Molecular Therapy
Methods & Clinical Development
Original Article

HIV-based lentiviral vectors: Origin and sequence differences

Nathan M. Johnson,[1] Anna Francesca Alvarado,[2] Trey N. Moffatt,[2] Joshua M. Edavettal,[2] Tarun A. Swaminathan,[2] and Stephen E. Braun[1,2]

[1]Division of Immunology, Tulane National Primate Research Center, Tulane University School of Medicine, Covington, LA 70433, USA; [2]Department of Pharmacology, Tulane University School of Medicine, New Orleans, LA 70112, USA

Three gene therapy strategies have received US Food and Drug Administration (FDA) approval; one includes HIV-1-based lentiviral vectors. These vectors incorporate features to provide long-term gene transfer and expression while minimizing generation of a replication-competent virus or pathogenicity. Importantly, the coding regions of viral proteins were deleted, and the *cis-acting* regulatory elements were retained. With the use of representative vectors developed for clinical/commercial applications, we compared the vector backbone sequences to the initial sources of the HIV-1. All vectors included required elements: 5′ long terminal repeat (LTR) through the Ψ packaging signal, central polypurine tract/chain termination sequence (cPPT/CTS), Rev responsive element (RRE), and 3′ LTR, including a poly(A) signal. The Ψ signaling sequence demonstrated the greatest similarity between all vectors with only minor changes. The 3′ LTR was the most divergent sequence with a range of deletions. The RRE length varied between vectors. Phylogenetic analysis of the cPPT/CTS indicated multiple sources, perhaps because of its later inclusion into lentiviral vector systems, whereas other regions revealed node clusters around the HIV-1 reference genomes HXB2 and NL4-3. We examine the function of each region in a lentiviral vector, the molecular differences between vectors, and where optimization may guide development of the lentiviral delivery systems.

INTRODUCTION

The core tenet of gene therapy as a treatment modality is the ability to deliver and express a transgene capable of imparting therapeutic benefit. Although a range of platforms have been developed to deliver such genetic material to target cell populations, viral vectors are a particularly effective and versatile tool. Although harnessing essential steps in the viral life cycle, these viral vectors separate the regulatory elements needed for transduction and transgene expression from the structural and enzymatic proteins needed for viral particle production. Thus, these replication-defective viruses are repurposed and utilized for their natural genome-modifying properties. Due to their ability to integrate into the host cell genome, the γ-retrovirus and lentivirus (LV) genera within the retroviridae family have received the most attention. The γ-retroviral vectors use the parental murine leukemia virus (MLV) benefit from simple design, high transduction rates, stable packing cell lines, and easy pseudotyping for broad tropism and can be made self-inactivating (SIN). The drawbacks of γ-retroviruses were first confronted in the initial clinical trials for X-linked severe combined immunodeficiency (X-SCID) in patients missing the γ common chain (γc) of the interleukin 2 (IL-2) receptor. Patients received autologous marrow-derived $CD34^+$ cells transduced with a therapeutic γ-retroviral vector driving γc from the long terminal repeat (LTR), and nine of ten patients developed a functional, adaptive immune system.[1] As γ-retroviral vectors tend to integrate near transcriptional start sites (TSSs),[2,3] it was recognized that vectors with the enhancers/promoters in a complete LTR had the potential for insertional mutagenesis. Theoretically, clonal expansion could select integrations that confer growth advantages. Therefore, as replication-defective vectors only complete a single round of infection, it was hypothesized that the chances of integration-activating proto-oncogenes or disrupting tumor suppressors were unlikely. However, clonal imbalances arose, and four patients developed T leukemia after the bone marrow transplantation in this clinical trial.[4] Further analysis revealed that in three of the four patients, integration near the LIM domain-only 2 (*LMO2*) proto-oncogene resulted in *cis*-activation.[1,4] Other studies revealed a preference for γ-retroviral vector integration at TSS and CpG islands.[5,6] In a separate trial for Wiskott-Aldrich syndrome (WAS) with a γ-retroviral vector expressing WAS protein (WASP),[7] all nine patients with successful engraftment of WASP-positive hematopoietic stem cells (HSCs) for least 1 year following infusion had integrations at known proto-oncogenes, including the *LMO2* region identified in the X-SCID trials.[8] Six patients proceeded to develop T cell acute lymphoblastic leukemia (ALL) between 16 months and 5 years after gene therapy, all of which showed a dominant *LMO2* clone.[8] These trials suggested that insertional mutagenesis at the *LMO2* proto-oncogene was not a feature unique to the γ_c-SCID vectors. Rather, they point to the γ-retroviral vector's integration properties.

Received 18 November 2020; accepted 23 March 2021;
https://doi.org/10.1016/j.omtm.2021.03.018.

Correspondence: Stephen E. Braun, PhD, Division of Immunology, Tulane National Primate Research Center, 18703 Three River Road, Covington, LA 70433, USA.
E-mail: stephen.braun@tulane.edu

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

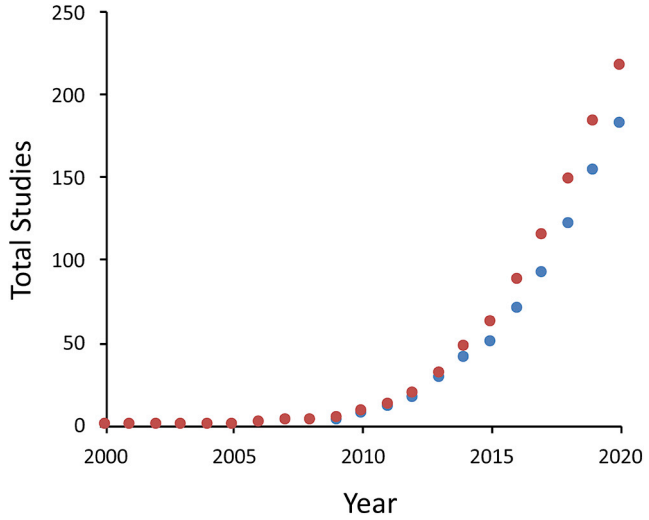


Figure 1. Growth of lentiviral vectors in clinical trials

Clinical trials using lentiviral or lentiviral vectors were identified at https://clinicaltrials.gov/ and accumulated from 2000 to 2020. Search terms were lentiviral (red) or lentiviral vector (blue).

Spurred by the shortcomings of these trials, focus was redirected (1) toward development of a safer γ-retroviral vector and (2) toward continued development of LV vectors. Replication-defective LV derived from parental HIV-1 clones, although similar in capacity for broad tropism and self-inactivation, are also able to transduce nondividing cells and importantly, integrate downstream of TSSs.[9] In contrast to the γ-retroviral vectors, LVs tether the preintegration complex (PIC) and facilitate integration into active chromatin regions through the action of integrase with host proteins such as LEDGF/p75.[5] In an interesting follow-up to the WAS trials, a LV SIN vector was used to express WASP under the endogenous *WAS* gene promoter to ensure physiologic protein expression in transduced autologous bone marrow-derived $CD34^+$ cells.[10] This allowed a more direct comparison, within the same disease background, of LV vectors to γ-retroviral vectors, specifically with regards to safety and integration patterns. A seen before, the LV insertion site analysis revealed a greater range of locations and a broader spectrum of gene classes being targeted.[10] Common insertion site analysis between patients revealed a much greater range of genomic preferences for the LV, none of which has been associated with clonal expansion.[10] Promoter choice played a role as well, as was discussed by the authors, to further enhance safety by reducing potential *cis*-activation and subsequent oncogenesis.[3,5] Although the SIN LV platform reduces the potential for oncogene activation via promoter insertion,[11,12] concerns have been raised regarding aberrant splicing events as a result of LV vector insertion favoring cryptic splice sites,[13–15] a risk that is also possible with MLV vectors,[16] and enhancer-mediated activation of oncogenes in a manner proportional to the strength of promoter.[13] In spite of these risks, LVs continue to display a high degree of safety in a range of clinical applications: WAS,[17] X-linked chronic granulomatous disease (CGD),[18] β-hemoglobinopathies,[19–21] and X-linked SCID,[22,23] among others.[2,24–28]

As with all gene therapy, the cost of LV therapy is significantly higher than other treatments such as HSC transplantation (HSCT), but this higher cost translates to fewer post-treatment complications.[29] Importantly, vector production comprises 48% of the total cost of gene therapy, and therefore, prices are likely to decrease both as the number of patients treated increases and as manufacturing evolves and improves.[29] In this aspect, MLV may hold some advantage due to ease of producing packaging cell lines relative to LV packaging cell lines. On the other hand, vesicular stomatitis virus G protein (VSV-G)-pseudotyped LV vectors can be concentrated, which allows for increased MOI and increased transduction of target cells.[30] LV vectors established a strong safety profile in early clinical trials[21,31] and demonstrated an excellent safety record.[25,26,32,33] The advantages of LV[10] are reflected in the rapid increase in clinical studies using LV during the early 2010s (Figure 1), culminating in the first US Food and Drug Administration (FDA)-approved LV-based gene-therapy treatment, tisagenlecleucel/Kymriah, in 2017.[34] To date, with the utilization of these properties, more than 25 different LV backbones have been used in over 200 clinical trials.

The HIV genome contains three structural genes and six accessory genes. During the normal life cycle, viral genomic RNA (vRNA) is expressed and packaged into viral particles, which bud from the cell surface. To ensure packaging of the full-length genomic RNA, the viral genome uses nuclear export domains to ferry the vRNA to the cytoplasm, and a Ψ packaging signal to incorporate the unspliced message into the viral particle. These viral particles bind to and infect target cells. In the target cell, the viral genome is converted to double-stranded DNA by HIV enzyme reverse transcriptase (RT). During RT, the U3 region of the viral genome is copied to the 5′ LTR in a complex, strand-jumping mechanism.[35] The PIC is transported to the nucleus where the 5′ and 3′ LTR facilitate integration into the host cell genome by the HIV enzyme integrase.

HIV-1-based LVs have deleted these protein-coding genes and only contain *cis-acting* regulatory elements required for the viral life cycle. These include the R/U5 to the Ψ domain, the central polypurine tract (cPPT), the Rev responsive element (RRE), and a truncated 3′ LTR. Other regulatory elements may be included to increase or regulate expression, for example, a heterologous internal promoter, a woodchuck hepatitis B virus post-transcriptional regulatory element (wPRE), internal ribosome entry site (IRES), or insulators. These studies describe the *cis-acting* regulatory domains derived from HIV-1, discuss the impact of each element in the LV vector, and compare the sources and origin of sequences in the HIV-based LV. We have focused on LV available from commercial sources and LV referenced in clinical applications (Table S1).

RESULTS

Evolution of LV packaging systems

Early first-generation LVs derived from HIV-1 were a three-plasmid expression system.[36] This 1st-generation packaging construct was a recombinant HIV genome that provided all required viral proteins in *trans* and facilitated viral particle assembly, while itself lacking

the regulatory *cis-acting* sequences necessary for replicating and producing infectious particles. It featured a heterologous human cytomegalovirus (CMV) immediate early 1 promoter to express HIV-1 Gag/Pol, while truncating the *env* gene and deleting the Vpu accessory protein. Other regulatory elements were modified; for example, the Ψ packaging signal was deleted, whereas preserving the splice donor (SD) site, and the insulin poly(A) signal replaced the 3′ LTR following the nef sequence.[36] An envelope gene, typically the highly stable VSV-G envelope, was included in a separate plasmid to provide receptor binding and membrane fusion characteristics to the viral particle. The third plasmid was the transfer vector with the Ψ packaging signal to allow assembly into the viral particles, the *cis-acting* sequences necessary for reverse transcription and integration, and an internal promoter to regulate transgene expression. The resulting viral-like particles are replication defective but can mediate one viral life cycle (binding, RT, integration, and expression). Second-generation packaging systems extended this work by attenuating the packaging construct to exclude Env, Vif, Vpr, Vpu, and Nef accessory proteins for improved safety considerations, particularly regarding use *in vivo*.[37] Importantly, these studies reported no difference in the efficiency of gene delivery.[37] Shortly thereafter, 3rd-generation vector systems added a constitutively active heterologous promoter upstream of the vector transcript in the 5′ LTR and reduced the packaging constructs to separate Gag/Pol and Rev expression plasmids.[38] Initially characterized using the pRRL and pCCL vectors included in this analysis, these advanced vectors featured the enhancer and promoter of Rous sarcoma virus (RSV) and CMV, respectively, regulating expression of the transfer vector.[1,38] Since the role of Tat in activating transcription of the genomic RNA from the LTR was replaced, eliminating Tat from the system will increase safety. Furthermore, the inclusion of Rev as a separate Rev plasmid reduced the probability of recombination that could lead to replication-competent LV (RCL). By combining these alterations with deletions in the 3′ LTR to confer self-inactivation,[11,12,38,39] these third-generation LVs brought unprecedented biosafety considerations to the platform and became a highly safe delivery system for gene therapy. Yet, within this generation of vectors, various transfer plasmids were created with differences among the numerous commercial vectors and vectors being explored in clinical trials.

LTR and Ψ packaging elements

In the HIV vector genomes, the first domain with homology to the HIV viral genome is the 5′ LTR through to the Ψ packaging signal. Most LVs use a heterologous promoter to regulate vRNA expression in packaging cells at the TSS (R/U5 region), so the 5′ LTR is incomplete and missing U3 in these vector constructs. During RT, the R domain in the LTR is involved in strand transfer and recreates the 5′ U3 from the 3′ LTR. Adjacent to the R/U5 region in the vRNA is the primer binding site (PBS), which initiates RT. In the 5′ untranslated region (UTR), the Ψ packaging signal is preceded by a SD and extends ~350 base pairs (bp) past the *gag* initiation codon. For efficient packaging of the full-length vRNA, the highly conserved 155-nucleotide (nt) RNA packaging signal, located in the 5′ leader of the vRNA, adopts one of two alternate RNA conformations. The nucleocapsid (NC) protein from Gag recognizes the Ψ packaging signal and efficiently assembles two unspliced viral genome copies into viral particles at the plasma membrane.[40,41] In the dense cellular milieu, the ability to discriminate and package its own unspliced dimeric viral RNA, from more than 40 spliced viral HIV RNA variants, relies on the Ψ packaging signal assuming a complex tandem three-way junction structure.[40,42] This conformation acts as a scaffolding to expose unpaired or weakly paired guanosines for high-affinity binding to the NC region of Gag.[40] Simultaneously, this structure sequesters the *gag* initiation codon (AUG) via base pairing to a portion of the upstream U5 region. Mutations to these guanosines or mutations that affect pairing of the AUG region to the U5 disrupt packaging by impairing either NC binding or formation of the tandem set of three-way junctions, respectively.[40,42]

The importance of these elements is underscored by studies exploring the requirements of Ψ and surrounding sequences in the context of efficient LV packaging and transduction. Although deletions or substitutions in Ψ do not affect the production of viral particles, it is absolutely required for the packaging of transgenic RNA into these particles.[43] Serial deletions of the stem loops and Ψ sequence past the *gag* initiation codon AUG (a region that has also been shown to be involved in paring with the U5)[42] have demonstrated that proper stem-loop formation is indispensable for encapsidation.[44] These studies confirmed, within the context of LV, that these domains are essential for packaging and efficient gene transfer. With the extension of the LV further into the *gag* gene, studies explored the effects of different lengths of 5′ *gag* sequence[45,46] by replacing wild-type (WT) sequence following the start of the Gag open reading frame (ORF) with a codon-optimized sequence.[45] They found an indirect role of this region in packaging, implicating this sequence in stabilizing RNA structures critical for encapsidation.[45,46] The portion between 60 nt[46] and 726 nt[45] probably affects RNA folding to facilitate Gag:RNA interactions that stabilize the dimer conformation, promoting faster dimerization, and resulting in more efficient packaging.[46] As the viral titer is the summation of viral expression, processing, and packaging, all of these functions are necessary to make a high-titer vector.[43–46]

To facilitate comparison between vectors, vector sequences were organized into groups defined by similar polymorphisms and length (Figure 2). Likely origins were inferred by generating phylogenetic tree distances between the vectors and the common HIV-1 reference sequences.[47] Representative vectors from each group and various HIV-1 sequences were aligned to reveal that all vectors cluster within two main reference sequence-containing nodes. Within the clinical vectors, groups iv, v, and vii clustered with the NL4-3 reference genome (Figure S1), whereas the remaining clinical trial groups (i, ii, and iii) clustered with HXB2 (the NCBI reference HIV-1 genome)[47] in close proximity with the broader node-containing groups vi and IIIB_LAI. Branches of this node with HXB2 also contained all commercial vectors (a–c). Two main clusters grouped closely with these reference sequences, suggesting that two main HIV-1 sequences were used to generate this region of the LV vectors. Alignment of predicted parental HIV reference genomes with LV

