



OPEN

## HIV-1 sequences in lentiviral vector genomes can be substantially reduced without compromising transduction efficiency

Helin Sertkaya<sup>1</sup>, Mattia Ficarelli<sup>1</sup>, Nathan P. Sweeney<sup>2</sup>, Hannah Parker<sup>1</sup>, Conrad A. Vink<sup>2</sup>✉ & Chad M. Swanson<sup>1</sup>✉

Many lentiviral vectors used for gene therapy are derived from HIV-1. An optimal vector genome would include only the viral sequences required for transduction efficiency and gene expression to minimize the amount of foreign sequence inserted into a patient's genome. However, it remains unclear whether all of the HIV-1 sequence in vector genomes is essential. To determine which viral sequences are required, we performed a systematic deletion analysis, which showed that most of the *gag* region and over 50% of the *env* region could be deleted. Because the splicing profile for lentiviral vectors is poorly characterized, we used long-read sequencing to determine canonical and cryptic splice site usage. Deleting specific regions of *env* sequence reduced the number of splicing events per transcript and increased the proportion of unspliced genomes. Finally, combining a large deletion in *gag* with repositioning the Rev-response element downstream of the 3' R to prevent its reverse transcription showed that 1201 nucleotides of HIV-1 sequence can be removed from the integrated vector genome without substantially compromising transduction efficiency. Overall, this allows the creation of lentiviral vector genomes that contain minimal HIV-1 sequence, which could improve safety and transfer less viral sequence into a patient's DNA.

Retroviral vectors are used for a wide variety of cellular therapies<sup>1</sup>. While they were originally based on gammaretroviruses, human immunodeficiency virus type 1 (HIV-1) has been adapted to produce lentiviral vectors<sup>2</sup>. Lentiviral vectors are particularly promising gene therapy tools due to their ability to facilitate efficient transduction and long-term stable transgene expression in both dividing and non-dividing cells<sup>2</sup>. They also have lower risk of genotoxicity due to insertional mutagenesis than gammaretroviral vectors. Lentiviral vectors are currently used for treatments that involve gene transfer into hematopoietic stem cells and lymphocytes. One example is Kymriah, an autologous chimeric antigen receptor (CAR) T cell immunotherapy that uses a HIV-1 based lentiviral vector to treat acute lymphoblastic leukemia<sup>3</sup>.

HIV-1 contains nine open reading frames: the structural proteins Gag, Pol and Env, the accessory proteins Vif, Vpr, Vpu and Nef that counteract innate and adaptive immunity and the regulatory proteins Tat and Rev that control viral gene expression<sup>4–6</sup>. In addition, the HIV-1 genome contains many linear or structural *cis*-acting RNA elements that regulate reverse transcription, RNA transcription, pre-mRNA splicing, intron-containing RNA nuclear export, translation and genomic RNA packaging<sup>6,7</sup>. The first generation lentiviral vector system used three plasmids to express the viral proteins required to make infectious particles and the genome<sup>8</sup>. The plasmid encoding the vector genome contained all of the signals necessary for its transcription, dimerization, encapsidation, reverse transcription and integration. It also had an internal promoter to express the transgene of interest. The second plasmid was the packaging vector containing intact reading frames for all of the HIV-1 proteins except Env to express Gag and Gag-Pol as well as the accessory and regulatory proteins<sup>8</sup>. The third plasmid expressed a viral protein to mediate entry into the target cell, often the vesicular stomatitis virus glycoprotein (VSV-G)<sup>8</sup>.

Since their initial development, considerable progress has been made in lentiviral vector constructs to reduce the number of viral proteins expressed by the packaging plasmids and the amount of viral sequence in the vector genome. Second generation packaging constructs eliminated Vif, Vpr, Vpu and Nef expression while maintaining Gag, Gag-Pol, Tat and Rev<sup>9</sup>. In addition, the woodchuck hepatitis virus post-transcriptional regulatory

<sup>1</sup>Department of Infectious Diseases, King's College London, London SE1 9RT, UK. <sup>2</sup>Cell & Gene Therapy Platform, Medicinal Science and Technology, GSK, Stevenage SG1 2NY, UK. ✉email: conrad.x.vink@gsk.com; chad.swanson@kcl.ac.uk

element (WPRE) was introduced at the 3' end of the vector genome<sup>10</sup>. While the function of this element remains unclear, it may improve vector titre and/or promote transgene expression by promoting RNA nuclear export or stability<sup>10–16</sup>. Third generation lentiviral vector systems further reduced the amount of viral sequence in the packaging and genome plasmids<sup>17</sup>. The requirement for Tat was eliminated by incorporating a constitutively active promoter in place of the HIV-1 5' U3 in the vector genome plasmid. Because Rev is required for Gag-Pol expression and vector titre, this protein was expressed *in trans* allowing a packaging construct containing only the *gag* and *pol* genes plus the *cis*-acting Rev-response element (RRE). These modifications reduced biosafety concerns, including the potential to produce replication-competent lentivirus (RCL)<sup>17,18</sup>. If recombination occurred between the packaging and genome constructs, RCL particles could theoretically be produced and spread beyond the intended target cells. Therefore, the third generation, four-plasmid system decreases the likelihood of potential recombination events between the vector genome and the *gag-pol* mRNA that could create RCL particles, though this is still a potential concern in large-scale production systems. Furthermore, a large portion of the HIV-1 3' U3 enhancer/promoter sequence was deleted in the vector genome to create self-inactivating (SIN) vectors<sup>18</sup>. This prevents transfer of this sequence to the 5' end during reverse transcription, thereby eliminating the enhancer and promoter elements in the integrated genome and reducing the potential for activating expression of surrounding genes in target cells.

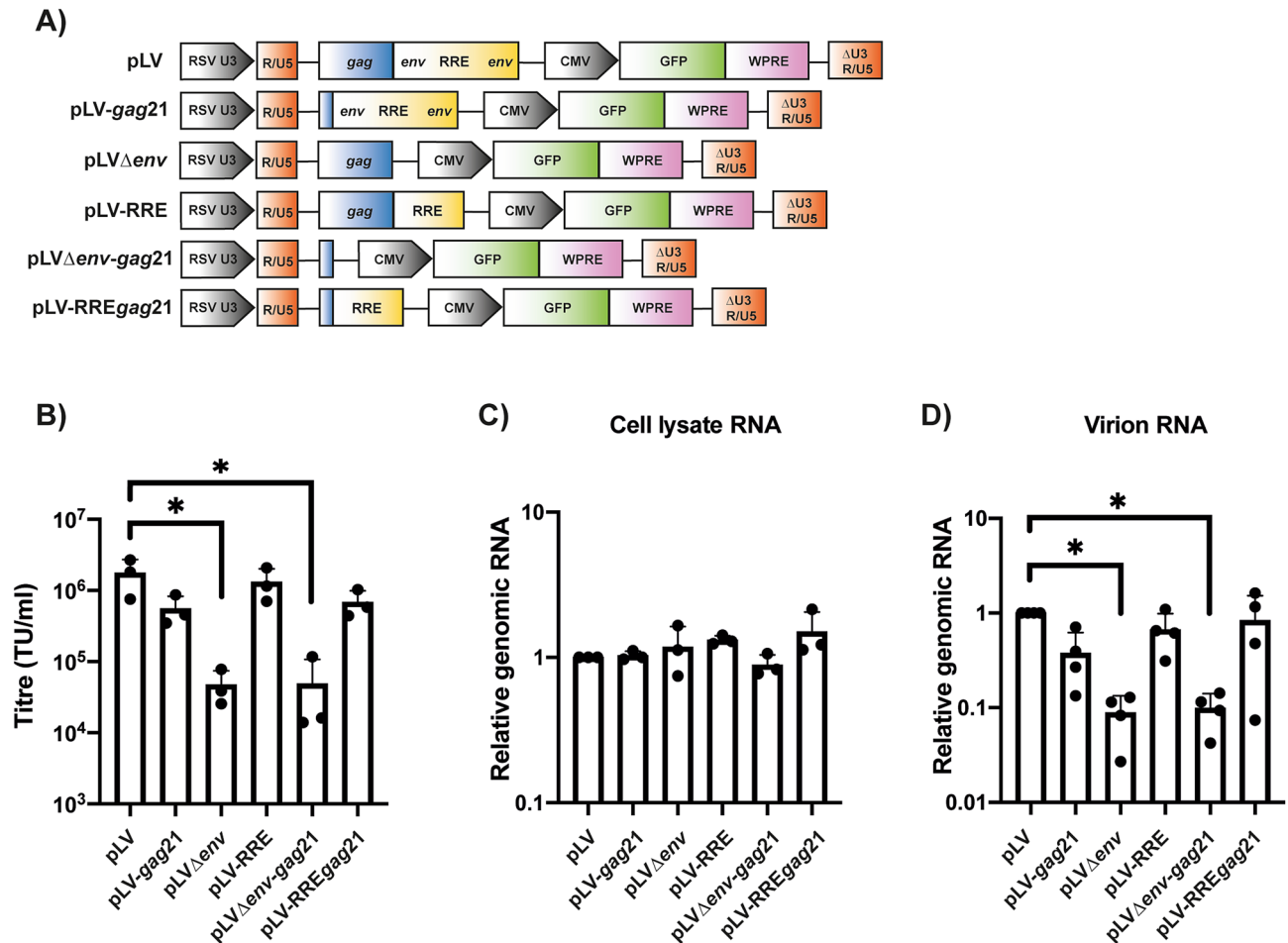
Importantly, vector genome constructs often still contain over 1.5 kilobase (kb) of HIV-1 sequence, including substantial regions of *gag* and *env*, and it is not clear whether all of this is required for transduction and transgene expression. To further decrease potential RCL production and potential interactions between the integrated provirus and host cell genome, vector genomes with Cre/loxP-mediated excision of HIV-1 sequences were designed<sup>19</sup>. However, this introduces the *cre* gene into the genomic RNA and substantially reduces the space available for promoter/transgene cassettes. Another approach was used for the LTR1 vector in which the genome was engineered so that the packaging sequences and RRE are present at the 3' end of the genomic RNA instead of the 5' end and are not reverse transcribed, thereby eliminating them from the provirus<sup>20</sup>. This reduced the HIV-1 sequence in the vector genome to only 4.8% of the full-length genome. However, optimal titre required a large intron derived from EF1 $\alpha$  to be inserted at the 5' end of the vector genome, which highlights that little is known about how splicing in the canonical vectors is regulated and how specific splice sites and nuclear export elements modulate vector titre.

