YMTHE, Volume *28*

Supplemental Information

Efficient Nuclease-Directed Integration of Lentivirus Vectors into the Human Ribosomal DNA Locus

Diana Schenkwein, Saira Afzal, Alisa Nousiainen, Manfred Schmidt, and Seppo Ylä-Herttuala

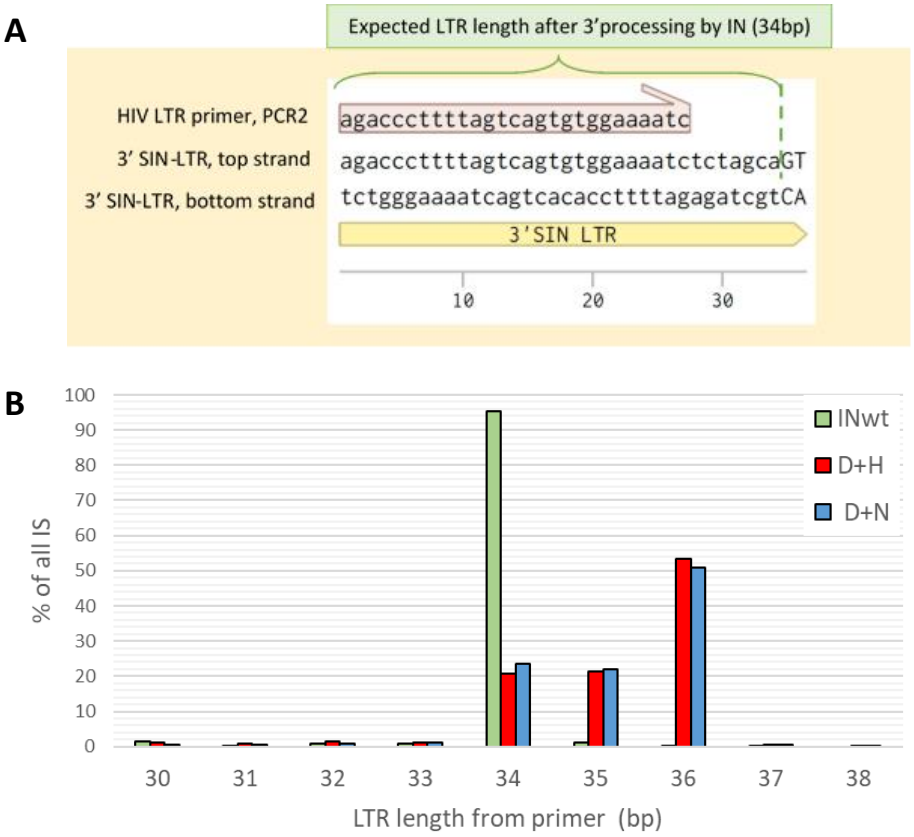

Figure S1: An illustration of the 3'LTR and analysis of LTR length in the different vector groups. A) An illustration of the 3'LTR present in the sequencing reads. The nested primer used in integration site extraction is shown on the top of the image. After IN-catalyzed processing of the LTRs and subsequent integration, the LTRs are expected to lack the terminal 3'GT-dinucleotide. After NHEJ-driven integration of the vector cDNA, the GT dinucleotide is frequently present. Vector integration after DNA repair through NHEJ is often accompanied with small insertions and deletions (indel mutations) at the cleaved site and at the termini of the inserted molecule, which can result in shorter and longer LTRs than expected. **B)** The proportions of LTRs with different lengths in the complete IS data of the three LV groups. LTR, long terminal repeat; SIN, self-inactivating LTR.