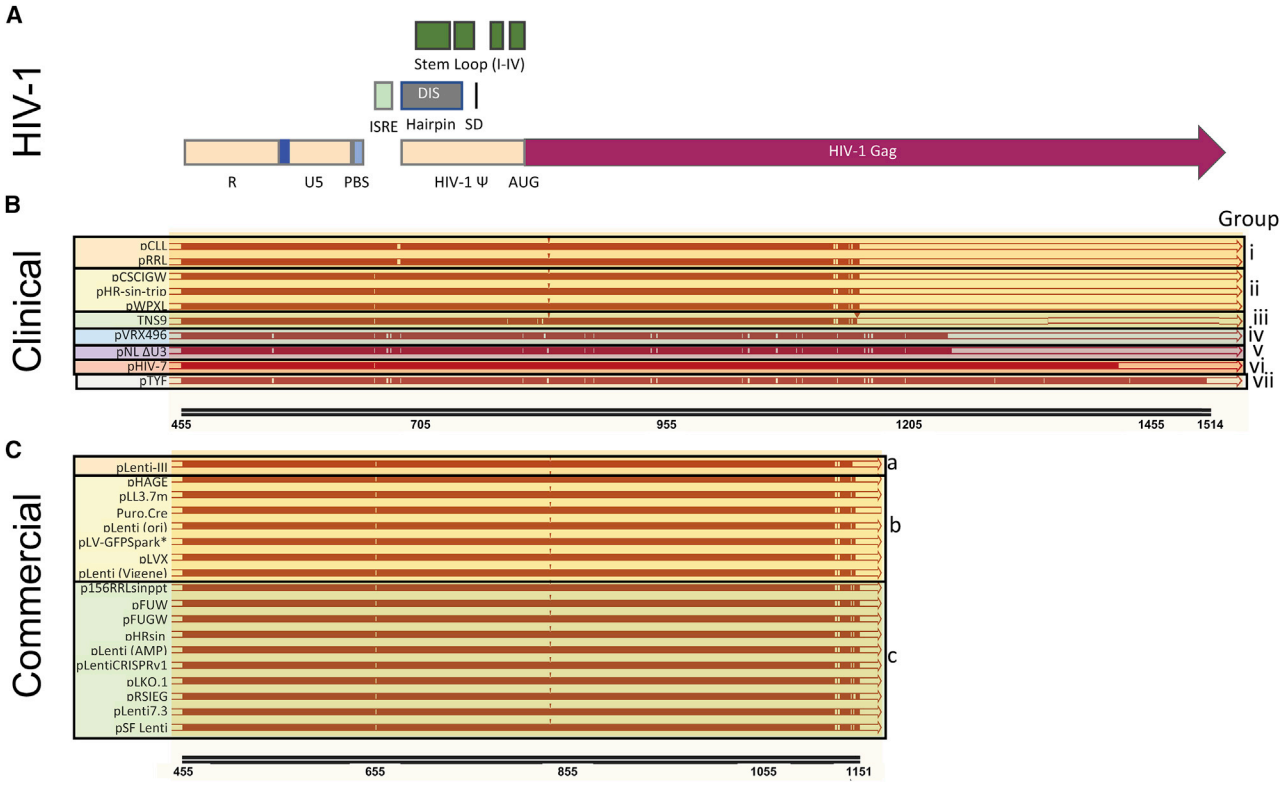


Figure 2. Alignment of the long terminal repeat (LTR) and Ψ packaging elements

(A) Diagram of HIV-1 functional domains. The *cis-acting* regulatory elements displayed are not exhaustive but reflect areas that are important for the function of lentiviral vectors. The region (blue box) in U5 (tan box) that binds to the AUG in the Ψ secondary structure (tan box) and participates in the hairpin formation. The primer binding site (PBS), interferon-stimulated response element (ISRE), dimerization-initiation sequence (DIS), and splice donor (SD) are identified. (B and C) Clinical (B) and commercial (C) vectors were aligned and grouped, i–vii for clinical and a–c for commercial vectors, according to homology and length.

groups clustering closest on the phylogenetic tree analysis shows scattered nt changes but with overall alignments in high agreement. In particular, NL4-3 aligns to groups iv, v, and vii with ≥99.5% sequence identity.

Furthermore, sequences were compared at functional domains, and a few notable differences were seen among the vectors. Among the distinguishing differences, starting with the cluster associated with NL4-3, the clinical pTYF, pNL ΔU3, and pVRX496 vectors have a TC → CA mutation in the poly(A) RNA hairpin, distal to the AATAAA poly(A) start signal yet prior to the U5 sequence involved in hairpin formation and altogether not in the region required for encapsidation.[40,41] The HXB2 cluster has a C → T mutation at position 654 in the PBS stem loop and immediately preceding the interferon-stimulated response element (ISRE).[48] Neither the ISRE nor the PBS hairpin is required for encapsidation. The pTYF, pNL ΔU3, and pVRX496 clusters have three additional substitutions within the ISRE. Although this sequence is important for efficient LTR-driven transcription, in SIN vectors with internal promoters, this is unlikely to have much effect. Perhaps ISRE-defective vectors would have less activation-induced expression once integrated, a trait that could be advantageous in approaches where strict transgene expression regulation is necessary. These vectors also have C → A substitution at 681, the first 5′ nt of the extended dimerization hairpin, likely leaving encapsidation unaffected. As expected from the phylogenetic analysis, both this mutation and the ISRE substitutions are seen in the NL4-3-predicted parental strain. The pCCL and pRRL have a 3-bp deletion just 5′ to the extended dimerization hairpin, similarly avoiding a region critical to NC encapsidation, leaving the dimerization-initiation sequence (DIS) hairpin intact to mediate initial vRNA dimerization via kissing-loop interaction.[40,49–51] The pCL20 vector is unique in the HXB2 vectors by using the HIV U3 basal transcriptional region (HXB2 375–454), terminating the *gag* gene 4 bp sooner, and aligning even more closely to HXB2 compared with other vectors in this group. Consistent with the essential role of Ψ in the natural HIV life cycle, mutations, substitutions, or deletions were not detected in any commercial or clinical vectors within the four Ψ stem loops (from HXB2 697 to 806), a region critical to the tandem three-way junction (Figure 3).

Several vectors incorporate safety features that disrupt the Gag ORF. The pTNS9 uniquely contains a T → C mutation at position 791 in

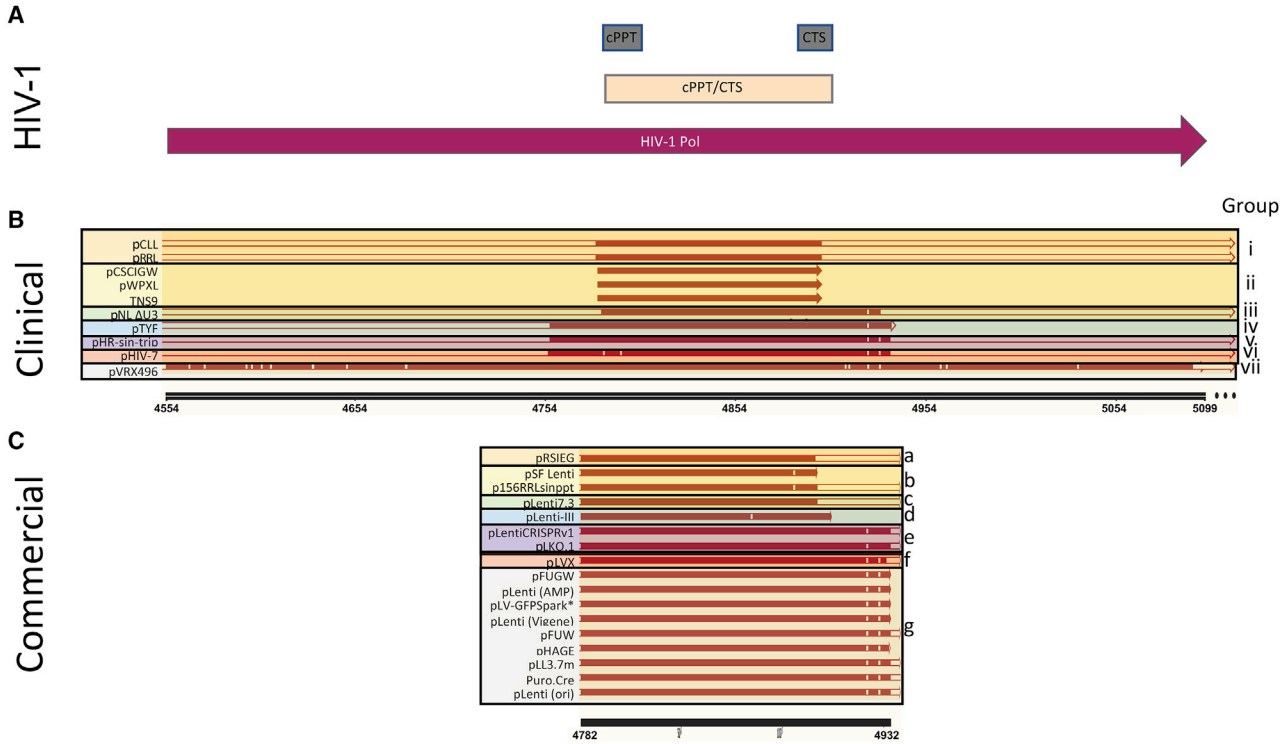


Figure 3. Alignment of the central polypurine tract and central termination sequence (cPPT/CTS)

(A) Diagram of HIV-1 functional domains in the polymerase gene. The extended cPPT/CTS regions are shown (tan box); the minimal cPPT and CTS are shown (gray boxes). (B and C) Clinical (B) and commercial (C) vectors were aligned and grouped, i–vii for clinical and a–g for commercial vectors, according to homology and length.

the ATG, which prevents *gag* translation. This mutation is also immediately preceding stem-loop IV within Ψ, replacing a non-canonical U:G pairing with a C:G pairing in the tandem loop between the AUG start codon and the LTR U5. Although this substitution is unlikely to disrupt interactions between these regions, it is unclear what impact sequestering this particular guanine with a strongly paired cytosine will have on Ψ binding to Gag. In contrast, all commercial and most clinical vectors (including pTNS9) incorporate a CG frameshift at position 833, which generates a premature stop codon after 21 amino acids (aa), 14 aa from Gag and 7 aa after the frameshift. VRX496 has a unique GA frameshift deletion at position 829, which also generates a small polypeptide. The clinical vectors TYF and HIV-7 leave the *gag* ORF intact. For TYF, the *gag* ORF leads to a 638-aa protein, which contains 240 aa from the *gag* sequence, 326 aa from *env*/RRE, and 61 aa from *pol* (cPPT/chain termination sequence [CTS]); in HIV-7, 631 bp of *gag* is open, which extends 179 bp into a sequence containing the RRE (210 aa of Gag plus 59 aa out of frame with *env*). In these vectors, a frameshift or truncation leaves the AUG and stem loop IV intact. In all vectors, *gag* sequences are extended well past the end of SL4, with the smallest inclusion being 353 nt of 5′ *gag* in pLenti-III, overall reflecting the importance of the 5′ portions of *gag* in genome packaging.[45,46] In support of these extended sequences into *gag*, studies that substituted the extended *gag* sequence with 350 bp of a multiple cloning site sequence find a 2- to 3-fold reduction in transduction efficiency.[43] This reduction in efficiency is likely attributable to the presence of a weak affinity binding site within this matrix sequence.[43] Further studies of the minimal *gag* sequence for non-viral RNAs to be efficiently packaged found that the 5′ half of the *gag* sequence confers 82% packaging efficiency, similar to the full-length *gag*, whereas scanning deletions showed a decrease in packaging efficiency.[45] Altogether, these studies suggest that the addition of *gag* sequence 3′ to the standard SL4 loop improves RNA packaging and bolsters overall transduction efficiency.

cPPT/CTS

One unique feature of the LV genus is the ability to infect nondividing cells. Although γ-retroviruses such as MLV require disruption of the nuclear membrane during mitosis for the PIC to gain access to the nucleus, LVs have been shown capable of mitosis-independent integration.[52] Early explorations into the HIV reverse transcription process revealed that, prior to migration into the nucleus, unintegrated HIV-1 DNA exists as a discontinuous plus strand, demarcated by the triplex DNA structure at a cPPT.[53,54] In addition to the polypurine tract (PPT) typically found proximal to the U3 region in other retroviruses, the LV cPPT in the integrase coding sequence functions as another initiation site for plus-strand synthesis.[55] This site is extremely sensitive to purine → pyrimidine mutations, such that mutations significantly delay viral growth and ablate this second plus-strand origin site.[55] This triplex creates a "DNA flap" of 99 nt, which when lost during mutation studies, showed the mutant DNA accumulating outside

of the nucleus and unable to pass the nuclear membrane.[54] These findings have been controversial with later studies questioning this model after using a chimeric virus that deleted the cPPT and replaced the HIV integrase with the MLV integrase.[56,57] These later studies reported that the chimeric virus was able to infect both dividing and non-dividing HeLa cells with similar efficiency, suggesting the LV cPPT containing the integrase sequence is not essential for nuclear transport.[56] However, more selective mutations of the HIV cPPT and CTS concluded that these regions were necessary for viral replication and maintained low levels (around 5%–15%) of WT replication, which was still capable of nuclear import.[58] Whereas the exact role of cPPT in integration remains ongoing, in the setting of high-titer LV gene-therapy vectors, this region is decidedly beneficial. Inclusion of this 118-bp region of the *pol* gene, which includes the cPPT and CTS, has resulted in an 85% increase in transduction of phytohemagglutinin and IL-2-stimulated T lymphocytes over cPPT-deficient vectors[59] and consistently produces a 2- to 10-fold increase in transduction efficiency by LV.[60,61] It has been shown that these specific sequences comprising the DNA flap are not sensitive to mutations.[60] Furthermore, a second cPPT, and therefore second DNA flap, adds nothing to transduction efficiency; an ectopic flap can compensate for the absence of a central flap.[60] With the integrase gene supplied on a separate packaging plasmid, mutations at this region in the LV vector do not need to conserve integrase activity.

Among the cPPT and CTS regions included to enhance transduction, all vectors are remarkedly similar, without any purine → pyrimidine mutations that would limit activity. Yet, we see this region is polymorphic (Figure 3) with varying sequences, varying lengths after the CTS, and varying positions within the LV vector. These differences define many unique groups of vectors, perhaps due to their more recent incorporation into LV.[62,63] Importantly, two vectors (pCL20 and VRX496) use cPPT (481 bp and 453 bp, respectively), which extend to the end of the integrase and into vif, whereas the modal length of extra sequence is 36 bp. Additionally, only these two vectors incorporate the cPPT before the RRE.

Phylogenetic tree analysis reveals a wide dispersal of these vector groups among various reference sequences. Commercial group c and the pCL20 cluster the closest with the reference genome HXB2 (Figure S2). Group b vectors, the only group to feature a deletion at the CTS region, cluster with several other commercial and clinical vectors on a branch with the HIV-1 isolate JRFL sequence. Group vi (pHIV-7) has two deletions in the cPPT sequence and is a branch closest to the node with NL4-3 and pIIIB_LAI. The wide clustering pattern of the groups within this tree and the different sequence lengths suggest different origins for this element in each clinical and commercial vector group. Alignment of each derived sequence to its predicted parent shows greater than 97.4% identity, with a modal identity of 100%. This implies that we have identified the HIV-1 source for most cPPT sequences.

Among the defining differences within the cPPT/CTS element, several commercial vectors show a single nucleotide deletion, position 4887, in the CTS. Since the purine composition does not change, and the PPT region seems to be more sensitive to the length of the purine stretch,[55] the effect of this mutation is unclear. The pHIV-7 has 2 deletions within the cPPT region, positions 4784 and 4793, but still maintains a stretch of 15 purines. Studies exploring the length of the purine stretch only start to see an effect on viral growth when the stretch is reduced to nine purines.[55] The other mutation seen in this element falls into the DNA flap region, an area not sensitive to mutations.[60] None of these LV mutations are seen in the predicted parental strains.

