Molecular Therapy
Advances
Original Article

Enhanced transcriptional insulation of lentiviral vectors using "sequence-upgraded polyA long terminal repeats" or "supA-LTRs"

Jordan Wright,[1] Ben M. Alberts,[1] Maria L. Martinez Quiles,[1] Alessandra Guarino,[2] Daniel Chipchase,[3] Nicholas G. Clarkson,[1] Kyriacos A. Mitrophanous,[1] and Daniel C. Farley[1]

[1]Oxford Biomedica, Oxford OX4 6LT, UK; [2]CYTOO, 38040 Grenoble, France; [3]Independent consultant, Scotland, UK

The safety and quality attributes of viral vector-based products are of paramount importance. Recent successful commercialized use of lentiviral vectors (LVs) *ex vivo* anticipates future, higher dose applications *in vivo*, leading to greater integration events per patient. Contemporary LVs employ "self-inactivating" (SIN) long terminal repeats (LTRs) that are deleted for enhancer-promoter sequences, which greatly minimizes the probability of insertional oncogenesis. However, such deletion also removes polyadenylation enhancer sequences, leading to reduced polyadenylation efficiency. Transcriptional read-in/out of integrated SIN-LVs typically results in interaction of LV sequences with the cellular transcriptome. Here we developed "sequence-upgraded polyA-LTRs" (supA-LTRs) that have ∼70-fold and ∼100-fold increased inherent polyadenylation activity in sense and antisense directions, respectively. This was achieved by redefinition of the polyA signal position with respect to the R region and optimal spacing of heterologous upstream and downstream enhancers, completely decoupling transcriptional termination from native HIV-1 polyadenylation sequences. The greatest block (>20-fold) to 5′ transcription read-in to the integrated LV cassette was observed when the supA-LTR was combined with LVs that have an inactivated major splice donor site ("SupA2KO-LVs"). SupA-LTRs also mediated an ∼2-fold increase in transgene expression in primary cells. SupA-LTRs represent a further step forward providing additional safety and utility to the next generation of LVs.

INTRODUCTION

Lentiviral vectors (LVs) based on HIV-1 are attractive delivery vehicles for gene therapy, but their principal functional sequences and configuration have remained unchanged for over two decades.[1,2] The defining safety features of 3rd-generation LVs are 3-fold. First, the use of a powerful constitutive promoter to transcribe the vector genomic RNA (vRNA) makes them independent of the auxiliary gene tat, which modulates numerous cellular functions.[3,4] Secondly, the separation of LV component DNA expression cassettes (genome, gagpol, rev, and envelope) greatly reduces the possibility of recombination to generate replication-competent LVs (RCLs).[5,6] Thirdly, the deletion of enhancer-promoter sequences from the native 3′ U3 promoter to generate "self-inactivating" (SIN) long terminal repeats (LTRs) negates potential LV-mediated activation of proto-oncogenes in somatic cells,[4] as well as further de-risking RCL formation. However, the U3 promoter, as well as the R-U5 sequences of the LTR, also harbors important overlapping polyadenylation-enhancer sequences to direct efficient pre-mRNA termination, cleavage, and polyadenylation of transcripts at the R/U5 junction.[7,8] These polyadenylation sequences within U3-R-U5 reflect the typical structure of an RNA polymerase II transcription termination sequence (see Figure 1A). The upstream enhancer (USE) is usually ∼20 nucleotides upstream of the polyadenylation signal (pAS) located in R. For HIV-1, and derivative LVs, the polyadenine tail is added at the RNA cleavage site (typically a "CA" dinucleotide) on the R/U5 boundary, located several nucleotides downstream of the pAS. A GU-rich downstream enhancer (DSE) is positioned ∼20 nucleotides downstream of the cleavage site (located in U5). Retroviruses typically employ relatively weak polyadenylation sequences because of the requirement to balance transcriptional activity driven from the 5′ LTR (i.e., avoiding premature termination) with active polyadenylation at the 3′ LTR, despite the LTRs being identical in sequence. This is achieved through adoption of functionally altered polyA RNA loop structures in 5′ and 3′ R regions, and the absence/presence of the USE at 5'/3′ ends of the viral RNA, respectively, owing to the U3 RNA sequence being present only at the 3′ LTR.[9,10] The 5′ pAS appears to have a role in HIV-1 latency since premature termination at the 5′ LTR is a feature of integrated HIV-1 genomes prior to activation.[11] Initially, stochastic low-level transcription from U3 becomes more productive if U1 snRNA is recruited at the emergent major splice donor (MSD) since this results in telescripting by the U1 snRNA and consequent suppression of the 5′ pAS.[12] This leads to production of small quantities of spliced mRNA, including tat mRNA. This initiates the tat-dependent feedforward mechanism, as tat binds to the 5′ transactivation response [TAR] loop, composing

Received 13 May 2025; accepted 18 December 2025;
https://doi.org/10.1016/j.omta.2025.201654.

Correspondence: Daniel C Farley, Oxford Biomedica, Oxford OX4 6LT, UK.
E-mail: d.farley@oxb.com

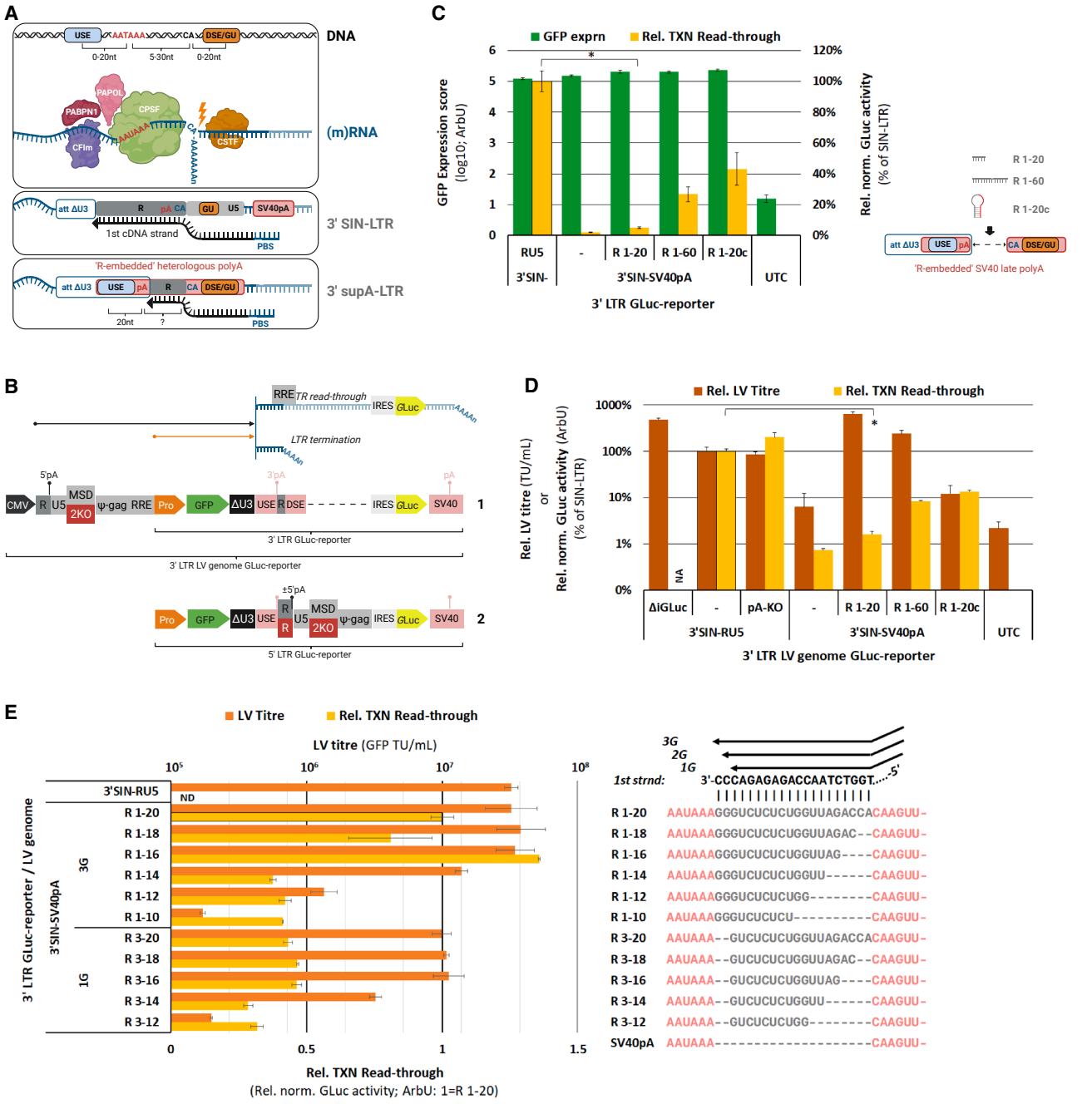

Figure 1. Defining the minimal length of HIV-1 R sequence with the R-embedded heterologous polyA

(A) Organizational structure of the 3′ supA-LTR (containing the R-embedded heterologous polyA) and 3′ SIN-LTR in relation to the requirements of both polyadenylation sequences/factors and 1st strand transfer during reverse transcription. The core polyadenylation factors shown are the mammalian cleavage factor I complex (CFIm; which binds to the upstream enhancer [USE]), polyadenylate-binding nuclear protein 1 (PABPN1), poly(A) polymerase A/G (PAPOLA/G), cleavage and polyadenylation specificity factor complex (CPSF), and cleavage stimulation factor complex (CSTF; which binds to the GU-rich downstream enhancer [DSE]). The self-inactivating (SIN) LTR is the att-ΔU3 region, followed by R-U5 and typically a back-up (SV40 late) polyA. The 3′ supA-LTR dispenses with all R-U5 except for a minimal R sequence incorporated into the "back-up" polyA. The 1st cDNA strand is derived from reverse transcription from a tRNA recruited to the primer binding site (PBS) at the 5′ end of the vRNA (not shown), and then "jumps" to the 3′ end of the vRNA via base-pairing with R sequence. (B) Luciferase-based transcription read-through reporters used in the study: cytomegalovirus (CMV) promoter, ψ-gag (core packaging signal), mutated major splice donor (MSD) is "2KO," all constructs contained the wPRE (not shown; see text). Constructs in [1] are 3′ read-out reporters, and the construct in [2] is the 5′ read-in reporter. (C) Initial testing of the impact of insertion of 20, 40 (R 1-20c), or 60 nucleotides downstream of SV40 late pAS on polyA activity measured by relative Gluc activity in transfected cell lysates compared with the 3′ SIN-LTR (data mean [SD], $n = 3$, $^*p = 0.002$. GFP Expression score (MFI x %

(legend continued on next page)

the first ~55 nucleotides of the viral RNA, leading to the recruitment of transcription elongation factors.[13–15]

The consequences of employing SIN-LTRs in standard 3rd-generation LVs is the inherent transcriptional read-out by the internal transgene cassette, which relies on the native polyA sequences for termination, unless the cassette is inverted and a heterologous polyA is supplied. Additionally, transcriptional read-in from upstream cell gene promoters occurs, likely made worse by some degree of suppression of the 5′ SIN-LTR pAS by U1 snRNA recruitment to the emergent MSD sequence as LV backbone sequences become transcribed.[16–20]

As part of our wider efforts to systematically improve LV genome sequences, here we have focused on creating a new generation of LVs that are transcriptionally insulated by novel engineering of LTR sequences. We show that these "sequence-upgraded polyA" (supA)-LTRs provide 70- and 100-fold increased termination on sense and antisense strands, respectively, in plasmid reporter assays compared with standard SIN-LTRs. We find that transcriptional read-in to the integrated LV from upstream of the 5′ LTR is reduced by 5- to 20-fold when onboarded onto MSD-mutated LV genomes. Additionally, transgene expression can be increased in a variety of cell lines and primary cells. This was achieved by optimal spacing of the USE-pAS-cleavage-DSE sequences via novel engineering of both 5′ R and 3′ SIN-LTR sequences within LV genome expression cassettes. Importantly, the supA-LTR harbors no transcriptionally active sequences meaning that they are classically and functionally "SIN." Additionally, since we did not alter attachment (att) sequences at 5′ U3/3′ U5 sites, we also observed a similar integration site (IS) distribution to SIN-LTR LVs.

RESULTS

Redefining the LV 3′ SIN-LTR by engineering relative polyA signal and R positioning

Due to the molecular gymnastics of reverse transcription of the LV vRNA, modifications of both 5′ and 3′ LTR sequences are required to bring about optimal spacing of enhancer sequences relative to the primary pAS. Moreover, the 97 nucleotide R sequence of HIV-1 is the basis of strand transfer of the first single-stranded DNA (cDNA) generated by reverse transcription from the tRNA primer-binding site (PBS) in the packaging sequence. This occurs by base-pairing to form an RNA:DNA complex to initiate priming for the minus strand (Figure 1A). Therefore, any reconfiguration of the LTR must ensure that sufficient R homology length exists at both 5′ and 3′ ends of the packaged vRNA. Typically, the powerful SV40 late polyA (SV40pA) is employed as a "back-up" due to poor transcription termination at the 3′ SIN-LTR during LV vRNA formation; this often adds a few hundred nucleotides to the vRNA between the 3′ U5 and the polyA tail. To improve termination during production, and reduce this length, others have positioned heterologous polyA sequences (full or just DSEs) much closer to the native 3′ LTR pAS.[21,22] We simply incorporated the entire back-up SV40pA as part of the improved SIN-LTR and thus utilize the SV40pA USE and DSE elements in their native positions relative to its own pAS. We deleted all 3′ R-U5 sequences and fused the att-ΔU3 SIN region directly to the core SV40 late polyA sequence (Figure 1A). We then took advantage of the fact that the sequence encoded between a given pAS and its cleavage site is retained in the resultant mRNA (or in this case vRNA) and theorized that a certain amount of R sequence could be placed between the SV40 pAS and its cleavage site. Therefore, this "R-embedded" heterologous polyA sequence effectively redefines the boundary of the R region within supA-LTRs with respect to the primary pAS being employed; the pAS itself now located within the SIN-ΔU3 region, not the R sequence. The first step was to empirically determine the shortest sequence of the R region that can be inserted between the pAS and cleavage region of a heterologous polyA sequence, without impacting two fundamental functionalities. First, there needed to be sufficient R region length to allow base-pairing with the 1st strand cDNA during reverse transcription to maintain LV titers, and secondly the R sequence needed to be sufficiently short to provide near-optimal spacing of pAS-cleavage site and DSE/GU-box sequences to obtain highly efficient polyadenylation during LV production.

We created *Gaussia* luciferase (Gluc)-reporter constructs containing a secreted Gluc gene and back-up polyA positioned downstream of test variant sequences. These were either full-length LV genome or 3′ LTR reporters to measure LV titers and transcription read-out of an LV cassette (Figure 1B, #1), or 5′ LTR reporters, to model read-in to an LV cassette (Figure 1B, #2). The GFP transgene was used upstream of polyA sequences for comparison/normalization controls in initial experiments. We generated two variants where R sequences of 1–20 or 1–60 were inserted directly after the SV40 polyA AAUAAA hexamer, and a third variant where the 1–20 sequence was appended with complementary sequence such that a stem loop (SL) is predicted to form (R 1-20c). R 1-20c was made to assess if bringing the pAS and DSE sequences spatially closer might be of benefit (Figure 1C). Adherent HEK293T cells were transfected with these reporters, alongside 3′ SIN-LTR and SV40pA (no R inserted) controls, and then GFP and Gluc expression scores generated. These initial results indicated that the SV40pA was ~50-fold stronger than the 3′ SIN-LTR and that insertion of R 1–20 into the SV40pA only minimally impacted polyadenylation being ~20-fold stronger than the 3′ SIN-LTR (Figure 1C). The R 1–60 and R 1-20c variants improved termination by up to 4-fold, suggesting

GFP cells; log10-transformed). (D) Similar to (C) but within LV genome reporters so that LV production titer as well as Luc activity is compared with the use of the 3′ SIN-LTR (data mean [SD], $n = 3$, $^*p < 0.001$); NA, not applicable. A control lacking IRES-Gluc sequence was included (ΔiGLuc). (E) Defining minimal R sequences within the R-embedded heterologous polyA that maximizes polyA activity and maintains LV titer. LV titer and Gluc activity as per (D). Assessment was made in the context of potential 1st strand cDNA transfer of vRNA variants containing 1, 2, or 3 5′ terminal "G" nucleotides (R 1-X) or just a single terminal G nucleotide (R 3-X) (data mean [SD], $n = 2$); ND – not done. Created in https://BioRender.com.

that pAS-DSE positioning may have been suboptimal, even despite the use of the SL variant (note that R 1–60 is able to form the TAR loop).