The aim of this study was to characterize the functional role of the viral sequences in the lentiviral vector genome and determine which portions could be removed to produce a smaller and potentially safer vector. We found that 850 nt of HIV-1 *gag* and *env* sequence could be removed from the vector genome without compromising transduction efficiency. This reduces the amount of HIV-1 sequence that could potentially recombine with the packaging vector to produce RCL particles. Because the splicing profile of lentiviral vector genomes are poorly characterized, we used Oxford Nanopore sequencing to determine the proportion of vector genomes that are spliced as well as the donor and acceptor sites that are utilized. This analysis showed that deleting 507 nt of *env*, which includes the HIV-1 canonical splice acceptor 7 (SA7), increased the proportion of unspliced genomic RNAs. Finally, we show that deleting most of *gag* can be combined with moving the RRE to 3' end of the genome so that it is not be reverse transcribed and integrated in the target cell. Overall, removing non-essential HIV-1 sequences from the vector genome may lead to a smaller, safer system that transfers less viral sequence into a patient's genome.

## Results

**850 nt of the HIV-1 *gag* and *env* regions can be deleted in the genomic RNA without compromising vector titre.** To determine which sequences derived from HIV-1 in the lentiviral vector genome plasmid (pLV) are required for efficient transduction, we analyzed the effect of deleting the *gag* and *env* regions either alone or in combination. Briefly, pLV<sup>21</sup> contains the Rous sarcoma virus enhancer/promoter<sup>17</sup>, the HIV-1 5' R-U5-leader region that contains sequences required for genomic RNA dimerization and encapsidation as well as splice donor 1 (SD1)<sup>8,22,23</sup>, a 364 nucleotide (nt) HIV-1 *gag* sequence with a frameshift mutation<sup>8</sup>, an 858 nt HIV-1 *env* sequence<sup>8</sup> that contains the 351 nt RRE<sup>6,24</sup> and the splice acceptor used for most fully spliced viral RNAs (SA7)<sup>23</sup>, the HIV-1 central polypurine tract (cPPT) that increases transduction efficiency<sup>25</sup>, a CMV-GFP reporter sequence, the WPRE<sup>10</sup> and the HIV-1 3' LTR with a 400-nucleotide deletion that abolished its promoter activity<sup>18</sup>. Sequences within *gag* have been proposed to regulate HIV-1 genomic RNA encapsidation, pre-mRNA splicing, nuclear export and reverse transcription<sup>23,26–29</sup>. Previous studies have reported that deletions in the *gag* region present in lentiviral vectors reduced titre<sup>30–32</sup>, though a 75 nt deletion in the 3' end of the region was tolerated<sup>33</sup>. However, we have shown that when nt 22–378 in *gag* were deleted in the context of full-length HIV-1, production of infectious VSV-G pseudotyped virus was similar to wild-type HIV-1<sup>34</sup>. While the amino acid sequence within this region is required for HIV-1 Env incorporation due to the interaction between Gag and the C-terminal tail of Env, this is dispensable when a heterologous envelope protein is expressed<sup>35–37</sup>, implying that viral RNA elements in *gag* outside of the dimer structure<sup>38</sup> are not required for infectious virus production. Therefore, we made pLV-*gag*21, in which 343 nt of the *gag* sequence is removed leaving only the first 21 nt containing a highly conserved sequence that is incorporated into the genomic RNA dimer structure (Fig. 1A)<sup>38–41</sup>. Deleting these 343 nt did not substantially decrease vector titre, intracellular genomic RNA expression or virion-associated genomes (Fig. 1B–D).

To determine the functional relevance of the RNA sequences in the *env* region, we first deleted it (pLV $\Delta$ *env*, Fig. 1A), which caused an ~40-fold decrease in vector titre and ~tenfold decrease in genomic RNA in virions (Fig. 1B,D). When 343 nt of *gag* and all of *env* were deleted (pLV $\Delta$ *env-gag*21), vector titre and genomic RNA in virions were also substantially reduced. When the 351 nt RRE was inserted in place of *env* to create pLV-RRE or pLV-RRE*gag*21, vector titre and genomic RNA abundance was similar to pLV (Fig. 1B–D). This indicates



**Figure 1.** Specific deletions in *gag* and *env* can be made without compromising transduction efficiency. **(A)** Schematic representation of the lentiviral genome vectors. **(B–D)** Lentiviral vectors were produced in HEK293T cells with the indicated genome, packaging construct pCMVΔR8.91 and envelope construct pVSV-G. **(B)** Infectious titres were determined in transduced HEK293T cells by flow cytometry for GFP-positive cells. The bar chart shows the average values of three independent experiments. Data are shown as mean  $\pm$  SD, \* $p < 0.05$  as determined by a one-way ANOVA test. **(C,D)** Relative genomic RNA abundance within the whole cell lysates **(C)** and virions **(D)** were determined by qRT-PCR. The bar charts show the average values of three **(C)** or four **(D)** independent experiments normalized to the value obtained for pLV. Data are shown as mean  $\pm$  SD, \* $p < 0.05$  as determined by a one-way ANOVA test.

that while the RRE potently promotes vector titre, other sequences in *env*, including SA7, are not required. This allows 507 nt to be deleted from the *env* region and 343 nt to be deleted from the *gag* region without substantially inhibiting transduction efficiency. In total, this removes 850 nt of HIV-1 sequence from the vector genome.

**The RRE and WPRE are required for optimal vector titre and transgene expression but do not control cytoplasmic genomic RNA abundance.** All lentiviral vectors derived from HIV-1 have the major splice donor (SD1) in the 5' leader region and this sequence is part of the genomic RNA dimer structure<sup>42</sup>. Many lentiviral vectors also have SA7 in the *env* sequence<sup>8</sup>. This leads to the vectors containing a short functional intron. Because intron-containing RNAs are usually retained in the nucleus and targeted for degradation, a nuclear export element is required for optimal cytoplasmic expression of the vector genome<sup>43</sup>. Lentiviral vectors often contain two viral RNA nuclear export elements that could function in the producer cell: the HIV-1 RRE, which interacts with Rev to mediate export through the CRM1 pathway, and the WPRE<sup>6,14,15,24</sup>. However, pLV-RRE does not contain the HIV-1 SA7 sequence and therefore does not have a canonical splice acceptor site. Therefore, we analyzed the effect of deleting the WPRE and the RRE on vector titre and genomic RNA abundance in this vector. To validate our protocol for analyzing cytoplasmic RNA abundance, we used a Rev-dependent subgenomic HIV-1 Gag-Pol-Vif construct<sup>44</sup> that contains SD1 and the splice acceptor (SA) sites used for *vif* and *vpr* mRNAs, SA1 and SA2. As expected, Gag expression from this vector required Rev (Supplementary Fig. S1A). In the absence of Rev, there was a large decrease in cytoplasmic *gag-pol* RNA abundance and a moderate decrease in nuclear RNA abundance, which reflects decreased intron-containing RNA nuclear export and stability (Supplementary Fig. S1B–D). Therefore, an HIV-1 construct that produces a transcript with one

**Figure 2.** Deletion of the RRE or WPRE reduces transduction efficiency but does not affect cytoplasmic genomic RNA abundance. **(A)** Schematic representation of the lentiviral vector genomes. **(B–E)** Lentiviral vectors were produced in HEK293T cells with the indicated genome, packaging construct pCMV $\Delta$ R8.91 and envelope construct pVSV-G. **(B)** Infectious titres were determined in transduced HEK293T cells by flow cytometry for GFP-positive cells. The bar chart shows the average values of three independent experiments. Data are shown as mean  $\pm$  SD, \* $p < 0.05$  as determined by a one-way ANOVA test. **(C)** The relative mean fluorescence intensity (MFI) of GFP-positive cells was determined, indicating the expression efficiency for the transgene. **(D)** Western blotting for the cytoplasmic marker,  $\alpha$ -tubulin, and the nuclear marker, lamin-B1, was performed on the HEK293T fractions as a control for fractionation for the whole cell lysate (WL), cytoplasmic fraction (CF) and nuclear fraction (NF) using anti- $\alpha$ -tubulin and anti-lamin-B1 antibodies for the pLV sample. The  $\alpha$ -tubulin and lamin-B1 blots are from independent gels. **(E)** Relative genomic RNA abundance within the cytoplasmic fraction was determined by qRT-PCR. **(F)** Relative genomic RNA abundance in the virions was determined by qRT-PCR. **(C,E,F)** The bar charts show the average values of three independent experiments normalized to the value obtained for pLV-RRE. Data are shown as mean  $\pm$  SD, \* $p < 0.05$  as determined by a one-way ANOVA test.

canonical viral splice donor and two splice acceptors requires Rev for cytoplasmic accumulation of the intron-containing RNA.