Alignments of this element also reveal differences regarding the length of sequences surrounding the cPPT/CTS. Groups i, ii, a, b, and c feature a minimal 118-bp cPPT/CTS, whereas the remainder of the groups begins 24 bp 5′ and terminates at a common site 36 bp 3′ to the CTS sequence. The pVRX496 is the exception to this, extending 227 bp to the 5′ of cPPT and an additional 201 bp 3′ through to the end of *pol* and integrase. Among the early work evaluating effects on transduction efficiency imparted by this element, studies that used a minimal 118-bp cPPT/CTS reported an increase from 36% to 75% with its inclusion,[59] whereas others, using a 178-bp fragment, found an increase from 15% to 50% and up to 80% depending on MOI.[64] Although direct comparisons have not yet been made, these works suggest minimal difference in transduction efficiency between the 118 bp and 178 bp elements; however, the effects of the 546 bp in pVRX496 were not determined.

RRE

Proper encapsidation of full-length genomic RNA requires the efficient export of intron-containing nascent HIV RNA transcripts from the nucleus into the cytoplasm; the 19-kDa Rev protein, an HIV accessory gene, provides this function. Rev consists of an arginine-rich region responsible for binding HIV RNA, oligomerization domains that bind additional Rev, and a short hydrophobic domain responsible for nuclear export. Rev interacts with the highly conserved 351 nt RRE domain (HXB2 7711−8061), a sequence in *env* overlapping the junction between gp120 and gp41.[65–67] Early studies on the RRE showed that a minimum of 234 bp was required;[65] however, the full-length RRE, extended on both ends by 58 bp and 59 bp, enhanced Rev activity by accommodating additional Rev proteins during oligomerization to the longer stem.[68] The extension of the RRE to a sequence surrounding the 351-nt region does not confer additional activity,[68] a finding that has held true in subsequent studies evaluating differing lengths of the RRE element in LV with regard to RNA encapsidation and viral titer.[46,69] The specific interaction between Rev and the RRE is mediated by Rev's arginine-rich motif with the stem IIB loop of the intricately branched RRE secondary structure (Figure 4D), which initiates the cooperative oligomerization of Rev necessary for nuclear export function.[67]

To ensure that the full-length vector genome is exported from the nucleus into the cytoplasm in the presence of Rev, the different LV vectors incorporate varying lengths of *env* sequence flanking the highly

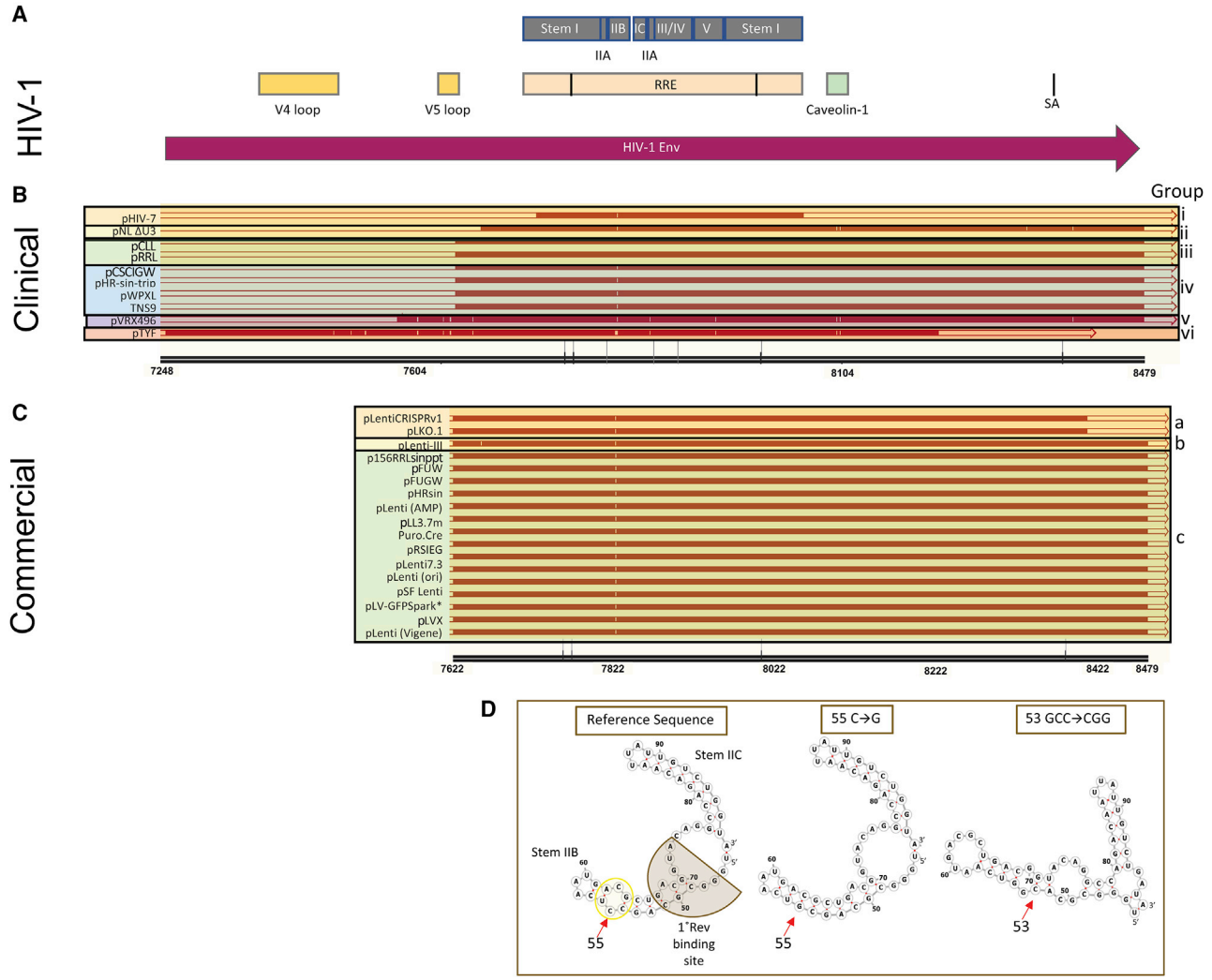


Figure 4. Alignment of the Rev response element (RRE)

(A) Diagram of HIV-1 functional domains in the *env* gene. The RRE is shown in tan; stems I–V are gray; the boarders of the minimal RRE and the 3′ splice acceptor (SA) site are depicted with lines; V4 and V5 domains of Env gp120 are yellow; the caveolin-1 binding site is orange. (B and C) Clinical (B) and commercial (C) vectors were aligned and grouped, i–vi for clinical and a–c for commercial vectors, according to homology and length. (D) RNA folding of the minimal RRE reference sequence (HXB2 7769–8002) and two mutant variations (red arrows). The Rev binding site is highlighted.

conserved 351-bp RRE element. In most vectors, the HIV-1 genome fragment containing the RRE is approximately 850 bp, follows the 5′ UTR SD site (HXB2 743), and incorporates the tat/rev splice acceptor (SA) site (HXB2 8379) from the *env* sequence.[36] Although pCL20 and HIV-7 RRE are shorter (768 bp and 365 bp), and VRX496 and TYF RRE are longer (926 bp and 964 bp, respectively), all vectors incorporate at least the full-length 351-bp RRE sequence (Figure 4). In viral production for LV vectors, the inclusion of the 859-bp RRE domain in the presence of Rev enhances infectious titers 6- to 37-fold and encapsidation efficiency up to 200-fold.[70]

Additionally, there is a high degree of conservation between the various LVs and the HIV-1 reference sequences. Phylogenetic groupings at this region place clinical group iii (Figure S3) and pCL20 closest to the HXB2 reference sequence. These are the only vectors without a substitution in the critical stem IIB structure that participates in initial Rev binding (Figure 4D). Groups ii, v, and vi, which cluster with NL4-3, have the same base pair changes as NL4-3 compared with the HXB2 reference sequence. Overall, the grouping of this region with reference sequences shows only a few main clusters. The alignment maps show almost all vectors with identical substitution in the stem IIB, and the nearly identical length element is incorporated into all commercial vectors. The relatively distant location of group i is likely due to its shorter length. Phylogeny supports that only a few different original sequences were used to generate this vector element.

Within the functional region, the most common substitution seen in commercial vectors and in 7 of 10 clinical trial vectors is a C → G at the 55-nt position of the minimal RRE (HXB2 position 7823). Although the consequences of this particular mutation have not been explored, this mutation is only a few nucleotides downstream of the high-affinity, purine-rich bubble of stem-loop IIb, the region responsible for initial Rev binding. The presence of duplexes or bulges adjacent to this region has been demonstrated to affect Rev oligomerization, whereby their presence increases the flexibility of the phosphate backbone, allows access to functional groups in the major groove, and accommodates bending of the RRE scaffolding. These changes facilitate the synergistic and cooperative nature of additional Rev recruitment to a greater extent than perfect duplex RNA structures, especially if in close proximity to the original binding site.[71] RNA folding predictions (Figure 4D) with a cytosine at position 55 show an additional bulge downstream of the primary Rev binding site, which is lost upon substitution with guanine. Underscoring the importance of adjacent bulges to Rev binding, the dynamic state of these RNA complex conformations, with interconverting conformational states occurring in the micro- and millisecond time scale, causes profound changes in biological activity of the RRE.[72] These different excited states have significant effects on the affinity of Rev binding. When considering a mutation in such close proximity to this region, its impact on the balance between excited state conformations and the availability of key nucleotides required in Rev binding must be considered.[72] Of the parental strains identified by phylogenetic tree analysis, only HXB2 lacks this substitution, and of the matched groups, only group iii vectors are homologous to the HXB2 parent sequence at this location. Notably, the pTYF clinical vector features a unique GCC → CGG substitution from position 53−55 (HXB2 7821−7823). As shown in the predicted RNA folding, this mutation appears to have a profound impact on the hairpin, altering the nucleotides that comprise the stem IIB structure itself. Such a mutation is highly likely to impact the initial Rev binding event. This mutation is not seen in the NL4-3-predicted parental strain. Further mutations seen in both pTYF and pVRX496, as well as their predicted parent NL4-3, are a G → A substitution at position 96 (HXB2 7864) and a G → A at position 178 (HXB2 7946) of the minimal RRE; these locate to stem IIC and the bulge of stem V, respectively. Although the aa change due to these mutations and its resulting impact on Env are irrelevant in the setting of a LV, the complex nature of the RRE structure can be significantly altered by single base pair changes.[73–75] As shown in several studies, minimal changes in either the RRE sequence or the Rev protein can alter the functional activity of Rev and the RREs.[73,74] Interestingly, these studies also identified a cognate Rev-RRE pair from a patient with only a few base pair changes that had increased activity.[73] These changes in a region of the RRE distant from the primary binding site promoted increased multimerization of Rev.[73] These data illustrate the sensitivity of Rev binding to the complex RRE secondary structure and how minor changes can potentially affect binding kinetics and subsequent nuclear export efficiency. Although the specific mutations seen in the LVs are different than those explored in the literature, changes in this highly intricate RNA scaffolding may impact binding kinetics and subsequent nuclear export efficiency. Moreover, the function of LV vectors could be further enhanced by incorporating optimal RRE/Rev variations.

3′ ΔLTR/PPT

The LTR region of an integrated HIV provirus is a duplicated ∼640-bp region critically important for transcriptional regulation of the HIV viral genome and subsequent reverse transcription and integration into infected cells.[76] It is divided into three regions (U3, R, and U5) with the 454-bp U3 region further subdivided into modulatory, enhancer, and core domains based on the transcription factor binding sites that regulate HIV gene expression.[77] To express the full-length viral RNA genome, transcriptional initiation begins at R/U5 in the 5′ LTR and continues until the poly(A) signal in U3/R of the 3′ LTR. Importantly, the promoter and enhancer sequences in the U3 only function to regulate HIV transcription when transferred to the 5′ LTR of progeny proviral DNA during reverse transcription.[78] For the integrated provirus, this region becomes responsible for regulating transcription of the viral RNA, and the wide range of regulatory elements it contains allows for changes in expression and viral replication in concert with the cellular environment. The U3 plays such a pivotal role that some mutations have been speculated to affect viral fitness and/or tropism. A third copy of the nuclear factor κB (NF-κB) site, as seen in subtype C HIV-1 strains, has been considered responsible for its rapid spread relative to other subtypes.[76,79] Furthermore, clonal analyses have identified an enrichment in a CCAAT-enhancer binding protein (C/EBP) site mutation, which confers enhanced affinity to the transcription factor among brain-derived isolates, suggesting increased fitness for CNS quasispecies.[76] The enhancers in this region are also strong *cis*-activators. Although LV vectors, compared with γ-retroviral vectors, have a decreased propensity to integrate into positions proximal to the promoter of a host gene, there remains a risk that the enhancer in the LTR will activate adjacent genes. For these reasons, these enhancers were removed from γ-retroviral and LV vectors, whereas preserving the polyadenylation signal, R domain, and U5. In these SIN vectors, deletions in the 3′ LTR are incorporated into the 5′ LTR during reverse transcription. Once integrated, the SIN vector does not possess the required enhancers to regulate transcription,[78] reducing the risk of insertional oncogenesis via *cis* or long-range enhancer activation, whereas restricting vectors to a single round of viral production.[78] Thus, these deletions add another layer of safety to the system. Additionally, SIN vectors show a reduced risk of mobilization by WT HIV-1.[39] With the HIV-1 polyadenylation signal located distal to the TATA box, and the integrase recognition and processing site at the 5′ upstream region of U3, LV can tolerate large deletions that remove the entire enhancer region, modulatory region, and even the TATA box without affecting viral titers.[11,38] For gene-therapy vectors, some disease models require tightly regulated expression, with full and deliberate control over transgene expression. In vectors with these deletions in the U3 region of the LTR, internal promoters will regulate transcription and allow for tissue-specific internal promoters.[80] In terms of efficacy, SIN vectors overcome the suboptimal RNA processing seen in early SIN vectors by including the wPRE.[81] When utilizing

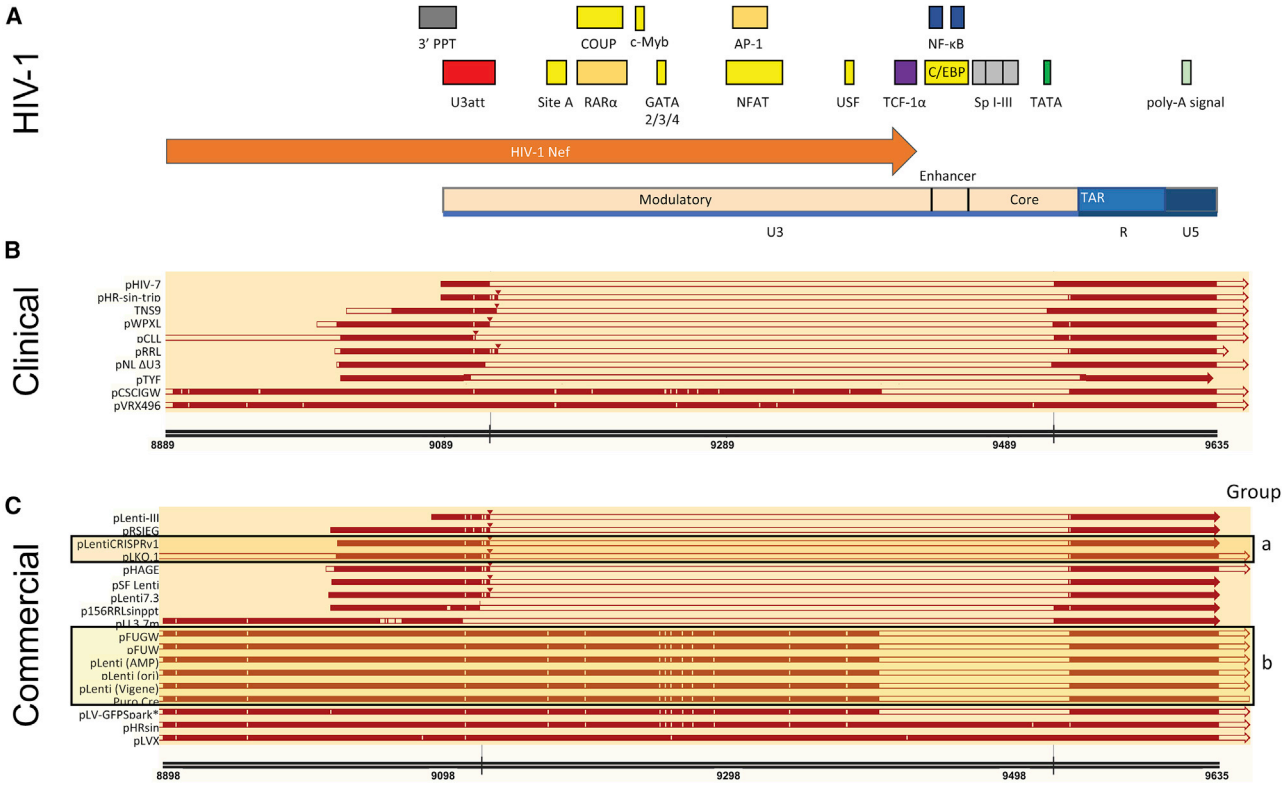


Figure 5. Alignment of the 3′ LTR

(A) Diagram of HIV-1 functional domains in 3′ LTR. The U3 is divided into modulatory, enhancer, and core promoter (tan box); the R/U5 (gray/blue) is divided, and the trans-activation response (TAR) element is shown. The HIV-1 Nef open reading frame is indicated (orange arrow). Various transcriptional binding sites are identified: retinoic acid receptor (RAR)α and AP-1 (orange), c-*myc*; GATA, NFAT, upstream stimulatory factor (USF), C/EBP (yellow), TCF-1α (purple), NF-κB (blue), SP1-III (gray), and TATA box (green); as well as other regulatory sequences: 3′ PPT (gray), U3att (red), site A, and COUP (yellow). (B and C) Clinical (B) and commercial (C) vectors were aligned and grouped (a and b), according to homology and length.

an external promoter such as RSV or CMV and enhancers such as SV40 in 3rd-generation vectors, the titers of infectious particles are comparable to LTR-driven vectors and efficiently express the transgene.[80,82,83]

Of the various viral components in the LV systems, the 3′ LTR shows the greatest phylogenetic variability between vectors, where clustering shows wide-branch dispersal. Yet, the number of main nodes with reference genomes is limited (Figure S4). Relative to the reference sequence HXB2, we observed the commercial vector pLVX in the closest branch, with the commercial pLL3.7 m and clinical pHIV-7 vectors in a close-out grouping. A small group of clinical vectors clustered with NL4-3 and the remainder were found as branches from a wide node containing reference pIIIB_LAI. Alignment of predicted parental reference strains with their derived LV vectors reveals a modal identity of 99.5% because the homology between these nodes is extensive (1 bp mismatch), beginning in the U3 and extending through R/U5 domains. Even with the high degree of variation between vectors, only three reference sequences contain all of the nodes, suggesting that a limited number of original sequences generated this element. Individual modifications of these main sequences are likely responsible for the subsequent wide branching of these nodes.

