



Open camera or QR reader and scan code to access this article and other resources online.

Analysis of HIV-1-Based Lentiviral Vector Particle Composition by PacBio Long-Read Nucleic Acid Sequencing

Saqlain Suleman,^{1–3} Mohammad S. Khalifa,¹ Serena Fawaz,¹ Sharmin Alhaque,¹ Yaghoub Chinaea,¹ and Michael Themis^{1,3,4,*}

¹Department of Life Sciences, Brunel University London, London, United Kingdom; ²Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, United Kingdom; ³Testavec Ltd., Queensgate House, Maidenhead, United Kingdom; and ⁴Division of Ecology and Evolution, Department of Life Sciences, Imperial College London, London, United Kingdom.

Lentivirus (LV) vectors offer permanent delivery of therapeutic genes to the host through an RNA intermediate genome. They are one of the most commonly used vectors for clinical gene therapy of inherited disorders such as immune deficiencies and cancer immunotherapy. One of the most difficult challenges facing their widespread application to patients is the large-scale production of highly pure vector stocks. To improve vector production and downstream purification, there has been a recent investment in the United Kingdom to establish good manufacturing process (GMP)-licensed centers for manufacture and quality control. Other requirements for these vectors include their target cell specificity and tropism, how to regulate gene expression of the therapeutic payload and their potential side effects. Comprehensive detail on the full nucleic acid content of LV is unknown, even though they have entered clinical trials. With potential adverse effects in mind, it is important to identify these contents to assess their safety and purity. In this study, we used highly sensitive PacBio long-distance, next-generation sequencing of reverse-transcribed vector component RNA to investigate the nucleic acid composition of recombinant HIV-1 particles generated by human 293T packaging cells. In this article, we describe our findings of nucleic acids other than the recombinant vector genome that exist, which could potentially be delivered during gene transfer, and suggest that removal of these unwanted components be considered before clinical LV application.

Keywords: PACBIO, lentivirus, HIV-1, HERV

INTRODUCTION

To date, more than 3,000 clinical trials have been initiated in the field of gene and cell therapy.^{1,2} Many use retrovirus (RV) or lentivirus (LV)-based vectors due to their excellent gene transfer capabilities and long-term and persistent therapeutic gene expression.^{3,4} To meet safety requirements and reach therapeutic potential for their use, significant modifications to these vectors have been made. As a result, RV produced by human embryonic kidney cells (HEK 293T) has been replaced by LV based on HIV-1,

which unlike RV can infect nondividing cells. Essential to LV production from HEK293T is the isolation of pure particles using methodologies that isolate vector from supernatant impurities such as producer cell debris and media components that may elicit an immune response by the host.⁵ In large-scale production, ultracentrifugation that is used for particle concentration is replaced by tangential flow filtration, benzoate treatment, and gradient purification or chromatography to remove cellular genomic DNA followed by elution and filtration.^{6,7}

*Correspondence: Prof. Michael Themis, Department of Life Sciences, Brunel University London, Hienz Wolff Building, Kingston Lane, Uxbridge, London UB8 3PH, United Kingdom, E-mail: michael.themis@brunel.ac.uk

© The Author(s) 2025. Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License [CC-BY] (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Other impurities, important to vector safety, include virus components that may enable replication competence to support propagation and spread within the host. Replication competence was observed in early generation RVs, and consequently, the genes coding for virus packaging have been removed by recombinant vector design. This uses genome splitting onto discrete plasmid constructs with reduced homology to generate vector particles carrying cis genomes and trans-packaging genes to avoid generating recombination and replication of competent virus. Historically, first-generation LV was produced by splitting the HIV 1 genome into three plasmids for transfection of virus-producer cells. The first plasmid carrying the vector genome included the expression cassette that comprised an internal promoter to drive therapeutic gene expression, a Rev response element, and a genome psi (ψ) packaging signal recognized by the LV GAG nucleoprotein to ensure only the RNA genome is packaged with native long terminal repeats (LTR) needed for vector integration. A second plasmid was used that included HIV transgenes essential for genome packaging into virus particles, and the third plasmid provided an envelope gene to generate a surface glycoprotein (*e.g.*, VSV-G) for virus attachment to the target cell for infection.⁸ Importantly, as neither second nor third plasmids carry ψ packaging signals, their packaging is prevented to circumvent virus propagation. To further reduce the potential for recombination leading to replication competence, second-generation vectors were engineered with even less plasmid homology by removal of accessory genes (*i.e.*, Vif, Vpu, Vpr, and Nef).^{9–11} Third-generation vectors include additional modifications by removal of Tat and the addition of Rev separately on a fourth plasmid to improve vector genome nuclear export and titer. Third-generation vectors also use self-inactivating design with a deletion in the 3' LTR containing the TATA box to abolish LTR promoter activity and reduce the potential for oncogene activation near the vector integration site.^{12,13}

It has been shown that the packaging of vector genomes into RV and LV particles is enriched by the presence and specificity of the psi (ψ) packaging signal generally located near the 5'-end of the viral RNA recognized by the GAG protein.¹⁴ It is unclear, however, to what extent nonviral RNA or DNA is packaged. These could include parts of the HEK293T genome, single genomes, human endogenous RV (HERV) such as sequences, spliced viral mRNAs, other cellular mRNAs, or the trans-plasmids used for vector production,¹⁵ and these should be avoided in gene delivery. In addition, alternative contaminants for gene transfer could originate from ectosomes and exosomes that pinch off from HEK293 producer cells. Importantly, ectosomes and exosomes that are not usually removed in LV production are known to contain nucleic acids, lipids, metabolites, and

proteins and have been reported to be associated with several cellular activities and cancer progression.