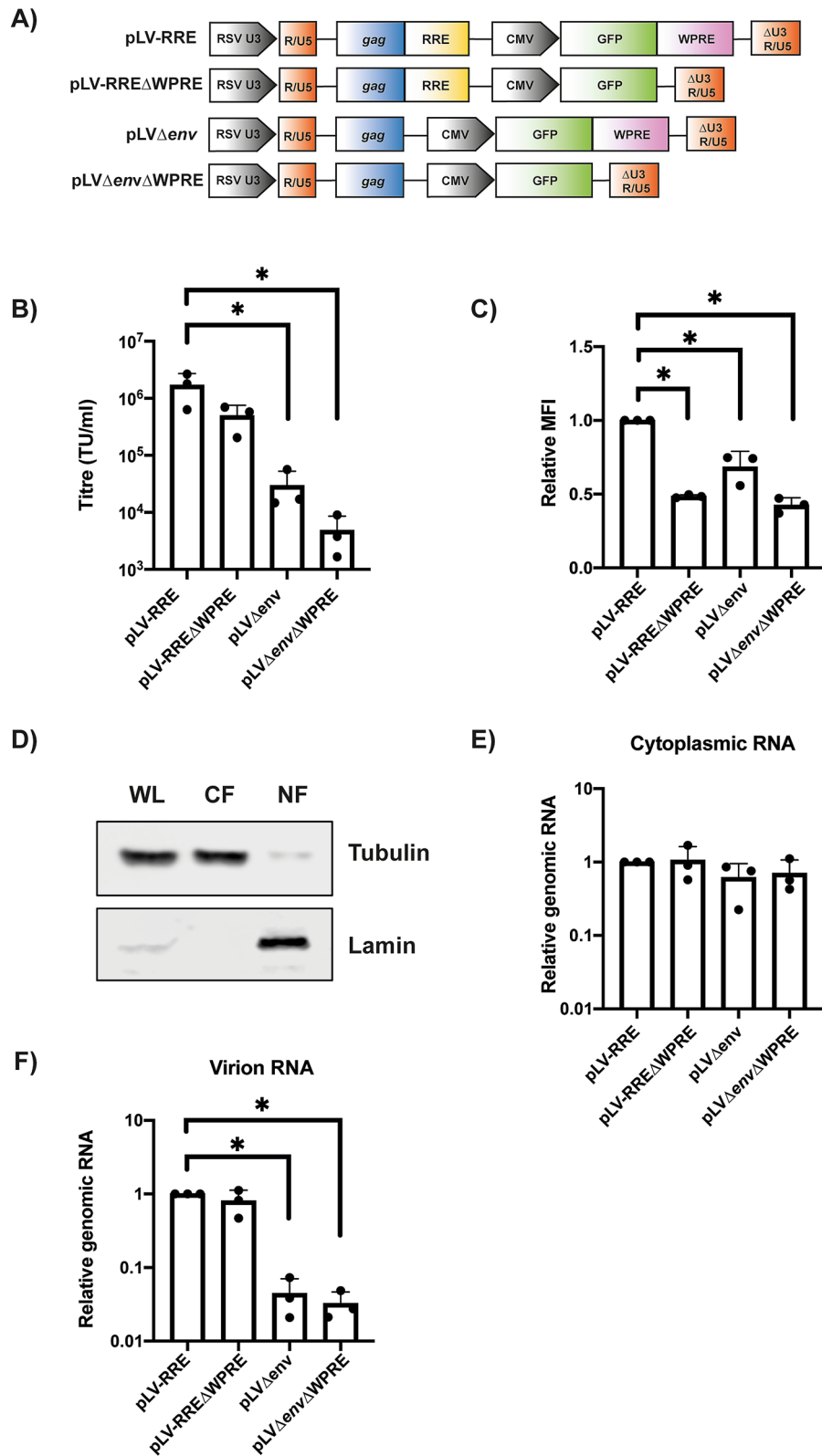
When the WPRE was deleted in the lentiviral vector (pLV-RRE $\Delta$ WPRE, Fig. 2A), there was a small, non-significant decrease in vector titre and a decrease in transgene expression in the target cell (Fig. 2B,C). However, there was no decrease in the amount of cytoplasmic or packaged genomic RNA (Fig. 2D–F), suggesting that this element does not contribute to the cytoplasmic accumulation of the viral RNA. Of note, low nuclear RNA levels prevented reproducible analysis of this compartment for this vector. When both the RRE and WPRE were deleted (pLV $\Delta$ env $\Delta$ WPRE), there was a > 300-fold decrease in vector titre, which correlated with a large decrease in the amount of genomic RNA in virions as well as a decrease in transgene expression in the target cell (Fig. 2B–E). Importantly, there was no significant decrease in cytoplasmic genomic RNA abundance implying that neither the RRE nor WPRE are required to promote cytoplasmic accumulation of this transcript. Interestingly, Rev has been reported to be required for genomic RNA packaging in addition to its role in intron-containing RNA nuclear export<sup>45–47</sup>, and our data is consistent with this additional function for Rev.

**Deleting 507 nt in *env* increases the proportion of unspliced genomic RNAs.** Next, we analyzed the titre for pLV, pLV-*gag*21, pLV-RRE and pLV-RRE*gag*21 (Fig. 3A–B) using a ddPCR assay in the HEK293Tsa cell line<sup>48</sup>, which produces high titre vectors. All four vectors had similar titres, further indicating that deleting regions in *gag* or *env* did not significantly impair vector production. Deletion of the entire *env* region in the vector genome (pLV $\Delta$ env) decreased titre approximately fourfold. This is a smaller effect size compared to similar experiments in HEK293T cells (Fig. 1D), indicating that the HEK293Tsa cells are less sensitive to alterations in the vector. However, it should be noted that this is a producer cell line used to make clinical-grade lentiviral vectors and most further experiments were performed in this cell line.

RNAs transcribed from lentiviral vector genome plasmids can be spliced using SD1 and SA7 as well as cryptic splice donors and acceptors. These events decrease the amount of full-length RNA available for packaging into virions and could lead to packaging of genomes containing internal deletions. Because there are splice sites or splicing enhancer/silencer elements in both *gag* and *env*<sup>23</sup>, we compared how deleting each region affected intracellular genomic RNA splicing in RNA isolated from transfected HEK293Tsa cells. Oxford Nanopore direct cDNA sequencing was used to characterize the splice variants because the long-read length allows full-length transcripts to be analyzed. Even though this sequencing approach initiates at the 3' end, > 500 near-full length transcripts (defined as reads containing at least one nucleotide upstream of SD1) for each vector were sequenced. These reads were used for further analysis, which allowed a comprehensive characterization of the splicing events and 3' cleavage site for each vector genome. A substantial number of reads (~ 30%) contained HIV-1 sequence beyond the 3' R (Figs. 4 and 5, Table 1, Supplementary Table S1), indicating that 3' end cleavage did not occur at the standard site for HIV-1 transcripts. This has been previously observed and is likely due to elimination of upstream poly(A) regulatory elements by the U3 deletion in the SIN vector<sup>19,49,50</sup>. Of note, there is a SV40 poly(A) signal downstream of the HIV-1 3' LTR in these vector plasmids, and 3' cleavage of the readthrough transcripts likely terminates there.

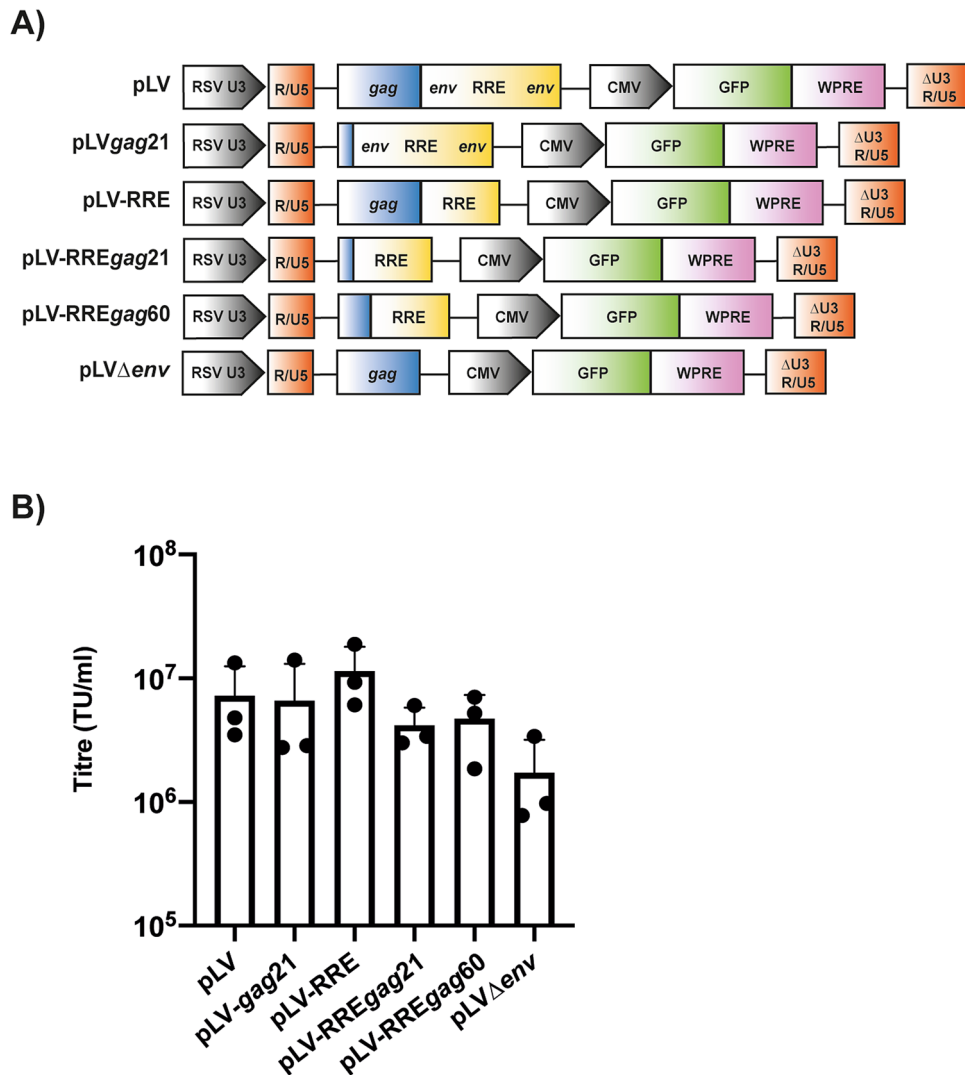
For pLV, pLV-*gag*21, pLV-RRE and pLV-RRE*gag*21, SD1 and 11 cryptic splice donors (CDs) were used in conjunction with SA7 or 29 cryptic splice acceptors (CAs) (Supplementary Table S1). Importantly, comparing pLV to pLV-RRE showed that deleting 507 nt from *env*, including SA7, decreased the number of splicing events per transcript and increased the proportion of unspliced genomic RNAs approximately threefold in the cell (Fig. 4, Table 1). Splicing events from SD1 to three predominant cryptic splice acceptors (CA6, CA9, CA10) were present in pLV-RRE, including a predominant cryptic acceptor (CA10) in GFP, but these were utilized less frequently than splicing from SD1 to SA7 plus cryptic splice acceptors for pLV. In addition, the internal splicing events from CD3 in pLV, which represent ~ 15% of the total splicing events, are not present in pLV-RRE because this sequence has been deleted. Therefore, deleting 507 nt of *env* decreases the number of splicing events and increases the proportion of unspliced transcripts.