Even though the phylogenic analysis showed only a few clusters, the 3′ LTR region showed substantial modifications, mostly in the retention of sequence proximal to the PPT and in the deletion of U3 enhancer elements necessary for the self-inactivation (Figure 5). First, just proximal to and overlapping with the 3′ LTR are the 3′ PPT and the HIV nef gene. The PPT functions as an initiation site for plus-strand synthesis,[55] typical of the retroviridae family. Since the different vectors retain various lengths of the 3′ PPT, various lengths of the partial Nef gene will be retained in these vectors without the ATG start codon. These variations between vectors distinguish the individual vectors within the branches. Despite these individual differences, the number of 3′ LTR clusters with reference genomes is limited (Figure S4).

Concerning the 3′ ΔLTR, not all vectors have generated SIN vectors. The first LV vector employed in human trials was VRX496, which used the full-length HIV-1 LTR and made the vector responsive to Tat activation. The commercial vectors, pHRsin and pLVX, also retained a full-length LTR. Many commercial vectors (group B and

LV-GFPSpark) and the clinical vector pCSCIGW have a 134-bp deletion in the U3 core promoter domain (Figure 5). Most clinical vectors and one-half of the commercial vectors extend the deletion to around 400 bp with variability between each vector for the precise deletion (Figure 5). These U3 deletions eliminate the modulatory, enhancer, and core promoter regions from the vector. These larger deletions may increase safety and decrease the potential for basal transcription of the integrated LTR. All vectors retain the critical U3 integrase attachment site (U3att), as this regulatory element is essential for LTR function in the viral life cycle. The clinical vectors pCL20 and pTYF delete the 15 bp in R domain after the poly(A) signal (AATAAA, HXB2 9612–9618) and then replace the U5 region with the heterologous poly(A) signal. Other vectors have the full-length R/U5 with or without an additional heterologous poly(A) signal.

When analyzing polymorphisms, the majority of vectors feature an A → C substitution in the domain (HXB2 position 9109) near the U3att site as seen in the pIIIb_LAI reference genome. Although it is unclear what effect this has on the integration process, it likely does not affect integration negatively, as this is one of the core functions of an LV. The loss of integration capacity would certainly have excluded vectors from further development. Within the U3 modulatory region, vectors have various size deletions in the overlapping regulatory elements, which likely impacts the transcription factor binding sites. Interestingly, most vectors have an A → G substitution just 5′ of the chicken ovalbumin upstream promoter (COUP) transcription factor binding site (HXB2 position 9167). Reference genomes pNL4-3 and pIIIb_LAI also contain this substitution, whereas HXB2 does not. This sequence was first described as "A site" when analyzing the protein binding sites in the modulatory region of U3, where the A site and "B site" interact with a binding protein.[84] Later, this B site was discovered to confer retinoic acid responsiveness and negatively regulate cells.[77,84] Other common changes include substitutions near the GATA-2/3/4 site, known to negatively regulate transcription. Only 50% of the commercial vectors maintain the NFAT and AP-1 regions, which align precisely with the reference genome. In contrast, only the clinical vectors pCSCIGW and pVRX496 contain these transcription factor binding sites, with a G → A change at position 9312, of unknown importance in pVRX496. Only two commercial vectors include the T cell-specific factor (TCF)-1α site, with seven vectors truncating the 3′ LTR five bases into this site, and the others terminating immediately after the U3att site. It is unclear why many commercial vectors use the TCF-1α to begin U3 deletion. Of the clinical vectors, only pCSCIGW begins its U3 deletion here.

Broadly speaking, we see three main strategies of handling the 3′ LTR. Clinical pHIV-7 and commercial pLenti-III have a fully stripped-down LTR, beginning shortly after the essential U3att region or even 5 bp from the end of this site as in pTYF and pLL3.7 m. These vectors recommence distal to the TATA box. Next, we see a large number of vectors, typified by many commercial vectors such as pFUGW, and only one clinical vector, pCSCIGW, where most of the modulatory U3 region remains intact, beginning deletion just 5′ to TCF-1α and resuming after the TATA box. The last variety is the few vectors with a full-length and unmodified 3′ LTR. In all cases, the R/U5 domains are similar and homologous to the reference sequence HXB2, with only minor changes in the 3′ end of the deletion.

DISCUSSION

With the rapid development and refinement of novel gene therapy techniques, including CRISPR-Cas9 and chimeric antigen receptors (CARs), safe and efficient delivery of these technologies is paramount. With the corroboration of the improvements in both LV and retroviral vectors, the most comprehensive safety study to date analyzed results from 17 clinical vector lots, 375 manufactured T cell products, and 308 infused patients to develop RCL or retrovirus (RCL/R) and integration-driven expansion.[26] This analysis supports the safety profile of these vectors in this application and their continued use in oncology, infectious disease, autoimmunity, and inherited genetic disorders, as well as encourages adoption by other disease fields. This discussion does not represent all vectors that have been developed or have progressed to clinical trials. We identified the LV vectors from https://clinicaltrials.gov/ or the supporting literature and collected the sequences of the LV backbone. Then, we analyzed the molecular differences between the vectors and the sources of the HIV-1 sequence. Important follow-up studies will include side-by-side comparisons of other regulatory elements in the LV vectors—such as the promoter or wPRE—in optimizing viral production, transduction, and transgene expression.

If optimizations produce a vector series with significantly better characteristics, the field might move toward standardized vectors that provide a "plug-and-play" system. Such standardization could expedite processing, evaluation, and approval by regulatory bodies and also have implications for the financial realities of good manufacturing practice (GMP) manufacturing. This move would still allow the development of multiple standardized vectors based on target population cell type, which would have different safety requirements and different therapeutic rationales. Although many approaches, including both the WAS and SCID-X1 trial, utilize gene transfer into stem cell populations, others, including most CAR approaches, transduce differentiated lymphocytes. The degree of cell-lineage differentiation has been shown to affect leukemia penetrance after insertional oncogenesis.[85] Stem cells, of necessity, have active signaling pathways in growth and self-renewal; therefore, they are more susceptible to oncogenic transformation. In contrast, mature cells have reduced these self-renewal pathways, which must be reactivated to facilitate unrestrained growth. These differences have implications for choice of the vector backbone, internal promoter, or other regulatory elements. How much transgene expression is necessary for therapeutic benefit and whether a particular cell type is suited to safely host such a enhancer/promoter should be explored to guide backbone choice and target cell population. This report, and its analysis of the different HIV-derived *cis-acting* elements, can be used to guide future research comparing the functional differences between LV vectors.

Importantly, variations between the vectors identified in this review may be regarded as regions of potential optimization or

standardization. Most HIV-based LV vectors used currently are 3rd-generation vectors with a heterologous external promoter regulating expression of the genomic RNA from R/U3. The 5′ LTR and Ψ element between all vectors have a remarkable degree of similarity, particularly at the Ψ region, likely reflecting the reliance of this region on its complex secondary structure for proper function. Even though the commercial vectors showed nearly identical sequences surrounding the Ψ region, the clinical vectors showed more variation, such as the length of *gag* sequence. Perhaps the most important difference in the vectors is the termination of the *gag* ORF. In most vectors, this is accomplished with a 2-nt frameshift insertion, which causes a downstream termination signal after 21 aa. In pTNS9, a T → C substitution mutates the ATG start codon. Since this sequence also participates in the Ψ hairpin formation, it would be interesting to study the effects on binding to Gag NC. The pHIV-7 and pTYF vectors do not have the frameshift mutation; therefore, 271 aa and 638 aa proteins, respectively, could potentially be translated. Any basal transcription of the provirus could express this heterologous, and potentially immunogenic, protein. The function of the ISRE sequence in some LV is unclear. Understanding whether ISRE-driven expression can occur during cell activation may be considered where tight regulation is required. How these differences in 5′ UTR affect packaging would be important to understanding the optimal packaging signal.

Within the cPPT/CTS region, the cPPT showed remarkably high similarity across the various LV; however, the CTS flap region and particularly the length of sequence included revealed a wide degree of diversity, even in the commercial vectors. Since this component was the most recent to be described, it may be that this element was incorporated into LV at multiple times. Of the indels with potential significance, it would be interesting to determine if the single nucleotide deletion seen in three of the CTS sequences affects the efficiency of chain termination. Furthermore, the inclusion of an additional sequence before, but mostly following, the CTS is of unclear function. If nuclear import and viral titers are shown to be unaffected by the presence of these sequences, general size considerations would support their removal.

The RRE is another region where function is reliant on secondary structure. The RNA-folding predictions suggest that substitutions in the critically important stem-loop regions of the minimal RRE, including the initial binding site stem IIB, may have significant effects on the folding structure. How these changes affect Rev binding and oligomerization will require additional studies. Based on the structural flexibility requirement, both for the initial binding event and the recruitment of additional Rev molecules, hypothetically, loss of a stem-loop bulge with the 55 C → G (HXB2 7823) might reduce flexibility and subsequent oligomerization, whereas 53 GCC → CGG (HXB2 7821) might add flexibility to the structure. The substitutions in three vectors occur in the nucleic acid sequence for the caveolin-1 binding site[86] and would not be expected to have a role in the vector. Some vectors extend the RRE sequence to the SA site for the tat/rev genes. With a SD in the 5′ UTR and the SA after RRE, the nuclear export function of Rev is critical for production of the LV vector. Length, again, is the final consideration in this region. Inclusion of the tat/rev SA site from *env* adds about 400 bp of sequence to the vector. The rationale for beginning the RRE at the 3′ end of the gp120 V5 loop in the majority of vectors is unclear, and benefits of including the gp120 or gp41 sequence could be considered in future studies.

Lastly, although the 3′ LTR region and its complex arrangement of regulatory elements are important for viral replication, the impact of various deletions in this region on the LV production process remains undefined, especially when using heterologous internal promoters to drive transgene expression. For safety considerations, vectors with the largest deletions should be favored; however, different cell types and differentiation states may regulate this concern. In cell types less susceptible to oncogenic transformations, maintaining more of the LTR could benefit expression or reverse transcription, depending on the desired therapeutic effect and whether safety can be confirmed. In cell types where epigenetic regulators could be useful, adding insulators or locus control regions might provide position-independent expression or block effects on neighboring genes. Of the elements discussed, the 3′ LTR showed a particularly wide range of phylogenetic branching within only a few select reference sequences. During the viral life cycle, the U3 region is largely responsible for viral RNA expression. Therefore, it would be interesting to compare viral replication of the different ΔU3 in the SIN LTR with the reference viruses in culture or in disease progression in patients. Higher steady-state viral loads or greater viral replication could correspond to a more active 3′ LTR, and their adaptation to gene-therapy vectors could require more extensive deletions.

Overall, this work shows a high degree of similarity between LV being used in gene therapy. All vectors include the same *cis-acting* regulatory elements to achieve a functional LV backbone. However, phylogenetic and forensic analysis suggests different sources for many of these vectors, different ranges in the deleted sequences, and different polymorphisms. A detailed analysis of these regions reveals potentially impactful polymorphisms, although the function remains to be determined. This review could be used in future studies focused on optimizing LV as a basis to begin element comparisons, and such work may help lead the field toward development of standardized vectors to streamline the process of developing delivery systems.

MATERIALS AND METHODS

Vectors

Vectors included the following: pLenti-III (Applied Biological Materials; cat # LV587);[87] pLentiCRISPR v.1 (Addgene; cat #52963);[88] p156RRLsinppt (Addgene; cat #42795);[89] pFUGW (Addgene; cat #14883);[90] pFUG (Addgene; cat #14882);[90] pHAGE (Addgene; cat #46793);[91] pHRsin (Addgene; cat #12265);[92] pLenti (AMP) (Addgene; cat #61422);[93] pLK0.1 (Addgene; cat #10878);[94] pLL3.7 m (Addgene; cat #89362);[95] Puro.cre (Addgene; cat #17408);[96] pRSIEG (Cellecta; SVSHU6EG-L); pLenti7.3 (Thermo Fisher Scientific; cat #V53406); pLenti (OriGene; cat #PS100109); pSF_Lenti (Sigma; cat

#OGS269); pLV-GFPSpark (Sinobiological; cat #LVCV-01); pLVX (Takara; cat #632164); and pLenti (Vigene; cat #P100020).

Vector analysis

Homologies between the HIV-1 genome and LV vector backbones used in both clinical and commercial applications were aligned with NCBI nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and plotted with the SnapGene sequence-alignment function (GSL Biotech). The HIV-1 reference genome HXB2 (GenBank: K03455.1) was divided into four main functional regions of interest: LTR and Ψ region (HXB2 1−1,514 bp), cPPT/CTS region (HXB2 4,554−5,099 bp), RRE region (HXB2 7,104−8,479 bp), and 3′ LTR (HXB2 8,896−9,717 bp). Each LV vector backbone sequence was aligned to each region to identify sequence homologies (Supplemental information). In the SnapGene figures, the filled-in red lines represent sequence alignment, whereas hollow regions represent mismatched or missing base pairs from the reference genome. The small arrowhead on the top of the vector represents an insertion.

RNA folding predictions

Analysis was done at RNAfold (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi). Only the minimal RRE portion was entered with the fold algorithm set to minimum free energy and partition function and to avoid isolated base pairs. RNA energy parameters were chosen according to the Turner model in 2004.[97] Centroid plain structure drawing viewed in Forna was taken for figures.

Phylogeny of vectors to reference sequences

Reference genomes included the following: the standard reference genome HXB2 (K03455.1), HIV-1 complete RNA genome (AF033819.3), pIIIB (EU541617.1), SF33 (AY352275.1), 89.6 (U39362.2), DH12 (AF069140.1), ACH320 (U34604.1), JRFL (U63632.1), and NL4-3 (M19921.2). The standard reference sequence HXB2 was entered as a query sequence in https://blast.ncbi.nlm.nih.gov/Blast.cgi. Reference sequences along with commercial and clinical vector sequences were entered in FASTA format as subject sequences. Alignments were optimized for highly similar sequences. BLAST was viewed as distance of tree results.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2021.03.018.