As the contents of packaged LV have not been fully characterized, we investigated in detail the nucleic acid composition of “full” genome and “empty” recombinant LV particles. For this study, particles not separated from ectosome and exosome vesicles were generated by three plasmid transfections of HEK293T cells and isolated for sensitive long-range PacBio RNA sequencing. PacBio sequencing, using single-molecule real-time (SMRT) sequencing, provides long-read sequences to identify complete genomes, through single-molecule sequencing.¹⁶ This has been performed previously using the genomes packaged within adeno-associated viral vectors to identify chimeric genomes with the packaging of aberrant genomes.^{17,18}

We describe our observations of a range of nucleic acids not of HIV-1 origin, in addition to the vector genome present in LV particles. Our data suggest that further investigation of the origin of these nucleic acids and their potential side effects on target cells should be considered and propose that removal of these contaminants would be a valuable step in manufacturing safer recombinant LV particles.

METHODS

Vector production and titration

The production of HR'SIN-cPPT-SEW-eGFP-W LV (pHR, full) and no transgene equivalent (empty) LV was carried out as previously described.¹⁹ pHR LV vector expresses eGFP under the internal promoter of SFFV. Infectious LV titer was calculated as previously reported.¹⁹ Briefly, 2×10^5 HEK293T cells were seeded and incubated at 37°C, 5% CO₂ overnight to adhere. Serial dilutions of the virus were prepared and incubated in a complete cell culture medium with 5 μ g/mL polybrene (Sigma Aldrich), for 20 min at room temperature before addition to cells; 72 h post-transfection, cells were harvested for green fluorescent protein (GFP) expression analysis via flow cytometry using ACEA Novocyte flow cytometer and NovoExpress software v1.2.5 (Agilent Technologies, Didsbury, England). Dilutions expressing 1–30% GFP expression were analyzed as accurate representations of viral titer, calculated as shown in the following equation:

$$\text{Titer (TU/mL)} = ((\text{Cell count} \times (\text{Percentage GFP expression}/100))/\text{Volume}) \times \text{DF}$$

pHR LV titer was calculated as 1.18×10^9 TU/mL.

Empty LV, carrying no reporter transgene, were titrated using Lenti-X™ p24 Rapid Titer Kit (Takara Bio), according to manufacturer's instructions. Absorbance was read on Elx808 absorbance reader at 450 nm and analyzed using Gen5 software (v2.06.10).

Nucleic acid harvesting

Viral RNA was harvested via QIAamp[®] viral RNA mini kit (Qiagen) according to the manufacturer's instructions. The nucleic acid concentration was determined using Nanodrop 2000C UV-Vis (v1.2.1) (Fisher Scientific). RNA samples were treated to remove gDNA contamination using RNase-Free DNase set and RNase kit (Qiagen) according to the manufacturer's instructions.

PacBio DNA sequencing

RNA samples were sequenced by the Center for Genomic Research, University of Liverpool. Sample preparation of involved RNA integrity using the Agilent 6000 Pico Bioanalyser and RNA pico chips; 1 μ g of RNA per sample was depleted of rRNA using a Ribozero kit followed using RNA clean AMPure beads and repeated Bioanalyser testing for diminished rRNA profiles. Treated with Cell Script plus poly-A polymerase tailing kit, undergoing 15 cycles of cDNA synthesis being once again purified with the AMPure beads and reanalyzed this with a DNA HS Bioanalyser chip.

The library was produced using the Pac Bio DNA template kit 1.0 followed by a final AMPure bead cleaning and bioanalyser control step. Samples were mixed with polymerase and diffusion loaded into LR cells and into a PacBio Sequel 2.1 chemistry sequencing machine. Settings were 1,200 min movie times and 240 min of preextension with an on plate load of 8 pM. The data were processed using the Iso-Seq workflow to retrieve CCS.

DNA contaminants identification

Prior to sequencing, RNA samples were analyzed for gDNA contamination prior to and subsequent to conversion to cDNA using GoScript[™] reverse transcriptase (Promega) according to manufacturer's instructions. Samples were amplified via PCR against β -actin and the viral LTR, shown below.

```
LTR-F GAGCTCTCTGGCTAACTAGG
LTR-R GCTAGAGATTTTCCACACTG
SY100216195-080 AAGAGAGGCATCCTCACCT
SY100216195-081 TACATGGCTGGGGTGTGAA
```

Screening for HERVS, cloning vectors, and packaging sequences

For all searches, the same search keys were used to enable comparison, sequence coverage had to be greater or equal to 80 bp, had to be derived from CCS of 100 bp or greater, and alignment had to be 10% or greater than the CCS length. The single best result was determined for each CCS where applicable.

To identify the relative presence of HERV-K113 in the samples, the prevalence of human genomic reads was calculated by aligning all CCS to the human reference

genome (GRCh38) and HERV-K113 genome and normalized per kilobase of genome per million reads.

RESULTS

Production of HIV-1 LV vectors

Replication defective second-generation SIN configuration HIV-1-based particles, HR'SIN-cPPT-SEW-eGFP-W (abbreviated to pHR LV in this article), or no vector genome (empty vector) (Fig. 1), were generated by three plasmid transient transfection of HEK293T virus-producer cells.²⁰ For this, packaging plasmids were used at identical concentrations plus or minus a third plasmid at the same concentration carrying the vector backbone. Viruses were harvested at 24, 48, and 72 h post-transfection and pooled for concentration. pHR LV particles were titrated via flow cytometry for GFP expression using infected HEK 293T cells. Vector titer for both pHR and empty LV was also determined using a p24 gag ELISA (Takara Bio). Each was diluted to 2.42×10^9 TU/mL to ensure that equal LV titers were used for subsequent investigation.