RNA transcribed from pLV and pLV-*gag*21 had similar splicing patterns with many splicing events from SD1 to SA7 as well as splicing events using cryptic donors and acceptors (Figs. 4A and 5A, Table 1). Both vectors had a similar number of splicing events per transcript and proportion of unspliced genomes in the cells (Table 1). However, deleting 507 nt of *env* in the context of the *gag* deletion in pLV-RRE*gag*21 decreased the



number of splicing events per transcript and increased the percentage of unspliced transcripts compared to pLV-*gag*21 (Fig. 5A,B, Table 1). This shows that deleting 850 nt of *gag* and *env* does not compromise the vector titre (Fig. 3B) and increases the proportion of unspliced genomic RNAs in cells (Fig. 5B). Because it has previously been reported that a stem-loop structure in nts 26–58 in *gag* helps stabilize the leader region structure for packaging<sup>46</sup>, we characterized a vector with this region of *gag* (pLV-RRE*gag*60). It had a similar titre and splicing pattern to pLV-RRE*gag*21 (Figs. 3B and 5C), which may indicate that another region of the pLV-RRE*gag*21 vector genomic RNA sequence has compensated for it. Overall, deleting 850 nt from the *gag* and *env* regions removes a





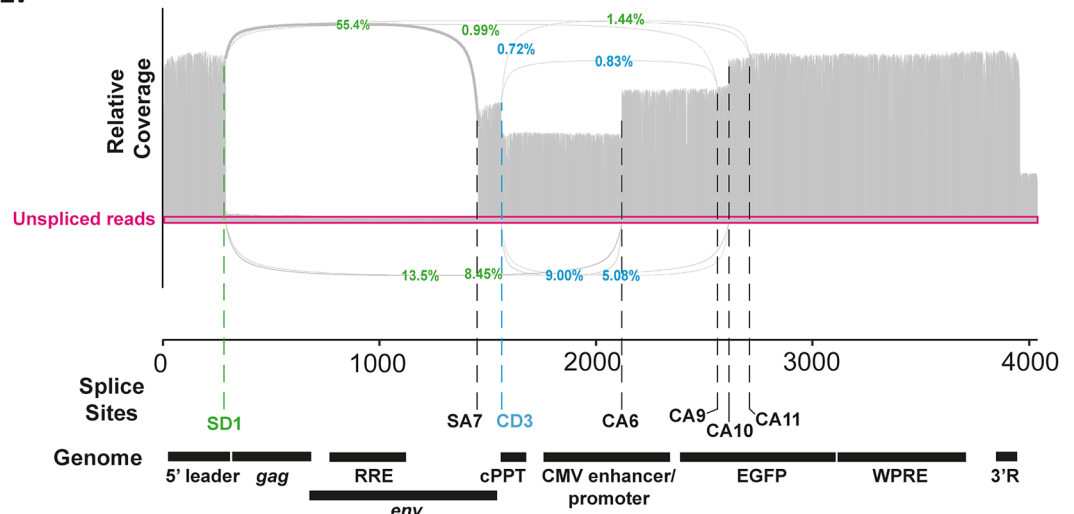
**Figure 3.** 850 nt can be eliminated from the lentiviral vector genome without decreasing titre in HEK293Tsa cells. **(A)** Schematic representation of the lentiviral vector genomes. **(B)** Lentiviral vectors were produced in HEK293Tsa cells with the indicated vector genome, packaging construct pCMVΔR8.91 and envelope construct pVSV-G. Infectious titres were determined in transduced HEK293Tsa cells by digital droplet PCR (ddPCR). The bar charts show the average values of three independent experiments. Data are shown as mean  $\pm$  SD. A one-way ANOVA test did not identify any samples as  $p < 0.05$ .

large portion of the HIV-1 sequence from the lentiviral vector genome without significantly reducing its titre and increases the proportion of unspliced transcripts available for packaging. However, vector titre is not increased in the context of a larger proportion of unspliced transcripts (Fig. 3B), possibly indicating that genome availability is not limiting for infectious virus production under these conditions, or the increase in the proportion of unspliced transcripts is not large enough to substantially increase titre (see Discussion).

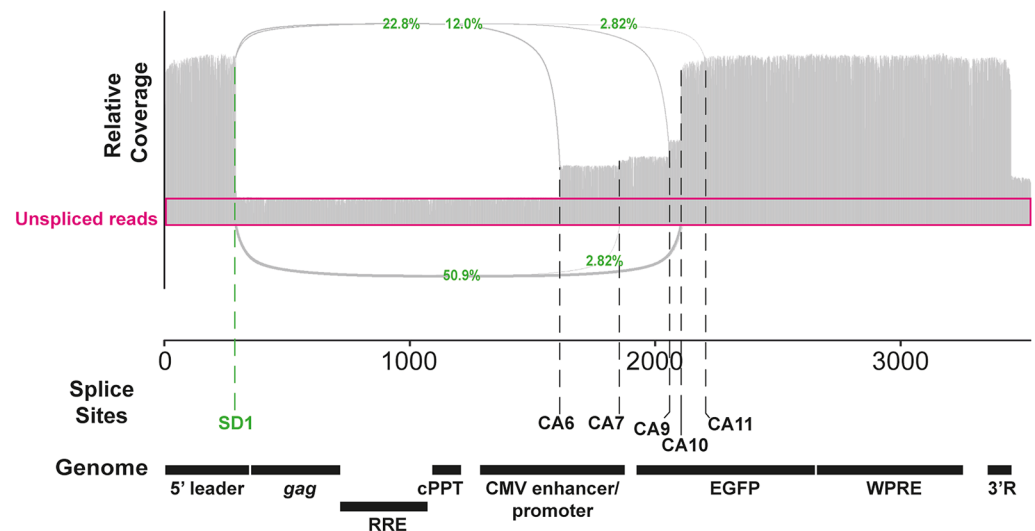
**The RRE can be moved to the 3' end of the genome so that it is not reverse transcribed and combined with the 343 nt deletion in gag without substantially compromising vector titre.** To determine if the amount of HIV-1 RNA in the vector genome could be further reduced, we replaced the RRE with four copies of the Mason-Pfizer monkey virus (M-PMV) constitutive transport element (CTE)<sup>51–53</sup> (pLV-4xCTE, Fig. 6A). This retroviral nuclear export element uses the NXF1 nuclear export pathway, which is the route that most cellular mRNA take, instead of the CRM1 pathway used by Rev-dependent HIV-1 intron-containing RNAs<sup>43</sup>. Of note, we have previously shown that four copies of the CTE promotes protein expression in human cells better than one copy<sup>53</sup>. The vector titre for this construct in HEK293T cells was lower than pLV-RRE but higher than pLVΔenv (Fig. 6B), indicating that the RRE is more efficient for lentiviral vector production than an RNA nuclear export element from a different retrovirus.

We also tested the effect of moving the RRE downstream of the HIV-1 3' R with the SV40 poly(A) signal in place of U5 (pLV-3'RRE, Fig. 6A). With this modification, the RRE is present in the genomic RNA in the producer

## A) pLV



## B) pLV-RRE

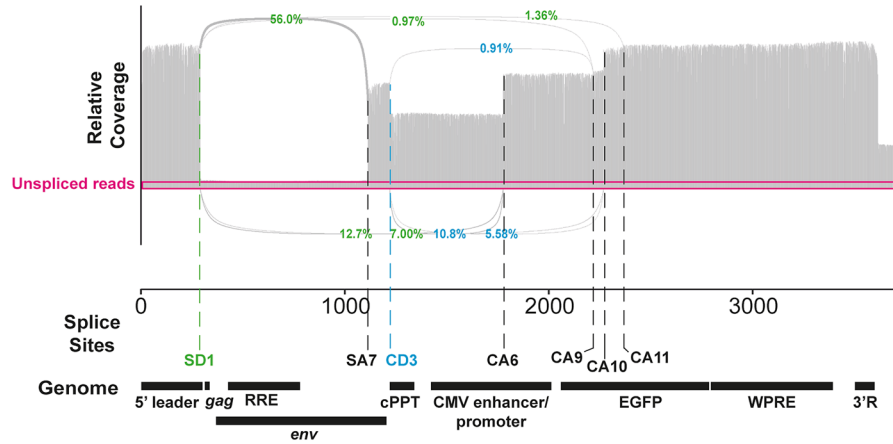


**Figure 4.** Deletion of 507 nt from the *env* region increases the proportion of unspliced genomic RNA in HEK293Tsa cells. The usage of canonical and cryptic splice donors and acceptors in (A) pLV and (B) pLV-RRE was determined by Oxford Nanopore sequencing. The percentage of total splicing events supporting the use of splice donor (SD1, green) or cryptic donor 3 (CD3, blue) is shown. The pink boxes represent relative unspliced read coverage. Only junctions supported by at least 10 independent reads are annotated in the sashimi plots. The genomic RNA features for each vector are depicted in the “genome” track.

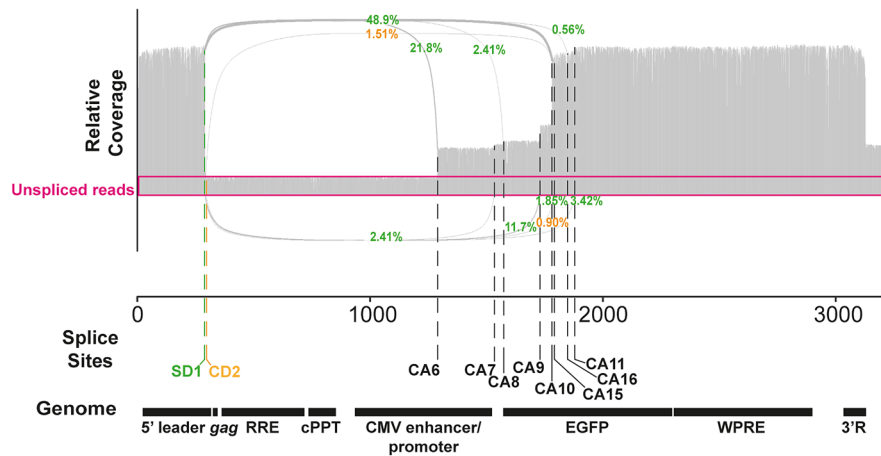
cell and virion but is not transferred to the integrated provirus in the target cell as it is excluded from reverse transcription<sup>20</sup>. This reduces the amount of viral sequence present in the transduced cell to potentially reduce interactions between the provirus and the genome of the host cell. It should be noted that this configuration differs from the previously described LTR1 vectors<sup>20</sup> that also have the RRE downstream of the 3' R in that: (1) pLV-3'RRE has the packaging signal at its standard position at the 5' end of the genome instead of downstream of the 3' R; (2) pLV-3'RRE does not contain a heterologous intron, which was required for efficient LTR1 vector titre; (3) pLV-3'RRE does not have the 3' U5 sequence and therefore the SV40 poly(A) sequence is the only complete poly(A) signal. When HEK293T cells were used as the producer cells, pLV-3'RRE had a similar vector titre compared to pLV or pLV-RRE (Fig. 6C), indicating that this configuration did not substantially inhibit infectivity.