To assess the impact of these variants on LV titer, we used the LV genome 3′ LTR reporters to perform a similar experiment, except that LV components were co-transfected to generate LV-GFP vector, which was titrated by flow cytometry of transduced adherent HEK293T cells. A DsRed-reporter plasmid was also spiked-in at 25% to normalized Gluc activity independently of the LV transgene GFP gene (we used this pDsRed normalization control for all further Gluc assays). This yielded broadly similar results, with the SV40pA and R 1–20 variant providing ∼100-fold improved termination compared with the 3′ SIN-LTR variant (Figure 1D). Surprisingly, a pAS knockout variant lost only 2-fold termination activity (in subsequent experiments a greater effect was observed); the titers of this "pA-KO" variant were also within 2-fold of the control vector. Strikingly, the R 1–20 variant appeared to impart exclusive polyadenylation at the desired pAS since resultant LV titers were similar to a non-Gluc reporter construct (ΔiGLuc) that employed the back-up SV40pA only. The titers of the R 1-20c variant were 10-fold lower than the 3′ SIN-LTR control, perhaps suggesting that certain secondary structure in this region of the SV40pA is not well tolerated.

We subsequently demonstrated that the minimal R region sequence could be reduced to just 14 nucleotides without drastically impacting on LV titers, and further (∼2-fold) improvements to polyA activity were observed as the R length reduced (Figure 1E). It has been shown that HIV-1 produces variant genomic RNAs, with 5′ terminal 1G or 2Gs or 3Gs via alternative transcription start site (TSS) usage, and that the 1G variant may be preferentially packaged and/or reverse transcribed.[23–26] We performed analysis on R-embedded polyA variants in which the R region modeled the production of LV vRNA containing three or a single "G" on its 5′ terminus. In other studies, we generated a novel hybrid CMV-RSV promoter to ensure generation of only 1G vRNA (data not shown); thus, in theory, allowing removal of the first two Gs within the R 1–20 embedded sequence. However, to date we have not observed an obvious difference in primary T cell transduction by LVs generated by CMV(3G) or CMV-RSV(1G) expression cassettes (data not shown). We observed a slight trend down in LV titers (HEK293T) for R 3-X variants when using CMV(3G) LV genome cassettes (Figure 1E). This indicates that any packaged 3G/2G vRNAs would generate mismatched overhangs after 1st strand transfer, possibly leading to abortive reverse transcription initiation events. Consequently, we chose to retain the native R region as 1–20 (i.e., 3Gs) within the R-embedded polyA sequence as a conservative measure, in the knowledge this configuration would be compatible with all packaged 1G/2G/3G LV vRNAs.

Engineering of a synthetic DSE/GU-rich box into the 5′ R region

The RNA downstream of the polyA cleavage site encoding the DSE derived from the R-embedded SV40pA is consequently not present within the resultant vRNA packaged into LV particles. Accordingly, without further modification, the USE-pAS-cleavage sequence of the resultant integrated LTR would rely on the native HIV-1 DSE present ∼100 nucleotides downstream in the U5 sequence (see schematic in Figure 2). To generate an optimally positioned DSE to service the "new" pAS, we sought to engineer the 5′ TAR SL to harbor a synthetic GU-rich box to function as a DSE (Figures 2A–2C). This approach is challenging, and previous attempts to insert a DSE/GU box at this position have failed.[22] Retroviruses appear to conserve a 5′ SL structure even when they naturally do not encode for a tat analogue. This was also true for a TAR-deleted (tat-independent) HIV-1 virus that acquired an artificial 5′ SL during infection studies, indicating conservative roles for a 5′ SL in RNA stability and/or packaging sequence formation.[27–29] Rous sarcoma virus (RSV) and mouse mammary tumor virus (MMTV) harbor their pAS immediately upstream of their TSS, and consequently employ GU-rich DSEs within their 5′ SL. First, we assessed the relative polyA strength of SIN-LTR Gluc-reporter constructs containing these two SL-DSEs (R 1–44/45) in the presence or absence of adjacent HIV-1 R-U5 sequence. These initial results indicated that the DSE within the MMTV R 1–44 sequence was not only as potent as the native SV40pA DSE but also retained this termination activity in the context of adjacent R-U5 sequence lacking a functional native HIV-1 pAS, unlike the RSV R 1–45 sequence (Figure 2D). Nevertheless, we designed several 5′ SLs based on both RSV and MMTV DSEs, as well as a synthetic polyA (SPA)[30] (Figures 2A and 2C). The first 20 nucleotides of these engineered SLs conserved the R 1–20 sequence necessary for 1st strand transfer and contained at least one CA dinucleotide inserted downstream to "offer" as a potential polyA cleavage site. These variants were tested for polyA activity within LTR Gluc reporters and assessed for LV production in adherent HEK293T cells (Figure 2E). The experiment revealed that three variants provided further improved transcription termination activity (R-GU5>R-GU2>R-GU4) over the USE-pA control that had no GU-rich DSE (R 1–97). However, the R-GU5 variant impacted LV output titers, and in follow-on experiments the R-GU4 variant appeared substantially dependent on the presence of the HIV-1 pAS retained downstream (data not shown). Therefore, given the R-GU2 variant provided the greatest reduction in transcription read-through while retaining high LV output titer, we focused further development of the supA-LTR based on R-GU2.

Further characterization of the 5′ R-GU2-based supA-LTR

Next, we assessed the relative importance of the sequence elements resident within the supA-LTR as present after reverse-transcription, as well as the reliance of polyA activity on the native HIV-1 pAS within the final R-U5 sequence. We generated a set of LTR Gluc reporters where one or more of the USE, pAS, R-GU2, and HIV-1 pAS were deleted and evaluated transcription read-through activity in HEK293T cells (Figure 3A). Deletion of the native HIV-1 pAS from a standard SIN-LTR resulted in a 9-fold increase in read-through (construct 2 vs. 1). Similarly, a supA-LTR deleted for all key sequences as well as the HIV-1 pAS (i.e., still containing other heterologous sequences from cloning) was 4-fold increased in

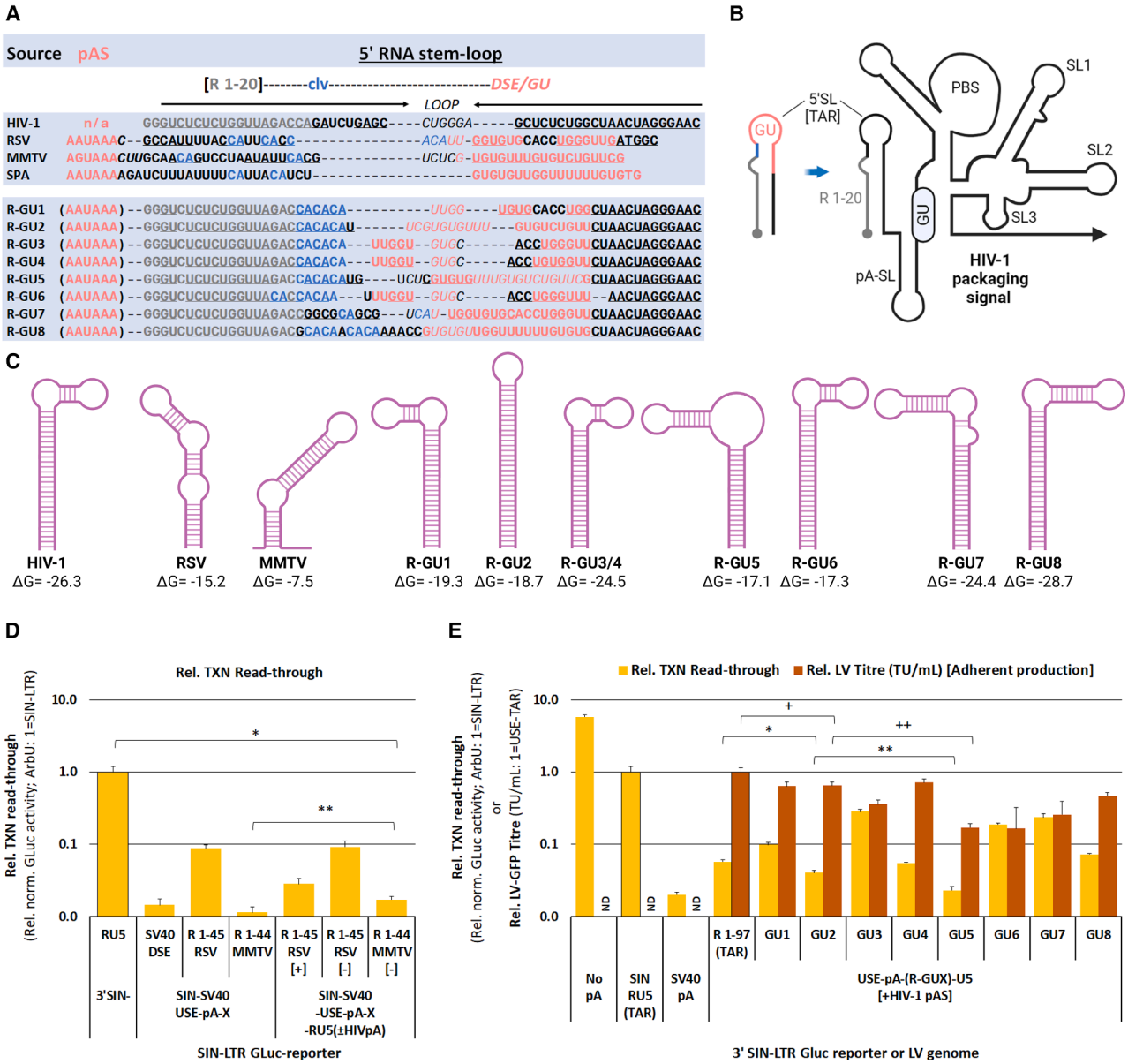

Figure 2. Engineering of a synthetic GU-rich DSE within the 5′ stem loop of the LV genomic RNA

(A) A schematic comparing wild-type virus (HIV-1, RSV, and MMTV) 5′ RNA genomic sequences that form a terminal stem loop (SL), and where potential polyA cleavage sites (CA) and relevant GU-rich DSE sequences are positioned for RSV/MMTV. The synthetic polyA-DSE variant "SPA" is also compared (see text for reference). Novel DSEs based on those of RSV, MMTV, and SPA were engineered into the HIV-1 TAR loop as part of the wider packaging signal shown in (B). R 1–20 and potential cleavage CA di-nucleotides are maintained in all variants R-GU1 to R-GU8. (C) Predicted 5′ SL folding and stability (ΔG) for each of the variants compared with virus counterparts. (D) Initial evaluation of RSV and MMTV DSE strength within LTR Gluc reporter constructs. All test constructs contained the native SV40 late USE-pAS with native virus R regions of stated length immediately downstream. One set had no HIV-1 R 1–20 or other R-U5 sequence present, whereas the second set included a fused R-U5 sequence where the presence/absence of the native HIV-1 pAS was also tested (data mean [SD], n = 3, $^{*}p$ = 0.012, $^{**}p$ = 0.028). (E) Evaluation of the polyA strength of the eight R-GUX variants within a luciferase LTR-reporter assay and the impact on LV production titers (data mean [SD], n = 3 [TXN Read-through], $^{*}p$ = 0.0038/$^{**}p$ = 0.003; n = 4 [LV titer], ^{+}p = 0.005/^{++}p < 0.001). ND, not done. Created in https://BioRender.com.

read-through compared with the SIN-LTR (construct 4 vs. 1). Restoring the native HIV-1 pAS in this construct improved polyA activity by 7.5-fold (construct 5 vs. 4) and 2-fold relative to the SIN-LTR (construct 5 v 1). In contrast, the insertion of the new pAS alone had no impact (construct 6 vs. 4), suggesting that the HIV-1 DSE present in U5 was only able to enhance the native pAS (constructs 5 and 7). Provision of the USE in combination with the new pAS produced a 4-fold increase in polyA activity in the

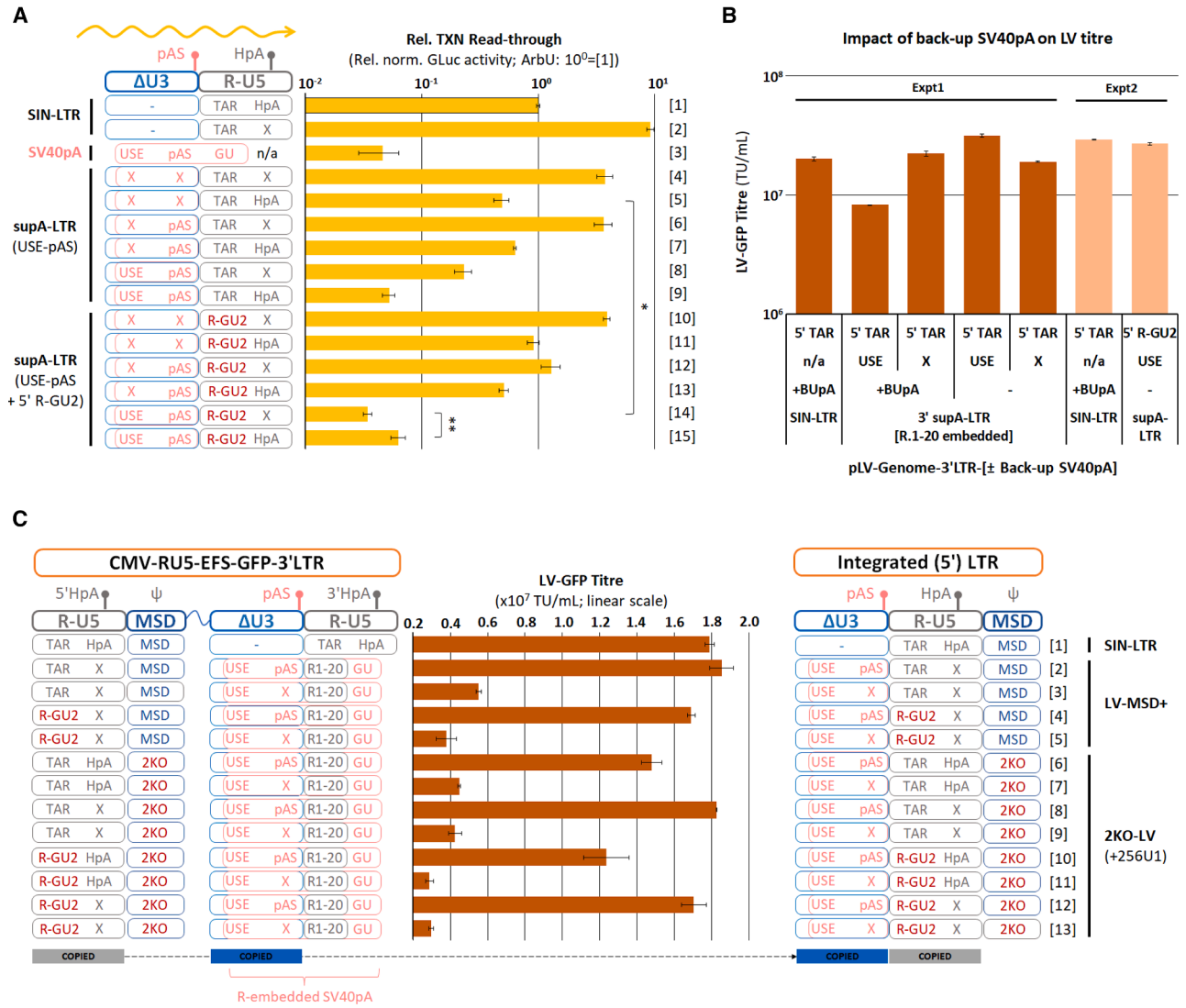

Figure 3. The impact of different functional sequences within the supA-LTR on polyadenylation and LV production

(A) Evaluation of SIN/supA-LTR configurations within 3′ LTR Gluc reporters in adherent HEK293T cells. The presence/absence of the USE, pAS, R-GU2, and the native HIV-1 pAS was tested in different variants [4–15] and compared with the SIN-LTR (+/− HIV-1 pAS; [1–2]) and the SV40 late polyA [3]. Normalized Gluc activity is plotted relative to that of the SIN-LTR [1] (data mean [SD], $n = 3$. *$p = 0.008$/**$p = 0.006$). (B) Assessment of the impact of the presence of two closely positioned USEs within pLV-Genome plasmids during production in suspension HEK293T cells. Presence of the USE in the supA-LTR and/or within the back-up SV40 polyA (+/− BUpA) is denoted (data mean [SD], $n = 2$). (C) The figure relates how the 5′ and 3′ LTR sequences in the pLV-Genome plasmid are reconfigured to LTRs flanking the integrated LV cassette and the production titers of LVs harboring the different variant sequences. Variants included the presence/absence of the 5′R-GU2 stem loop in the context of an intact/mutated 5′ HIV-1 pAS and/or an intact/mutated major splice donor (MSD+/2KO) (data mean [SD]; $n = 2$). LVs were produced in suspension HEK293T cells; only the SIN-LV plasmid contained the back-up polyA due to the effects reported in (C) (see text). Created in https://BioRender.com.

absence of the HIV-1 pAS relative to the SIN-LTR (construct 8 vs. 1), which was further increased to 20-fold when the HIV-1 pAS was restored (construct 9 vs. 1), indicating that the HIV-1 pAS-DSE was behaving as a back-up polyA signal. When the synthetic R-GU2 DSE was employed, the greatest reduction in read-through activity of 29-fold was observed in the absence of the HIV-1 pAS (construct 14 vs. 1). Overall, the USE appeared to provide the greatest enhancement to polyadenylation at the new pAS, and the R-GU2 DSE provided complete independence from downstream native HIV-1 sequences.