Collected particle supernatants were treated by DNase I and RNase digestion to remove external contaminants. Viral RNA, purified from lysed particles, was treated with and without reverse transcription and analyzed for the absence of HEK293T DNA by polymerase chain reaction (PCR) of the β -actin gene and vector LTRs (Fig. 2). The absence of bands for β -actin in viral samples but their presence in control cellular nucleic acid demonstrated the vector to be free of gDNA contamination at this level of detection. Batches of vectors were also subjected to more sensitive quantitative PCR that also showed no presence of contaminating HEK293T gDNA (data not shown). PCR of the HIV-1 LTR in samples without reverse transcription identified weak bands suggesting these could have arisen from reverse transcription within HIV-1 particles that can occur during maturation. Positive LTR bands in the empty virus particles suggest that the plasmid DNA carrying the HIV-1 trans-acting components used to generate virus particles was present. This corresponds with the data analyzed from PacBio sequencing of vector RNA (Fig. 2).

Analysis of LV particle contents following long-distance PacBio (IsoSeq) sequencing

Extracted DNase-I treated viral RNA samples were examined for RNA integrity using an Agilent 6000 Pico Bioanalyser and RNA pico chips. RNA was depleted for rRNA, then cleaned and reanalyzed for diminished rRNA profiles. cDNA libraries were synthesized and purified before examination with a DNA HS Bioanalyser chip. PacBio sequencing of the extracted samples identified various reads and the sequence metrics of each prep.

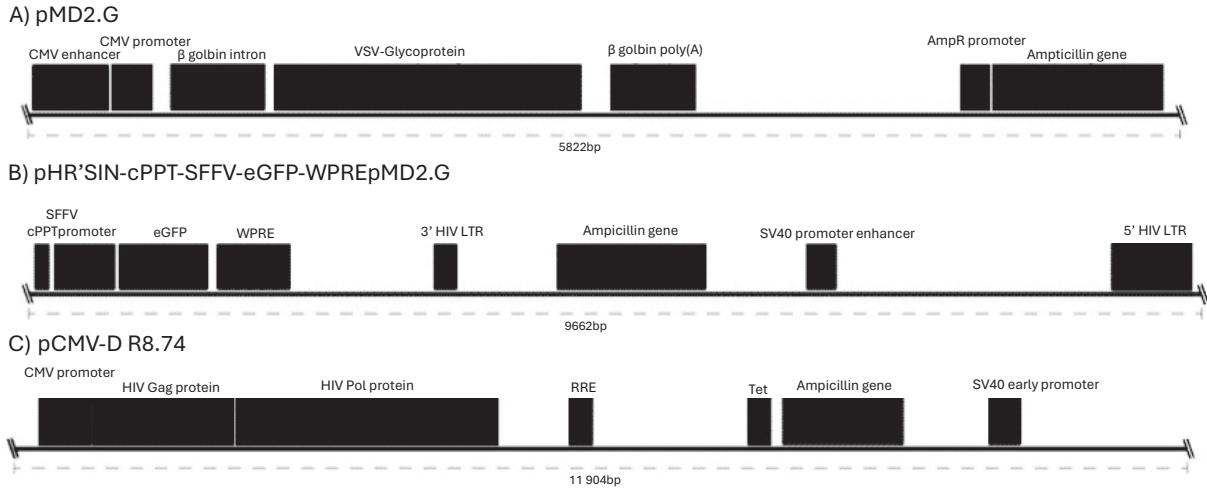


Figure 1. Vector used for LV production. **(A–C)** Each plasmid construct in a linear form. **(A)** pMD2.G carried a CMV enhancer, CMV promoter, Beta-globin intron 1, VSV-Glycoprotein gene, Beta-globin poly(A), and components for growth and selection in bacteria. **(B)** pHR'SIN-cPPT-SFFV-eGFP-WPRE contains the cPPT region of HIV-1, an SSFV promoter, the eGFP reporter gene, WPRE, 3'HIV SIN LTR, the promoter/enhancer from SV40 and the 5' SIN HIV LTR. **(C)** pCMV-dR8.74 contains a CMV promoter to drive the expression of the HIV-1 gag and pol genes, RRE, tet, Ampicillin gene, and the SV40 early promoter. LTR, long terminal repeat; LV, lentivirus.

As expected, the pHR LV vector had 3,956,793 greater number of subreads than the empty, nongene-carrying, LV preparation (6,617,371 vs. 2,660,578). The sequence data were processed using the IsoSeq pipeline within the SMRT Link software suite (v5.1.0.26412). As the primary aim of this experiment was to sensitively detect transcripts, the final IsoSeq output files were not used, as these involved iterative clustering of sequences resulting in reduced sensitivity. Instead, the intermediate Circular Consensus Reads (CCS) generated as part of the IsoSeq pipeline were used. These reads were generated by creating a consensus sequence from multiple sequencing passes of the same template

molecule that improved sequence quality and reduced errors. As the sensitivity was critical, no lower limit to the minimum number of sequencing passes was specified.

When processed, a similar number of CCS reads were identified between the vector particles; 657,675 CCS reads were identified in full viral prep (pHR LV) and 691,011 CCS reads in empty particle prep (Table 1). The mean CCS length was similar between full (1,272 bp) and empty (1,823 bp) LV vectors. The distribution of CCS read length in both pHR and empty LV particles followed a double exponential curve with the slope flatter for empty LV preps, as expected (Fig. 3).