Finally, we analyzed whether combining the 343 nt *gag* deletion with moving the RRE downstream of 3' R was compatible with high titre lentiviral vector production in the HEK293Tsa cell line using a ddPCR assay (Fig. 6D). There is a small (twofold) decrease in titre using this assay when the RRE is moved downstream of 3' R, though there is still  $> 1 \times 10^7$  TU/ml for pLV-3'RRE without concentration by ultracentrifugation. Importantly, the deletion in *gag* does not further reduce vector titre. Therefore, 850 nt of HIV-1 sequence can be removed from the *gag* and *env* regions and the RRE can be moved downstream of the 3' R without substantially reducing vector titre. This reduces the HIV-1 sequence present in the provirus after reverse transcription by 1,201 nt.

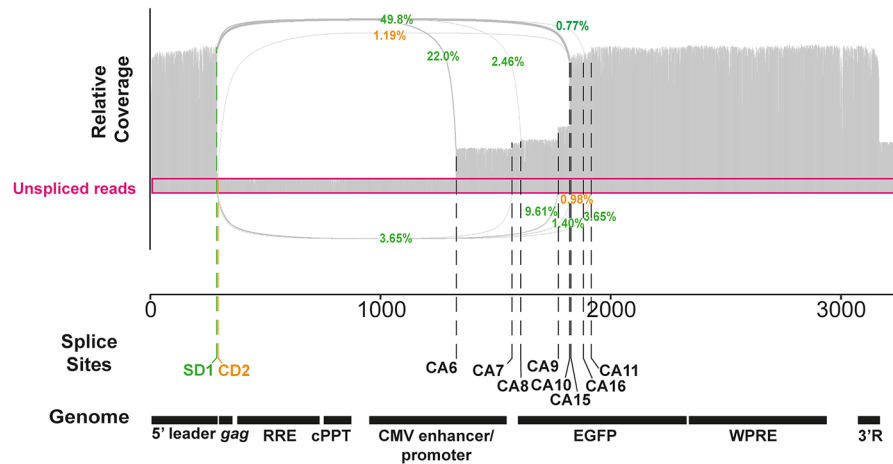
**A) pLV-gag21**



**B) pLV-RREgag21**



**C) pLV-RREgag60**

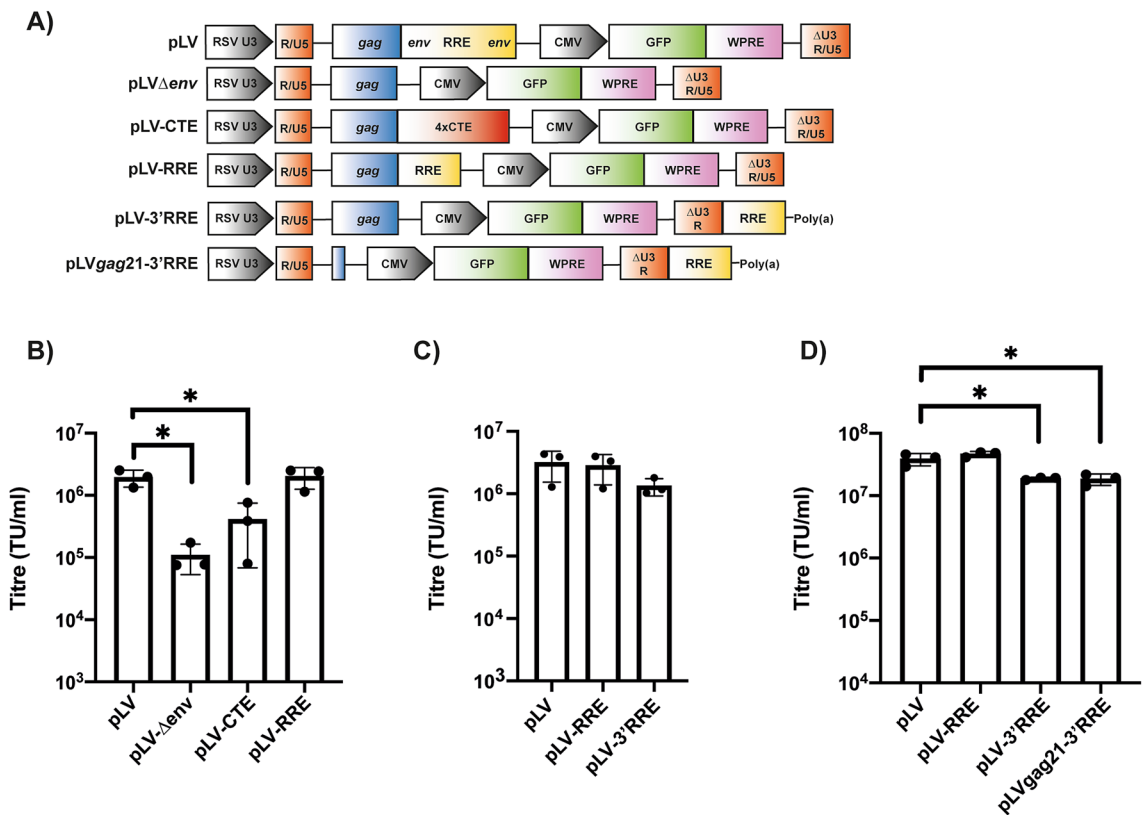


**Figure 5.** Deletions in *gag* can be combined with the 507nt deletion in *env* to increase the proportion of unspliced genomic RNA in HEK293Tsa cells. The usage of canonical and cryptic splice donors and acceptors in (A) pLV-*gag21*, (B) pLV-RRE*gag21* and (C) pLV-RRE*gag60* was determined using Oxford Nanopore sequencing. The percentage of total splicing events supporting the use of SD1 (green) and CD2 (orange) and CD3 (blue) is shown. The pink boxes represent relative unspliced read coverage. Only junctions supported by a minimum of 10 independent reads are annotated in the sashimi plots. The genomic RNA features for each vector are depicted in the “genome” track.



	pLV	pLV-RRE	pLV-gag21	pLV-RREgag21	pLV-RREgag60
% Unspliced transcripts	4.90	15.3	5.71	12.1	10.2
% Readthrough 3' R	29%	28%	30%	34%	34%
Splice events/transcript	1.11	0.83	1.13	0.89	0.91
<b>% of total splicing events</b>					
SD1-A7	55.4	NA	56.0	NA	NA
SD1-CA6	13.5	22.8	12.7	21.8	22.0
SD1-CA9	0.99	12.0	0.97	11.70	9.61
SD1-CA10	8.45	50.9	6.95	48.9	49.8
CD3-CA6	9.00	NA	10.8	NA	NA
CD3-CA10	5.08	NA	5.58	NA	NA

**Table 1.** Splicing events in lentiviral vector genomic RNAs.



**Figure 6.** The RRE can be eliminated from the provirus in target cells by moving it downstream of the 3' R. (A) Schematic representation of the lentiviral genome vectors. (B,C) Lentiviral vectors were produced in HEK293T cells with the indicated vector genome, packaging construct pCMVΔR8.91 and envelope construct pVSV-G. Infectious titres were determined in transduced HEK293T cells by flow cytometry for GFP + cells. (D) Lentiviral vectors were produced in HEK293Tsa cells with the indicated vector genome, packaging construct pCMVΔR8.91 and envelope construct pVSV-G. Infectious titres were determined in transduced HEK293Tsa cells by digital droplet PCR (ddPCR). The bar charts show the average values of three independent experiments. Data are shown as mean ± SD, \*p < 0.05 as determined by a one-way ANOVA test.

### Discussion

Herein, we have conducted a systematic analysis of the viral sequences present in a lentiviral vector genome to determine which regions are not essential for transduction efficiency and could potentially be eliminated. There are several HIV-1 sequences in lentiviral vector genomes including the 5' leader region that regulates dimerization, encapsidation and reverse transcription<sup>22</sup>, a 5' region of *gag*<sup>8</sup>, the RRE<sup>24</sup> and splicing signals<sup>23</sup> embedded within the long stretch of *env* sequence as well as other sequences required for efficient reverse transcription<sup>25,54</sup>. There is also the woodchuck hepatitis virus WPRE in many lentiviral vectors<sup>10</sup>. However, the full functional relevance of many of these sequences for vector transduction efficiency remains unclear.

Most importantly, we found that 850 nt of HIV-1 sequence in the *gag* and *env* regions of the vector genome can be deleted without compromising titre. While there are several reported *cis*-acting RNA elements in these regions, only the RRE is required. Deleting this 850 nt of HIV-1 sequence has several potential advantages. First, it decreases the sequence identity between the genome and packaging vectors that could recombine and produce RCLs. Second, since transduction efficiency decreases as the genome length increases<sup>55</sup>, reducing the HIV-1 portion of the vector genome could increase the capacity for promoters and transgenes. Third, deleting the non-RRE portion of the *env* sequence simplifies the splicing profile of the vector. We also showed that the RRE can be moved distal to the 3' R region and combined with the large deletion in *gag* without substantially decreasing transduction efficiency. Moving the RRE to this position eliminates it from the vector genome after reverse transcription in target cells, thereby reducing the amount of viral sequence in the integrated provirus<sup>20</sup>.