During this work we transitioned from the 3′ LTR LV genome Gluc reporters, which can generate LVs for titration while simultaneously reporting on LTR polyA activity, to simplified LV genome plasmids, by simple deletion of the IRES-Gluc sequence (see Figure 1B). Consequently, this provided the supA-LTR variants with the

SV40pA as a back-up polyA in close proximity (as typically used in SIN-LTR LV genome cassettes), even though we demonstrated that the back-up is redundant with the 3′ supA-LTR. Surprisingly, however, when we moved from adherent to suspension HEK293T cells, we observed a modest drop in LV titer relative to the standard SIN-LTR containing LV (data not shown). To probe this further, we generated pGenome plasmids containing the 3′ supA-LTR with or without the USE element, and with or without the SV40pA back-up, and produced LVs in suspension HEK293T cells (Figure 3B). These data indicated that the presence of two closely positioned USE elements (one in the R-embedded SV40pA and the other in the back-up SV40pA ∼200 nucleotides downstream) was detrimental to LV titers, presumably due to some kind of interference in suspension HEK293T cells. A similar effect has been reported elsewhere, although in adherent HEK293T cells.[31] Therefore, all further supA-LTR pGenome plasmids did not employ the back-up SV40pA.

Toward "SupA2KO" LVs: Pairing supA-LTRs and MSD-mutated LV genomes

In previous work, we have developed MSD-mutated LV genomes to ablate aberrant splicing from the packaging signal into the transgene cassette during LV production, which we refer to as "2KO-LVs."[32] For 3rd-generation LVs (MSD+), such aberrantly spliced vRNAs typically express the transgene during production and can be detected in LV product and converted into episomal cDNAs in target cells. We have also observed that 2KO-LVs are subject to 2- to 5-fold less 5′ SIN-LTR transcriptional read-in in transduced primary cells and cell lines (Figure S1); we were interested to assess if this was of benefit in combination with the supA-LTR. 2KO-LVs require the co-expression of a modified U1 snRNA ("256U1") that binds to the LV vRNA packaging region, which negates the attenuating effect of MSD-mutation on production titers. The supA-LTR was transitioned into the 2KO-LV genome to assess production in suspension (serum-free) HEK293T cells and the impact of deletion of several functional sequences (Figure 3C). The results show that a 3rd-generation (MSD+) or 2KO-LV expression cassette harboring just the 3′ supA-LTR sequence (R 1–20 embedded SV40pA) could generate the same titer as that of the standard SIN-LTR LV, independently of the native HIV-1 pAS (constructs 2 and 8 vs. 1). Deletion of the new pAS and native 5′ HIV-1 pAS resulted in no polyA signals being present in the final integrated LV, and a 4-to-6 fold reduction in titer was observed (constructs 3, 5, 9, 13 vs. 2). The use of the 5′ R-GU2 DSE did not impact titers (construct 4, 10, and 12 vs. 2), although there appeared to be a slight benefit in mutation of the native 5′ pAS when using a 2KO-LV (construct 10 vs. 12). Interestingly, the employment of the USE alone could not restore titers of a 2KO-LV containing essentially a standard SIN-LTR (i.e., with native HIV-1 pAS; construct 7 vs. 9). The USE-only modification had a similar impact on production titers of an MSD+ LV, shown in a follow-on experiment (Figure S2A; construct 4 vs. 2). Finally, we demonstrated that the 5′ R-GU2 DSE could be used at the 3′ supA-LTR as part of the R-embedded SV40pA, entirely replacing the SV40pA DSE element (Figures S2A and S2B). In summary, we established that the supA-LTR sequences could be utilized within both 3rd-generation (MSD+) and 4th-generation 2KO-LV cassettes, without impacting on output titers.

Engineering of a bi-directional "supA-2pA-LTR" by insertion of an inverted pAS-DSE into the 3′ SIN-U3 region

We sought to modify the supA-LTR further to encode functional polyA sequences on the antisense strand to protect the integrated LV cassette from transcriptional read-in from outside the 3′ LTR, to give a "supA-2pA" ("sooper-tooper") LTR. This would also provide a back-up polyA for the internal heterologous polyA typically employed with an inverted transgene cassette, since these sequences would be copied to the 5′ LTR after reverse transcription. Neither the HIV-1 LTR nor standard 3rd-generation LV SIN-LTR employ polyA sequences on the antisense strand, and so there is potential for antisense (exonic) cellular read-in transcripts to form dsRNA with the transgene mRNA. The core SV40 "late" polyA USE sequence used in 5′-to-3′ direction within the supA-LTR also encodes for the two native SV40 "early" pAS AAUAAA hexamers on the opposite strand. Consequently, we tested several inverted GU-rich DSE sequences immediately downstream of the second SV40 early pAS within LTR Gluc reporter constructs in suspension HEK293T cells (Figure 4A). The GU-rich DSEs were from the native SV40 early polyA (iGU1), the DSE used within the 5′ R-GU2 SL (iGU2), rabbit β-globin (HBB) polyA (iGU3), Herpes Simplex Virus thymidine kinase (hsvTK) polyA (iGU4), a modified MMTV polyA DSE (iGU5), the Melanocortin 4 receptor (MC4R) polyA (iGU6), and a modified version of the DSE within the SPA (iGU7) (see Figure S3). Normalized luciferase activity was plotted relative to an inverted LTR containing the USE-pAS sequence but with a native R-U5 harboring a mutated HIV-1 pAS (construct 2), which alone provided 10-fold less transcription read-in compared with no LTR (construct 1). Deletion of the SV40 early twin polyA signals resulted in ablated transcription termination (construct 3 vs. 1). Of the variants containing a GU-rich DSE downstream of the twin polyA signals, iGU1, iGU2, iGU5, and iGU7 provided ≥100-fold reduced transcriptional read-in compared with no LTR. The use of the R-GU2 sequence as part of the sense strand supA-LTR did not generally impact this activity, except when employed with the same GU2 sequence within iGU2 (construct 9), where its activity was effectively negated. Given the ability of the two GU2 sequences separated by only 112 nucleotides to base-pair, we theorize that such an interaction inhibits polyadenylation.

We then assessed the impact of these inverted DSE sequences on sense strand transcription termination by (re-)inverting these supA-2pA LTRs into Gluc reporters and testing in suspension HEK293T cells (Figure 4B). For iGU1 and iGU2 we included an additional control where the native HIV-1 pAS was re-instated to ascertain whether inverted GU-rich sequences might impact on supA-LTR in-/dependence on this signal. Normalized luciferase activity was plotted relative to an LTR containing the USE-pAS sequence but with a native R-U5 (construct 3), which alone provided 10-fold less transcription read-in compared with a standard SIN-LTR (construct 2). Instating a supA-LTR with or without the native HIV-1 pAS provided ∼37-fold greater transcriptional

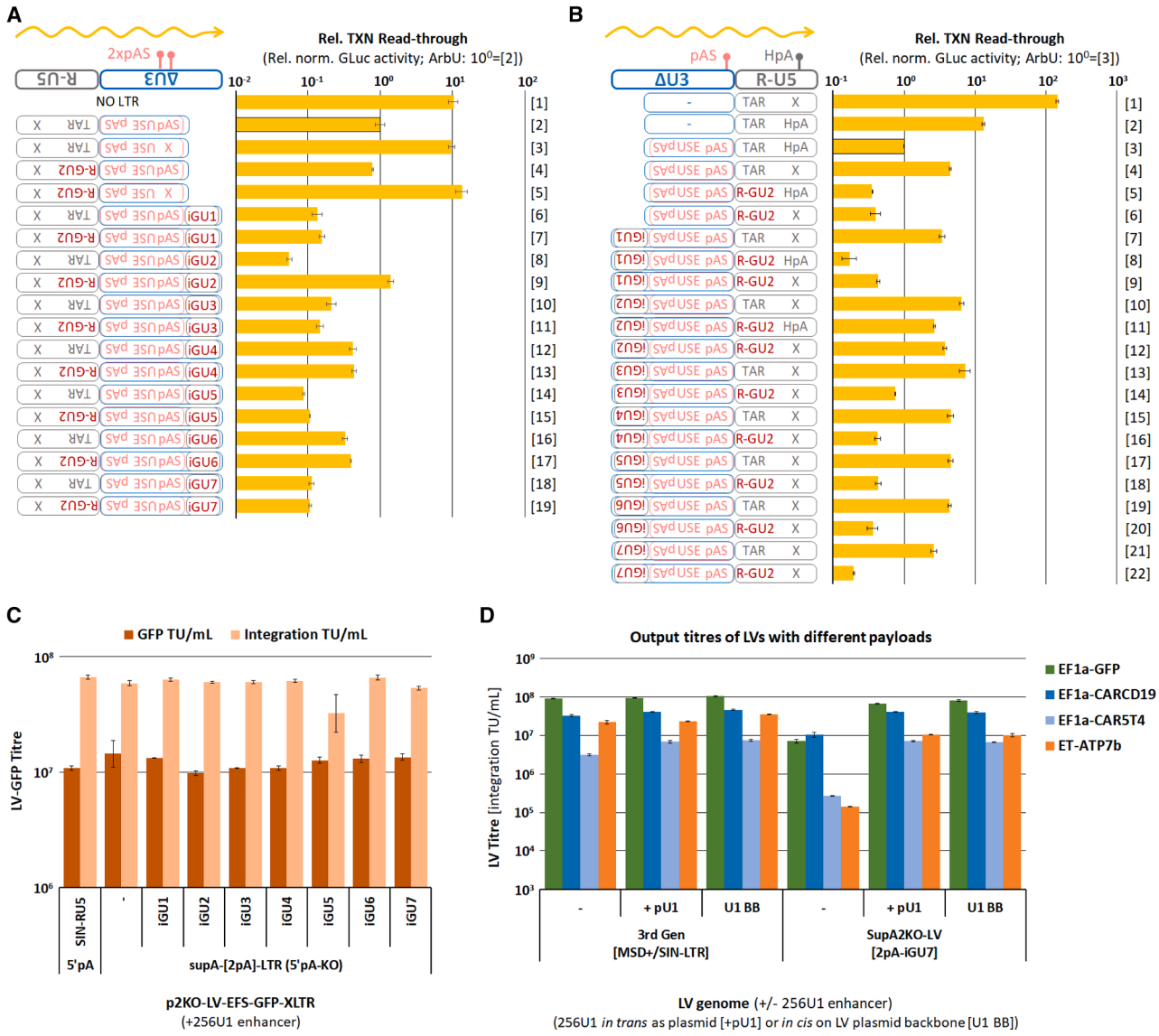

Figure 4. Further development of the supA-LTR to include polyA sequences on the antisense strand

(A) Evaluation of transcription read-through of inverted 3′ LTRs using Gluc reporter plasmids. All inverted LTRs harbored the USE-pAS sequence (of SV40 late polyA) and contained the inverted R-U5 lacking the HIV-1 pAS, and either the inverted TAR or R-GU2 stem loop. On the top strand downstream of the two SV40 early polyA signals (2xpAS) the test GU-rich DSEs 1–7 were inserted (see Figure S3 for sequences). Normalized Gluc activity is plotted relative to that of construct [2] (data mean [SD], $n = 2$). (B) Evaluation of transcription read-through of "re"-inverted LTR sequences from (A), including the test sequences iGU1-7. Additionally, variants iGU1 and iGU2 were tested with the native HIV-1 pAS present [8] and [11], respectively. Normalized Gluc activity is plotted relative to that of construct [3] (data mean [SD], $n = 2$). (C) Production titers for 2KO-LVs harboring variant supA(2pA)-LTRs iGU1-7 compared with a 3rd-generation SIN-LV (data mean [SD], $n = 2$[GFP TU/mL], $n = 4$[Integration TU/mL]). (D) Production (Integration) titers of SupA2KO-LVs containing the iGU7 variant with different payloads and sizes (EF1α-GFP-wPRE = 2.6kb; EF1α-CARCD19-wPRE = 3.3kb; EF1α-CAR5T4-wPRE = 3.4kb; ET-ATP7b-wPRE = 7.8kb). Production of SupA2KO-LVs is dependent on co-expression of the 256U1 snRNA enhancer molecule that rescues the effect of MSD-mutation by the 2KO feature. 256U1 can also enhance production of some 3rd-generation LVs. The 256U1 expression cassette is most simply employed by insertion in *cis* within pLV genome backbones (BB). Use of different promoters and GOI transgene cassettes are denoted: EF1α and synthetic Enh1mTTR (ET) promoter; two chimeric antigen receptor (CAR) against CD19 or 5T4; and ATPase copper transporting beta (ATP7b) (data mean [SD], $n = 2$). Created in https://BioRender.com.

insulation than the SIN-LTR, consistent with previous results (constructs 5, 6 vs. 2). The presence of the inverted DSE within the iGU1 variant appeared to improve transcriptional termination of the supA-LTR, but only when the native HIV pAS was present (construct 8 vs. 9). A similar effect was observed with iGU2; however, the overall polyadenylation activity of an iGU2-containing supA-LTR was ∼10-fold lower than the supA-LTR (constructs 11, 12 vs. 5, 6), presumably due to a similar potential mechanism of

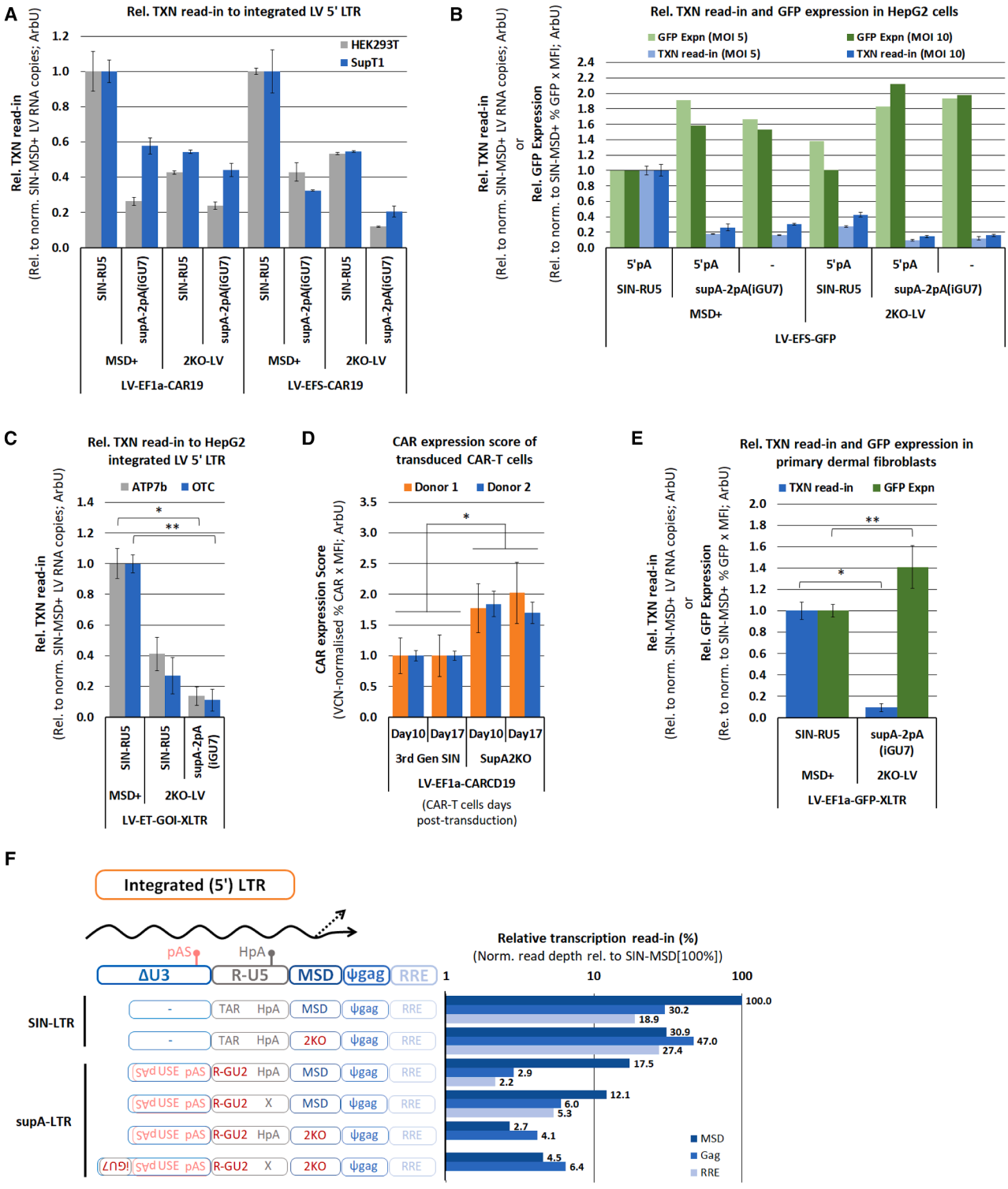

Figure 5. Quantification of transcription insulation and transgene expression from integrated LVs harboring SIN or supA LTRs

Transgene expression and/or transcription read-in to 5′ LTRs of integrated 3rd-generation LVs (SIN/MSD+) or SupA2KO-LVs containing different transgene cassettes in a variety of cell lines and primary cells. (A) to (C) and (E) Transcriptional read-in was quantified by RT-qPCR (for amplicon position see Figure 6A) of total RNA extracted 10 days post transduction; data (mean [SD]) was normalized to integrated vector copy number, and plotted relative to 3rd-generation LV (set to 1). GFP Expression scores (%GFP-

(legend continued on next page)

base-pairing between GU2 sequences. SupA-2pA-LTR variants 4, 5, and 6 had similar polyadenylation activities as the supA-LTR (constructs 16, 18, 20 vs. 6), whereas iGU7 appeared to provide a slight improvement to sense strand transcription termination over the "original" supA-LTR (construct 22 vs. 6).