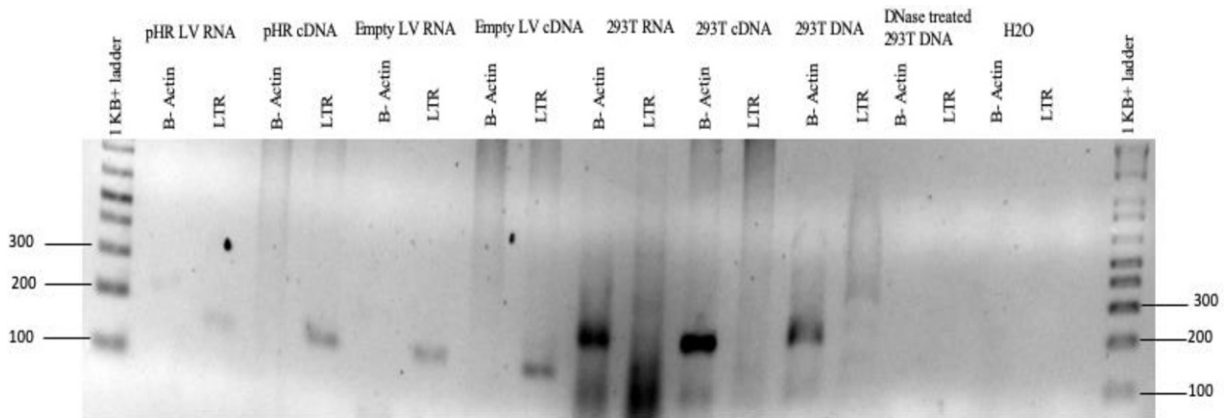


Figure 2. Agarose gel electrophoresis of the viral vector RNA following PCR amplification to identify 293T cell components in virus preps. Virus particles were treated with DNase I and then lysed. Purified RNA was reverse transcribed before PCR amplification. PCR primers identifying b-actin were used to identify HEK293T genomic DNA or cDNA. The absence of b-actin in viral samples and the presence of cellular nucleic acid indicate the vector to be free of gDNA contamination. Positive bands were identified for reverse-transcribed LTRs. Weak LTR bands for nonreverse transcribed samples suggest reverse transcription within particles by HIV-1 RT during maturation. Positive LTR bands in the empty virus particles suggest the presence of contaminating trans-plasmid used to generate virus particles. Samples were run on a 2% agarose-TBE gel at 70V for 35 min. HEK, human embryonic kidney; PCR, polymerase chain reaction.

Table 1. Read Metrics for PacBio Sequenced Viral RNA Samples

Metric	pHR LV	Empty LV
Subreads	6,617,371	2,660,578
Subread bases (bp)	6,947,142,447	3,737,276,349
CCS reads	657,675	691,011
CCS bases (bp)	836,782,922	1,259,747,933
Mean CCS passes	8	2
Mean CCS read length (bp)	1,272	1,823

Metrics for pHR and empty LV RNA samples are shown.
CCS, Circular Consensus; LV, lentivirus.

Taxonomic assignment based on filtered BLAST analysis shows sequences present of multiple viral origins

To determine whether library DNA sequences were from RNA contained in virus and not packaged from the HEK293T cell transcriptome, BLAST was used to determine the highest probable matches. To confirm this, BLAST searches were made against the pHR viral backbone, human genome, and virus RefSeq genome library. These sequences were found to align with HIV-1 abundantly as well as other viruses.

To investigate the coverage of sequences for full genomic sequences, viral sequences were separated by species and mapped and CCS reads were determined for each. Samples were taxonomically assigned by using them as a query for nucleotide BLAST²¹ against the viral RefSeq database (<https://www.ncbi.nlm.nih.gov/genome/viruses/>). Alignments were filtered to retain the single best hit for each CCS read, hits of 80 bp or longer, hits derived from reads that are 100 bp or longer, and hits where the alignment matched over at least 10% of the CCS read length. Postfiltering, 23.6% of reads from pHR were taxonomically assigned. However, a much smaller proportion of empty LV reads were taxonomically assigned (4.3%) due to the alignment lengths generally

being extremely short and because there is much lower content in this sample. The top 10 species identified based on filtered reads alone are shown (Table 2). A full representation of the alignments is also provided (Supplementary Table S1). Note that the Woodchuck hepatitis virus and SFFV reads represent the presence of these elements in the pHR LV vector and may also represent the presence of the transfected cloning vector (HR'SIN-cPPT-SEW-eGFP-W). As this was not used in the production of empty particles, this sequence was not identified in the empty particle preparation. Furthermore, two aligned taxonomical classes identified in empty viruses (Bovine viral diarrhea virus 3 Th/04 KhonKaen and Cafeteria roenbergensis virus BV-PW1) were not identified in the pHR LV prep. Conversely, 23 taxonomical classes were identified in full viral particles but not in empty viruses, suggesting increased aberrant packaging of sequences may be triggered by vector genome production and packaging.