While the proportion of unspliced RNA was increased up to threefold by removing 507 nt of the *env* region, several splicing events remain, and vector titre was not substantially increased. A possible reason that vector titre did not increase could be that the cytoplasmic abundance of the genomic RNA is in excess of the amount required for efficient packaging. While we did these experiments using small scale three plasmid transfection of HEK293T or HEK293Tsa cells, it will be interesting in the future to compare the titre of the pLV and pLV-RRE genomes in a more clinically relevant setting, such as the recently described bacterial artificial chromosome system<sup>56</sup>, in which the relative abundance of the vector genome and viral proteins may be different. Another possible reason that vector titre did not substantially increase for pLV-RRE compared to pLV is that the threefold increase in the proportion of the unspliced RNA is not large enough to raise the titre. Even in the absence of a canonical HIV-1 splice acceptor, ~85% of the vector transcripts are spliced for pLV-RRE. It would be ideal to eliminate all splicing in the genomic RNA since the only relevant transcript for transduction is the unspliced RNA. However, this may be a challenge. While the deletions in *env* to produce pLV-RRE and pLV-RRE*gag21* eliminate CD3, which contributes to removing an internal portion of the genome in ~15% of pLV transcripts, many splicing events that utilize SD1 remain. Several cryptic splice acceptors can be used in the SD1 splicing events, including acceptor sites in the transgene as exemplified by CA10 in the GFP open reading frame. The SD1 sequence is part of the three-way junction in the complex leader RNA structure that mediates dimerization and encapsidation<sup>42</sup>. Therefore, mutations that eliminate SD1 would have to be made in a context that does not affect this structure and cryptic donor sites may be utilized if SD1 is eliminated. As we have shown, Oxford Nanopore sequencing is an excellent tool to determine the complete splicing profile of the genomic RNA due to its long-read length. The specific splicing profile is likely to differ depending on the transgene cassette and should be evaluated in the context of specific therapeutic constructs. Using Oxford Nanopore sequencing or similar technologies may be useful for evaluating the effect of potential cryptic splice sites in transgenes or regulatory elements introduced into the vector genome.

Of note, there are potential safety risks for integrated lentiviral vectors due to interactions between the vector and host cell genomes. Splice sites within the lentiviral vector have been reported to be used in combination with splice sites in cellular genes to produce fusion transcripts between vector and cellular RNAs, which could contribute to vector-mediated genotoxicity<sup>20,57,58</sup>. In particular, splicing between a lentiviral vector transcript and *HMG2A2* transcripts resulted in the dysregulation of *HMG2A2* expression in a patient treated for  $\beta$ -thalassaemia, resulting in benign clonal expansion of hematopoietic progenitors<sup>59</sup>. While we have not evaluated whether eliminating HIV-1 sequences containing cryptic splice sites affects interactions between the vector and cellular transcripts, this is a potential advantage of the minimal pLV*gag21*-3'RRE vector. Therefore, this vector is an alternative to the previously described LTR1 constructs<sup>20</sup> in that both have the *env* region eliminated from the integrated vector genome but differ in that pLV*gag21*-3'RRE contains the HIV-1 SD1 while LTR1 has a heterologous intron.

Furthermore, in HIV-1 infected patients, interactions between the vector and wild type virus can lead to the vector becoming mobilized by HIV proteins provided in *trans*. This could result in the packaging of vector RNA in HIV-1 particles and spread transduction beyond the intended target cells or to newly infected individuals<sup>60,61</sup>. While the enhancer and promoter elements are eliminated in integrated SIN vectors<sup>18</sup>, mobilization of the integrated vector genome is not totally eliminated<sup>61</sup> and moving the RRE downstream of the 3' R reduces the risk of this problem in HIV-1 infected cells<sup>20</sup>. Recombination between the vector genome and wild type viral genome could also occur. Thus, for immunotherapies based on lentiviral vector transduction of CD4+ T cells that may be delivered to HIV-1 positive patients, using lentiviral vectors with substantially reduced amounts of HIV-1 sequence may be beneficial.

Interestingly, while the RRE is essential for nuclear export of HIV-1 intron-containing RNA<sup>24</sup>, the most prominent effect of deleting this region in the context of the pLV-RRE vector is on the amount of genomic RNA packaged into virions. Therefore, in the context of a genome that does not have multiple canonical HIV-1 splice sites, the RRE appears to be more important for packaging the genome into virions than promoting cytoplasmic RNA abundance. Because pLV-RRE only contains one canonical splice donor (SD1) in contrast to HIV-1, which contains at least four defined splice donors and eight splice acceptors<sup>23</sup>, its RNA may have a lower requirement for Rev-RRE-mediated nuclear export than the full-length virus. It also does not contain many of the *cis*-acting repressive sequences (CRSs)/instability sequences (INSS) found in HIV-1 RNAs that require Rev for efficient expression<sup>24</sup>. Furthermore, the WPRE has been reported to be an RNA nuclear export element<sup>13–15</sup>, which could make it functionally redundant with the RRE. However, deletion of the RRE, WPRE or both elements together did not substantially affect cytoplasmic genomic RNA abundance, indicating that they are not required for nuclear export of this transcript. The hepatitis B virus post-transcriptional regulatory element and WPRE have recently been shown to promote gene expression by recruiting the TENT4-ZCCHC14 complex<sup>16</sup>. This leads to a 'mixed tail' at the 3' end of mRNAs, in which non-adenosine nucleotides are inserted into the poly(A) tail, that protects the mRNA from degradation and may be the mechanism by which the WPRE promotes gene expression from lentiviral vector genomes. Rev and the RRE have previously been reported to be required for efficient genomic RNA encapsidation<sup>45–47</sup> and our data supports those observations. The specific mechanism by which Rev and

the RRE regulate HIV-1 or lentiviral vector genomic RNA packaging remains unclear and is an important area for further investigation.

In summary, we have shown that it is possible to decrease the length of lentiviral vector genomes by deleting non-essential regions in *gag* and *env* and moving the RRE downstream of 3' R, which eliminates it from the integrated provirus. This may improve vector design by increasing transgene capacity, reducing potential RCL formation and decreasing the amount of splicing that occurs in the vector transcripts. This could also potentially reduce genotoxicity by decreasing the number of fusion transcripts created by splicing between the integrated vector and cellular RNAs, though this would have to be validated in the context of a therapeutic model system. While we used a GFP transgene in the experiments in this report, analyzing long, complex transgenes such as next generation CAR or T cell receptor (TCR) constructs in the pLV*gag*21-3'RRE construct would help determine its clinical potential. These transgene cassettes express proteins in addition to the CAR or TCR to improve T cell expansion and persistence in vivo, render transduced T cells resistant to the immunosuppressive tumor microenvironment or enable clinical intervention to selectively eliminate transduced cells in vivo if severe toxicity occurs<sup>62,63</sup>. The size and complexity of these and other therapeutic transgenes that can be delivered using lentiviral vectors may benefit from a minimal vector genome with decreased splicing potential. Overall, removing non-essential HIV-1 elements from lentiviral vector genomes may enhance vector performance and reduce the amount of viral sequence transduced into the genome of a patient's cells during gene therapy.

## Materials and methods

**Plasmids.** The previously described lentiviral vector genome constructs used in this study was pLV (pRRL-PPT-CMV-GFP-WPRE)<sup>21</sup>. The *gag* and *env* deletions were made in pLV using overlapping PCR. The WPRE was removed by cutting at EcoRI sites flanking the WPRE sequence. The 4xCTE sequence transferred from pGPV-4xCTE<sup>53</sup> using the NotI and EcoRI sites into pLV $\Delta$ *env* to create pLV-CTE. For pLV-3'RRE, the RRE and SV40 poly(A) sequences were synthesized using GenScript fragment synthesis and inserted into the KpnI and AvrII sites in pLV $\Delta$ *env*. pCMV $\Delta$ R8.91<sup>9</sup>, pGag-Pol-Vif (pGPV)<sup>44</sup>, pVSV-G<sup>64</sup>, pRev<sup>53</sup> and pGFP<sup>65</sup> have previously been described.

**Cells.** HEK293T and HEK293Tsa<sup>48</sup> cells were grown in Dulbecco's Modified Eagle Medium (DMEM) plus Gluta-Max (Life Technologies) supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were maintained in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C.

**Vector production.** HEK293T or HEK293Tsa cells were seeded 24 h prior to transfection at a density of 10<sup>6</sup> cells per well in a six-well plate. Plasmids for transfection was prepared in Opti-MEM. Each well was co-transfected with 1  $\mu$ g vector genome, 1  $\mu$ g packaging plasmid and 0.5  $\mu$ g pVSV-G using PEI at a DNA/PEI ratio of 1:3. Media was changed 6 h post-transfection and the supernatant containing viral vector particles were filtered through a 0.45  $\mu$ m Millex-HA filter (Millipore) 48 h post-transfection.

**Quantification of vector titre by FACS.** HEK293T or HEK293Tsa were plated in 96-well plates as target cells for transduction. A serial dilution of supernatant containing viral vectors was prepared and added to the wells. Cells were harvested 48-h post-transduction, fixed in 2% paraformaldehyde, and re-suspended in 1  $\times$  phosphate-buffered saline (PBS). Samples were run on a flow cytometer to detect GFP positive cells which was used to calculate the titer (transducing units (TU)/ml).

**Quantitative RT-PCR.** Cells were washed with 1xPBS and the RNA was extracted using the RNeasy mini kit (Qiagen) following the manufacturer's instructions. To extract virion RNA, the supernatant was centrifuged at 20,000 $\times$ g for 2 h through a 20% sucrose cushion in 1  $\times$  PBS to pellet the vector particles. Virion RNA was extracted using the QIAamp Viral mini kit (Qiagen). 1  $\mu$ g of cellular RNA and 20  $\mu$ L of virion RNA was reverse transcribed using the High Capacity cDNA Reverse-Transcription kit (Applied Biosystems). Quantitative PCR was performed using the Taqman Universal PCR mix and the QuantiStudio 5 System (Thermo Fisher). Absolute quantification was determined using a standard curve of the lentiviral vector DNA plasmid. The genomic RNA primers were TCTCGACGCAGGACTCG/TACTGACGCTCTCGCACC (forward/reverse), and the probe was FAM-ATCTCTCTCCTTCTAGCCTC-TAMRA. Of note, this primer-probe set spans SD1 and only detects unspliced genomic RNA.

**RNA isolation.** A protocol previously optimized to analyze HIV-1 cytoplasmic abundance was used to isolate cytoplasmic and nuclear viral transcripts<sup>44</sup>. Briefly, 48 h post transfection, the cells were lysed on ice in 400  $\mu$ L of cold, NB buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.5% NP-40), and centrifuged at 500 $\times$ g for 10 min to pellet nuclei. 200  $\mu$ L of the cytoplasmic supernatant was added to 600  $\mu$ L of RLT buffer (Qiagen). The nuclear pellet was washed twice in cold NB buffer, resuspended in 400  $\mu$ L of RLT buffer (Qiagen) and spun through a Qias shredder column (Qiagen). The cytoplasmic and nuclear RNA was extracted using the RNeasy Mini kit (Qiagen).