ACKNOWLEDGMENTS

We thank Dr. Ken Cornetta (Indiana University School of Medicine) for helpful discussions and reading the manuscript, Robin Rodriguez (TNPRC) for help with images, and Nancy Busija for thoroughly editing the manuscript. Research reported in this publication was supported by the National Center for Research Resources and Office of Research Infrastructure Programs (ORIP) at the National Institutes of Health (NIH) through grant P51 OD011104 (TNPRC); National Institute of Allergy and Infectious Diseases of the NIH under award numbers NIAID AL110158 (S.E.B.) and NIAID F30AI150452 (N.M.J.); Department of Pharmacology (S.E.B.); and Tulane University School of Medicine (S.E.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, S.E.B.; methodology, N.M.J., A.F.A., T.N.M., and S.E.B.; investigation, N.M.J., A.F.A., T.N.M., J.M.E., and T.A.S.; analysis, N.M.J., A.F.A., T.N.M., and S.E.B.; curation, A.F.A., T.N.M., J.M.E., and T.A.S.; writing − original draft, N.M.J. and S.E.B.; writing − review & editing, N.M.J. and S.E.B.; visualization, N.M.J., A.F.A., T.N.M., J.M.E., T.A.S., and S.E.B.; funding acquisition, N.M.J. and S.E.B.; supervision and administration, S.E.B.

DECLARATION OF INTERESTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting agencies. The authors declare no competing interests.

REFERENCES

1. Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415–419.

2. Sinn, P.L., Sauter, S.L., and McCray, P.B., Jr. (2005). Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors–design, biosafety, and production. Gene Ther. *12*, 1089–1098.

3. Cattoglio, C., Pellin, D., Rizzi, E., Maruggi, G., Corti, G., Miselli, F., Sartori, D., Guffanti, A., Di Serio, C., Ambrosi, A., et al. (2010). High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood *116*, 5507–5517.

4. Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. *118*, 3132–3142.

5. Engelman, A., and Cherepanov, P. (2008). The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. PLoS Pathog. *4*, e1000046.

6. Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749–1751.

7. Boztug, K., Schmidt, M., Schwarzer, A., Banerjee, P.P., Díez, I.A., Dewey, R.A., Böhm, M., Nowrouzi, A., Ball, C.R., Glimm, H., et al. (2010). Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N. Engl. J. Med. *363*, 1918–1927.

8. Braun, C.J., Witzel, M., Paruzynski, A., Boztug, K., von Kalle, C., Schmidt, M., and Klein, C. (2014). Gene therapy for Wiskott-Aldrich Syndrome-Long-term reconstitution and clinical benefits, but increased risk for leukemogenesis. Rare Dis. *2*, e947749.

9. Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. *2*, E234.

10. Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science *341*, 1233151.

11. Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L., and Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J. Virol. *72*, 9873–9880.

12. Iwakuma, T., Cui, Y., and Chang, L.J. (1999). Self-inactivating lentiviral vectors with U3 and U5 modifications. Virology *261*, 120–132.

13. 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.

14. Cesana, D., Sgualdino, J., Rudilosso, L., Merella, S., Naldini, L., and Montini, E. (2012). Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. J. Clin. Invest. *122*, 1667–1676.

15. Moiani, A., Paleari, Y., Sartori, D., Mezzadra, R., Miccio, A., Cattoglio, C., Cocchiarella, F., Lidonnici, M.R., Ferrari, G., and Mavilio, F. (2012). Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J. Clin. Invest. *122*, 1653–1666.

16. Anthony-Gonda, K., Bardhi, A., Ray, A., Flerin, N., Li, M., Chen, W., Ochsenbauer, C., Kappes, J.C., Krueger, W., Worden, A., et al. (2019). Multispecific anti-HIV duoCAR-T cells display broad in vitro antiviral activity and potent in vivo elimination of HIV-infected cells in a humanized mouse model. Sci. Transl. Med. *11*, eaav5685.

17. Hacein-Bey Abina, S., Gaspar, H.B., Blondeau, J., Caccavelli, L., Charrier, S., Buckland, K., Picard, C., Six, E., Himoudi, N., Gilmour, K., et al. (2015). Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome. JAMA *313*, 1550–1563.

18. Kohn, D.B., Booth, C., Kang, E.M., Pai, S.Y., Shaw, K.L., Santilli, G., Armant, M., Buckland, K.F., Choi, U., De Ravin, S.S., et al.; Net4CGD consortium (2020). Lentiviral gene therapy for X-linked chronic granulomatous disease. Nat. Med. *26*, 200–206.

19. Ferrua, F., Cicalese, M.P., Galimberti, S., Giannelli, S., Dionisio, F., Barzaghi, F., Migliavacca, M., Bernardo, M.E., Calbi, V., Assanelli, A.A., et al. (2019). Lentiviral haemopoietic stem/progenitor cell gene therapy for treatment of Wiskott-Aldrich syndrome: interim results of a non-randomised, open-label, phase 1/2 clinical study. Lancet Haematol. *6*, e239–e253.

20. Magrin, E., Miccio, A., and Cavazzana, M. (2019). Lentiviral and genome-editing strategies for the treatment of β-hemoglobinopathies. Blood *134*, 1203–1213.

21. Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. Nature *467*, 318–322.

22. Mamcarz, E., Zhou, S., Lockey, T., Abdelsamed, H., Cross, S.J., Kang, G., Ma, Z., Condori, J., Dowdy, J., Triplett, B., et al. (2019). Lentiviral Gene Therapy Combined with Low-Dose Busulfan in Infants with SCID-X1. N. Engl. J. Med. *380*, 1525–1534.

23. Cavazzana, M., Six, E., Lagresle-Peyrou, C., André-Schmutz, I., and Hacein-Bey-Abina, S. (2016). Gene Therapy for X-Linked Severe Combined Immunodeficiency: Where Do We Stand? Hum. Gene Ther. *27*, 108–116.

24. Lwin, S.M., Syed, F., Di, W.-L., Kadiyirire, T., Liu, L., Guy, A., Petrova, A., Abdul-Wahab, A., Reid, F., Phillips, R., et al. (2019). Safety and early efficacy outcomes for lentiviral fibroblast gene therapy in recessive dystrophic epidermolysis bullosa. JCI Insight *4*, e126243.

25. McGarrity, G.J., Hoyah, G., Winemiller, A., Andre, K., Stein, D., Blick, G., Greenberg, R.N., Kinder, C., Zolopa, A., Binder-Scholl, G., et al. (2013). Patient monitoring and follow-up in lentiviral clinical trials. J. Gene Med. *15*, 78–82.

26. Marcucci, K.T., Jadlowsky, J.K., Hwang, W.T., Suhoski-Davis, M., Gonzalez, V.E., Kulikovskaya, I., Gupta, M., Lacey, S.F., Plesa, G., Chew, A., et al. (2018). Retroviral and Lentiviral Safety Analysis of Gene-Modified T Cell Products and Infused HIV and Oncology Patients. Mol. Ther. *26*, 269–279.

27. Olbrich, H., Slabik, C., and Stripecke, R. (2017). Reconstructing the immune system with lentiviral vectors. Virus Genes *53*, 723–732.

28. Ginn, S.L., Amaya, A.K., Alexander, I.E., Edelstein, M., and Abedi, M.R. (2018). Gene therapy clinical trials worldwide to 2017: An update. J. Gene Med. *20*, e3015.

29. Coquerelle, S., Ghardallou, M., Rais, S., Taupin, P., Touzot, F., Boquet, L., Blanche, S., Benaouadi, S., Brice, T., Tuchmann-Durand, C., et al. (2019). Innovative Curative Treatment of Beta Thalassemia: Cost-Efficacy Analysis of Gene Therapy Versus Allogenic Hematopoietic Stem-Cell Transplantation. Hum. Gene Ther. *30*, 753–761.

30. Braun, S.E., Wong, F.E., Connole, M., Qiu, G., Lee, L., Gillis, J., Lu, X., Humeau, L., Slepushkin, V., Binder, G.K., et al. (2005). Inhibition of simian/human immunodeficiency virus replication in CD4+ T cells derived from lentiviral-transduced CD34+ hematopoietic cells. Mol. Ther. *12*, 1157–1167.

31. Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., et al. (2009). Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science *326*, 818–823.

32. Cesani, M., Plati, T., Lorioli, L., Benedicenti, F., Redaelli, D., Dionisio, F., Biasco, L., Montini, E., Naldini, L., and Biffi, A. (2015). Shedding of clinical-grade lentiviral vectors is not detected in a gene therapy setting. Gene Ther. *22*, 496–502.

33. 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.

34. Sadelain, M. (2017). CD19 CAR T Cells. Cell *171*, 1471.

35. Hu, W.-S., and Hughes, S.H. (2012). HIV-1 reverse transcription. Cold Spring Harb. Perspect. Med. *2*, a006882.

36. Naldini, L., Blömer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263–267.

37. Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat. Biotechnol. *15*, 871–875.

38. Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471.

39. 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.

40. Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., et al. (2015). RNA structure. Structure of the HIV-1 RNA packaging signal. Science *348*, 917–921.

41. Heng, X., Kharytonchyk, S., Garcia, E.L., Lu, K., Divakaruni, S.S., LaCotti, C., Edme, K., Telesnitsky, A., and Summers, M.F. (2012). Identification of a minimal region of the HIV-1 5′-leader required for RNA dimerization, NC binding, and packaging. J. Mol. Biol. *417*, 224–239.

42. Rein, A. (2019). RNA Packaging in HIV. Trends Microbiol. *27*, 715–723.

43. Kim, S.H., Jang, S.I., Park, C.Y., and You, J.C. (2009). Investigation of requirements for efficient gene delivery using the HIV-1 based lentiviral transduction system. Biochem. Biophys. Res. Commun. *383*, 192–197.

44. Kim, S.H., Jun, H.J., Jang, S.I., and You, J.C. (2012). The determination of importance of sequences neighboring the Psi sequence in lentiviral vector transduction and packaging efficiency. PLoS ONE *7*, e50148.

45. Liu, Y., Nikolaitchik, O.A., Rahman, S.A., Chen, J., Pathak, V.K., and Hu, W.S. (2017). HIV-1 Sequence Necessary and Sufficient to Package Non-viral RNAs into HIV-1 Particles. J. Mol. Biol. *429*, 2542–2555.

46. Kharytonchyk, S., Brown, J.D., Stilger, K., Yasin, S., Iyer, A.S., Collins, J., Summers, M.F., and Telesnitsky, A. (2018). Influence of gag and RRE Sequences on HIV-1 RNA Packaging Signal Structure and Function. J. Mol. Biol. *430*, 2066–2079.

47. NCBI. Human immunodeficiency virus 1 (ID: 10319), Genome. https://www.ncbi.nlm.nih.gov/genome/?term=Human+immunodeficiency+virus+1.

48. Battistini, A., Marsili, G., Sgarbanti, M., Ensoli, B., and Hiscott, J. (2002). IRF regulation of HIV-1 long terminal repeat activity. J. Interferon Cytokine Res. *22*, 27–37.

49. Bernacchi, S., Abd El-Wahab, E.W., Dubois, N., Hijnen, M., Smyth, R.P., Mak, J., Marquet, R., and Paillart, J.C. (2017). HIV-1 $Pr55^{Gag}$ binds genomic and spliced RNAs with different affinity and stoichiometry. RNA Biol. *14*, 90–103.

50. Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J., and Kjems, J. (2004). RNA interactions in the 5′ region of the HIV-1 genome. J. Mol. Biol. *336*, 369–379.

51. Laughrea, M., Shen, N., Jetté, L., and Wainberg, M.A. (1999). Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. Biochemistry *38*, 226–234.

52. Weinberg, J.B., Matthews, T.J., Cullen, B.R., and Malim, M.H. (1991). Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. J. Exp. Med. *174*, 1477–1482.

53. Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H., and Clavel, F. (1994). HIV-1 reverse transcription. A termination step at the center of the genome. J. Mol. Biol. *241*, 651–662.

54. Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L., and Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. Cell *101*, 173–185.

55. Charneau, P., Alizon, M., and Clavel, F. (1992). A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. J. Virol. *66*, 2814–2820.

56. Yamashita, M., and Emerman, M. (2005). The cell cycle independence of HIV infections is not determined by known karyophilic viral elements. PLoS Pathog. *1*, e18.

57. Dvorin, J.D., Bell, P., Maul, G.G., Yamashita, M., Emerman, M., and Malim, M.H. (2002). Reassessment of the roles of integrase and the central DNA flap in human immunodeficiency virus type 1 nuclear import. J. Virol. *76*, 12087–12096.

58. Iglesias, C., Ringeard, M., Di Nunzio, F., Fernandez, J., Gaudin, R., Souque, P., Charneau, P., and Arhel, N. (2011). Residual HIV-1 DNA Flap-independent nuclear import of cPPT/CTS double mutant viruses does not support spreading infection. Retrovirology *8*, 92.

59. Manganini, M., Serafini, M., Bambacioni, F., Casati, C., Erba, E., Follenzi, A., Naldini, L., Bernasconi, S., Gaipa, G., Rambaldi, A., et al. (2002). A human immunodeficiency virus type 1 pol gene-derived sequence (cPPT/CTS) increases the efficiency of transduction of human nondividing monocytes and T lymphocytes by lentiviral vectors. Hum. Gene Ther. *13*, 1793–1807.

60. De Rijck, J., and Debyser, Z. (2006). The central DNA flap of the human immunodeficiency virus type 1 is important for viral replication. Biochem. Biophys. Res. Commun. *349*, 1100–1110.

61. Follenzi, A., Ailles, L.E., Bakovic, S., Geuna, M., and Naldini, L. (2000). Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. Nat. Genet. *25*, 217–222.

62. Sirven, A., Pflumio, F., Zennou, V., Titeux, M., Vainchenker, W., Coulombel, L., Dubart-Kupperschmitt, A., and Charneau, P. (2000). The human immunodeficiency virus type-1 central DNA flap is a crucial determinant for lentiviral vector nuclear import and gene transduction of human hematopoietic stem cells. Blood *96*, 4103–4110.

63. Mautino, M.R., Ramsey, W.J., Reiser, J., and Morgan, R.A. (2000). Modified human immunodeficiency virus-based lentiviral vectors display decreased sensitivity to trans-dominant Rev. Hum. Gene Ther. *11*, 895–908.

64. Dardalhon, V., Herpers, B., Noraz, N., Pflumio, F., Guetard, D., Leveau, C., Dubart-Kupperschmitt, A., Charneau, P., and Taylor, N. (2001). Lentivirus-mediated gene transfer in primary T cells is enhanced by a central DNA flap. Gene Ther. *8*, 190–198.

65. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature *338*, 254–257.

66. Kjems, J., Brown, M., Chang, D.D., and Sharp, P.A. (1991). Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. Proc. Natl. Acad. Sci. USA *88*, 683–687.

67. Fernandes, J., Jayaraman, B., and Frankel, A. (2012). The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. RNA Biol. *9*, 6–11.

68. Mann, D.A., Mikaélian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Butler, P.J.G., Gait, M.J., and Karn, J. (1994). A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. J. Mol. Biol. *241*, 193–207.

69. D'Costa, J., Brown, H.M., Kundra, P., Davis-Warren, A., and Arya, S.K. (2001). Human immunodeficiency virus type 2 lentiviral vectors: packaging signal and splice donor in expression and encapsidation. J. Gen. Virol. *82*, 425–434.

70. Grewe, B., Ehrhardt, K., Hoffmann, B., Blissenbach, M., Brandt, S., and Uberla, K. (2012). The HIV-1 Rev protein enhances encapsidation of unspliced and spliced, RRE-containing lentiviral vector RNA. PLoS ONE *7*, e48688.

71. Zemmel, R.W., Kelley, A.C., Karn, J., and Butler, P.J.G. (1996). Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. J. Mol. Biol. *258*, 763–777.

72. Chu, C.-C., Plangger, R., Kreutz, C., and Al-Hashimi, H.M. (2019). Dynamic ensemble of HIV-1 RRE stem IIB reveals non-native conformations that disrupt the Rev-binding site. Nucleic Acids Res. *47*, 7105–7117.

73. Jackson, P.E., Tebit, D.M., Rekosh, D., and Hammarskjold, M.L. (2016). Rev-RRE functional activity differs substantially among primary HIV-1 isolates. AIDS Res. Hum. Retroviruses *32*, 923–934.

74. Sloan, E.A., Kearney, M.F., Gray, L.R., Anastos, K., Daar, E.S., Margolick, J., Maldarelli, F., Hammarskjold, M.-L., and Rekosh, D. (2013). Limited nucleotide changes in the Rev response element (RRE) during HIV-1 infection alter overall Rev-RRE activity and Rev multimerization. J. Virol. *87*, 11173–11186.

75. Legiewicz, M., Badorrek, C.S., Turner, K.B., Fabris, D., Hamm, T.E., Rekosh, D., Hammarskjöld, M.L., and Le Grice, S.F.J. (2008). Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. Proc. Natl. Acad. Sci. USA *105*, 14365–14370.