Next, we generated 2KO-LV-EFS-GFP vectors with standard SIN-LTR (+Back-up polyA), supA-LTR, or the iGU-appended supA-2pA-LTRs in suspension HEK293T cells and titrated clarified harvest supernatants by both flow cytometry and integration assay (Figure 4C). As expected, both the integration and GFP titers were generally similar. Since the iGU7 variant provided the greatest transcriptional termination activity on both sense (∼70-fold) and antisense (∼100-fold) strands, and enabled high titer LV production, we moved forward with this variant.

We then evaluated the output titers of SupA2KO(iGU7)-LV genomes containing payloads with different promoters and transgenes, compared with standard 3rd-generation LVs (Figure 4D). We compared LV production in the absence/presence of the 256U1 snRNA enhancer (essential for 2KO-LV production) by provision of the 256U1 expression cassette either *in trans* as a separate plasmid or *in cis* within the pLV genome plasmid (adds just ∼700 nt to plasmid size). These data demonstrate that SupA2KO-LVs can be produced as efficiently as 3rd-generation LVs, independently of the payload. The "net" change in vRNA size when considering both the insertion of novel polyA sequences and the mechanism of vRNA formation is minimal. While technically the SupA2KO-LV vRNA genome (R-to-R) is 9 nucleotides longer, when also considering inefficient polyadenylation at the 3′ SIN-LTR of the 3rd-generation LV (that necessitates the back-up SV40 polyA), in practice the SupA2KO-LV vRNA is slightly shorter by 210 nucleotides compared with our 3rd-generation LV. We do not expect such minimal size differences to be biologically meaningful, and our empirical LV titer data demonstrate that SupA2KO-LV vRNA is similarly structured (Psi forms correctly), is of similar stability, and undergoes cDNA formation/integration equivalently to 3rd-generation LVs. Interestingly, we have observed that output titers of SupA2KO-LVs can be further enhanced by addition of a novel chemical inducer Ingenol 3-angelate[33] and now routinely use this in SupA2KO-LV production.

Finally, we applied the supA-2pA-LTR into LV genomes containing subtly different SIN-LTR sequences from widely available sources and produced MSD+/2KO-LV-EFS-GFP vectors in suspension HEK293T cells (Figure S4).[34] These differed in the amount of retained Nef sequence upstream of the 3′ ppt and the extent of U3 sequence deletion (Figure S4A). The SIN-LTRs employed in this study contain the least amount of native HIV-1 sequence compared with other sources, resulting in a 412-nucleotide deletion in U3 ("OXB"). Only 92 nucleotides of Nef sequence are retained, which encompasses the 3′ ppt and just 38 nucleotides into U3. This aspect is retained within supA-2pA-LTRs, with the 99 nucleotides of the polyA (iGU7) sequences between the att-ΔU3 region and the new pAS. These data indicate that longer retained HIV-1 sequence upstream of the polyA sequences within supA-2pA-LTRs do not impact SupA2KO-LV or standard MSD+ LV output titers, but only that maximally deleted U3 regions may be preferred generally (Figure S4B).

Reduced transcriptional read-in and improved transgene expression of integrated SupA2KO-LVs

Having demonstrated the improvements to transcriptional termination observed for the supA-2pA-LTR in preliminary transient transfection (Gluc) assays, we then focused on performance within integrated LVs. We sought to quantify 5′ transcription insulation of integrated SupA2KO-LVs by RT-qPCR, as well as transgene expression in different contexts. We generated 3rd-generation SIN-LVs (MSD+) and SupA2KO-LVs harboring several different transgene cassettes, titrated them by integration assay (HEK293T cells), and then used these stocks to transduce target cells at matched MOIs. Transduced cells were passaged for 10 days to ensure non-integrated LV cDNA was lost from cells, as we have demonstrated previously.[32,35] This enabled assessment of transcription read-in to the 5′ LTR by use of an RT-qPCR assay of total extracted RNA that employed an RT step using a mix of oligo-dT primers and random hexamers to generate cDNA. The amplicon primer-probe set was just upstream of the MSD (see Figure 6A). Thus, transcribed 5′ LV backbone sequences originating from transcription initiation from upstream cellular promoters could be quantified (Figures 5A–5C and 5E). In some settings we measured GOI expression by (immuno-)flow cytometry. Data were normalized by integrated vector-copy number (VCN) analysis (qPCR to Psi-gag). These results show that transcriptional read-in to the 5′ supA-2pA-LTR of integrated LV-EF1a/EFS-CAR19 cassettes in HEK293T or SupT1 cells was reduced by 2- to 5-fold when employed in MSD+ LV genomes and was reduced by up to 10-fold when combined with a 2KO-LV (Figure 5A). Transduction of HepG2 cells at MOI 5 or 10 by 3rd-generation SIN-LVs or SupA2KO-LVs encoding EFS-GFP provided a similar pattern

positive x median fluorescence intensity) were generated by flow cytometry and normalized/plotted in the same way. (A) LV-EF1α/EFS-CAR19 transduction of HEK293T and SupT1 cell lines (n = 2). (B) LV-EFS-GFP transduction of HepG2 cell line at stated MOIs; presence/absence of HIV-1 polyA signal denoted (n = 2). (C) LV-ET-ATP7b and LV-ET-OTC transduction of HepG2 cell line, n = 4 $^{*}p < 0.001$, $^{**}p < 0.001$. (D) LV-EF1α-CAR19 transduction of primary T cells with relative CAR expression at day 10 or 17 post transduction, n = 4, $^{*}p < 0.001$ [donor 1], $p < 0.001$ [donor 2]. (E) LV-EF1α-GFP transduction of primary human dermal fibroblasts (see Figure 6)—three independent transductions across five MOIs; transcription read-in, n = 15 $^{*}p < 0.001$; GFP Expression score, n = 15 $^{**}p < 0.001$. (F) Evaluation of transcription read-in to 5′ LTR of integrated LV-EFS-GFP cassettes in HEK293T cells by RNA-seq. LVs varied in MSD status (MSD+ or 2KO), presence/absence of the native HIV-1 pA (HpA), and use of either SIN or supA LTRs. Number of VCN-normalized RNA reads mapping to the MSD, ψ-gag, and RRE regions (i.e., mobilized vRNA) are plotted relative to the number of MSD region reads for the LV-MSD-SIN vector. When the MSD is present, splicing from this position to elsewhere (i.e., over ψ-gag/RRE) leads to lower ψ-gag/RRE reads. Created in https://BioRender.com.

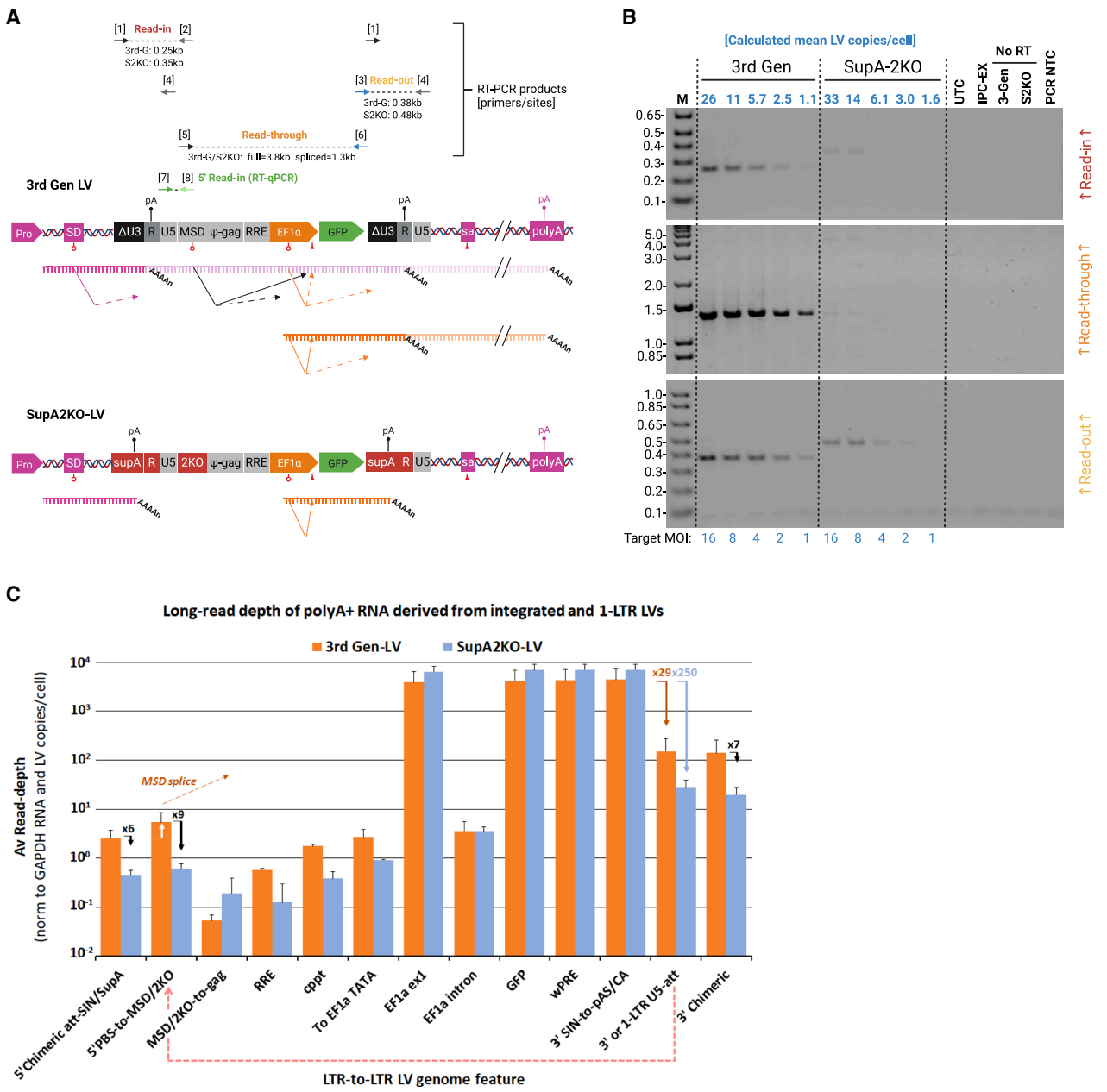

Figure 6. Assessment of transcription insulation of SIN-LVs and SupA2KO-LVs in primary human dermal fibroblasts

(A) A schematic showing the structure of 3rd-generation LV or SupA2KO-LV encoding an EF1α-GFP-wPRE cassette (wPRE not shown) integrated into target cells. The integrated LVs are shown in the context of a theoretical insertion site within a 5′ intron of a cellular transcription unit; consequently, these flanking sequences can interact with LV sequences (see references within text). For 3rd-generation LVs, where the polyadenylation sequences within the SIN-LTRs are weak, transcriptional read-through the integrated LV can result in splicing from the major splice donor (MSD) to internal splice acceptors, or potentially cell gene splice acceptors downstream of the 3′ SIN-LTR (LVs tend to integrate into introns). SupA-LTRs greatly reduce transcriptional read-in, and the absence of MSD activity imparted by the 2KO feature means that no further splicing interactions are possible. The positions and size of the three RT-PCR amplicons utilized in (B) are shown, flanked by the position of primers ([1] to [6], represented by arrows; see methods for sequence details and Figure S8 for annealing contexts). The position of the 5′ read-in (RT)qPCR amplicon, used for quantification analysis in Figures 5A–5C and 5E, is defined by primers [7] and [8]. (B) Agarose gel images of amplicon products generated by the three RT-PCRs to assess read-in, read-through, and read-out denoted in (A), within total extracted RNA from pools of human primary dermal fibroblasts transduced with stated LVs at denoted MOIs. Actual (calculated) mean vector copies/cell was done by qPCR to ψ-gag on extracted genomic DNA. Comparison of amplicon product yields is possible between 3rd-generation LVs and SupA2KO-LVs within each RT-PCR assay type, but not between RT-PCR assays due to primer/PCR differences. The positive correlation between actual (calculated) mean LV copies/cell and amplicon yield indicates that these comparisons are meaningful. Note that a proportion of amplicon signal is from RNA derived from non-integrated 1-/2-LTR episomal

(legend continued on next page)

whereby supA-2pA-LTRs reduced transcription read-in by up to 5-fold in MSD+ LV genomes and by 10-fold when employed within a 2KO-LV (Figure 5B). Assessment of GFP expression (VCN-normalized GFP Expression score: MFI x %-positive cells) in these transduced populations indicated that GOI expression was enhanced by up to 2-fold when using SupA2KO-LV genomes (Figure 5B). A further set of HepG2 transductions, this time with LVs encoding Copper-transporting ATPase 2 (ATP7b) or Ornithine transcarbamylase (OTC) driven by the semi-synthetic Enh1mTTR ("ET") promoter, again demonstrated substantially less 5′ transcriptional read-in to the SupA2KO-LV cassette, by 7- and 9-fold, respectively, compared with the standard 3rd-generation LV harboring a SIN-LTR (Figure 5C). Additionally, we generated CAR-T cells by transducing activated PBMCs with either 3rd-generation LV or SupA2KO-LV encoding CAR-CD19 and measured CAR expression at days 10 and 17 post transduction. CAR expression scores (% positive x MFI) were normalized to VCN and showed a consistent ∼2-fold greater CAR expression in CAR-T cells produced by SupA2KO-LV transduction (Figure 5D). A 140% increase in transgene expression imparted by the SupA-LTR was observed in human primary dermal fibroblasts (hDFs), which was accompanied by a 10-fold reduction in 5′ transcriptional read-in (Figure 5E); see later for greater details regarding transductions and expanded analyses for the hDF study.

Finally, we performed RNA-seq (Illumina) on total extracted RNA from HEK293T cells transduced (MOI 1) with MSD+, 2KO-LV, and supA/supA-2pA-LTR vectors harboring an EFS-GFP cassette. This analysis reported on percentage RNA read-depth just upstream of the MSD, the Psi-gag region (Gag), and the RRE sequence, to provide a measure of transcription read-in to the 5′ LTR (Figure 5F). These data were normalized to VCN, determined by qPCR against the packaging signal DNA, and then plotted relative to the read-depth for the standard 3rd-generation SIN-LV at the upstream-MSD position (set to 100%). The reason for doing this was that, for MSD+ LVs, the MSD read-depth was decoupled from that of the downstream Gag and RRE positions, generally showing a greater detection of transcribed RNA upstream of the MSD, which was in agreement with MSD-mediated splicing-out of gag/RRE sequences to splice acceptors in the downstream cassette (see later), or possibly *trans*-splicing out to adjacent cellular exons.[18] As expected, for integrated 2KO-LV genomes, the relative amount of the MSD/gag/RRE sequences were more similar to each other, since there is no splice donor to partake in splicing events. When the 2KO-LV was paired with the supA/supA-2pA-LTRs the detection of transcribed LV backbone sequences was reduced by > 20-fold compared with the 3rd-generation SIN-LV. Indeed, for SupA2KO-LVs, RRE sequences were not detected at all, indicating that the variability of low read number below 4% may have been close to the level of detection of the RNA-seq approach. We attempted to analyze potential chimeric reads (i.e., cellular RNA transcripts fused to LV backbone sequences) in the short-read dataset but could not overcome the technical challenge of mis-mapping shared SV40 polyA sequences between the supA-LTR and the endogenous SV40 T-Antigen cassette present within HEK293T cells. Instead, we performed the in-depth study in primary cells (lacking endogenous SV40 polyA sequences) reported in the following sections.