Screening for packaging sequences and cloning vectors

No sequences carrying the HIV-1 ψ packaging sequence were identified in empty LV particles, as expected. In full HIV-1 virions, however, 3,370 CCS reads contained significant hits to this sequence. Of these, 3,347/3,370 (99.32%) were identified with pHR, HIV, and human reference databases; 3,289 (98.27%) sequences align with pHR; however, only 26 (0.78%) were associated with HIV and 32 (0.96%) with the human reference databases suggesting that the ψ packaging sequences were derived from pHR and the RNA used in this analysis did not derive from a human source. As only 0.51% of all CCS reads (657,675) were identified carrying the ψ packaging sequence, this suggests the majority of sequences packaged both in full and empty particles were aberrantly selected. When mapping the sequences identified in full and empty viral preps

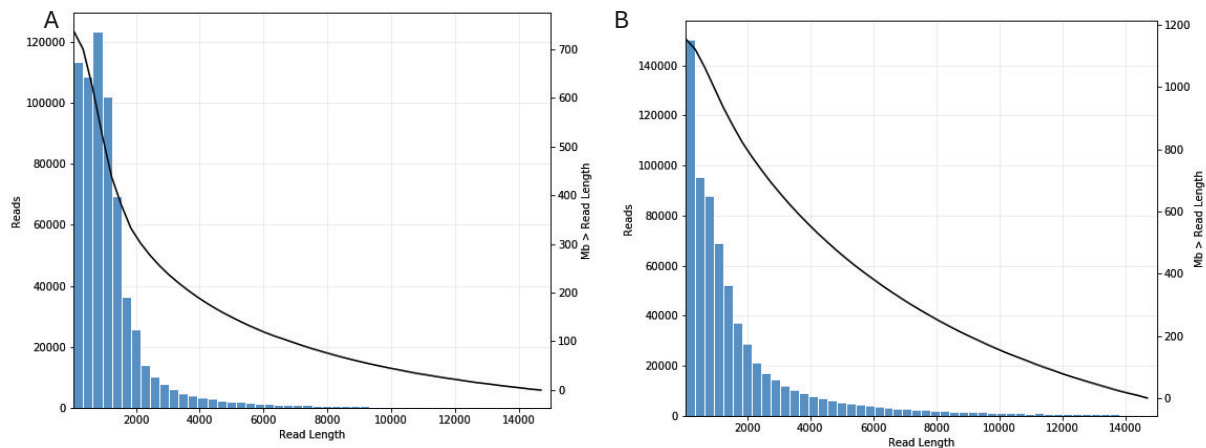


Figure 3. Read length (bp) distribution for pHR and empty vector LV RNA. Read length (bp) distribution for pHR (A) and empty (B) LV RNA, respectively. Both show a double exponential curve, though pHR LV carrying the vector genome shows greater numbers of large read lengths and reads between 0 and 2,000 bp.

Table 2. PacBio Reads of pHR LV and Empty LV

Species	TaxonId	pHR LV hits	Empty LV hits
A			
BeAn 58058 virus	67,082	946	274
Chrysochromulina ericina virus	455,364	120	3
Friend murine leukemia virus	11,795	9,164	0
Human adenovirus 5	28,285	190	43
Human endogenous retrovirus K113	166,122	7	4
Human immunodeficiency virus 1	11,676	51,539	2,446
Macaca mulatta polyomavirus 1	1,891,767	21,023	160
Salmonella virus SP6	194,966	1,875	21
Spleen focus-forming virus	11,819	132	0
Vesicular stomatitis Indiana virus	11,277	574	17
Woodchuck hepatitis virus	35,269	69,357	0

A—Taxonomic assignment based on filtered BLAST hits. Hits show the number of reads aligned to pHR and empty virus samples.

to the pHR plasmid or the trans-acting gag/pol carrying plasmid pCMVR8.74, this identified numerous sequence alignments suggesting the plasmid vectors used in vector production are packaged at low level (Fig. 4).

Detection of HERV contamination of virus preparations

We noticed that HERV-K113 were detected at low levels in both pHR (seven hits with a relative abundance of 4.15×10^{-5}) and empty particles (four hits with a relative abundance of 1.33×10^{-3}), suggesting these are also packaged upon virus production. The identity of these reads is provided (Table 3). By identifying the relative abundance of human and HERV sequences, we sought to determine whether there is a disproportionately high amount of HERV, or equivalent amounts, as the latter would suggest that the HERV identified could be from 293T human genomic DNA contamination.

To address this, all CCS reads were aligned to the human reference genome, filtered in the same way as described earlier, and counts were calculated. Following

this, the counts to human and HERV-K113 sequences were normalized based on the size of the respective genomes, and as a frequency per million reads, that is, reads per kilobase of genome per million reads (RPKM). In the pHR LV prep, the relative abundance of human sequences was calculated as 37.0 RPKM compared to HERV-K113 at 739.2 RPKM. In empty LV preps, human and HERV sequence relative abundance was calculated as 11.6 and 422.4 RPKM, respectively. This identifies significantly greater reads for HERV-K113 in proportion to genome size and suggests HERV-K113 sequences are not derived from human genomic DNA contamination. However, it is important to note that the number of HERV-K113 reads is very low for each sample.

Additionally, in all cases, the HERV-associated reads aligned better to the human reference sequence than to the viral database sequences, specifically, the length and BLAST alignment score of matches were better in all cases, and the percentage identity was better in most cases. This suggests the reads actually would be derived from the human genome, which would be a good indication that there are no nonintegrated HERV sequences. A summary of the key alignment metric differences between HERV and human alignments is shown (Supplementary Table S2).