76. Krebs, F.C., Hogan, T.H., Quiterio, S., Gartner, S., and Wigdahl, B. (2001). Lentiviral LTR-directed Expression, Sequence Variation, and Disease Pathogenesis. HIV Seq. Compend. *2001*, 29–70, https://www.hiv.lanl.gov/content/sequence/HIV/COMPENDIUM/2001/partI/Wigdahl.pdf.

77. Pereira, L.A., Bentley, K., Peeters, A., Churchill, M.J., and Deacon, N.J. (2000). A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic Acids Res. *28*, 663–668.

78. 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.

79. Rodenburg, C.M., Li, Y., Trask, S.A., Chen, Y., Decker, J., Robertson, D.L., Kalish, M.L., Shaw, G.M., Allen, S., Hahn, B.H., and Gao, F.; UNAIDS and NIAID Networks for HIV Isolation and Characterization (2001). Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. AIDS Res. Hum. Retroviruses *17*, 161–168.

80. Elsner, C., and Bohne, J. (2017). The retroviral vector family: something for everyone. Virus Genes *53*, 714–722.

81. Kraunus, J., Schaumann, D.H.S., Meyer, J., Modlich, U., Fehse, B., Brandenburg, G., von Laer, D., Klump, H., Schambach, A., Bohne, J., and Baum, C. (2004). Self-inactivating retroviral vectors with improved RNA processing. Gene Ther. *11*, 1568–1578.

82. Schambach, A., Bohne, J., Chandra, S., Will, E., Margison, G.P., Williams, D.A., and Baum, C. (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. Mol. Ther. *13*, 391–400.

83. Schambach, A., Mueller, D., Galla, M., Verstegen, M.M.A., Wagemaker, G., Loew, R., Baum, C., and Bohne, J. (2006). Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. Gene Ther. *13*, 1524–1533.

84. Orchard, K., Perkins, N., Chapman, C., Harris, J., Emery, V., Goodwin, G., Latchman, D., and Collins, M. (1990). A novel T-cell protein which recognizes a palindromic sequence in the negative regulatory element of the human immunodeficiency virus long terminal repeat. J. Virol. *64*, 3234–3239.

85. Newrzela, S., Cornils, K., Li, Z., Baum, C., Brugman, M.H., Hartmann, M., Meyer, J., Hartmann, S., Hansmann, M.-L., Fehse, B., and von Laer, D. (2008). Resistance of mature T cells to oncogene transformation. Blood *112*, 2278–2286.

86. Hovanessian, A.G., Briand, J.P., Said, E.A., Svab, J., Ferris, S., Dali, H., Muller, S., Desgranges, C., and Krust, B. (2004). The caveolin-1 binding domain of HIV-1 glycoprotein gp41 is an efficient B cell epitope vaccine candidate against virus infection. Immunity *21*, 617–627.

87. Wang, R., Li, Y., Tsung, A., Huang, H., Du, Q., Yang, M., Deng, M., Xiong, S., Wang, X., Zhang, L., et al. (2018). iNOS promotes $CD24^{+}CD133^{+}$ liver cancer stem cell phenotype through a TACE/ADAM17-dependent Notch signaling pathway. Proc. Natl. Acad. Sci. USA *115*, E10127–E10136.

88. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

89. Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184.

90. Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868–872.

91. Wilson, A.A., Murphy, G.J., Hamakawa, H., Kwok, L.W., Srinivasan, S., Hovav, A.H., Mulligan, R.C., Amar, S., Suki, B., and Kotton, D.N. (2010). Amelioration of emphysema in mice through lentiviral transduction of long-lived pulmonary alveolar macrophages. J. Clin. Invest. *120*, 379–389.

92. Cudré-Mauroux, C., Occhiodoro, T., König, S., Salmon, P., Bernheim, L., and Trono, D. (2003). Lentivector-mediated transfer of Bmi-1 and telomerase in muscle satellite cells yields a duchenne myoblast cell line with long-term genotypic and phenotypic stability. Hum. Gene Ther. *14*, 1525–1533.

93. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

94. Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283–1298.

95. Zhou, X.X., Fan, L.Z., Li, P., Shen, K., and Lin, M.Z. (2017). Optical control of cell signaling by single-chain photoswitchable kinases. Science *355*, 836–842.

96. Kumar, M.S., Erkeland, S.J., Pester, R.E., Chen, C.Y., Ebert, M.S., Sharp, P.A., and Jacks, T. (2008). Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc. Natl. Acad. Sci. USA *105*, 3903–3908.

97. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA *101*, 7287–7292.

OMTM, Volume *21*

Supplemental information

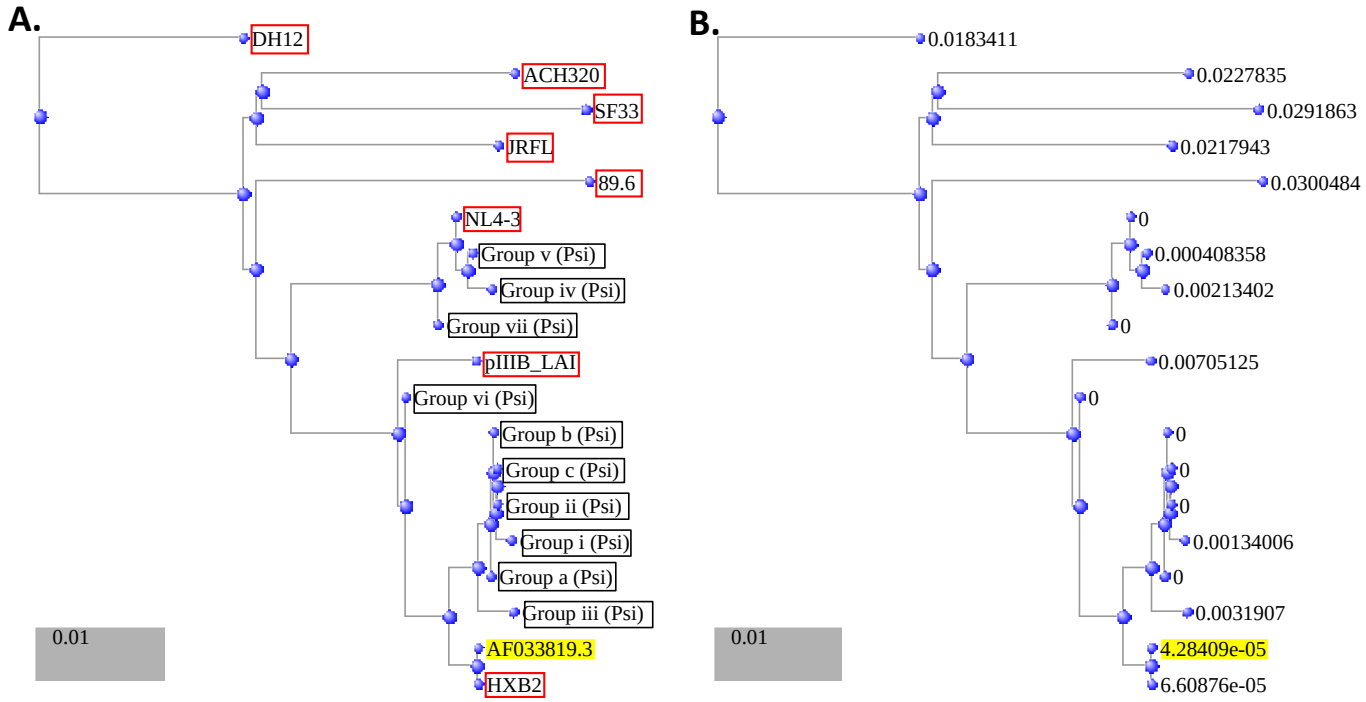
HIV-based lentiviral vectors: Origin and sequence differences

Nathan M. Johnson, Anna Francesca Alvarado, Trey N. Moffatt, Joshua M. Edavettal, Tarun A. Swaminathan, and Stephen E. Braun

Clincal Trial	Backbone	Source	PMID:	PMID:	Comments
NCT01560182, NCT01560182, NCT01515462, NCT00991224, NCT02247843, NCT01381003 – 2012, NCT01545323, NCT02247843, NCT02234934, NCT01380990, NCT02757911 – 2016, NCT02453477, NCT03488394, NCT01560182	pCCL, pCCLsin	https://www.addgene.org/81071/		25588820	
NCT02999984 - 2016	EFS - (LV EFS ADA)	pCCL	24256635	22434141	
NCT01545323	pEF-DEST51-LEKTI	pCCL	24329107		
NCT01515462, NCT01029366 - 2009	pRRL, pRRLsin	https://www.addgene.org/12252/	23845947, 9765382	19384291	pRRL came from Naldini lab (JVI 98). R. Zufferey from Trono lab inserted the SIN in 98 and then WPRE in 99. Zennou et al finally inserted the cPPT.
NCT03585764	pCLPS	pRRL with cPPT from NL4-3	19211796	12816995	"pCLPS is a derived from pRRL.sin.wPRE, a 3rd generation LV vector (34, 35) in which the cPPT of HIV-1 (NL4-3) (36, 37) was inserted im-mediately upstream of the CMV promoter. "
NCT03645486, NCT03727555, NCT03645460, NCT03217617	TYF, pTYF	https://www.addgene.org/19983/	10441560		
NCT00569985	rHIV7-shI-TAR-CCR5RZ	John Rossi	20555022		
NCT02378922 - 2015	LVsh5/C46 (Cal-1)	NA	closed		
NCT00295477, NCT02135406, NCT00295477	VRX496 (or pN1cptASenv?)	VRX494 as surrogote for VRX496.	17090675	16168713	
NCT01380990 - 2011	pWPXL (EFS-ADA LV CD34+)	https://www.addgene.org/12257/	(RV=16905365		
NCT03315078, NCT01306019	pCL20	Nienhuis AW	23075105		
NCT01639690	TNS9	Michel Sadelain (Sloan Kettering)	24429337		
	pCSCIGW	Ken Cornetta (NGTB)			
	pHRsin.trip	Richard C. Mulligan (Harvard Gene Therapy Institute)			
	pNL.ΔU3	Jakob Reiser	10849549		Includes cPPT

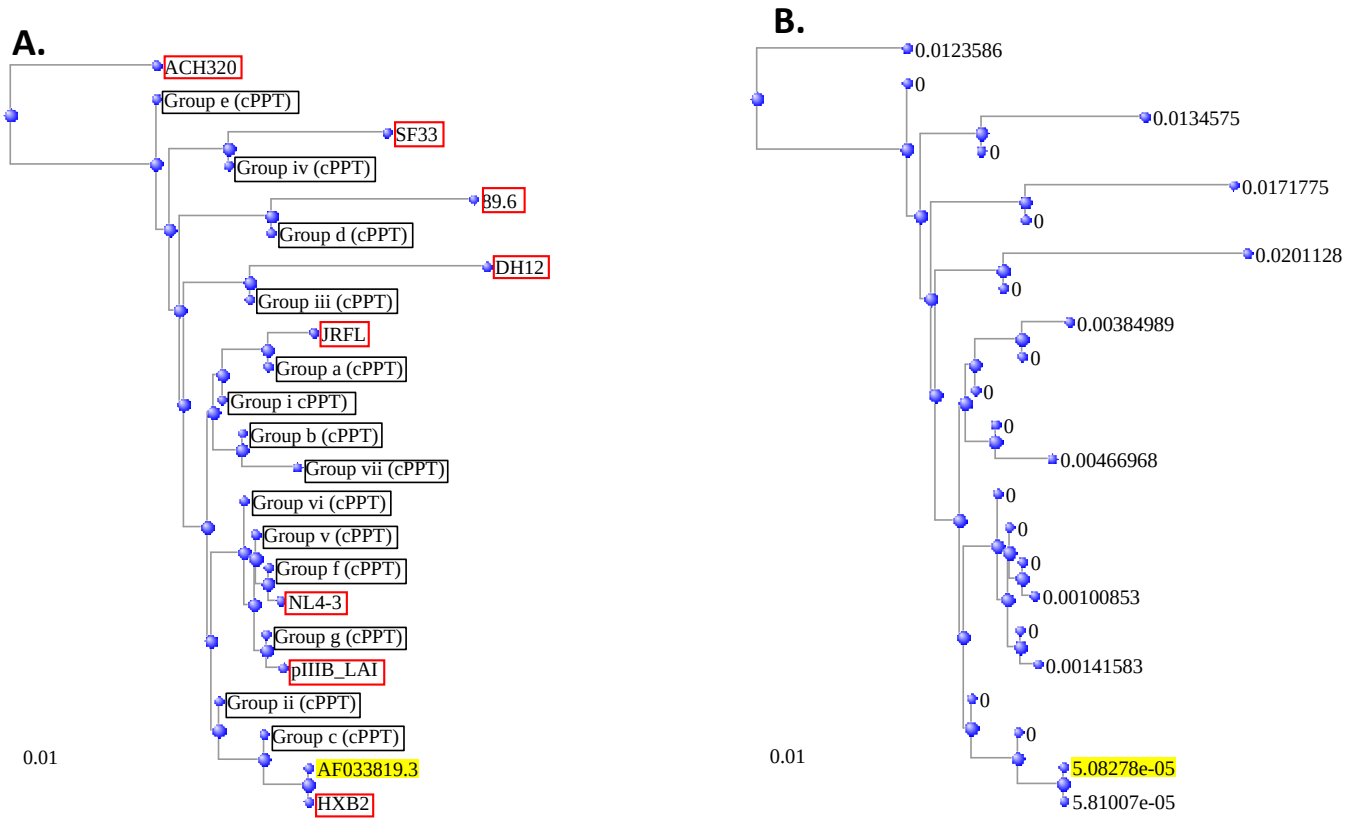
	pHIV (pSico)	https://www.addgene.org/21373/	18371425		
NCT02193191	BB305	NA	29669226	11743206	beta thalassemia
NCT02140554, NCT03207009	HPV569, pHPV569	NA			beta thalassemia with insulators
NCT02030847	pCDCAR1, scFv-CD28-41BB-CD3ζ	NA			https://www.creative-biolabs.com/car-t/pcdcar1-cd19-h-28bb%CE%B6-t-95.htm
NCT02051257	anti-CD19CAR-4-1BB-CD3zeta-EGFRt	NA	27118452		
NCT03156101, NCT03265106, NCT04204161, NCT03638193, NCT03638206, NCT03941626	anti-X CAR TCR-ζ/4-1BB (BinD19)	NA			https://clinicaltrials.gov/ct2/show/NCT03156101
NCT03157804 - 2017	PGK-FANCA.Wpre	NA	20001454		
NCT02559830	CG1711 hALD (MND-ALD)		19892975		http://science.sciencemag.org/content/sci/suppl/2009/11/05/326.5954.818.DC1/Cartier.SOM.pdf
	pGC-E1 vector with hHGF insert	NA	23484149		
NCT03454893, NCT04145037	AVR-RD-01, AVAR-RD-02	NA			http://www.avrobio.com/technology/
NCT03282656	LeGO-SFFV-shRNAmiR5 (BCL11A)	NA	33283990	30195795	DOI: 10.1056/NEJMoa2029392/suppl_file/nejmoa2029392_protocol.pdf
NCT03321123	pLTG1563	NA			https://clinicaltrials.gov/ct2/show/NCT03321123