Assessment of transcriptional insulation of SIN-LVs and SupA2KO-LVs in transduced primary cells

We produced stocks of a 3rd-generation SIN-LV (MSD+) and a SupA2KO-LV encoding an EF1α-GFP cassette and titrated them by integration assay on adherent HEK293T cells. We dosed cultures of hDFs with increasing MOIs (1 to 16) of both stocks of LVs, passaged for 10 days, before measuring GFP expression (see Figure 5E), and extracting genomic DNA and total RNA. Copy-number analysis of genomic DNA revealed slightly higher-than-target transduction in general, and ∼125% by the SupA2KO-LV stock compared with the 3rd-generation LV. We performed three sets of RT-PCRs on total RNA using primer pairs displayed in Figure 6A (RT primer was a mix of oligo-T and random hexamers). These amplicons were designed to assess read-in and read-through from upstream cellular promoters, as well as transgene read-out passed the 3′ polyA cleavage site and then on to the 3′ U5 sequence. Since the total RNA substrate was derived from pools of transduced cells representing a plurality of ISs (and associated position effects), the RT-PCR results report on the inherent properties of the LTRs of the two LV types. Despite the semi-quantitative nature of RT-PCR analysis, there was a very good positive correlation between amplicon yield and the actual (calculated) copies/cell achieved (Figure 6B). The read-in amplicon (Figure 6B, upper panel) utilized reverse primer [2], which binds just upstream of the MSD, and substantially more product was generated by the 3rd-generation LV in a dose-dependent manner. The SupA2KO-LV also generated less read-out (Figure 6B, lower panel), although, since the reverse primer [4] binds to the U5 sequence, we cannot exclude the possibility that this signal represents GFP pre-mRNA immediately prior to cleavage and polyA addition. The most striking result from this analysis was the difference in the "read-through" RT-PCR amplicon (Figure 6B, middle panel). This amplicon is derived from forward primer [5] upstream of the MSD and reverse primer [6], which binds to the wPRE (not shown in schematic; between GFP and 3′ LTR). The size of this amplicon (1.3kb) correlates to the product of the MSD-to-EF1α splice acceptor site (Figure 6A; solid black arrow), rather than a

cDNA (see Figure S6). Controls included untransduced cell control (UTC), in-process & RNA extraction control (IPC-EX), no reverse-transcriptase control to demonstrate absence of contaminating pDNA within LV stocks at transduction (No RT), and a no-template PCR mix control (PCR NTC). (C) Long-read (nanopore) analysis of polyA+ RNA from human primary dermal fibroblasts transduced with stated LVs. Data are mean [SD] read-depth across the integrated LV cassettes from both sets of samples from MOI 2 and MOI 16 transductions [$n = 2$]. The pink dotted arrow on the x axis indicates the "read-around" transcript derived from 1-LTR circular episomes (present at ∼30% total LV DNA; see Figure S6) and potential splicing from the MSD to the EF1α splice acceptor. The white arrow indicates that potentially up to 50% of aberrant MSD spliced product may be coming from "read-around" 1-LTR episome transcription, although 5′ chimeric read-in may be under-detected.

"full-length" amplicon (3.8kb). This implies that transcriptional read-into the 5′ SIN-LTR continues across the entire 3rd-generation LV cassette, and the promiscuous MSD splices to the EF1α splice acceptor site (as it does in LV production[32]), resulting in a transcript that is readily detectable at ∼1 integrated LV copy/cell. The integrated SupA2KO-LV did not generate this spliced product (since the MSD is inactivated), nor was there any evidence of stable production of the full-length transcription read-through 3.8kb transcript. We conducted a similar, supportive experiment in the hDFs using LVs encoding a CMV-driven (intron-less) cassette encoding BMP2 (Bone Morphogenetic Protein-2[36]) presented in Figure S5. Here, an ∼5-fold reduction in 5′ transcriptional read-in to the SupA2KO-LV was observed. The apparent lack of substantial read-through product from the 3rd-generation LV cassette in this case seems to be due to the MSD not "finding" a strong splice-acceptor site in the internal transgene sequence. We sequenced the read-through RT-PCR amplicon to find that the MSD had spliced to a weak cryptic splice acceptor within the transgene 5′ UTR (data not shown).

Quantification of transcriptional insulation of SIN-LVs and SupA2KO-LVs in transduced primary cells

We performed long-read (nanopore) sequencing of polyA+ RNA from the LV-EF1α-GFP transduced primary human dermal fibroblasts, extracted from whole cell samples 10 days post transduction at MOI 2 and 16. We compared relative read-depth (normalized to GAPDH RNA input and VCN) for both 3rd-generation LV and SupA2KO-LV backbone sequences. Figure 6C reflects this quantified assessment and generally supports the RT-PCR analysis in Figure 6B, but with important caveats (see below). Data are plotted as average read-depth at key positions across the LV cassettes (mean SD of MOI 16 and 2 independent analyses). The read-depth across the internal EF1α-GFP cassette was very similar for both LV types, with clear loss of the EF1α intron, and reflective of high mRNA production from the EF1α promoter. The read-depths at the 3′ U5 attachment sequence (3′ att) were 29-fold and 250-fold lower for 3rd-generation and SupA2KO LVs, respectively, compared with 3′ LTR sequences upstream of the polyA cleavage site, which was again in-line with the RT-PCR analysis. The number of 3′ chimeric reads (transcription read-out) was of a similar magnitude to the read-depth at the 3′ att position and reflected a similar ∼7-fold reduction for the SupA2KO-LV.

A similar difference (∼6-fold) between the two vector types was observed for 5′ chimeric reads, even though the number of reads was ∼50-fold lower compared with 3′ chimeric reads (reflecting power of the internal EF1α promoter). However, we noticed a slight disconnect between the number of 5′ chimeric reads and the read-depth across the PBS-to-MSD region of the 3rd-generation LV, which were ∼2-fold higher. While this difference was possibly not statistically significant, it led us to perform further analysis to assess whether non-integrated LV (i.e., 1-/2-LTR circles) had been present in the transduced hDFs samples and had contributed to transcribed RNA that mapped to LV backbone sequences. Transcription read-out by the internal EF1α promoter from an LTR circle effectively also becomes transcription read-in, since RNA pol-II will then continue through the "upstream" packaging signal, back through the internal backbone sequences, and again into the internal transgene cassette (see Figure S6A). To assess the level of 1-/2-LTR "read-around," we conducted a control RT-PCR using a forward primer upstream of the 3′ ppt and a reverse primer within the packaging signal upstream of the MSD; such RT-PCR products can only come from read-through 1-/2-LTRs within circular episomes. These data reported in Figure S6B indicates a substantial amount of read-around 1-LTR RNA for the 3rd-generation LV compared with the SupA2KO-LV. A lower amount of RNA derived from read-through of 2-LTR circle episomes was detected for the 3rd-generation LV and was undetectable for the SupA2KO-LV. For the 3rd-generation LV containing the MSD, similarly to integrated LV cassettes, an EF1α-derived read-around transcript from an LTR circle episome has the capacity to utilize the MSD and splice to internal splice acceptors. No such spliced product is possible for the SupA2KO-LV, since its modified SL2-loop within the packaging sequence lacks a splice donor. To assess this directly, we analyzed read-depth across all splice junctions mapping within the LV cassettes (Figure S7). These data confirm that for both LV types the primary spliced transcript was the EF1α-derived GFP mRNA, with very minor splice variants derived from this internal transcription unit. In addition, analysis confirmed the 3rd-generation LV cassette produced the MSD-to-EF1α splice acceptor product at a similar magnitude as indicated by read-depth upstream of the MSD in Figure 6C. Given that, in general, relatively few 5′ chimeric reads were detected from integrated LVs, this could be interpreted that up to half (i.e., explaining the ∼2-fold discrepancy of 5′ chimeric to MSD read-depth) of this spliced RNA (from the MSD to the EF1α splice acceptor) was derived from 1-/2-LTR circle episome read-around. However, we reanalyzed extracted DNA from the transduced hDFs to find that all samples from both LV types had 65-75% integrated copies compared with 25–35% non-integrated episomal copies (Figures S6C and S6D). Gene expression from LTR circles is generally reported to be lower than for integrated LVs.[37] In addition, it is possible that the long-read approach may have underestimated the number of 5′ read-in events in general, due to tail-off of sequence coverage toward the 5′ end. This would presumably also impact the ability to detect (and map) read-around transcripts that had circled multiple times, although the stability of such mRNAs is questionable.

SupA2KO-LVs and 3rd-generation SIN-LVs share integration site distribution in human cells

The supA-LTRs generated in this study retain the minimal integration attachment sequence (att) of HIV-1 and SIN-LVs (see Figure S3)[38] and so are expected to retain the general integration profile observed for SIN-LVs. To empirically determine whether the additional sequences engineered into the supA-LTRs impact on IS preference, we produced 3rd-generation SIN-LVs and SupA2KO-LVs encoding EFS-GFP and transduced triplicate adherent HEK293T and suspension SupT1 cell cultures at matched MOIs, before pooling extracted genomic DNA to single samples. Sample analysis was carried out by LAM-PCR/Illumina NGS (Azenta) in triplicate and IS analysis

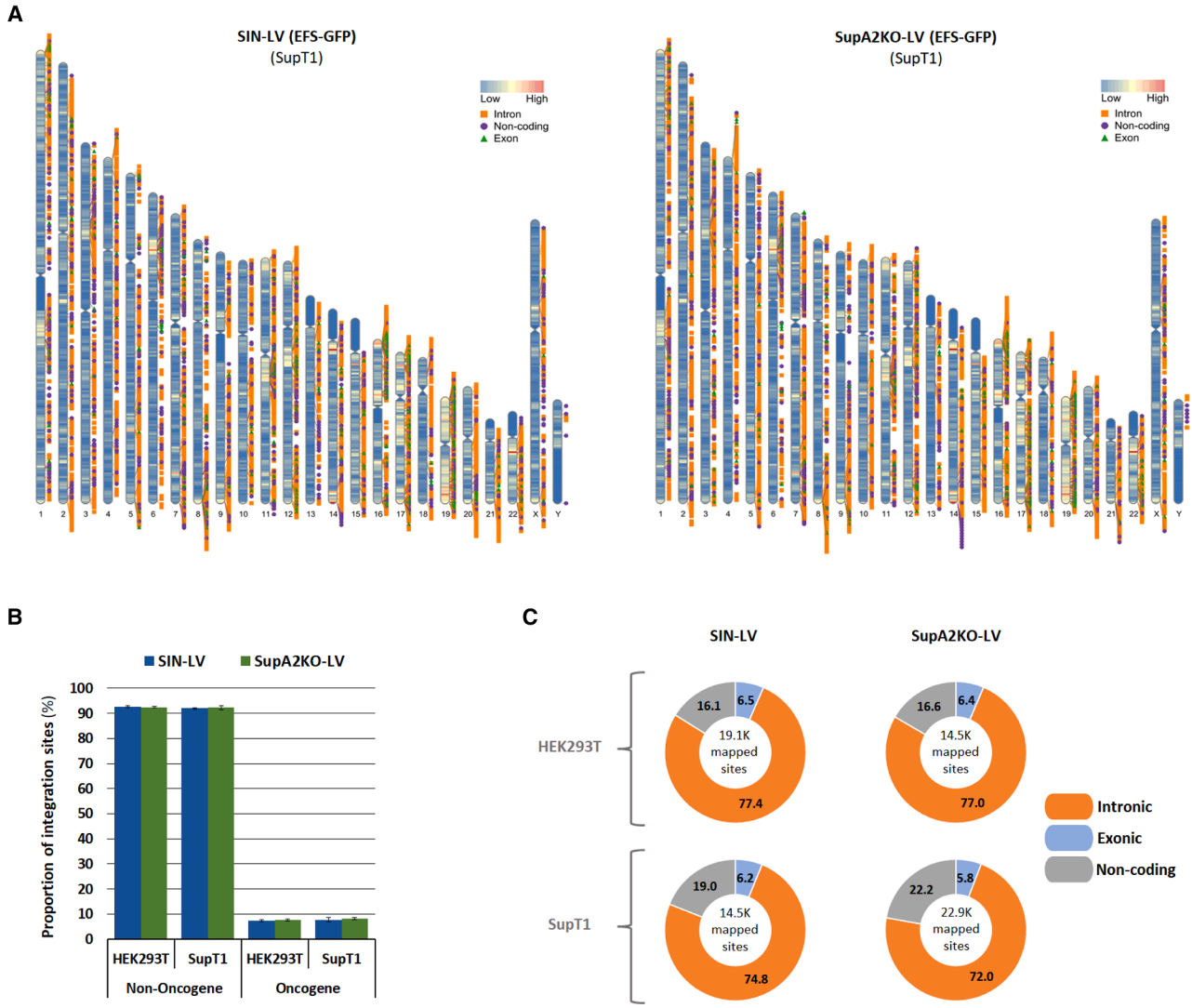

Figure 7. Integration site analysis of transduced HEK293T and SupT1 cells with LV-EFS vectors bearing SIN or SupA LTRs

(A) Example karyograms of LV integration sites for SIN-LV or SupA2KO-LV transduced SupT1 cells (produced in R/RIdeogram). (B) Proportion of total integration sites of SIN-LVs and SupA2KO-LVs within 100Mb of non-oncogenes or oncogenes in HEK293T and SupT1 cells (data mean [SD]; see methods). (C) Relative proportion of all mapped integration sites in intronic, exonic, or non-coding regions in HEK293T and SupT1 cells transduced with SIN-LV or SupA2KO-LV. Created in https://BioRender.com.

performed (in-house; common site, distribution, and cancer gene proximity). Example karyograms for these analyses are displayed in Figure 7A (SupT1) and summary IS analysis in Figures 7B and 7C. These data are very similar for both HEK293T and SupT1 cells and indicate no difference between SIN-LV and SupA2KO-LV distribution in exons, introns, non-coding regions, or proximity to known oncogenes. The IS distribution for the SIN-LV in HEK293T cells was in good agreement with others' published work.[39]

DISCUSSION

HIV-1 based LVs remain attractive gene delivery vehicles owing to the excellent safety data associated with the features of 3rd-generation LVs. The self-inactivating LTR was first developed within γ-retroviral vectors (RVs) and used for hematopoietic stem cell (HSC) therapy to minimize insertional mutagenesis.[40–42] The SIN feature removes powerful enhancer-promoter sequences that may transactivate proto-oncogenes laying close to ISs or distally via chromatin looping.[43] Further work demonstrated that use of powerful retroviral enhancer/promoters internally to drive the transgene should also be avoided for HSC therapy even within SIN-LVs, which are generally less genotoxic than SIN-RVs due to differences in IS preferences.[44–48] A more recent advance has been the identification and use of short chromatin insulators to shield the recipient genome from the transactivation potential of the transgene enhancer/promoter.[49]

A less pervasive but well characterized mechanism of insertional mutagenesis by SIN-LVs is via transcriptional interaction of LV sequences within or adjacent to cellular genes.[18,19,43,50] This has been made more acute by the fact that the native HIV-1 polyA sequences retained within LVs are weak and that the HIV-1 USE is also deleted from the SIN-LTRs. This mainly occurs via transcriptional read-in to the LV cassette and potential splicing with the MSD or other internal splice sites. Read-in to the 5′ SIN-LTR appears to continue to at least the 3′ SIN-LTR, as evidenced by the splicing from the MSD to internal splice acceptors. If transcription continues through the 3′ SIN-LTR, and there are no strong/cryptic internal splice acceptors, then potentially the MSD could splice to the next available splice acceptor (e.g., within the cellular gene), as has been reported.[18] Very low-level transcription of full-length integrated SIN-LV genomes has been reported, which was seemingly due to internal cryptic enhancer/promoter activity associated with an Sp1 site within the core packaging signal sequences.[51,52] Interestingly, the 2KO modification ("2KOm5"; a replacement of SL2) partially abolishes this Sp1 site, which might contribute to reduced LV backbone mobilization, in addition to the other mechanisms discussed here.

The SupA2KO-LV backbone theoretically provides a further block to RCL formation during production, due to reduction in potential for vRNA mobilization, owing to the improved transcriptional insulation and particularly the dependency on the 256U1 enhancer molecule. The hallmark of LV development has been the division of increasing number of vector components on to separate expression cassettes to minimize the probability of recombination leading to RCL formation. We have generated a mathematical model to attempt to quantify the risk of RCL formation from a standard 4-component system by transient transfection (Farley et al.., 2025[53]). In this model, the probability of RCL formation from a 5-component system such as SupA2KO-LV (necessitating a recombination event to place the 256U1 cassette between the LTRs) is > 1,000-fold less than that of the 4-component system.

