22.62
	3794 BD-A7	Env 1974-2021	5	15.43
	1841 BD-A3	Pol 551-600	6	30.17
	1885 BD-A5	Tat 70-118	8	18.77
	BD-A8	Nef 110-157	33	12.52
HIV-1 sequences (Subtype B)	Binding domain	HIV-1 target coordinates in HXB2	Number of mutations in max % of sequences	Max % of sequences
	1341 BD1-D4	Vpu 84-147	7	20.44
	1341 BD2-D4	Vpu 137-183	27	20.16
	1563 BD-A7	Env 1974-2021	3	19.25
	773 BD-A3	Pol 551-600	4	34.44
	802 BD-A5	Tat 70-118	4	21
	1627 BD-A8	Nef 110-157	9	14.55

Supplementary Table 2. Summary of sequence variant analysis of HIV-1 binding domain target sequences. Sequence variants of HIV-1 binding domain target sequences were analysed using AnalyzeAlign as described in Material and Methods.

Supplementary Table 3.

Name of primer	F or R	Primer sequence	Usage
3'ER reverse	R	GGTTCTCCTCCAGTCTC	to detect t.s junction by conventional PCR/sequencing
5'ER forward	F	TACCGACGATCTGCGAC	to detect t.s junction by conventional PCR/sequencing
5'ER BD-A3 reverse	R	TCGAGTAACGCCTATTCTG	to detect t.s junction by conventional PCR
5'ER BD-A7 reverse	R	ACCAATATTTGAGGGCTTC	to detect t.s junction by conventional PCR/sequencing
5'ER BD-A8 reverse	R	TTATCTGGCTCAACTGGTACTA	to detect t.s junction by conventional PCR/sequencing
5'ER BD-A5 reverse	R	TGCTTTGATAGAGAAGCTTGATGAG	to detect t.s junction by conventional PCR/sequencing
3'ER BD1/2-D4 forward	F	GAGACAGCGACGAAGAGCTCAT	to detect RNA trans-splicing transcripts by qRT-PCR, conventional PCR, sequencing
3'ER reverse	R	CAGGGTTCTCCTCCAGTCTC	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER BD-A7 reverse	R	CACCTTCTTCTTATTCTTCGG	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER BD-A3 reverse	R	CCTCTGTCGAGTAACGCCTATTCT	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER BD-A5 reverse	R	AGTCTGACTGTTCTGATGAGCTCTTC	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER BD-A8 reverse	R	TCTACCTTATCTGGCTCAACTGGTACTA	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER forward	F	TGGCGGCACGTTTG	to detect RNA trans-splicing transcripts by qRT-PCR
3'ER probe	Probe	6FAM-GAAGCGGAGCTACTAACTCAGCCTGCTGA-BHQ1	to detect RNA trans-splicing transcripts by qRT-PCR
5'ER probe	Probe	6FAM-CGGGAGATGGGGGAGGCGAA-BHQ1	to detect RNA trans-splicing transcripts by qRT-PCR

Supplementary Table 3. Primers used for PCR, qRT-PCR and sequencing.