Figure S1. LTR and Psi Phylogenetic Tree



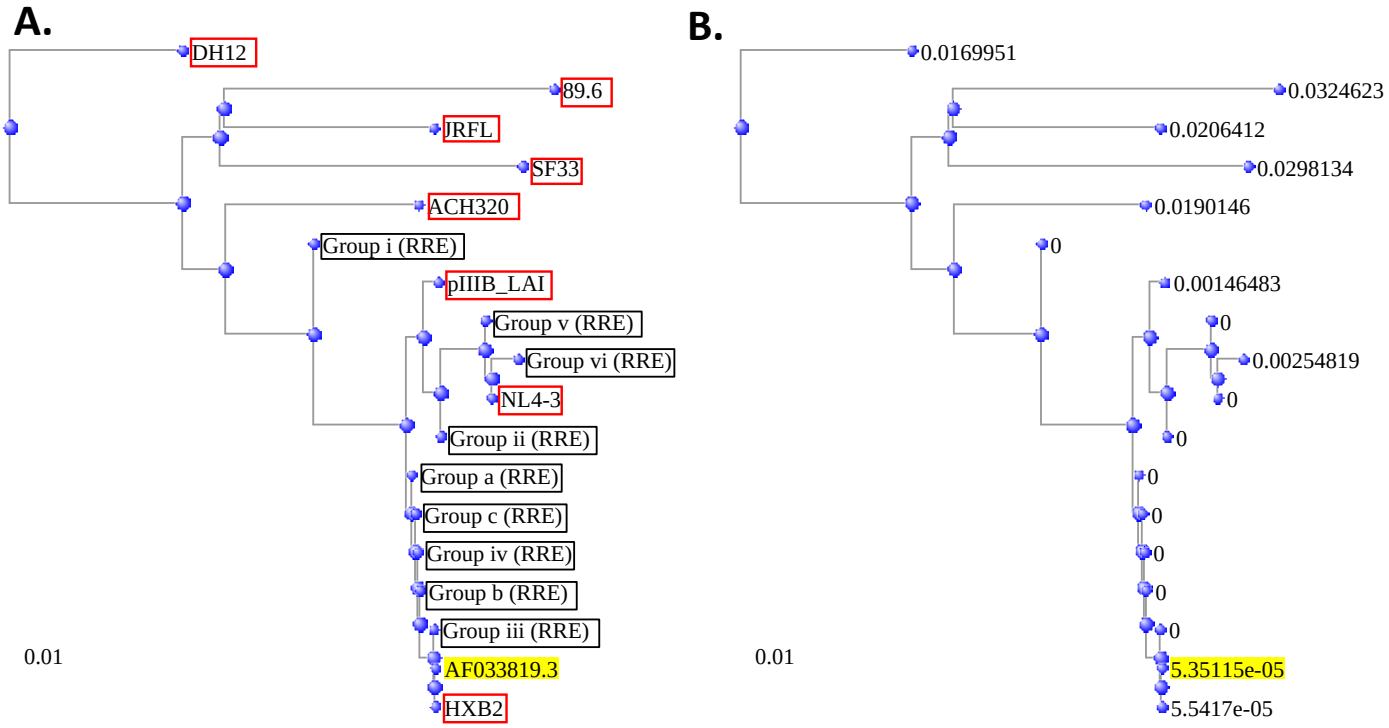
A. Red boxes indicate commonly used reference genomes, the reference HIV-1 sequence is highlighted in yellow. Groups correspond to those in Figure 2, with the Clinical vector Group i (pCCL and pRRL) Group ii (pCSCIGW, pHR-SIN-Trip, and pWPXL),Group iii (pTNS9) Group iv (pVRX496), Group v (pNL ΔU3), Group vi (pHIV-7) Group vii (pTYF) and Commercial vectors Group a (pLenti-III) Group b (pHAGE, pLL3.7m, Puro.Cre, pLenti (Origene), pLV-GFPSpark®, pLVX, and pLenti (Vigene)) Group c (p156RRLsinppt, pFUW, pFUGW, pHRsin, pLenti(AMP), pLentiCRISPR v1, pLKO.1, pRSIEG, pLenti7.3, and pSF Lenti) **B.** Distance to the node.

Figure S2. Central Poly-purine Tract Phylogenetic Tree



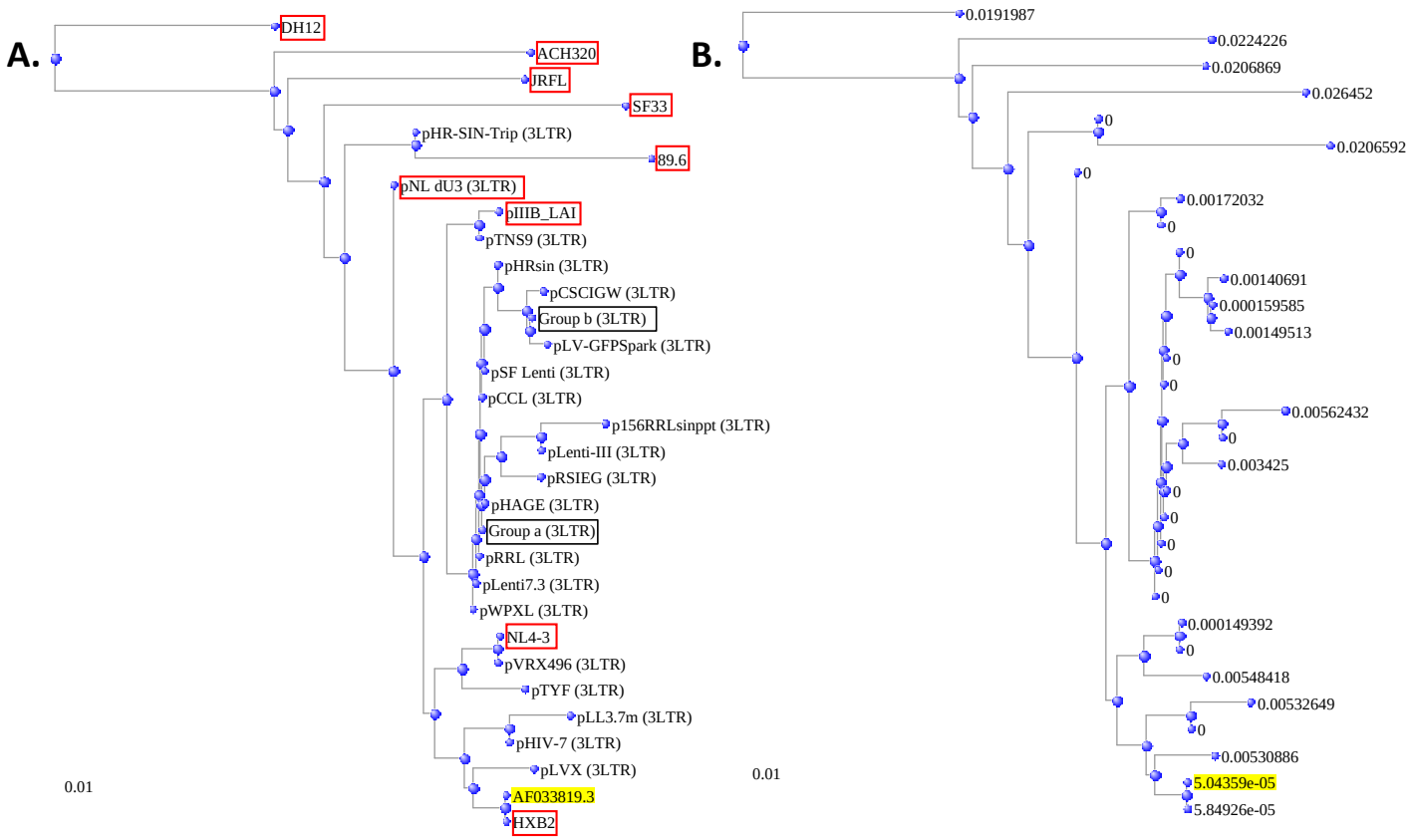
A. Red boxes indicate commonly used reference genomes, the reference HIV-1 sequence is highlighted in yellow. Groups correspond to those in Figure 3, with the Clinical vector Group i (pCCL and pRRL) Group ii (pCSCIGW, pWPXL, and pTNS9), Group iii (pNL ΔU3) Group iv (pTYF), Group v (pHR-SIN-Trip), Group vi (pHIV-7) Group vii (pVRX496) and Commercial vectors Group a (pRSIEG) Group b (pSF Lenti and p1556RRLsinppt) Group c (pLenti7.3) Group d (pLenti-III) Group e (pLentiCRISPR v1 and pLKO.1) Group f (pLVX) Group g (pFUGW, pLenti(AMP), pLV-GFPSpark®, pLenti (Vigene), pFUW, pHAGE, pLL3.7m, Puro.Cre, and pLenti (Origene) **B.** Distance to the node.

Figure S3. Rev Response Element Phylogenetic Tree



A. Red boxes indicate commonly used reference genomes, the reference HIV-1 sequence is highlighted in yellow. Groups correspond to those in Figure 4, with the Clinical vector Group i (pHIV-7) Group ii (pNL ΔU3) Group iii (pCCL and pRRL) Group iv (pCSCIGW, pHR-SIN-Trip, pWPXL, and pTNS9), Group v (pVRX496), Group vi (pTYF) and Commercial vectors Group a (pLentiCRISPR v1 and pLKO.1) Group b (pLenti-III) Group c (p156RRLsinppt, pFUW, pFUGW, pHAGE, pHRsin, pLenti(AMP), pLL3.7m, Puro.Cre, pRSIEG, pLenti7.3, pLenti (Origene), pSF Lenti, pLV-GFPSpark®, pLVX, and pLenti (Vigene)) **B.** Distance to the node.

Figure S4. 3'LTR Phylogenetic Tree



A. Red boxes indicate commonly used reference genomes, the reference HIV-1 sequence is highlighted in yellow. Groups correspond to those in Figure 5, with the Commercial vectors Group a (pLentiCRISPR v1 and pLKO.1) and Group b (pFUGW, pFUW, pLenti(AMP), pLenti (Origene), pLenti (Vigene), and Puro.Cre. **B.** Distance to the node.

LTR and Ψ

Job Title:LTR and ψ
Program: BLASTN

Query: HXB2 ID: lcl|Query_19948(dna) Length: 1511

Subject #1:Group iii (Psi) ID: lcl|Query_19950 Length: 699
Subject #2:Group i (Psi) ID: lcl|Query_19951 Length: 696
Subject #3:Group ii (Psi) ID: lcl|Query_19952 Length: 699
Subject #4:Group v (Psi) ID: lcl|Query_19953 Length: 794
Subject #5:Group vii (Psi) ID: lcl|Query_19954 Length: 1057
Subject #6:Group iv (Psi) ID: lcl|Query_19955 Length: 788
Subject #7:Group vi (Psi) ID: lcl|Query_19956 Length: 966
Subject #8:Group a (Psi) ID: lcl|Query_19957 Length: 690
Subject #9:Group b (Psi) ID: lcl|Query_19958 Length: 693
Subject #10:Group c (Psi) ID: lcl|Query_19959 Length: 699

Sequences producing significant alignments:

Description	Name	Taxid	Max Score	Total Score	Query cover	E Value	Per. Ident	Acc. Len	Accession
Group vii (Psi)		0	1786	1786	69%	0.0	97.16	1057	Query_19954
Group vi (Psi)		0	1779	1779	63%	0.0	99.90	966	Query_19956
Group v (Psi)		0	1317	1317	52%	0.0	96.60	794	Query_19953
Group iv (Psi)		0	1308	1308	52%	0.0	96.58	788	Query_19955
Group c (Psi)		0	1240	1240	46%	0.0	98.71	699	Query_19959
Group ii (Psi)		0	1240	1240	46%	0.0	98.71	699	Query_19952
Group b (Psi)		0	1234	1234	45%	0.0	98.99	693	Query_19958
Group a (Psi)		0	1234	1234	45%	0.0	98.99	690	Query_19957
Group i (Psi)		0	1225	1225	46%	0.0	98.43	696	Query_19951
Group iii (Psi)		0	1223	1223	46%	0.0	98.28	699	Query_19950

Alignments:

```
Query        455  GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGAC     604
Query_19954  1    ..........................................................................................CA..........................................................     150
Query_19956  1    ..........................................................................................................................................................     150
Query_19953  1    ..........................................................................................CA..........................................................     150
Query_19955  1    ..........................................................................................CA..........................................................     150
Query_19959  1    ..........................................................................................................................................................     150
Query_19952  1    ..........................................................................................................................................................     150
Query_19958  1    ..........................................................................................................................................................     150
Query_19957  1    ..........................................................................................................................................................     150
Query_19951  1    ..........................................................................................................................................................     150
Query_19950  1    ..........................................................................................................................................................     150

Query        605  CCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACCTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCC     754
Query_19954  151  ..............................................T.............TA..G.........A.....................................................     300
Query_19956  151  ..............................................T.....................................................................................     300
Query_19953  151  ..............................................T.............TA..G.........A.....................................................     300
Query_19955  151  ..............................................T.............TA..G.........A.....................................................     300
Query_19959  151  ..............................................T.....................................................................................     300
Query_19952  151  ..............................................T.....................................................................................     300
Query_19958  151  ..............................................T.....................................................................................     300
Query_19957  151  ..............................................T.....................................................................................     300
Query_19951  151  ...........................................................---.....................................................................     297
Query_19950  151  ..............................................T.....................................................................................     300
```

LTR and Ψ

```
Query         755  AAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGAT--CGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATTAAAACATATAGTATGGGCAAGCA    904
Query_19954   301  ..................................................G.....................--AA.....................................C.........C.....................    450
Query_19956   301  ..................................................................--.....................................................................    450
Query_19953   301  ..................................................G.....................--AA.....................................C.........C.....................    450
Query_19955   301  ..................................................G.................--..--.....................................C.........C.....................    448
Query_19959   301  ..................................................................CG.....................................................................    452
Query_19952   301  ..................................................................CG.....................................................................    452
Query_19958   301  ..................................................................CG.....................................................................    452
Query_19957   301  ..................................................................CG.....................................................................    452
Query_19951   298  ..................................................................CG.....................................................................    449
Query_19950   301  .................................C.............................T....A....CG.....................................................................    452

Query         905  GGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGC     1054
Query_19954   451  ................................T.....G.....................................................................A.....GT.............     600
Query_19956   451  ..................................................................................................................     600
Query_19953   451  ................................T.....G.....................................................................A.....GT.............     600
Query_19955   449  ................................T.....G.....................................................................A.....GT.............     598
Query_19959   453  ..................................................................................................................     602
Query_19952   453  ..................................................................................................................     602
Query_19958   453  ..................................................................................................................     602
Query_19957   453  ..................................................................................................................     602
Query_19951   450  ..................................................................................................................     599
Query_19950   453  ..................................................................................................................     602

Query        1055  ATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCCTATAGTGCAGAACATCC     1204
Query_19954   601  ............TG.................C.....T.................................G.........................A...A..GC.................................C...     750
Query_19956   601  ..................................................................................................................     750
Query_19953   601  ............TG.................C.....T.................................G.........................A...A..GC.................................C...     750
Query_19955   599  ............TG.................C.....T.................................G.........................A...A..GC.................................C...     748
Query_19959   603  ..................................................................CC.CC...........G..C.....                                                         699
Query_19952   603  ..................................................................CC.CC...........G..C.....                                                         699
Query_19958   603  ..................................................................CC.CC...........                                                                  690
Query_19957   603  ..................................................................CC.CC...........                                                                  690
Query_19951   600  ..................................................................CC.CC...........G..C.....                                                         696
Query_19950   603  ..................................................................CC.CC...........G..C.....                                                         699

Query        1205  AGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTGATACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAA     1354
Query_19954   751  ..................................................................................A.....................................................T.........     900
Query_19956   751  ..................................................................................................................     900
Query_19953   751  ........................................                                                                                                          794
Query_19955   749  ....................................                                                                                                              788

Query        1355  ACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAA     1504
Query_19954   901  ..................................................................T.................................................................     1050
Query_19956   901  ....................................................................                                                                              966

Query        1505  CTACTAG                                                                                                                                           1511
Query_19954  1051  .......                                                                                                                                           1057
```