Others have shown that the internal promoter strength can also modulate the degree of transcription read-in, although this was assessed within MSD-containing LVs.[20] In our study, a similar reduction in transcription read-in to integrated SupA2KO-LVs was observed irrespective of promoter strength. Transcription of the 5′ ψ-gag sequence leads to RNA encoding the partial gag sequence, which is why historically the open reading frame (ORF) of the partial gag is typically disrupted by frameshift mutation in contemporary LV genomes. The other reason for mutating gag is to avoid expression of partial gag protein during LV production that might otherwise lead to defective particles. As a side note, it is important to highlight that some (more recently available) 3rd-generation LV genome plasmids do not contain any mutation of the partial gag ORF, and in some, versions this ORF extends in-frame into downstream RRE and cppt sequences, encoding short env and integrase sequences, respectively. Transcriptional read-out of the integrated LV cassette occurs via transgene transcription through the 3′ SIN-LTR into adjacent intronic/exonic sequences, although the use of post-transcriptional regulatory elements (PREs) such as from Woodchuck hepatitis virus (wPRE) has been reported to reduce this.[54] Of note is that PREs (including wPRE) have more recently been shown to act by introducing mixed polyAs to enhance mRNA stability acting after polyA site choice.[55–57] The wPRE (mutated for protein X expression) was present by default in all LTR reporters and LV genomes in this study since it is typically employed in 3rd-generation LVs. The supA-LTR can impart modest but useful increases in GOI expression and reduced 3′ LTR read-out even in the presence of the wPRE.

Regulatory authorities are emphasizing the need to improve on the safety/quality aspect of viral vector gene therapy products in general.[58] The work reported here is part of our wider efforts to systematically improve LV genome sequences for the next wave of LV-based therapies, particularly in readiness for high(er)-dose, *in vivo* applications. This has culminated in the TetraVecta System, which incorporates the supA-LTR and the MSD-mutated 2KO-LV genome, as well as (optionally) the TRiP System for transgene repression during LV production.[59,60] Maximal transcriptional insulation of integrated LV cassettes is achieved when these two features are combined within "SupA2KO-LV" genomes. Using long-read NGS, RNA-seq, and RT-qPCR, we were able to show that 5′ transcriptional read-in to integrated SupA2KO LVs in different cell lines and primary cells are subject to 5- to 20-fold less 5′ transcription read-in. The absence of the MSD means that splicing from this promiscuous splice donor to transgene or cellular RNAs is not possible. Indeed, the inability to recruit an endogenous U1 RNP (via an active MSD) to the nascent incoming transcript means that U1-mediated suppression of polyadenylation at the 5′ supA-LTR cannot occur.[12,61]

Our study of transcription insulation of the two LV types in the primary hDF cells revealed some interesting findings. In attempting to augment the sensitivity of detection of potential chimeric transcripts we used high MOIs (up to 16) to increase the number of integration events per cell. While hDF cells are useful in their ability to be (adherently) cultured easily, and maintain a steady cell division rate, they do not divide as rapidly as the other cell lines used in our work. Consequently, by day 10 post transduction, these cells had slowed dramatically, and this led to sufficiently high levels (∼30%) of non-integrated LV cDNA that made the long-read analysis more complex. Hence also, the hDF analysis by RT-(q)PCR reflects transcription insulation as a "sum" of both integrated LVs and episomal LV cDNA. The proportions of non-integrated LV cDNA in target cells has been reported to be 20|9|1 for linear|1-LTR|2-LTR episomes, respectively.[62] The linear forms are primarily cytoplasmic, and so we focused on expression from the LTR circles.[63] We consider the implications of these findings to be of direct relevance for gene therapy by LVs, since the presence of episomal LV cDNA and associated transcripts has previously been raised as a concern by work from Tal Kafri's lab.[64,65] Our results confirm this previous limitation of SIN-LVs in their capability of generating packageable vRNA that is derived from the internal promoter, resulting from failure to terminate transcription at the single LTR of the episome. As a solution, the authors employed an inverted internal cassette as a safety

feature, such that any potential (rare) production of packaging-signal encoding vRNA in the transduced cell would base-pair with the internal transgene transcript, leading to formation of dsRNA and PKR-mediated cell shutdown. Unfortunately, active inverted transgene LVs also require obviation of the PKR response during LV production, and it is our experience that LV titers tend to be lower in cells where PKR has been attempted to be knocked down/out (J.W. and D.C.F., unpublished data). Our analyses in transduced hDFs provides confirmation that read-around transcription from 1-/2-LTR circle episomes not only occurs in a relevant setting but can also result in utilization of the MSD, leading to splicing and stabilization of aberrant transcripts. Our comparison of LV transduction of hDFs using an intronless transgene cassette indicates that the presence of a strong splice acceptor to which the MSD can partner is a key driver for stabilization of mRNA derived from both read-through (integrated LVs) and read-around (LTR circle episomes) transcription. Implementation of the SupA-LTR and the 2KO modification provide substantial transcriptional insulation in both settings. This has relevance to mediation of gene delivery into terminally differentiated or slowly growing primary cells (e.g., brain, eye) where a level of episomal LV cDNA will be generated post transduction and may remain long-term.

All transcript reads mapping across the 5′ LTR and into the integrated LV backbone must contain a polyA tail derived from the 3′ LTR, but the lack of cellular sequence detection may have been due to diminishment in read efficiency toward the 5′ end of the LV cassettes. The comparison of SIN-LV and SupA2KO-LV (indicating an ∼6-fold reduction for the latter) should be considered preliminary, given that the 5′ chimeric reads detected in transduced hDFs were ∼50-fold lower than for 3′ chimeric reads. We were unable to detect 5′ chimeric reads terminating at the 5′ polyA sites, and this may indicate that such transcripts are generally very unstable and therefore not represented in the analysis. Conceptionally, it seems reasonable to assume that implementation of stronger 5′ transcriptional terminal might lead to greater levels of premature cellular mRNA transcripts, but this is not supported by this preliminary analysis. Even if SupA2KO-LVs yield greater numbers of prematurely terminated cell gene mRNAs in other cell types, we think this is preferable over the read-through and stabilizing effects of splicing from the MSD mediated by integrated 3rd-generation SIN-LVs. The resultant stable chimeric transcripts associated with 3rd-generation LVs could also be potential gain-of-function ones (i.e., transgene-related), in addition to being functionally ablated in the cell gene into which integration occurred.

By definition, it is impossible to avoid insertional impact of cellular genes when using an integration-proficient LV. The success of HIV-1 based LV gene therapy is built on the fact that there are two copies of each cell gene and there is typically a great deal of gene redundancy in cellular gene pathways/circuitry. If this was not the case, LV-based gene therapy would simply not be operable. Empirically, the early concerns that integrating viral vectors might only need to disrupt 1–2 genes to push a significant number of somatic cells into an oncogenic state have not been realized. It is worth noting that HIV-1 containing its native LTR—and therefore a more potent polyadenylation sequence compared with SIN-LVs—has not been considered primarily genotoxic based on capacity to induce premature cell transcript termination. Rather, the active U3 promoter (with potential upregulation by tat) and the MSD are the primary drivers in these very rare cases, where transcription read-out and upregulation of cell gene expression is observed (see Rist et al. 2025[66] and references therein). Indeed, just the HIV-1 LTR-MSD sequence (delivered by CRISPR-Cas9) has been used in isolation to mediate similar effects of HIV-1 upregulation of BACH2.[67] Notably, the U3 promoter is absent from 3rd-generation LVs, and now SupA2KO-LVs add further safety due to the absence of the MSD and improved polyadenylation sequences.

The core innovation of the supA-LTR is the decoupling of transcription termination from the native HIV-1 pAS within the LV genome and introduction of more active heterologous counterparts with optimized positioning. The primary defining feature of the polyA sequence—the pAS hexamer—is effectively moved into the ΔU3-SIN region upstream of R. In contrast to lentiviruses, a similar configuration is used by other retroviruses, including RSV (alpharetrovirus), MMTV (betaretrovirus), and HTLV-1 (deltaretrovirus), where the pAS is present between the promoter TATA box and the transcription start site. Here, this was achieved in LVs by creating the "R-embedded" heterologous (SV40 late) polyA sequence at the 3′ end of the vRNA expression cassette for production. In principle, other strong heterologous polyA sequences such as from the β-globin gene could also be used. We have empirically shown that insertion of just the first 14 nucleotides of R downstream of the heterologous pAS results in sufficient R sequence between the pAS and the RNA cleavage site to allow first strand transfer, as evidenced by the maintenance in LV titers. Since the polyadenylation activities of variants R 1–20 and R 1–14 were similar, we moved forward with the R 1–20 variant as a conservative measure. Follow-up analysis of integrated LV shows that for 90% of transgene mRNAs the polyA tail is added to the second of the three CA dinucleotides, with the first (i.e., R 19–20) and third CA dinucleotides used equally as sole alternatives (data not shown). To date, the shortest length of homology shown to allow efficient first strand transfer has been demonstrated in murine leukemia virus, being 12 nucleotides of R region at both 5′ and 3′ ends of the vRNA.[68] For HIV-1 replication, a study was performed assessing the wild-type 3′ R region length (97 nt) and truncated versions (37 and 15 nt), with both truncated versions resulting in progressively greater attenuated virus growth kinetics of 50% and 5%, respectively, compared with wild-type virus.[69] Previously, others reduced the size of the 3′ R region to 47 nucleotides within an LV vRNA expression cassette by employing heterologous polyA sequences (DSEs) immediately downstream, without negative impact on titers.[22] Importantly, the authors also recognized that, due to the nature of the HIV-1 LTR copying process, these heterologous polyA sequences would not be retained within the final integrated vector. To address this, they attempted to insert the β-globin-derived DSE downstream of the native HIV-1 5′ PAS within the 5′ R-U5

sequence such that 19 nucleotides of the 5′ R sequence were deleted. However, this reduced production titers, and transcriptional read-through was actually made worse. These findings, together with our observations that residual HIV-1 polyA sequences could impact on activity of heterologous polyA sequences, justified our approach to optimally space heterologous sequences such that they dominate the residual native counterparts. Since the *att* sequence of the 5′ U3 sequence is retained, we did not expect differences in IS preference for SupA2KO-LVs compared with 3rd-generation SIN-LVs, and this was empirically verified.

Overall, our results support the general model of efficient polyadenylation at a given pAS positionally defined by DSE strength and the speed/strength of processing modulated by factors recruited at the USE.[70] Here we used the strong and well-defined SV40 late polyA USE sequence in its native context as part of the R-embedded heterologous polyA, allowing the removal of the back-up polyA. Schambach and colleagues improved polyadenylation of SIN-LVs by inserting two copies of the core USE (44 nts) from SV40 late polyA into the ΔU3-SIN region, although these were designed within an SL to perhaps make them more accessible to the auxiliary polyA protein factors (personal communication).[31] The USE from the β-globin polyA has also been implemented in a similar way.[71] In these cases, polyadenylation is still dependent on the native HIV-1 pAS and DSE some ∼100 nucleotides downstream of the USE sequences. We were able to develop the supA-LTR containing an optimal USE-pAS-DSE configuration through a series of rationally designed 5′ SLs containing GU-rich sequences based on the strong MMTV polyA DSE, while also maintaining production titers. The modified 5′ SL retains the first 20 nucleotides of R required for transfer of the 1st strand cDNA to the complementary R 1–20 sequence that is immediately upstream of the polyA tail present after processing at the R-embedded heterologous polyA sequence. The R-GU2 variant contains additional G/U sequences that enable it to function as an efficient DSE for the new pAS, while also maintaining a stable 5′ SL that does not appear to impact on correct folding of the packaging signal. The final supA-LTR variant (supA-2pA-iGU7) can function independently of the native HIV-1 pAS and DSE sequences. Interestingly, the production of high-titer supA-LTR LVs with a mutated HIV-1 5′ pA seems to argue against a role for the native pAS in LV vRNA packaging, as reported for HIV-1 genomic RNA by others.[72] The supA-2pA-LTR harbors pASs on the antisense strand to mitigate against cellular read-in from outside the 3′ end of the integrated LV, as well as provide a back-up polyA for an inverted transgene cassette if utilized. A similar approach to antisense strand polyA site duplication within integrated LTRs has been reported previously as an attempt to limit GOI expression during production.[73] For simplicity, we now refer to "SupA-LTR" as being the supA-2pA(iGU7) variant.

Here we utilized a luciferase-based plasmid reporter assay to test/optimize polyA sequences, where 70- to 100-fold increases in termination efficiencies for sense and antisense strand evaluations against the standard SIN-LTR were observed, respectively. However, it should be noted that this assay may be limited in modeling relevant levels of transcription read-through due to the high transfection efficiency of HEK293T cells, leading to non-physiological numbers of active transcription units within the cell.[74] This approach likely overestimates the problem of transcription read-through weaker polyA sequences due to potential squelching of protein factors and transcriptional read-around circular plasmid templates. Nevertheless, it allows for a stringent assessment of polyA sequences for further testing in transduced cells, as we have demonstrated here.

In summary, SupA2KO-LVs represent a step-change in LV-derived RNA biogenesis in regards to: [1] the predictable and efficient production of full-length vRNA during LV production (owning to MSD mutation), [2] reduced likelihood of RCL formation (due to MSD mutation and dependency on the 256U1 enhancer), [3] the transcriptional insulation of the transgene cassette in transduced cells (reduced read-through and mRNA stabilization by MSD mutation), and [4] increased transgene expression resulting from improved polyadenylation sequences.

MATERIALS AND METHODS

Plasmids

HIV-1 LV genomes and packaging plasmids were constructed as described previously.[75–77] Mutations to the 5′ and 3′ LTRs were made by gene synthesis and subsequent cloning using NdeI, PspXI or NheI, KpnI restriction sites, respectively. Different promoter/transgene cassettes were synthesized and inserted into LV vector genome plasmids using MluI and NheI restriction sites. The 256U1 expression cassette was based on the endogenous human U1 gene (GenBank: J00318.1; nucleotides 41–724). Nucleotides 3–11 of the first 11nt comprising the native splice donor complementary sequence was replaced with sequence complementary to the HIV-1 nucleotides 256–271. The resulting 256U1 cassette, including endogenous promoter and termination signals, were then inserted into a re-derived KanR backbone that was termed pRKH-256U1. The 3′ LTR polyadenylation reporter plasmid was first made through deletion of vector sequences from upstream of the external promoter through to upstream of the transgene cassette of an LV-EF1a-tbs-GFP genome plasmid to generate the pEF1a-tbs-GFP-wPRE-SIN intermediate plasmid. SIN-pA-IRES-*Gaussia* Luciferase sequences were then synthesized and cloned into the 3′UTR of the intermediate plasmid using SalI and AatII restriction sites to generate the pEF1a-tbs-GFP-wPRE-SINLTR-IRES-Gluc reporter plasmid. All supA-LTR variants were synthesized as fragments then cloned into this reporter plasmid using BamHI and AvrII restriction sites.

Cell lines

Vector production was performed in adherent HEK293T cells (ATCC; LGC Standards, UK) or in-house suspension-adapted "HEK293T-1.65s" cells, which were derived from an adherent HEK293T cell GMP bank. Adherent HEK293T cell lines have been assessed by PCR single-locus technology (Eurofins Medigenomix Forensik GmbH) and verified against the DSMZ database. Adherent HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (Sigma #D5671) supplemented with 2mM L-glutamine (Sigma

#G7513), 1% (v/v) non-essential amino acids (Sigma #M7145), and 10% fetal bovine serum (Gibco #10101145). SupT1 cells were maintained in RPMI Medium (Gibco #11875093) supplemented with 10% (v/v) 10% fetal bovine serum. HepG2 cells (human hepatocytes) were maintained in minimal essential medium (Gibco #32360026) supplemented with 2mM L-glutamine, 1% (v/v) non-essential amino acids, and 10% fetal bovine serum. 92BR cells (donkey testis fibroblasts) were maintained in Dulbecco's Modified Eagle Medium supplemented with 2mM L-glutamine, 1% (v/v) non-essential amino acids, and 10% fetal bovine serum. Human dermal fibroblasts (Themo Fisher Scientific, #C0135C) were maintained in human fibroblast expansion basal medium (Gibco #M106500) supplemented with 2% low serum growth supplement (Gibco #S00310). Adherent cells were kept at 37°C in 5% CO_2 in a static incubator. Suspension HEK293T-1.65s cells were maintained in Freestyle-293 serum-free medium (Invitrogen #12338-026) supplemented with 0.1% cholesterol (Invitrogen # 12531018) and kept at 37°C in 5% CO_2 while shaking at 200 r.p.m.