DISCUSSION

RV and LV offer efficient and permanent delivery of therapeutic genes for gene therapy. Several improvements to vector manufacture that include high titer particle production and purity in addition to improved safety design have resulted in their successful use in the clinic for rare genetic diseases and cancer immunotherapy.²² Most efforts to generate LV for gene therapy have used cell lines or transient transfection protocols of producer cells to yield pseudotyped particles. To remove contaminants,

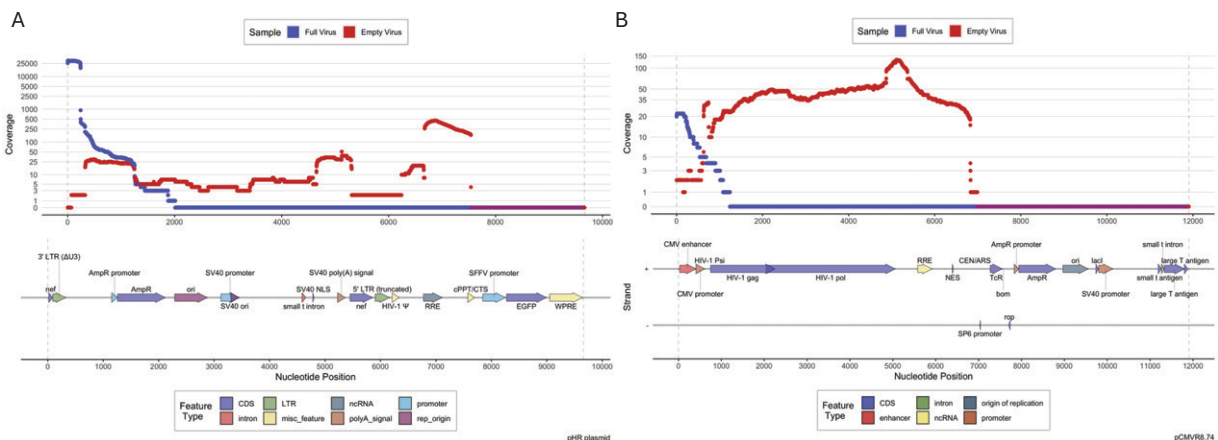


Figure 4. PacBio sequence reads aligned construct plasmids. Reads identified in viral preps were aligned to the pHR plasmid construct (A) or pCMVR8.74 (B). These identify differences between the sequence identities of viral preps.

Table 3. PacBio Reads of pHR LV and Empty LV

B	Read	% Identity		Hit length		Bit score	
		HERV	Human	HERV	Human	HERV	Human
pHR LV	m54118_181207_113652/15794380/ccs	90.08	97.99	383	498	436.611	886.605
	m54118_181207_113652/5832778/ccs	83.78	98.36	641	1,039	337.493	1808.4
	m54118_181207_113652/63308073/ccs	84.34	91.67	249	696	111.505	702.246
	m54118_181207_113652/46530831/ccs	83.67	97.98	490	1,041	291.899	1792.54
	m54118_181207_113652/63505012/ccs	91.96	95.94	833	1,379	987.705	2111.7
	m54118_181207_113652/12779821/ccs	86.43	96.29	258	916	168.993	1364.35
Empty LV	m54118_181207_113652/47579377/ccs	89.5	98.22	543	731	555.552	1271.18
	m54118_181208_080006/23855664/ccs	88.04	98.55	92	415	67.8929	737.929
	m54118_181208_080006/32702952/ccs	88.68	88.11	106	513	67.8929	341.458
	m54118_181208_080006/42599406/ccs	86.61	83.59	224	585	180.887	202.693
	m54118_181208_080006/22610651/ccs	93.39	94.96	257	694	359.299	926.252

B—HERV sequence identity in pHR and empty LV samples. HERVK113 reads identified in viral samples were mapped to the HERVK113 and human genomes, respectively, to identify percentage identity. Higher percentage identity is found in the human genome, determining this as the source of the transcripts. A full list of hits and abundances are provided in Supplementary Table S1.

such as producer cell debris that may be toxic or elicit a host immune response, particles can be concentrated by ultracentrifugation.^{5,23} More recently, more refined downstream purification protocols have generated LV particles of high quality.^{6,7} Although these technologies provide vector batches that have external impurities removed, it is unclear to what extent impurities exist within particles other than the therapeutic genome.

Following infection of the host, two copies of the vector RNA genome are used for successful reverse transcription before genome integration as therapeutic cDNA. RV and to a lesser extent LV genotoxicity are known to be associated with the vector genome following integration as highlighted by several *in vitro* and *in vivo* nonclinical applications and clinical trials to correct genetic disorders.²⁴ Understanding the cause of vector-associated side effects on the host is, therefore, of utmost importance, and models have been developed to identify and measure vector genome/host interactions that alter cancer gene expression and may lead to oncogenesis. Risk assessment has been restricted to the vector genome as it has been assumed that packaging is specific and determined by recognition of the packaging ψ site present on the vector backbone.^{25–28} It has, however, been shown that RVs can also package spliced virus RNA, several tRNAs, and varying amounts of cellular mRNAs in a concentration-dependent manner.^{14,15}

It has also been shown that HERVs devoid of a ψ site can be packaged into virions by human producer cells at low levels.¹⁵ HERVs being of RV origin exist as vestigial elements and represent $\sim 8\%$ of the human genome. They have had a profound influence on human evolution and have been modified over time to prevent their activity that including their propagation and influence on gene expression.²⁹ Unlike RVs, most endogenous RV elements cannot transfer between cells, and most HERVs exist in deleted forms. However, an example of a preserved

element with full-genome sequence is HERV-K113 that belongs to the HERV-K family,³⁰ which retains its RV structure with gag, pol, and env genes bordered on either side by LTRs. Importantly, HERV LTRs can inhibit or upregulate nearby promoter/enhancer activity and can promote novel splice activity, and under specific conditions such as cancer and autoimmune disease, can become active.³¹ During HIV-1 infection, HERV-K113 expression has been shown to increase to the point where its RNA can be detected in patient plasma.^{32,33} Although little complementation activity exists between HIV and HERVs,^{29,34} RV-specific microarray has identified HIV-1 derived packaging cells able to package HERV transcripts.^{35,36}