cPPT

```
Job Title:HXB2 cPPT
Program: BLASTN
Query: HXB2    ID: lcl|Query_27334(dna) Length: 547

Subject #1:Group i cPPT)      ID: lcl|Query_27336 Length: 118
Subject #2:Group ii (cPPT)    ID: lcl|Query_27337 Length: 118
Subject #3:Group v (cPPT)     ID: lcl|Query_27338 Length: 178
Subject #4:Group iii (cPPT)   ID: lcl|Query_27339 Length: 146
Subject #5:Group iv (cPPT)    ID: lcl|Query_27340 Length: 180
Subject #6:Group vi (cPPT)    ID: lcl|Query_27341 Length: 177
Subject #7:Group vii (cPPT)   ID: lcl|Query_27342 Length: 549
Subject #8:Group a (cPPT)     ID: lcl|Query_27343 Length: 116
Subject #9:Group b (cPPT)     ID: lcl|Query_27344 Length: 116
Subject #10:Group c (cPPT)    ID: lcl|Query_27345 Length: 117
Subject #11:Group d (cPPT)    ID: lcl|Query_27346 Length: 124
Subject #12:Group e (cPPT)    ID: lcl|Query_27347 Length: 153
Subject #13:Group f (cPPT)    ID: lcl|Query_27348 Length: 151
Subject #14:Group g (cPPT)    ID: lcl|Query_27349 Length: 153

Sequences producing significant alignments:
                      Max    Total   Query   Per.    E       Per.    Acc.
Description           Taxid  Score   Score   cover   Value   Ident   Len     Accession
Group vii (cPPT)      0      920     920     99%     0.0      97.07  549     Query_27342
Group v (cPPT)        0      318     318     32%     1e-90    98.88  178     Query_27338
Group iv (cPPT)       0      315     315     32%     2e-89    98.33  180     Query_27340
Group vi (cPPT)       0      307     307     32%     3e-87    97.77  177     Query_27341
Group e (cPPT)        0      278     278     27%     2e-78    99.35  153     Query_27347
Group g (cPPT)        0      272     272     27%     1e-76    98.69  153     Query_27349
Group f (cPPT)        0      268     268     27%     1e-75    98.68  151     Query_27348
Group iii (cPPT)      0      265     265     26%     2e-74    99.32  146     Query_27339
Group d (cPPT)        0      224     224     22%     3e-62    99.19  124     Query_27346
Group ii (cPPT)       0      219     219     21%     1e-60    100.00 118     Query_27337
Group i cPPT)         0      219     219     21%     1e-60    100.00 118     Query_27336
Group c (cPPT)        0      217     217     21%     5e-60    100.00 117     Query_27345
Group a (cPPT)        0      215     215     21%     2e-59    100.00 116     Query_27343
Group b (cPPT)        0      209     209     21%     8e-58    99.15  116     Query_27344


Alignments:
Query        4554  CCAGTAAAAACAATACATACTGACAATGGCAGCAATTTCACCGGTGCTACGGTTAGGGCCGCCTGTTGGTGGGCGGGAATCAAGCAGGAATTTGGAATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTA  4703
Query_27342  1     ............G.......A.....................A..A....A....A.....................G.................C.............................A.....................  150


Query        4704  AAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAgggggggattgggggTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATa  4853
Query_27342  151   ..........................................................................................................................................  300
Query_27338  1                                                    ..........................................................................................  97
Query_27340  1                                                    ..........................................................................................  97
Query_27341  1                                                    ..........................-.....-..........................................................  96
Query_27347  1                                                                              ..........................................................................  72
Query_27349  1                                                                              ..........................................................................  72
Query_27348  1                                                                              ..........................................................................  72
Query_27339  1                                                                                ........................................................................  70
Query_27346  1                                                                              ..........................................................................  72
Query_27337  1                                                                              ..........................................................................  73
Query_27336  1                                                                              ..........................................................................  73
Query_27345  1                                                                              ..........................................................................  72
Query_27343  1                                                                              ..........................................................................  72
Query_27344  1                                                                              ..........................................................................  72
```

cPPT


Query 4854 caaactaaagaattacaaaaacaaattac-aaaaattcaaaattTTCGGGTTTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAA 5003
Query_27342 301 .....-.....C.C.....G.....G.....G..C..... 450
Query_27338 98 .....-.....G.....G..... 178
Query_27340 98 .....T.....G..... 180
Query_27341 97 .....-.....G.....G..... 177
Query_27347 73 .....-.....G..... 153
Query_27349 73 .....-.....G.....G..... 153
Query_27348 73 .....-.....G.....G... 151
Query_27339 71 .....-.....G..... 146
Query_27346 73 .....C.....-..... 124
Query_27337 74 .....-..... 118
Query_27336 74 .....-..... 118
Query_27345 73 .....-..... 117
Query_27343 73 .....-..... 116
Query_27344 73 .....-....-..... 116

Query 5004 GTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAGAAC 5099
Query_27342 451 .....C..... 546

RRE

RID: 0XA025VK114
Job Title:HXB2 RRE
Program: BLASTN
Query: HXB2 ID: lcl|Query_50555(dna) Length: 1232

Subject #1:Group i (RRE) ID: lcl|Query_50557 Length: 333
Subject #2:Group iii (RRE) ID: lcl|Query_50558 Length: 858
Subject #3:Group ii (RRE) ID: lcl|Query_50559 Length: 826
Subject #4:Group vi (RRE) ID: lcl|Query_50560 Length: 964
Subject #5:Group v (RRE) ID: lcl|Query_50561 Length: 931
Subject #6:Group iv (RRE) ID: lcl|Query_50562 Length: 858
Subject #7:Group a (RRE) ID: lcl|Query_50563 Length: 784
Subject #8:Group b (RRE) ID: lcl|Query_50564 Length: 857
Subject #9:Group c (RRE) ID: lcl|Query_50565 Length: 858

Sequences producing significant alignments:

Description		Max Score	Total Score	Query cover	E Value	Per. Ident	Acc. Len	Accession
Group vi (RRE)	0	1703	1703	78%	0.0	98.55	964	Query_50560
Group v (RRE)	0	1657	1657	75%	0.0	98.82	931	Query_50561
Group iii (RRE)	0	1585	1585	69%	0.0	100 0	858	Query_50558
Group c (RRE)	0	1580	1580	69%	0.0	99.88	858	Query_50565
Group iv (RRE)	0	1580	1580	69%	0.0	99.88	858	Query_50562
Group b (RRE)	0	1572	1572	69%	0.0	99.77	857	Query_50564
Group ii (RRE)	0	1498	1498	67%	0.0	99.39	826	Query_50559
Group a (RRE)	0	1443	1443	63%	0.0	99.87	784	Query_50563
Group i (RRE)	0	610	610	27%	1e-177	99.70	333	Query_50557

Alignments:

```
Query         7261  CTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTTAATTGTGGAGGGGAATTTTTCTACTGTAATTCAACACAACTGTTTAATAGTACTTGGT   7410
Query_50560   1     ..........................................................................................................................................................   150

Query         7411  TTAATAGTACTTGGAGTACTGAAGGGTCAAATAACACTGAAGGAAGTGACACAATCACCCTCCCATGCAGAATAAAACAAATTATAAACATGTGGCAGAAAGTAGGAAAAGCAATGTATGCCCCTCCCATCAGTGGACAAATT-AGATGTT  7560
Query_50560   151   ...........................................................A.....................T.................G.............................................-.......  300
Query_50561   1                                                                                                                                                ....A.......  12

Query         7561  CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAG   7710
Query_50560   301   .............T................................A........G............................C.....................................................................   450
Query_50561   13    .............T................................A........G............................C.....................................................................   162
Query_50558   1                                                                                      ..................................................................................   89
Query_50565   1                                                                                      ..................................................................................   89
Query_50562   1                                                                                      ..................................................................................   89
Query_50564   1                                                                                      ...................................-..............................................................   88
Query_50559   1                                                                                                                        ..........................................................   57
Query_50563   1                                                                                      ..................................................................................   89
```

RRE

```
Query          7711  TAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGT   7860
Query_50560    451  ..............................................................................................................CGG.....................................   600
Query_50561    163  ..............................................................................................................G.....................................   312
Query_50558     90  ..............................................................................................................G.....................................   239
Query_50565     90  ..............................................................................................................G.....................................   239
Query_50562     90  ..............................................................................................................G.....................................   239
Query_50564     89  ..............................................................................................................G.....................................   238
Query_50559     58  ..............................................................................................................G.....................................   207
Query_50563     90  ..............................................................................................................G.....................................   239
Query_50557      1  ..............................................................................................G...............................................       138

Query          7861  CTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTGGGGATTT   8010
Query_50560    601  ...A..........................................................................A.....................................   750
Query_50561    313  ...A..........................................................................A.....................................   462
Query_50558    240  ..............................................................................................................................   389
Query_50565    240  ..............................................................................................................................   389
Query_50562    240  ..............................................................................................................................   389
Query_50564    239  ..............................................................................................................................   388
Query_50559    208  ..............................................................................................................................   357
Query_50563    240  ..............................................................................................................................   389
Query_50557    139  ..............................................................................................................................   288

Query          8011  GGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAA   8160
Query_50560    751  ..............................................................................A...T.....................................   900
Query_50561    463  ..............................................................................A...T.....................................   612
Query_50558    390  ..............................................................................................................................   539
Query_50565    390  ..............................................................................................................................   539
Query_50562    390  ..............................................................................................................................   539
Query_50564    389  ..............................................................................................................................   538
Query_50559    358  ..............................................................................A...T.....................................   507
Query_50563    390  ..............................................................................................................................   539
Query_50557    289  ..........................................          333

Query          8161  TTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAA   8310
Query_50560    901  ..............................................................          964
Query_50561    613  ..............................................................................................................................   762
Query_50558    540  ..............................................................................................................................   689
Query_50565    540  ..............................................................................................................................   689
Query_50562    540  ..............................................................................................................................   689
Query_50564    539  ..............................................................................................................................   688
Query_50559    508  ..............................................................................................................................   657
Query_50563    540  ..............................................................................................................................   689

Query          8311  GAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATagaagaagaaggtggagagagagacagagacagaTCCA   8460
Query_50561    763  ..........................................................................T...................................   912
Query_50558    690  ..............................................................................................................................   839
Query_50565    690  ..............................................................................................................................   839
Query_50562    690  ..............................................................................................................................   839
Query_50564    689  ..............................................................................................................................   838
Query_50559    658  ......................G.......................................................T...................................   807
Query_50563    690  ..............................................................................................     784
```

RRE

```
Query          8461  TTCGATTAGTGAACGGATC      8479
Query_50561    913   ...................      931
Query_50558    840   ...................      858
Query_50565    840   ...................      858
Query_50562    840   ...................      858
Query_50564    839   ...................      857
Query_50559    808   ...................      826
```

PPT & 3’ΔLTR sin

RID: 0XHENTK5114
Job Title:HXB2 PPT & 3’LTR
Program: BLASTN
Query: HXB2 ID: lcl|Query_40404(dna) Length: 823

Subject #1:pHIV-7 (3’LTR) ID: lcl|Query_40406 Length: 234
Subject #2:pHR-SIN-Trip (3’LTR) ID: lcl|Query_40407 Length: 234
Subject #3:pTNS9 (3’LTR) ID: lcl|Query_40408 Length: 281
Subject #4:pWPXL (3’LTR) ID: lcl|Query_40409 Length: 362
Subject #5:pRRL (3’LTR) ID: lcl|Query_40410 Length: 304
Subject #6:pCCL (3’LTR) ID: lcl|Query_40411 Length: 304
Subject #7:pNL ‘U3 (3’LTR) ID: lcl|Query_40412 Length: 304
Subject #8:pTYF (3’LTR) ID: lcl|Query_40413 Length: 104
Subject #9:pCSCIGW (3’LTR) ID: lcl|Query_40414 Length: 690
Subject #10:pVRX496 (3’LTR) ID: lcl|Query_40415 Length: 823
Subject #11:pLenti-III (3’LTR) ID: lcl|Query_40416 Length: 234
Subject #12:pRSIEG (3’LTR) ID: lcl|Query_40417 Length: 305
Subject #13:Group a (3’LTR) ID: lcl|Query_40418 Length: 300
Subject #14:pHAGE (3’LTR) ID: lcl|Query_40419 Length: 301
Subject #15:pSF Lenti (3’LTR) ID: lcl|Query_40420 Length: 304
Subject #16:pLenti7.3 (3’LTR) ID: lcl|Query_40421 Length: 306
Subject #17:p156RRLsinppt (3’LTR) ID: lcl|Query_40422 Length: 305
Subject #18:pLL3.7m (3’LTR) ID: lcl|Query_40423 Length: 402
Subject #19:Group b (3’LTR) ID: lcl|Query_40424 Length: 689
Subject #20:pLV-GFPSpark## (3’LTR) ID: lcl|Query_40425 Length: 689
Subject #21:pHRsin (3’LTR) ID: lcl|Query_40426 Length: 822
Subject #22:pLVX (3’LTR) ID: lcl|Query_40427 Length: 738

Sequences producing significant alignments:

Description	Taxid	Max Score	Total Score	Query cover	E Value	Per. Ident	Acc. Len	Accession
pVRX496 (3’LTR)	0	1476	1476	100%	0.0	99.03	823	Query_40415
pHRsin (3’LTR)	0	1435	1435	99%	0.0	98.18	822	Query_40426
pLVX (3’LTR)	0	1330	1330	89%	0.0	99.19	738	Query_40427
Group b (3’LTR)	0	854	1202	83%	0.0	97.41	689	Query_40424
pCSCIGW (3’LTR)	0	850	1198	83%	0.0	97.21	690	Query_40414
pLV-GFPSpark## (3’LTR)	0	848	1197	83%	0.0	97.21	689	Query_40425
pTNS9 (3’LTR)	0	377	510	33%	1e-107	100.0	281	Query_40408
pNL ‘U3 (3’LTR)	0	372	563	36%	6e-106	100.0	304	Query_40412
pLL3.7m (3’LTR)	0	368	683	49%	8e-105	100.0	402	Query_40423
pHIV-7 (3’LTR)	0	368	434	28%	8e-105	100.0	234	Query_40406
pRSIEG (3’LTR)	0	364	550	38%	1e-103	98.55	305	Query_40417
pWPXL (3’LTR)	0	364	559	37%	1e-103	99.50	362	Query_40409
p156RRLsinppt (3’LTR)	0	363	543	37%	4e-103	99.50	305	Query_40422
pLenti7.3 (3’LTR)	0	363	556	37%	4e-103	99.50	306	Query_40421
pSF Lenti (3’LTR)	0	363	552	36%	4e-103	99.50	304	Query_40420
pHAGE (3’LTR)	0	363	547	36%	4e-103	99.50	301	Query_40419
Group a (3’LTR)	0	363	545	36%	4e-103	99.50	300	Query_40418
pLenti-III (3’LTR)	0	363	363	25%	4e-103	98.09	234	Query_40416
pCCL (3’LTR)	0	363	552	36%	4e-103	99.50	304	Query_40411
pRRL (3’LTR)	0	363	552	36%	4e-103	99.50	304	Query_40410
pHR-SIN-Trip (3’LTR)	0	363	363	24%	4e-103	99.50	234	Query_40407
pTYF (3’LTR)	0	152	152	9%	8e-40	100.0	104	Query_40413

PPT & 3’ΔLTR sin

Alignments:

```
Query        8897  CTCGAGACCTGGAAAAACATGGAGCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAG  9046
Query_40415  1     ..........A.....................................A.....................................A.....................................................  150
Query_40426  1      .........A.....................................A.....................................A.....................................................  149
Query_40427  1      .........A.....................................A.....................................A.....................................................  149
Query_40424  1      .........A.....................................A.....................................A.....................................................  149
Query_40414  1     .....C....A.....................................A.....................................A.....................................................  150
Query_40425  1      .........A.....................................A.....................................A.....................................................  149
Query_40412  1                                                                                                                                                  32
Query_40423  1      .........A.....................................A.....................................A.....................................................  149
Query_40417  1                                                                                                                                                  32
Query_40409  1                                                                                                                                                  34
Query_40422  1                                                                                                                                                  32
Query_40421  1                                                                                                                                                  33
Query_40420  1                                                                                                                                                  31
Query_40419  1                                                                                                                                                  28
Query_40418  1                                                                                                                                                  27
Query_40411  1                                                                                                                                                  31
Query_40410  1                                                                                                                                                  31
```

```
Query        9047  CTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTAGCAGAACTACACACCAGGGCCAGGGGTCA  9196
Query_40415  151   ..................................................................................................................G.............................  300
Query_40426  150   ..................................................................C.................................................G.....................A...  299
Query_40427  150   ...............................A..................C.....................................................................................  299
Query_40424  150   ..................................................................C.....................................................G.....................A...  299
Query_40414  151   ..................................................................C.....................................................G.....................A...  300
Query_40425  150   ..................................................................C.....................................................G.....................A...  299
Query_40408  1          ..........................................................C................                                                          74
Query_40412  33    ..................................................................                                                                          103
Query_40423  150   ...AC.--.G.-.--.--.....................................                                                                              203
Query_40406  1                                    ..................................                                                                       35
Query_40417  33    ..........................................................C.....T.....                                                                  106
Query_40409  35    ..........................................................C..........                                                                  108
Query_40422  33    ...................................................GC.....C..........                                                                  106
Query_40421  34    ..........................................................C..........                                                                  107
Query_40420  32    ..........................................................C..........                                                                  105
Query_40419  29    ..........................................................C..........                                                                  102
Query_40418  28    ..........................................................C..........                                                                  101
Query_40411  32    ..........................................................C..........                                                                  105
Query_40410  32    ..........................................................C..........                                                                  105
```

```
Query        9197  GATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTTGAGCCAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGA  9346
Query_40415  301   ..................................................G.....................................................A...........T.....................  450
Query_40426  300   ..............................................A...G...G.......A......G.............C.....................................A.........  449
Query_40427  300   ..................................................G.....................................................................................  449
Query_40424  300   ..............................................A...G...G.......A......G.............C.....................................A.........  449
Query_40414  301   ..............................................A...G...G.......A......G.............C.....................................A.........  450
Query_40425  300   ..............................................A...G...G.......A......G.............C.....................................A.........  449
```

PPT & 3’ΔLTR sin