Peripheral blood mononuclear cell culture

Peripheral blood mononuclear cells (PBMCs) from two donors were sourced from Cellular Technology Ltd. PBMCs were cultured for 48 h pre-transduction with CD3/CD28 T cell Expander Dynabeads (Thermo Fisher Scientific # 11141D) in X-Vivo complete medium (Lonza #BEBP02-061Q) supplemented with 10% human AB serum (Sigma) and 10 ng/mL IL-7 (Miltenyi Biotec #130-095-363) and IL-15 (Miltenyi Biotec #130-095-765) at 2×10^6 cells/beads per mL. Concentrated LVs were then added to the cells at a matched MOI. Following transduction, cells were maintained for 17 days, with cells maintained between 6.0 x 10^5 and 2 x 10^6 viable cells by splitting or media/cytokine supplementation every 48–72 h.

At 10 and 17 days post transduction, cells were fixed and stained for CD19-CAR expression by flow cytometry using a PE-labelled Human CD19 antibody (ACROBioSystems #CD9-HP2H3-200). CD3-positive T cells were selected using an FITC anti-human CD3 antibody (BioLegend #317306). Cell viability was assessed using a Zombie NiR viability antibody (BioLegend #423106). In addition to staining, cell pellets were maintained on days 10 and 17 for assessment of integrated VCN by qPCR (detailed below).

Lentiviral vector production

The standard scale production of LVs in adherent HEK293T cells was in 10cm dishes under the following conditions (all conditions were scaled by area when performed in other formats): HEK293T cells were seeded at 3.5x10^6 cells/plate in 10mL complete HEK293T medium (DMEM containing 10% heat-inactivated fetal bovine serum, 2mM L-glutamine, 1x non-essential amino acids) and incubated at 37°C in 5% CO2 throughout production. Approximately 24 h later, cells were transfected using the following mass ratios of plasmids: 4.5μg LV genome, 1.4μg pGagPol, 1.1μg pRev, 0.7μg pVSVG. Transfection was mediated by mixing DNA with Lipofectamine 2000CD (Thermo Fisher Scientific #12566014) in Opti-MEM (Invitrogen #11058-021) according to manufacturer's protocol. About 16-20h post transfection, sodium butyrate (Merck #1371270250) was added to the culture to a final concentration of 10mM. Six hours later, the medium was removed from the plate and replaced with fresh, complete DMEM. A single harvest was performed 24h later where supernatants were filtered (0.22μm) and frozen at −80°C.

The production of LVs in small-scale suspension HEK293T-1.65s cells was performed using the following mass ratios of plasmid per 1.6x10^6 cells in 1mL of Freestyle-293 media: 0.95μg LV genome, 0.1μg pGagPol, 0.07μg pVSVG, and 0.06μg pRev. Where appropriate, 0.14μg of p256U1 plasmid was included. Transfection was mediated by mixing cDNA with Lipofectamine 2000CD in Freestyle media according to the manufacturer's protocol. About 16-20h later, sodium butyrate was added to a final concentration of 10mM, then cultures were returned to the incubator. A single harvest was performed 24h later where supernatants were filtered (0.22μm) before being stored at −80°C.

For larger-scale production of LVs, suspension HEK293T-1.65s cells were transfected at 40 mL volume in E125 shake flasks using the same plasmid ratios and transfection reagent as detailed above. Cells were supplemented with 10 mM sodium butyrate 16–20 h post transfection. Twenty-three hours later, cells were supplemented with 4 mM $MgCl_2$ and 7 U/mL Benzonase (Merck #1016950001) and further cultured for 1 h at 37°C, 5% CO_2, 200 r.p.m. Following nuclease treatment, a single vector harvest was performed as previously detailed. The filtered vector, 38.5 mL, was concentrated by ultracentrifugation at 20,000 r.p.m for 1.5 h. The pelleted vector was resuspended in 0.3mL TSSM and aliquoted, and integration titers were determined by integration assay (see below).

Vector titration by flow cytometry

Ninety-six well plates were seeded with 1.2x10^4 HEK293T cells/well in 100μL DMEM. Twenty-four hours later, the medium was removed and replaced with 50μL/well of serially diluted vector in the presence of 8μg/mL polybrene. Three hours post transduction, an additional 50μL of fresh DMEM was added per well. Three days post transduction, the medium was removed and cells were treated with Tryp-LE (Sigma) for 5 min to disrupt the monolayer. Cells were then resuspended in DMEM containing a working concentration of 1/1000 of SYTOX AADvanced Dead Cell Stain (Thermo Fisher Scientific, #S10349). Cells were then assayed by flow cytometry using a 488nm laser on an Attune NxT flow cytometer (Thermo Fisher Scientific). GFP expression was scored in 10,000 live cells, as determined by SYTOX staining. Data were analyzed with associated Attune NxT and/or FlowJo software. The percentage of GFP-positive cells was then used to calculate vector titer of the starting material.

For CAR staining, transduced cells were trypsinized and transferred to a deep-well 96-well plate. PBS, 1 mL, was added per well then cells were centrifuged at 1000 x g for 5 min. Cell pellets were then incubated with 100μL fluorescent CAR-specific antibody (Stratech #115-584-072) diluted in 5% goat serum FACS stain buffer (BD Biosciences #554656) for 20 min at room temperature. Upon incubation cells were centrifuged at 1,000 x g for 5 min and the supernatant was

discarded. Cell pellets were then resuspended in 33% Cytofix buffer in PBS (BD Biosciences #554655). Fluorescence was assayed on an Attune NxT Flow Cytometer and data were analyzed as described above.

Transduction of cells for transcriptional read-in and gene expression studies

Cells were seeded in 24-well plates (0.5mL/well) in appropriate media 24h prior to or on the day of transduction SupT1 [on the day] and HepG2 at $1x10^5$ cell/well; HDF and HEK293T cells at $4.5x10^4$ cell/well. Cell counts were performed on the day of transduction and vector diluted in complete DMEM containing 8μg/mL polybrene to achieve the desired MOI. The medium was replaced with 0.25mL of diluted vector. Three hours later an additional 0.25mL of complete medium was added per well. In all instances transduced cells were passaged three times over 10 days before being harvested. Transgene expression was assessed on day 10 by flow cytometry. The remaining cells were harvested and split into two for genomic DNA extractions (to assess VCN) and RNA extraction (for transcriptional read-in assessment).

Vector transduction and titration by qPCR

For LV titration by qPCR, transductions of adherent HEK293T cells were performed at the 12 well plate scale, with cell densities and volumes scaled up accordingly from the 96 well plate transduction method outlined above. Transduced cells were passaged three times over 10 days before being harvested and genomic DNA was extracted using an automated QIAcube DNA extraction system (Qiagen). Genomic DNA was then subjected to TaqMan qPCR using primers and probes directed to the extended packaging sequence of HIV-1 (and the cellular RPPH1 gene) with TaqMan 2X Universal PCR Master Mix (Thermo Fisher Scientific, #4304437; see Figure S1). qPCR assays were performed on an Applied Biosystems QuantStudio 7 thermocycler (Thermo Fisher Scientific). The average number of integrated copies per cell was used to back-calculate the integrating titers of the original vector material.

To assess the levels of integrated and episomal DNA in transduced cells, QIAcube-extracted DNA was subjected to size exclusion using a PippinHT (Sage Science HTP0001) 6-10kb 0.75% Agarose cassette (Sage Science HPE7510) according to manufacturer instructions. Briefly, 20 μL extracted DNA was mixed with 5 μL loading solution, added directly to the 0.75% Agarose cassette and run alongside the provided 75E Marker (Sage Science HPE7504). Eluted genomic DNA of >10kb was subsequently collected and subjected to qPCR as detailed above. All size excluded samples were run alongside the QIAcube-extracted DNA to ensure comparability. Episomal DNA was calculated as Total psi copies per cell (QIAcube-extracted DNA) minus PippinHT size excluded DNA.

Gaussia Luciferase assay

Adherent HEK293Ts cells were transfected in triplicate with pEF1a-tbs-GFP-wPRE-[LTR]-IRES-Gluc or pLV-pEF1a-tbs-GFP-wPRE-[LTR]-IRES-Gluc reporter plasmids at the 96 well plate scale using lipofectamine. pTK-dsRed Express reporter plasmid was included as a control for transfection efficiency. Forty-eight hours post transfection, media were removed from wells and transferred to a fresh plate. Remaining cell monolayers were assayed for GFP and dsRed expression using an Attune NxT flow cytometer as described above. Media were assayed for luciferase activity using the Gaussia Luciferase Flash Assay Kit (Thermo Fisher Scientific, #16159) according to the manufacturer's instructions. Luminescence was measured using a SpectraMax i3e plate reader (Molecular Devices). For each well, Gaussia Luciferase activity was normalized to the corresponding GFP reporter expression to calculate the relative Gaussia Luciferase activity. All wells were normalized to dsRed expression to account for differences in transfection efficiency. Transfections of suspension HEK293T-1.65s cells were performed at the 24 well plate scale using the same timings as listed above. At harvest, an aliquot of cell suspension was assayed for GFP and dsRed expression as above, and the remaining cell suspension centrifuged for 5 min at 300 x g to pellet the cells. The supernatant was transferred to a fresh tube and assayed for *Gaussia* luciferase activity as described above.

RNA extraction and qRT-PCR

Cells were washed once with PBS then total RNA was extracted using the RNeasy extraction kit (Qiagen #74104) according to the manufacturer's instructions. RNA, 250 ng, was DNAse treated (ezDnase, Thermo Fisher Scientific #11766051) according to the manufacturer's instructions. For the qRT-PCR analyses, 250ng total cell RNA was DNAse treated as above and subjected to reverse transcription using the SSIV VILO system (Thermo Fisher Scientific #11766050). cDNA was then subjected to TaqMan qPCR (Thermo Fisher Scientific #4304437) using primers to detect transcripts of interest. Endpoint PCR was performed on cDNA using CloneAmp polymerase (Takara #639298) with the primers stated. See Table S1 for primer/probe sequences.

Read-in RT-qPCR assay

Cells were harvested 10 days post transduction. Adherent cell monolayers were released by treatment with trypsin and resuspension in an appropriate medium before being split into two aliquots. Transduced suspension cell cultures were divided into two cultures and cells pelleted by centrifugation for 5 min at 300 x g. For a given transfection both genomic DNA and total RNA were extracted from the two aliquots according to the methods stated above. Genomic DNA was subjected to TaqMan qPCR to determine vector integrant copies, and total RNA was subjected to TaqMan RT-qPCR using primers targeting sequences within the proviral Psi sequence and cellular GAPDH (see Table S1). To generate a transcriptional "read-in" score Psi RNA copies were first normalized to GAPDH RNA copies to adjust to RNA input. Adjusted Psi RNA copies were then normalized to the average number of integrated vector DNA copies per cell (VCN) to give a normalized read-in value.

High-throughput short-read sequencing of integrated LV RNA (RNA-seq)

For analysis of read-in transcripts cellular RNA was extracted from HEK293T cells transduced with a panel of LV-GFP vectors as per the methods outlined above. RNA sequencing was performed by Azenta Life Sciences. Briefly, polyA-selected RNA was enriched prior

to reverse transcription. cDNA was then subjected to 2x150bp Illumina sequencing using the NovaSeq 6000 platform to a read-depth of 50M read pairs. Sequence data were then aligned to reference vector genomes using STAR-Arriba. A custom Bash script utilizing SAMtools[78] was then used to calculate read-depth in the integrated LV genomes over Psi (including MSD), truncated *gag*, and RRE sequences. Read-depth at these coordinates was normalized to the library size (as determined by number of reads in proper pairs mapping to the integrated LV genome) and integrated VCN, as independently assayed by qPCR (see above).

High-throughput long-read sequencing of integrated LV RNA

For analysis of read-in transcripts via Nanopore long read sequencing, RNA was extracted from Human Dermal Fibroblasts 10 days after transduction with a panel of LV-EF1α-GFP vectors as per methods outlined above. Long read sequencing was performed on 500 ng of total RNA using a cDNA-PCR Sequencing V14-Barcoding Kit (Nanopore SQK-PCB114.24) according to manufacturer instructions. Briefly, full-length RNA was initially reverse transcribed following addition of a poly-T RT adapter. Full-length cDNA was then PCR amplified and barcoded prior to being pooled and sequenced using a MinION MK1b (flow cell R10.4.1). Samples were sequenced for 72 h to allow sufficient number of reads to be captured for each sample. Sequencing data were aligned to a combined reference genome containing the human genome (hg38) and the corresponding LV genome using minimap2, with parameters -ax splice and - -secondary=no. SAMtools depth with the parameters -d 0 and -a was then used to calculate LV and GAPDH sequence read-depths. To isolate the chimeric reads LV-human genome, first the sequencing data were aligned to the corresponding LV genome using minimap2, with parameters -ax splice and - -secondary=no. Reads mapping to the 5′ LTRs and 3′ LTRs were extracted into separate files and subsequently aligned to the hg38 genome to quantify the number of chimeric reads. To quantify splicing in the RNA-seq datasets, junction reads were extracted using the RegTools junction extract tool on the BAM file containing only positive strand alignments that mapped to the LV genome.[79] The splice junctions were then annotated using the RegTools junctions annotate command, using FASTA and GTF files containing the Lenti genome sequence and transcript information, respectively.

LV integration site analysis

The insertion site analysis service from Azenta Life Sciences was used to assess the location and frequency of the lentivector integration. The results from the Azenta analysis were validated using an in-house pipeline. Briefly, to prepare the raw sequencing data for analysis, FastQC[80] was used for initial quality control checks and Cutadapt[81] was used to remove Illumina sequencing adapters. To align the sequencing reads, a combined reference genome was generated; this genome combined the human reference genome (hg38) with the lentivector genome, effectively treating the lentivector genome as an additional "chromosome." Reads were aligned to the combined genome using BWA-MEM[82] with the flag -Y. SAMtools[78] was then used to exclude the unmapped reads (-F4) and those with low alignment quality score (-q20). The remaining mapped reads were filtered again with SAMtools to select the reads that mapped any of the hg38 chromosomes and/or the LTRs within the lentivector chromosome in the combined genome. A customized python script was then used to select only the reads that mapped both the hg38 and the LTRs. Coverage files were generated using BEDtools[83] via the genomecov command specifying the -bg parameter. The ISs were annotated using the BED Annotation tool (GitHub: https://github.com/vladsavelyev/bed_annotation). These datasets were used to generate ideogram figures in R using the RIdeogram package.[84]

Statistical analysis

For comparison of means, an F-test for variance was performed to determine whether an equal or unequal, two-tailed t test was performed. For Figure S1, an ANOVA single factor assessment of different means was performed. All alpha values were set to 0.05. Specific data were compared within stated figures to highlight significant differences between key constructs.

DATA AND CODE AVAILABILITY

Data can be made available on request.

ACKNOWLEDGMENTS

This work was fully funded by OXB. We thank Rebecca Astley for her support in nanopore analysis.

AUTHOR CONTRIBUTIONS

J.W. and B.M.A. conducted the majority of experimental work and should be considered joint first authors. M.L.M.Q. generated an internal data pipeline analysis tool and generated the integration site analysis. A.G. contributed to experimental work. D.C. generated an internal data pipeline analysis tool and generated the RNA-seq analysis. J.W., N.G.C., and K.A.M. provided input into writing of the manuscript. J.W. and D.C.F. conceived of the concepts of work. D.C.F. wrote the bulk of the manuscript.

DECLARATION OF INTERESTS

J.W., B.M.A., M.L.M.Q., N.G.C., K.A.M., and D.C.F. are current employees of OXB and hold OXB stock and/or stock options. D.C. is a previous OXB employee and still holds OXB stock options. J.W. and D.C.F. are inventors on patent applications related to technology associated with the work reported: WO2021160993A1, WO2023062365A2, WO2021229242A1.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omta.2025.201654.

REFERENCES

1. Elisa, V. (2000). Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. J. Gene Med. *2*, 308–316.
2. Valentina, P. (2021). Designing Lentiviral Vectors for Gene Therapy of Genetic Diseases. Viruses *13*, 1526.
3. Clark, E., Nava, B., and Caputi, M. (2017). Tat is a multifunctional viral protein that modulates cellular gene expression and functions. Oncotarget *8*, 27569–27581.
4. Hiroyuki, M. (1998). Development of a Self-Inactivating Lentivirus Vector. J. Virol. *72*, 8150–8157.
5. Cornetta, K., Yao, J., Jasti, A., Koop, S., Douglas, M., Hsu, D., Couture, L.A., Hawkins, T., and Duffy, L. (2011). Replication-competent Lentivirus Analysis of Clinical Grade Vector Products. Mol. Ther. *19*, 557–566.