**Western Blotting.** 48 h post-transfection, whole cell or the cytoplasmic and nuclear fractions were lysed in radioimmunoprecipitation (RIPA) buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate) and 2 $\times$ loading buffer (60 mM Tris-HCl (pH 6.8), 10%  $\beta$ -mercaptoethanol, 10% glycerol, 2% sodium dodecyl sulfate (SDS), 0.1% bromophenol blue). Cell lysates were resolved by SDS-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. Protein

bands were detected using the LI-COR infrared imaging system (LI-COR UK LTD). The antibodies used were 1:50 HIV-1 anti-p24<sup>Gag</sup> (183-H12-5C)<sup>66</sup>, 1:1000 anti-Hsp90 (Santa Cruz Biotechnology, sc7947), 1:5000 anti- $\alpha$ -Tubulin antibody (Abcam, ab7291), 1:1000 anti-Lamin B1 (Abcam, ab8982), 1:10,000 Dylight™ 800-conjugated anti-mouse/rabbit secondary antibodies (Cell Signalling Technology). The Rev-specific monoclonal antibody Rev-6 (1:500) was provided by Michael Malim and was generated by immunizing BALB/c mice with purified hexahistidine-tagged HIV-1<sub>HXB3</sub> Rev, boosting three times and recovering the spleen to generate hybridomas with the mouse plasmacytoma cell line SP2/OAg. Hybridoma culture supernatants were screened by enzyme linked immunosorbent assay for reactivity against plastic-adhered antigen.

**Quantification of vector titre by Digital Droplet PCR.** HEK293Tsa cells<sup>48</sup> were transduced with 50, 5 and 0.5  $\mu$ L of supernatant containing viral vectors. 72 h post-transduction, cells were harvested, and genomic DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen). Approximately 50 ng of extracted genomic DNA was digested using MluI-HF (New England Biolabs). From the digestion mixture, 5  $\mu$ L of digested genomic DNA was added to a PCR reaction mix containing 2 $\times$  ddPCR Supermix for Probes (no dUTP) (Bio-Rad), a primer/probe set for the RNaseP reference gene (TaqMan Copy Number Reference Assay, Applied Biosystems) and a primer/probe set for the target, which spans the SD1 and therefore specifically detects unspliced genomic RNA (TCTCGACGCAGGACTCG/CGCTCTCGACCCATCTC (forward/reverse) and probe FAM-CTCCTTCTAGCCTCCGCTAG-BHQ1), at a final concentration of 0.9  $\mu$ M of each primer and 0.25  $\mu$ M of each probe. 20  $\mu$ L of the final mix was added to a 96-well plate, and droplets were generated using the Automated Droplet Generator (Bio-Rad) following the manufacturer's instructions. PCR was performed using the C1000 Touch Thermal Cycler (Bio-Rad) and droplets were read using a QX200 droplet reader. The average copy number per cell was used to calculate titre (TU/ml).

**Analysis of lentiviral vector splicing using Oxford Nanopore sequencing.** HEK293Tsa cells<sup>48</sup> were transfected with pLV, pLV-gag21, pLV-RRE, pLV-RREgag21 or pLV-RREgag60, along with pCMV $\Delta$ R8.91 and pVSV-G. 48 h post-transfection, RNA was extracted using the RNeasy Mini kit (Qiagen). Polyadenylated mRNA was isolated from approximately 10  $\mu$ g of total RNA using the Dynabeads mRNA purification kit (ThermoFisher Scientific) following the manufacturer's instructions. The quality and quantity of mRNA was assessed using the TapeStation 4200 (Agilent). Nanopore libraries were prepared with 100 ng of poly(A) + RNA using the Direct cDNA Sequencing kit (SQK-DCS109) with Native Barcoding (EXP-NB104) following the manufacturer's instructions (Oxford Nanopore Technologies). The final library was loaded onto an R9.4.1 MinION flow cell (FLO-MIN111, Oxford Nanopore Technologies) and sequenced for 72 h.

The barcoded raw sequence data (FAST5 files) generated by Oxford Nanopore sequencing were basecalled and debarcoded using GUPPY basecaller version 3.1.5 to output decoded FASTQ files. The debarcoded FASTQ files for each vector were selected using samtools version 1.10. The reads were then aligned to the human genome (hg38) and the lentiviral vector genome simultaneously using minimap2<sup>67</sup>. Near-full length vector genome reads (reads that contain at least 1 nucleotide upstream of SD1) were selected using samtools. Lentiviral vector-mapping junction-spanning reads were isolated using regtools (<https://github.com/griffithlab/regtools>) to allow per-junction read counting. The percentage of HIV-1 junction-spanning reads was calculated by dividing the number of reads for each junction by the total number of junction-spanning reads in the library. Sashimi plots were generated using ggsashimi<sup>68</sup>.

**Statistical analysis.** Statistical significance was determined using a one-way ANOVA test. Data are represented as mean  $\pm$  SD. Significance was ascribed to p values p < 0.05.

**Ethics declarations.** No human tissue samples, human subjects or animals were used in this study.

## Data availability

The Oxford Nanopore RNA-seq datasets are available at GSE165829.

Received: 4 February 2021; Accepted: 17 May 2021

Published online: 08 June 2021

## References

- Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* <https://doi.org/10.1126/science.aan4672> (2018).
- Naldini, L., Trono, D. & Verma, I. M. Lentiviral vectors, two decades later. *Science* **353**, 1101–1102. <https://doi.org/10.1126/science.aah6192> (2016).
- Sadelain, M. CD19 CAR T cells. *Cell* **171**, 1471. <https://doi.org/10.1016/j.cell.2017.12.002> (2017).
- Sundquist, W. I. & Krausslich, H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924. <https://doi.org/10.1101/cshperspect.a006924> (2012).
- Malim, M. H. & Bieniasz, P. D. HIV restriction factors and mechanisms of evasion. *Cold Spring Harb. Perspect. Med.* **2**, a006940. <https://doi.org/10.1101/cshperspect.a006940> (2012).
- Karn, J. & Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med.* **2**, a006916. <https://doi.org/10.1101/cshperspect.a006916> (2012).
- Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* <https://doi.org/10.1101/cshperspect.a006882> (2012).
- Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267. <https://doi.org/10.1126/science.272.5259.263> (1996).
- Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.* **15**, 871–875. <https://doi.org/10.1038/nbt0997-871> (1997).



10. Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.* **73**, 2886–2892. <https://doi.org/10.1128/JVI.73.4.2886-2892.1999> (1999).
11. Schambach, A. *et al.* Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. *Gene Ther.* **13**, 641–645. <https://doi.org/10.1038/sj.gt.3302698> (2006).
12. Hlavaty, J. *et al.* Effect of posttranscriptional regulatory elements on transgene expression and virus production in the context of retrovirus vectors. *Virology* **341**, 1–11. <https://doi.org/10.1016/j.virol.2005.06.037> (2005).
13. Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J. Virol.* **72**, 5085–5092. <https://doi.org/10.1128/JVI.72.6.5085-5092.1998> (1998).
14. Popa, I., Harris, M. E., Donello, J. E. & Hope, T. J. CRM1-dependent function of a cis-acting RNA export element. *Mol. Cell. Biol.* **22**, 2057–2067. <https://doi.org/10.1128/mcb.22.7.2057-2067.2002> (2002).
15. Pocock, G. M. *et al.* Diverse activities of viral cis-acting RNA regulatory elements revealed using multicolor, long-term, single-cell imaging. *Mol. Biol. Cell* **28**, 476–487. <https://doi.org/10.1091/mbc.E16-08-0612> (2017).
16. Kim, D. *et al.* Viral hijacking of the TENT4-ZCCHC14 complex protects viral RNAs via mixed tailing. *Nat. Struct. Mol. Biol.* **27**, 581–588. <https://doi.org/10.1038/s41594-020-0427-3> (2020).
17. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471. <https://doi.org/10.1128/JVI.72.11.8463-8471.1998> (1998).
18. Zufferey, R. *et al.* Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–9880. <https://doi.org/10.1128/JVI.72.12.9873-9880.1998> (1998).
19. Fang, Y., Gong, X., Xu, M., Zeng, F. & Zhang, J. A self-deletion lentiviral vector to reduce the risk of replication-competent virus formation. *J. Gene Med.* **15**, 102–112. <https://doi.org/10.1002/jgm.2700> (2013).
20. Vink, C. A. *et al.* Eliminating HIV-1 packaging sequences from lentiviral vector proviruses enhances safety and expedites gene transfer for gene therapy. *Mol. Ther.* **25**, 1790–1804. <https://doi.org/10.1016/j.ymthe.2017.04.028> (2017).
21. Jarrosson-Wuilleme, L. *et al.* Transduction of nondividing human macrophages with gammaretrovirus-derived vectors. *J. Virol.* **80**, 1152–1159. <https://doi.org/10.1128/JVI.80.3.1152-1159.2006> (2006).
22. Keane, S. C. & Summers, M. F. NMR Studies of the structure and function of the HIV-1 5'-leader. *Viruses* **8**, 338. <https://doi.org/10.3390/v8120338> (2016).
23. Sertznig, H., Hillebrand, F., Erkelenz, S., Schaal, H. & Widera, M. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology* **516**, 176–188. <https://doi.org/10.1016/j.virol.2018.01.011> (2018).
24. Pollard, V. W. & Malim, M. H. The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**, 491–532. <https://doi.org/10.1146/annurev.micro.52.1.491> (1998).
25. Zennou, V. *et al.* The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain. *Nat. Biotechnol.* **19**, 446–450. <https://doi.org/10.1038/88115> (2001).
26. Najera, I., Krieg, M. & Karn, J. Synergistic stimulation of HIV-1 rev-dependent export of unspliced mRNA to the cytoplasm by hnRNP A1. *J. Mol. Biol.* **285**, 1951–1964. <https://doi.org/10.1006/jmbi.1998.2473> (1999).
27. Ooms, M., Abbink, T. E., Pham, C. & Berkhout, B. Circularization of the HIV-1 RNA genome. *Nucleic Acids Res.* **35**, 5253–5261. <https://doi.org/10.1093/nar/gkm564> (2007).
28. Beerens, N. & Kjems, J. Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. *RNA* **16**, 1226–1235. <https://doi.org/10.1261/rna.2039610> (2010).
29. Luban, J. & Goff, S. P. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* **68**, 3784–3793. <https://doi.org/10.1128/JVI.68.6.3784-3793.1994> (1994).
30. Buchsacher, G. L. Jr. & Panganiban, A. T. Human immunodeficiency virus vectors for inducible expression of foreign genes. *J. Virol.* **66**, 2731–2739. <https://doi.org/10.1128/JVI.66.5.2731-2739.1992> (1992).
31. Parolin, C., Dorfman, T., Palu, G., Gottlinger, H. & Sodroski, J. Analysis in human immunodeficiency virus type 1 vectors of cis-acting sequences that affect gene transfer into human lymphocytes. *J. Virol.* **68**, 3888–3895. <https://doi.org/10.1128/JVI.68.6.3888-3895.1994> (1994).
32. Kotsopoulou, E., Kim, V. N., Kingsman, A. J., Kingsman, S. M. & Mitrophanous, K. A. A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. *J. Virol.* **74**, 4839–4852. <https://doi.org/10.1128/jvi.74.10.4839-4852.2000> (2000).
33. Koldej, R. M. & Anson, D. S. Refinement of lentiviral vector for improved RNA processing and reduced rates of self inactivation repair. *BMC Biotechnol.* **9**, 86. <https://doi.org/10.1186/1472-6750-9-86> (2009).
34. Antzin-Anduetza, I., Mahiet, C., Granger, L. A., Odendall, C. & Swanson, C. M. Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* **14**, 49. <https://doi.org/10.1186/s12977-017-0374-1> (2017).
35. Mammano, F., Kondo, E., Sodroski, J., Bukovsky, A. & Gottlinger, H. G. Rescue of human immunodeficiency virus type 1 matrix protein mutants by envelope glycoproteins with short cytoplasmic domains. *J. Virol.* **69**, 3824–3830. <https://doi.org/10.1128/JVI.69.6.3824-3830.1995> (1995).
36. Freed, E. O. & Martin, M. A. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *J. Virol.* **69**, 1984–1989. <https://doi.org/10.1128/JVI.69.3.1984-1989.1995> (1995).
37. Reil, H., Bukovsky, A. A., Gelderblom, H. R. & Gottlinger, H. G. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. *EMBO J.* **17**, 2699–2708. <https://doi.org/10.1093/emboj/17.9.2699> (1998).
38. Lu, K. *et al.* NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **334**, 242–245. <https://doi.org/10.1126/science.1210460> (2011).
39. Ficarelli, M. *et al.* CpG dinucleotides inhibit HIV-1 replication through zinc finger antiviral protein (ZAP)-dependent and -independent mechanisms. *J. Virol.* **94**, e01337-01319. <https://doi.org/10.1128/JVI.01337-19> (2020).
40. Mayrose, I. *et al.* Synonymous site conservation in the HIV-1 genome. *BMC Evol. Biol.* **13**, 164. <https://doi.org/10.1186/1471-2148-13-164> (2013).
41. Ngandu, N. K. *et al.* Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. *Virol. J.* **5**, 160. <https://doi.org/10.1186/1743-422X-5-160> (2008).
42. Keane, S. C. *et al.* RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917–921. <https://doi.org/10.1126/science.aaa9266> (2015).
43. Rekosh, D. & Hammarskjold, M. L. Intron retention in viruses and cellular genes: Detention, border controls and passports. *Wiley Interdiscip. Rev. RNA* **9**, e1470. <https://doi.org/10.1002/wrna.1470> (2018).
44. Sherer, N. M. *et al.* Evolution of a species-specific determinant within human CRM1 that regulates the post-transcriptional phases of HIV-1 replication. *PLoS Pathog* **7**, e1002395. <https://doi.org/10.1371/journal.ppat.1002395> (2011).
45. Brandt, S. *et al.* Rev proteins of human and simian immunodeficiency virus enhance RNA encapsidation. *PLoS Pathog* **3**, e54. <https://doi.org/10.1371/journal.ppat.0030054> (2007).
46. Kharytonchyk, S. *et al.* Influence of gag and RRE sequences on HIV-1 RNA packaging signal structure and function. *J. Mol. Biol.* **430**, 2066–2079. <https://doi.org/10.1016/j.jmb.2018.05.029> (2018).
47. Anson, D. S. & Fuller, M. Rational development of a HIV-1 gene therapy vector. *J. Gene Med.* **5**, 829–838. <https://doi.org/10.1002/jgm.415> (2003).

48. Biffi, A. *et al.* Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **341**, 1233158. <https://doi.org/10.1126/science.1233158> (2013).
49. Zais, A. K., Son, S. & Chang, L. J. RNA 3' readthrough of oncoretrovirus and lentivirus: Implications for vector safety and efficacy. *J. Virol.* **76**, 7209–7219. <https://doi.org/10.1128/jvi.76.14.7209-7219.2002> (2002).
50. Yang, Q., Lucas, A., Son, S. & Chang, L. J. Overlapping enhancer/promoter and transcriptional termination signals in the lentiviral long terminal repeat. *Retrovirology* **4**, 4. <https://doi.org/10.1186/1742-4690-4-4> (2007).
51. Bray, M. *et al.* A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc. Natl. Acad. Sci. U S A* **91**, 1256–1260. <https://doi.org/10.1073/pnas.91.4.1256> (1994).
52. Wodrich, H., Schambach, A. & Krausslich, H. G. Multiple copies of the Mason-Pfizer monkey virus constitutive RNA transport element lead to enhanced HIV-1 Gag expression in a context-dependent manner. *Nucleic Acids Res.* **28**, 901–910. <https://doi.org/10.1093/nar/28.4.901> (2000).
53. Swanson, C. M., Puffer, B. A., Ahmad, K. M., Doms, R. W. & Malim, M. H. Retroviral mRNA nuclear export elements regulate protein function and virion assembly. *EMBO J.* **23**, 2632–2640. <https://doi.org/10.1038/sj.emboj.7600270> (2004).
54. Le Grice, S. F. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. *J. Biol. Chem.* **287**, 40850–40857. <https://doi.org/10.1074/jbc.R112.389056> (2012).
55. Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. Systematic determination of the packaging limit of lentiviral vectors. *Hum. Gene Ther.* **12**, 1893–1905. <https://doi.org/10.1089/104303401753153947> (2001).
56. Chen, Y. H. *et al.* Rapid lentiviral vector producer cell line generation using a single DNA construct. *Mol. Ther. Methods Clin. Dev.* **19**, 47–57. <https://doi.org/10.1016/j.omtm.2020.08.011> (2020).
57. Moiani, A. *et al.* Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J. Clin. Invest.* **122**, 1653–1666. <https://doi.org/10.1172/JCI61852> (2012).
58. Bokhoven, M. *et al.* Insertional gene activation by lentiviral and gammaretroviral vectors. *J. Virol.* **83**, 283–294. <https://doi.org/10.1128/JVI.01865-08> (2009).
59. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* **467**, 318–322. <https://doi.org/10.1038/nature09328> (2010).
60. Bukovsky, A. A., Song, J. P. & Naldini, L. Interaction of human immunodeficiency virus-derived vectors with wild-type virus in transduced cells. *J. Virol.* **73**, 7087–7092. <https://doi.org/10.1128/JVI.73.8.7087-7092.1999> (1999).
61. Hanawa, H., Persons, D. A. & Nienhuis, A. W. Mobilization and mechanism of transcription of integrated self-inactivating lentiviral vectors. *J. Virol.* **79**, 8410–8421. <https://doi.org/10.1128/JVI.79.13.8410-8421.2005> (2005).
62. Petersen, C. T. & Krenciute, G. Next generation CAR T cells for the immunotherapy of high-grade glioma. *Front. Oncol.* **9**, 69. <https://doi.org/10.3389/fonc.2019.00069> (2019).
63. Brandt, L. J. B., Barnkob, M. B., Michaels, Y. S., Heiselberg, J. & Barington, T. Emerging approaches for regulation and control of CAR T cells: A mini review. *Front. Immunol.* **11**, 326. <https://doi.org/10.3389/fimmu.2020.00326> (2020).
64. Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U. & Malim, M. H. HIV-1 infection of non-dividing cells: Evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *EMBO J.* **16**, 4531–4539. <https://doi.org/10.1093/emboj/16.15.4531> (1997).
65. Swanson, C. M., Sherer, N. M. & Malim, M. H. SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. *J. Virol.* **84**, 6748–6759. <https://doi.org/10.1128/JVI.02526-09> (2010).
66. Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: Definition of critical amino acids involved in cell tropism. *J. Virol.* **66**, 6547–6554. <https://doi.org/10.1128/JVI.66.11.6547-6554.1992> (1992).
67. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> (2018).
68. Garrido-Martin, D., Palumbo, E., Guigo, R. & Breschi, A. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360. <https://doi.org/10.1371/journal.pcbi.1006360> (2018).

## Acknowledgements

We thank the members of Chad Swanson's, Michael Malim's and Stuart Neil's labs for helpful discussions. We also thank Caroline Goujon and Michael Malim for generously providing reagents. The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 p24 Hybridoma (183-H12-5C) (Cat# 1513) from Dr. Bruce Chesebro. These studies were funded by Medical Research Council grant MR/M019756/1 and MR/S000844/1 to CMS. MR/S000844/1 is part of the EDCTP2 programme supported by the European Union. MF is supported by the UK Medical Research Council (MR/R50225X/1) and is a King's College London member of the MRC Doctoral Training Partnership in Biomedical Sciences. HS is supported by the BBSRC Industrial CASE Partnerships (ICP) training grant (BB/P504609/1) with GSK. This work was also supported by the Department of Health via a National Institute for Health Research Comprehensive Biomedical Research Centre award to Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust.

## Author contributions

Conceptualization, C.A.V. and C.M.S.; Methodology, N.P.S.; Investigation, H.S., M.F, H.P; Visualization, H.S., M.F; Writing –Original Draft, C.M.S.; Writing –Review & Editing, H.S., M.F, N.P.S, C.A.V, C.M.S; Funding Acquisition, C.A.V. and C.M.S.; Supervision, C.A.V and C.M.S.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-91309-w>.

**Correspondence** and requests for materials should be addressed to C.A.V. or C.M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021