After we demonstrated the removal of HEK 293 genomic DNA and unpackaged RNA from culture supernatants using DNaseI and RNase treatment and PCR and q-RT PCR analysis of the β -actin gene and vector LTRs, we chose the highly sensitive Illumina PacBio long-range RNA sequence platform to determine the nucleic acid composition of LV, with and without vector genomes. Interestingly, PacBio sequencing identified incomplete HERV-K113 sequences packaged in empty and to a lesser extent full LV particles at relative frequencies of 1.33×10^{-3} and 4.51×10^{-5} . However, longer HERV-K113 of more than 1 kb were identified in pHR LV. We assume competition between HERV-K113 and pHR genomes for integration would be biased to pHR favored by the HIV-1 integrase, which should exclude HERV sequences that have no HIV-1 homology. In a clinical application, with an increased viral load of greater than 1×10^{10} TU, HERV-K113 sequences may become more common, especially when delivered by empty LV particles.³⁷ Importantly, though, we confirmed that the HERV-K113 sequences found were at values greater than background human reads with values at 36.4 RPKM times more prevalence than human genome reads in empty virus and 19.9 RPKM times for full virus

samples. This difference in the relative quantity and RPKM in empty viruses is worth investigating to identify whether this difference is the result of sequencing bias or due to less competition for HERV-K113 sequence packaging in LV particles. An investigation as to whether HERV K113 gene transfer to nonhuman cells by LV could be beneficial using sensitive PCR, whole-genome next generation sequencing (NGS) analysis, TES, inverse PCR, or LAM PCR. As yet, to what extent HERV-K113 gene transfer presents a potential genotoxic hazard is unknown.

Aberrantly packaged nucleic acid sequences represented by the alignment of PacBio reads to the virus database appeared with more frequent reads in the full virus compared to empty particles and suggested potential promotion of contaminating sequence packaging alongside the vector genome. Sequencing of pHR and “empty” particles identified the Woodchuck hepatitis virus sequences only in the pHR vector used for optimal RNA export during vector production by HEK 293 cells; however, the presence of HIV-1 virus sequences in both pHR and empty LV suggests identification of sequences remaining in the plasmid vector used for virus packaging. Interestingly, sequences belonging to the Friend murine leukemia virus and Spleen focus-forming virus were only found in pHR and sequences of Macaca mulatta polyomavirus 1, Salmonella virus SP6, BeAn 58058 virus, Vesicular stomatitis Indiana virus, human adenovirus 5, Chrysochromulina ericina virus, and HERV K113 were present in both vectors. Unwanted packaged sequences would be expected to reduce effective viable vector titer, even though read length was mainly short with reads of up to 9 kb rarely identified. To improve LV batch quality, it would be interesting to remove empty particles during vector production. In addition, production

without ectosomes and exosomes could improve effective vector and safety if, indeed, the contaminants identified in this study were provided by these vesicles.³⁸ Methodologies to remove these contaminants from LV preparations exist and should be considered.³⁹

In conclusion, by using sensitive PacBio NGS, we have identified multiple aberrantly packaged nucleic acid species in full genome carrying and empty LV particles. Work to determine whether these contaminants are integrated would be useful, as would the exosome contribution to gene transfer and side effects. We propose to reach a more characterized clinical grade LV, some of the potential issues raised in this report be considered with the aim to improve the safety profile and titer of LV vectors.

ACKNOWLEDGMENTS

The authors would like to thank Centre for Genomic Research, University of Liverpool, for the PacBio sequencing service platform used in this study and a GlaxoSmithKline (GSK) studentship funding award.

AUTHOR DISCLOSURE

The authors have no conflicts of interest to declare.

FUNDING INFORMATION

GSK funded the studentship.

SUPPLEMENTARY MATERIAL

Supplementary Table S1

Supplementary Table S2

REFERENCES

- Ginn SL, Amaya AK, Alexander IE, et al. Gene therapy clinical trials worldwide to 2017: An update. *J Gene Med* 2018;20(5):e3015.
- Wiley. Gene Therapy Clinical Trials Worldwide. 2023; Available from: <https://a873679.fmphost.com/fmi/webd/GTCT> [Last accessed: March 1, 2023].
- Lundstrom K. Viral vectors in gene therapy. *Diseases* 2018;6(2); doi: 10.3390/diseases602042
- Qiao J, Moreno J, Sanchez-Perez L, et al. VSV-G pseudotyped, MuLV-based, semi-replication-competent retrovirus for cancer treatment. *Gene Ther* 2006;13(20):1457–1470.
- Baekelandt V, Eggermont K, Michiels M, et al. Optimized lentiviral vector production and purification procedure prevents immune response after transduction of mouse brain. *Gene Ther* 2003;10(23):1933–1940.
- Milone MC, O'Doherty U. Clinical use of lentiviral vectors. *Leukemia* 2018;32(7):1529–1541.
- Bandeira V, Peixoto C, Rodrigues AF, et al. Downstream processing of lentiviral vectors: Releasing bottlenecks. *Hum Gene Ther Methods* 2012;23(4):255–263. Au
- Burns JC, Friedmann T, Driever W, et al. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: Concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci U S A* 1993;90(17):8033–8037.
- Zufferey R, Nagy D, Mandel RJ, et al. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nat Biotechnol* 1997; 15(9):871–875.
- Fouchier RA, Simon JH, Jaffe AB, et al. Human immunodeficiency virus type 1 Vif does not influence expression or virion incorporation of gag-, pol-, and env-encoded proteins. *J Virol* 1996;70(12):8263–8269.
- Dull T, Zufferey R, Kelly M, et al. A third-generation lentivirus vector with a conditional packaging system. *J Virol* 1998;72(11): 8463–8471.
- Cornetta K, Yao J, Jasti A, et al. Replication-competent lentivirus analysis of clinical grade vector products. *Mol Ther* 2011;19(3):557–566.