6. Farley, D.C., McCloskey, L., Thorne, B.A., Tareen, S.U., Nicolai, C.J., Campbell, D.J., Bannister, R., Stewart, H.J., Pearson, L.J., and Moyer, B.J. (2015). Development of a replication-competent lentivirus assay for dendritic cell-targeting lentiviral vectors. Mol Ther Methods Clin Dev *2*, 15017.
7. DeZazzo, J.D., Kilpatrick, J.E., and Imperiale, M.J. (1991). Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3′ end formation. Mol. Cell Biol. *11*, 1624–1630.
8. Yang, Q., Lucas, A., Son, S., and Chang, L.-J. (2007). Overlapping enhancer/promoter and transcriptional termination signals in the lentiviral long terminal repeat. Retrovirology *4*, 4.
9. Klasens, B.I., Thiesen, M., Virtanen, A., and Berkhout, B. (1999). The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. Nucleic Acids Res. Suppl. *27*, 446–454.
10. Das, A.T., Klaver, B., and Berkhout, B. (1999). A Hairpin Structure in the R Region of the Human Immunodeficiency Virus Type 1 RNA Genome Is Instrumental in Polyadenylation Site Selection. J. Virol. *73*, 81–91.
11. Zhang, Z., Klatt, A., Henderson, A.J., and Gilmour, D.S. (2007). Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. Genes Dev. *21*, 1609–1614.
12. Ashe, M.P., Griffin, P., James, W., and Proudfoot, N.J. (1995). Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. Genes Dev. *9*, 3008–3025.
13. Ashe, M.P., Pearson, L.H., and Proudfoot, N.J. (1997). The HIV-1 5′ LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J. *16*, 5752–5763.
14. Ashe, M.P., Furger, A., and Proudfoot, N.J. (2000). Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA *6*, 170–177.
15. Jens, B. (2004). Mutation of the major 5′ splice site renders a CMV-driven HIV-1 proviral clone Tat-dependent: connections between transcription and splicing. FEBS Lett. *563*, 113–118.
16. Anne-Kathrin, Z. (2002). RNA 3′ Readthrough of Oncoretrovirus and Lentivirus: Implications for Vector Safety and Efficacy. J. Virol. *76*, 7209–7219.
17. Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of Retroviral Integration In Hematopoietic Cells. Mol. Ther. *13*, 1031–1049.
18. Daniela, C. (2012). Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. J. Clin. Investig. *122*, 1667–1676.
19. Almarza, D., Bussadori, G., Navarro, M., Mavilio, F., Larcher, F., and Murillas, R. (2011). Risk assessment in skin gene therapy: viral–cellular fusion transcripts generated by proviral transcriptional read-through in keratinocytes transduced with self-inactivating lentiviral vectors. Gene Ther. *18*, 674–681.
20. Scholz, S.J., Fronza, R., Bartholomä, C.C., Cesana, D., Montini, E., von Kalle, C., Gil-Farina, I., and Schmidt, M. (2017). Lentiviral Vector Promoter is Decisive for Aberrant Transcript Formation. Hum. Gene Ther. *28*, 875–885.
21. Iwakuma, T., Cui, Y., and Chang, L.-J. (1999). Self-Inactivating Lentiviral Vectors with U3 and U5 Modifications. Virology *261*, 120–132.
22. Koldej, R.M., and Anson, D.S. (2009). Refinement of lentiviral vector for improved RNA processing and reduced rates of self inactivation repair. BMC Biotechnol. *9*, 86.
23. Siarhei, K. (2016). Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. Proc. Natl. Acad. Sci. USA *113*, 13378–13383.
24. Brown, J.D., Kharytonchyk, S., Chaudry, I., Iyer, A.S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S.H., Singh, K., et al. (2020). Structural basis for transcriptional start site control of HIV-1 RNA fate. Science *368*, 413–417.
25. Nikolaitchik, O.A., Liu, S., Kitzrow, J.P., Liu, Y., Rawson, J.M.O., Shakya, S., Cheng, Z., Pathak, V.K., Hu, W.-S., and Musier-Forsyth, K. (2021). Selective packaging of HIV-1 RNA genome is guided by the stability of 5′ untranslated region polyA stem. Proc. Natl. Acad. Sci. USA *118*, e2114494118.
26. Nikolaitchik, O.A., Islam, S., Kitzrow, J.P., Duchon, A., Cheng, Z., Liu, Y., Rawson, J.M.O., Shao, W., Nikolaitchik, M., and Kearney, M.F. (2023). HIV-1 usurps transcription start site heterogeneity of host RNA polymerase II to maximize replication fitness. Proc. Natl. Acad. Sci. USA *120*, e2305103120.
27. Das, A.T., Harwig, A., Vrolijk, M.M., and Berkhout, B. (2007). The TAR Hairpin of Human Immunodeficiency Virus Type 1 Can Be Deleted When Not Required for Tat-Mediated Activation of Transcription. J. Virol. *81*, 7742–7748.
28. Trubetskoy, A.M., Okenquist, S.A., and Lenz, J. (1999). R Region Sequences in the Long Terminal Repeat of a Murine Retrovirus Specifically Increase Expression of Unspliced RNAs. J. Virol. *73*, 3477–3483.
29. Das, A.T., Harwig, A., and Berkhout, B. (2011). The HIV-1 Tat Protein Has a Versatile Role in Activating Viral Transcription. J. Virol. *85*, 9506–9516.
30. Levitt, N., Briggs, D., Gil, A., and Proudfoot, N.J. (1989). Definition of an efficient synthetic poly(A) site. Genes Dev. *3*, 1019–1025.
31. Schambach, A., Galla, M., Maetzig, T., Loew, R., and Baum, C. (2007). Improving Transcriptional Termination of Self-inactivating Gamma-retroviral and Lentiviral Vectors. Mol. Ther. *15*, 1167–1173.
32. Wright, J., Alberts, B.M., Hood, A.J.M., Nogueira, C., Miskolczi, Z., Vieira, C.R., Chipchase, D., Lamont, C.M., Goodyear, O., Moyce, L.J., et al. (2025). Improved production and quality of lentiviral vectors by major-splice-donor mutation and co-expression of a novel U1 snRNA-based enhancer. Heliyon *11*, e43732.
33. Moore-Kelly, C., Reddem, R., Alberts, B.M., Wright, J., Evans, T., Kulkarni, A., Clarkson, N.G., Farley, D.C., Mitrophanous, K.A., and Saraiva Raposo, R.A. (2025). Enhancing titers of therapeutic lentiviral vectors using PKC agonists. Mol. Ther. Methods Clin. Dev. *33*, 101484.
34. Johnson, N.M., Alvarado, A.F., Moffatt, T.N., Edavettal, J.M., Swaminathan, T.A., and Braun, S.E. (2021). HIV-based lentiviral vectors: Origin and sequence differences. Mol. Ther. Methods Clin. Dev. *21*, 451–465.
35. Ellis, S., Fong-Wong, L., Iqball, S., Thoree, V., Mitrophanous, K.A., and Binley, K. (2012). Assessment of Integration-defective HIV-1 and EIAV Vectors In Vitro and In Vivo. Mol. Ther. Nucleic Acids *1*, e60.
36. Ishihara, A., Zekas, L.J., Weisbrode, S.E., and Bertone, A.L. (2010). Comparative efficacy of dermal fibroblast-mediated and direct adenoviral bone morphogenetic protein-2 gene therapy for bone regeneration in an equine rib model. Gene Ther. *17*, 733–744.
37. Apolonia, L. (2020). The Old and the New: Prospects for Non-Integrating Lentiviral Vector Technology. Viruses *12*, 1103.
38. Takao, M. (1998). Specific and Independent Recognition of U3 and U5 att Sites by Human Immunodeficiency Virus Type 1 Integrase In Vivo. J. Virol. *72*, 8396–8402.
39. van Haasteren, J., Munis, A.M., Gill, D.R., and Hyde, S.C. (2021). Genome-wide integration site detection using Cas9 enriched amplification-free long-range sequencing. Nucleic Acids Res. *49*, e16.
40. Yu, S.F., von Rüden, T., Kantoff, P.W., Garber, C., Seiberg, M., Rüther, U., Anderson, W.F., Wagner, E.F., and Gilboa, E. (1986). Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. Proc. Natl. Acad. Sci. USA *83*, 3194–3198.
41. Hawley, R.G., Covarrubias, L., Hawley, T., and Mintz, B. (1987). Handicapped retroviral vectors efficiently transduce foreign genes into hematopoietic stem cells. Proc. Natl. Acad. Sci. USA *84*, 2406–2410.
42. Wu, C., and Dunbar, C.E. (2011). Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity. Front. Med. *5*, 356–371.
43. Cesana, D., Volpin, M., Secanechia, Y.N.S., and Montini, E. (2017). Safety and Efficacy of Retroviral and Lentiviral Vectors for Gene Therapy. In Safety and Efficacy of Gene-Based Therapeutics for Inherited Disorders (Springer International Publishing), pp. 9–35.
44. Zychlinski, D., Schambach, A., Modlich, U., Maetzig, T., Meyer, J., Grassman, E., Mishra, A., and Baum, C. (2008). Physiological Promoters Reduce the Genotoxic Risk of Integrating Gene Vectors. Mol. Ther. *16*, 718–725.
45. Modlich, U., Navarro, S., Zychlinski, D., Maetzig, T., Knoess, S., Brugman, M.H., Schambach, A., Charrier, S., Galy, A., Thrasher, A.J., et al. (2009). Insertional Transformation of Hematopoietic Cells by Self-inactivating Lentiviral and Gammaretroviral Vectors. Mol. Ther. *17*, 1919–1928.
46. Cesana, D., Ranzani, M., Volpin, M., Bartholomae, C., Duros, C., Artus, A., Merella, S., Benedicenti, F., Sergi Sergi, L., Sanvito, F., et al. (2014). Uncovering and Dissecting the Genotoxicity of Self-inactivating Lentiviral Vectors In Vivo. Mol. Ther. *22*, 774–785.

47. Eugenio, M. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J. Clin. Investig. *119*, 964–975.

48. David, R.M., and Doherty, A.T. (2017). Viral Vectors: The Road to Reducing Genotoxicity. Toxicol. Sci. *155*, 315–325.

49. Papayanni, P.-G., Psatha, N., Christofi, P., Li, X.-G., Melo, P., Volpin, M., Montini, E., Liu, M., Kaltsounis, G., Yiangou, M., et al. (2021). Investigating the Barrier Activity of Novel, Human Enhancer-Blocking Chromatin Insulators for Hematopoietic Stem Cell Gene Therapy. Hum. Gene Ther. *32*, 1186–1199.

50. Arianna, M. (2012). Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J. Clin. Investig. *122*, 1653–1666.

51. Logan, A.C., Haas, D.L., Kafri, T., and Kohn, D.B. (2004). Integrated Self-Inactivating Lentiviral Vectors Produce Full-Length Genomic Transcripts Competent for Encapsidation and Integration. J. Virol. *78*, 8421–8436.

52. Hanawa, H., Persons, D.A., and Nienhuis, A.W. (2005). Mobilization and Mechanism of Transcription of Integrated Self-Inactivating Lentiviral Vectors. J. Virol. *79*, 8410–8421.

53. Farley, D., Stockdale, S., Moore-Kelly, C., Miskin, J., Reiser, J., and Mitrophanous, K. (2025). Risks of replication-competent retro/lentivirus from associated vector systems: Is it time for a roadmap toward reduced testing? Mol. Ther. Methods Clin. Dev. *33*, 101601.

54. Higashimoto, T., Urbinati, F., Perumbeti, A., Jiang, G., Zarzuela, A., Chang, L.-J., Kohn, D.B., and Malik, P. (2007). The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors. Gene Ther. *14*, 1298–1304.

55. Dongwan, K. (2020). Viral hijacking of the TENT4–ZCCHC14 complex protects viral RNAs via mixed tailing. Nat. Struct. Mol. Biol. *27*, 581–588.

56. Young-suk, L. (2024). Deadenylation kinetics of mixed poly(A) tails at single-nucleotide resolution. Nat. Struct. Mol. Biol. *31*, 826–834.

57. Mouzannar, K., Schauer, A., and Liang, T.J. (2024). The Post-Transcriptional Regulatory Element of Hepatitis B Virus: From Discovery to Therapy. Viruses *16*, 528.

58. Eric, B. (2018). Regulating the gene-therapy revolution. Nature *564*, S20–S22.

59. Maunder, H.E., Wright, J., Kolli, B.R., Vieira, C.R., Mkandawire, T.T., Tatoris, S., Kennedy, V., Iqball, S., Devarajan, G., Ellis, S., et al. (2017). Enhancing titres of therapeutic viral vectors using the transgene repression in vector production (TRiP) system. Nat. Commun. *8*, 14834.

60. Farley, D. (2018). Standardizing viral vector manufacture: maximizing production with the TRiP SystemTM. Cell Gene Ther. Insights *4*, 983–996.

61. Ran, Y., Deng, Y., and Yao, C. (2021). U1 snRNP telescripting: molecular mechanisms and beyond. RNA Biol. *18*, 1512–1523.

62. Butler, S.L., Hansen, M.S., and Bushman, F.D. (2001). A quantitative assay for HIV DNA integration in vivo. Nat. Med. *7*, 631–634.

63. Mandal, D., and Prasad, V.R. (2009). Analysis of 2-LTR Circle Junctions of Viral DNA in Infected Cells. Methods Mol. Biol. *485*, 73–85.

64. Ma, H., and Kafri, T. (2004). A single-LTR HIV-1 vector optimized for functional genomics applications. Mol. Ther. *10*, 139–149.

65. Hu, P., Bi, Y., Ma, H., Suwanmanee, T., Zeithaml, B., Fry, N.J., Kohn, D.B., and Kafri, T. (2018). Superior lentiviral vectors designed for BSL-0 environment abolish vector mobilization. Gene Ther. *25*, 454–472.

66. Rist, M., Kaku, M., and Coffin, J.M. (2025). Ex vivo HIV DNA integration in STAT3 drives T cell persistence—A model of HIV-associated T cell. lymphoma.PLoS Pathog. *21*, e1013087.

67. Christian, M.L., Dapp, M.J., Scharffenberger, S.C., Jones, H., Song, C., Frenkel, L.M., Krumm, A., Mullins, J.I., and Rawlings, D.J. (2022). CRISPR/Cas9-Mediated Insertion of HIV Long Terminal Repeat within BACH2 Promotes Expansion of T Regulatory–like Cells. J. Immunol. *208*, 1700–1710.

68. Que, D. (2001). Effects of Homology Length in the Repeat Region on Minus-Strand DNA Transfer and Retroviral Replication. J. Virol. *75*, 809–820.

69. Berkhout, B., van Wamel, J., and Klaver, B. (1995). Requirements for DNA Strand Transfer During Reverse Transcription in Mutant HIV-1 Virions. J. Mol. Biol. *252*, 59–69.

70. Neve, J., Patel, R., Wang, Z., Louey, A., and Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biol. *14*, 865–890.

71. Cannon, P.M., and Ngiam, C. WO2005056057A1 - Minimal lentiviral vector system. https://patents.google.com/patent/WO2005056057A1.

72. Smyth, R.P., Smith, M.R., Jousset, A.-C., Despons, L., Laumond, G., Decoville, T., Cattenoz, P., Moog, C., Jossinet, F., Mougel, M., et al. (2018). In cell mutational interference mapping experiment (in cell MIME) identifies the 5′ polyadenylation signal as a dual regulator of HIV-1 genomic RNA production and packaging. Nucleic Acids Res. *46*, e57.

73. Bagnis, C., Zwojszczyki, G., Chiaroni, J., and Bailly, P. (2014). Off-on Polyadenylation Strategy as a Supplemental Mechanism for Silencing Toxic Transgene Expression During Lentiviral Vector Production. Biotechniques *56*, 311–318.

74. Cohen, R.N., van der Aa, M.A.E.M., Macaraeg, N., Lee, A.P., and Szoka, F.C. (2009). Quantification of plasmid DNA copies in the nucleus after lipoplex and polyplex transfection. J. Control. Release *135*, 166–174.

75. Kim, V.N., Mitrophanous, K., Kingsman, S.M., and Kingsman, A.J. (1998). Minimal Requirement for a Lentivirus Vector Based on Human Immunodeficiency Virus Type 1. J. Virol. *72*, 811–816.

76. Kotsopoulou, E., Kim, V.N., Kingsman, A.J., Kingsman, S.M., and Mitrophanous, K.A. (2000). A Rev-Independent Human Immunodeficiency Virus Type 1 (HIV-1)-Based Vector That Exploits a Codon-Optimized HIV-1 GagPol Gene. J. Virol. *74*, 4839–4852.

77. Farley, D.C., Iqball, S., Smith, J.C., Miskin, J.E., Kingsman, S.M., and Mitrophanous, K.A. (2007). Factors that influence VSV-G pseudotyping and transduction efficiency of lentiviral vectors - in vitro and in vivo implications. J. Gene Med. *9*, 345–356.

78. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

79. Cotto, K.C., Feng, Y.-Y., Ramu, A., Richters, M., Freshour, S.L., Skidmore, Z.L., Xia, H., McMichael, J.F., Kunisaki, J., Campbell, K.M., et al. (2023). Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. Nat. Commun. *14*, 1589.

80. Bittencourt, A.S. (2010). FastQC: a quality control tool for high throughput sequence data (Babraham Bioinformatics). http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

81. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10.

82. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics *26*, 589–595.

83. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

84. Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., and Chen, J. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. *6*, e251.