13. Zufferey R, Dull T, Mandel RJ, et al. Self-inactivating lentivirus vector for safe and efficient *in vivo* gene delivery. *J Virol* 1998;72(12): 9873–9880. De
14. Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 2011;410(4):609–633.
15. Patience C, Takeuchi Y, Cosset FL, et al. Packaging of endogenous retroviral sequences in retroviral vectors produced by murine and human packaging cells. *J Virol* 1998;72(4): 2671–2676.
16. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 2015;13(5):278–289.
17. Tai PWL, Xie J, Fong K, et al. Adeno-associated virus genome population sequencing achieves full vector genome resolution and reveals human-vector chimeras. *Mol Ther Methods Clin Dev* 2018;9:130–141.
18. Xie J, Mao Q, Tai PWL, et al. Short DNA hairpins compromise recombinant adeno-associated virus genome homogeneity. *Mol Ther* 2017; 25(6):1363–1374.
19. Khonsari H, Schneider M, Al-Mahdawi S, et al. Lentivirus-mediated frataxin gene delivery reverses genome instability in Friedreich ataxia patient and mouse model fibroblasts. *Gene Ther* 2016;23(12):846–856.
20. Suleman S, Fawaz S, Roberts T, et al. Optimised protocols to generate high titre lentiviral vectors using a novel transfection agent enabling extended HEK293T culture following transient transfection and suspension culture. *J Virol Methods* 2024;325:114884.
21. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;215(3): 403–410.
22. Nobles CL, Sherrill-Mix S, Everett JK, et al. CD19-targeting CAR T cell immunotherapy outcomes correlate with genomic modification by vector integration. *J Clin Invest* 2020;130(2): 673–685.
23. McNally DJ, Darling D, Farzaneh F, et al. Optimised concentration and purification of retroviruses using membrane chromatography. *J Chromatogr A* 2014;1340(100):24–32.
24. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia. *Nature* 2010;467(7313):318–322.
25. Nowrouzi A, Cheung WT, Li T, et al. The fetal mouse is a sensitive genotoxicity model that exposes lentiviral-associated mutagenesis resulting in liver oncogenesis. *Mol Ther* 2013;21(2): 324–337.
26. Chandler RJ, LaFave MC, Varshney GK, et al. Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J Clin Invest* 2015;125(2):870–880.
27. Themis M, Waddington SN, Schmidt M, et al. Oncogenesis following delivery of a nonprimate lentiviral gene therapy vector to fetal and neonatal mice. *Mol Ther* 2005;12(4):763–771.
28. Rulli SJJ, Hibbert CS, Mirro J, et al. Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol* 2007;81(12): 6623–6631.
29. van der Kuyl AC. HIV infection and HERV expression: A review. *Retrovirology* 2012;9:6–6.
30. Beimforde N, Hanke K, Ammar I, et al. Molecular cloning and functional characterization of the human endogenous retrovirus K113. *Virology* 2008;371(1):216–225.
31. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet* 2008; 42:709–732.
32. Contreras-Galindo R, Lopez P, Velez R, et al. HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) *in vitro*. *AIDS Res Hum Retroviruses* 2007; 23(1):116–122.
33. Contreras-Galindo R, Kaplan MH, Markovitz DM, et al. Detection of HERV-K(HML-2) viral RNA in plasma of HIV type 1-infected individuals. *AIDS Res Hum Retroviruses* 2006;22(10): 979–984.
34. Ogata T, Okui N, Sakuma R, et al. Integrase of human endogenous retrovirus K-10 supports the replication of replication-incompetent Int-human immunodeficiency virus type 1 mutant. *Jpn J Infect Dis* 1999;52(6):251–252.
35. Zeilfelder U, Frank O, Sparacio S, et al. The potential of retroviral vectors to cotransfer human endogenous retroviruses (HERVs) from human packaging cell lines. *Gene* 2007;390(1–2): 175–179.
36. Sakai H, Siomi H, Shida H, et al. Functional comparison of transactivation by human retrovirus rev and rex genes. *J Virol* 1990;64(12): 5833–5839.
37. Martinez-Molina E, Chocarro-Wrona C, Martinez-Moreno D, et al. Large-scale production of lentiviral vectors: Current perspectives and challenges. *Pharmaceutics* 2020;12(11): 1051; doi: 10.3390/pharmaceutics12111051
38. Aslan C, Kiaie SH, Zolbanin NM, et al. Exosomes for mRNA delivery: A novel biotherapeutic strategy with hurdles and hope. *BMC Biotechnol* 2021;21(1):20.
39. Do Minh A, Star AT, Stupak J, et al. Characterization of extracellular vesicles secreted in lentiviral producing HEK293SF cell cultures. *Viruses* 2021;13(5); doi: 10.3390/v13050797

Received for publication July 29, 2024;
accepted after revision December 19, 2024.

Published online: February 20, 2025.