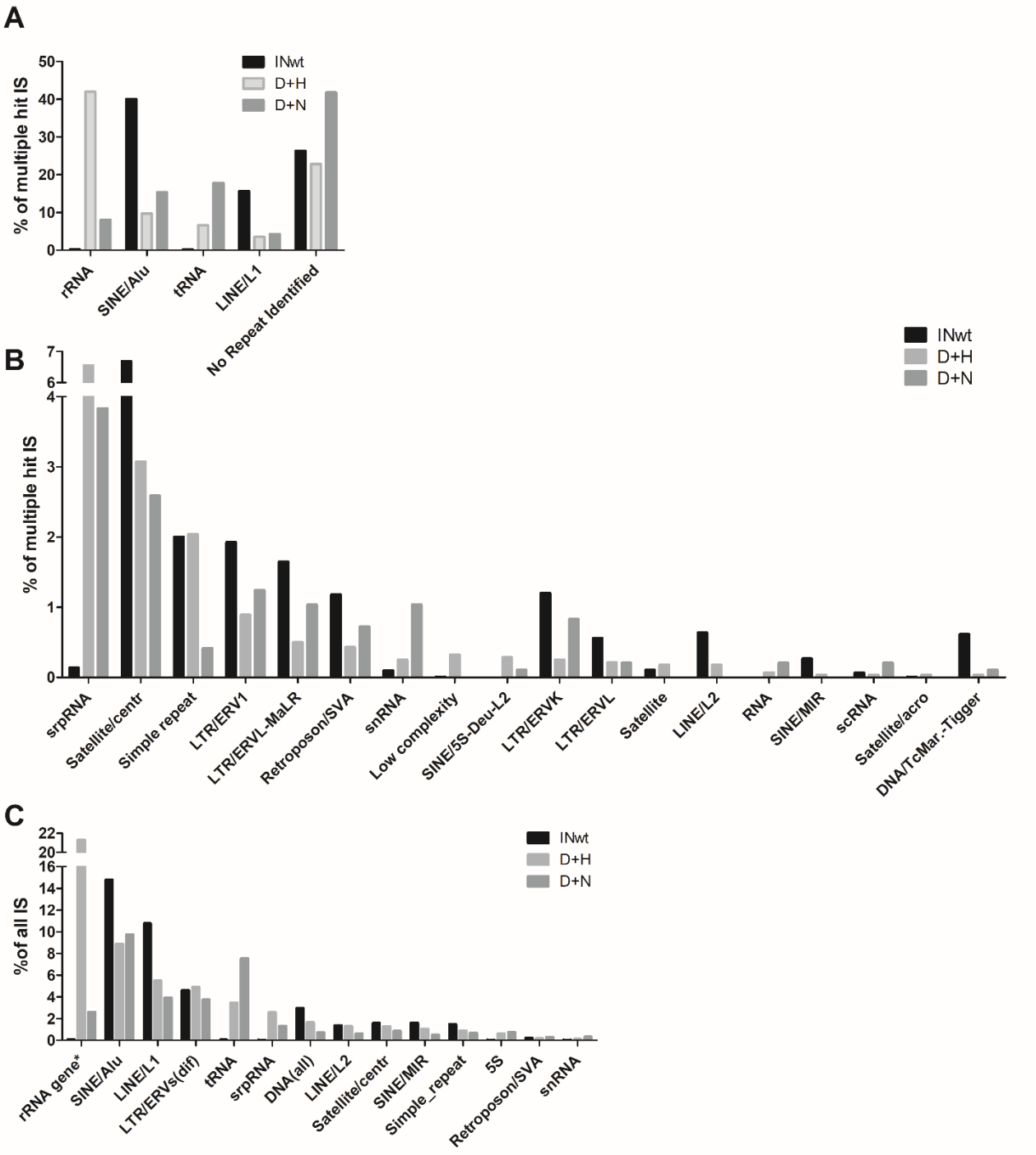

Figure S2: Localization of multiple hit (MH)-integration sites in different repeat classes and the frequency of integration into different repeats in the total IS data. **A**: Integration frequency in different repeat types in the MH-IS data (high frequency integration); **B**: Integration frequency in different repeat types of the MH-IS data (low frequency integration). **C**: Integration frequency within different repeat types in the total IS data of the LVs. In A and B the repeat identification and annotations from RepeatMasker were used. In C, rRNA gene repeats were annotated manually (marked with *) due to the inability of the RepeatMasker to identify other rRNA gene features than the genes encoding for the molecules incorporated into mature ribosomes (See Figure 1 for pre-rRNA gene composition).

Figure S3: Illustration of the genomic region where targeted integration was detected with ddPCR. The I-PpoI site on chromosome 21 (Chr21:8444914-8444917) is shown with a green label and a blue box, and the primer annealing 32 bp downstream of the I-PpoI site is visible on the lowest sequence row. The cleavage sites for the BsuRI restriction enzyme, used to digest the genomic DNA prior to ddPCR, are show. The length of the genomic region shown corresponds to the maximum length of the ddPCR product based on the extension time of the program.

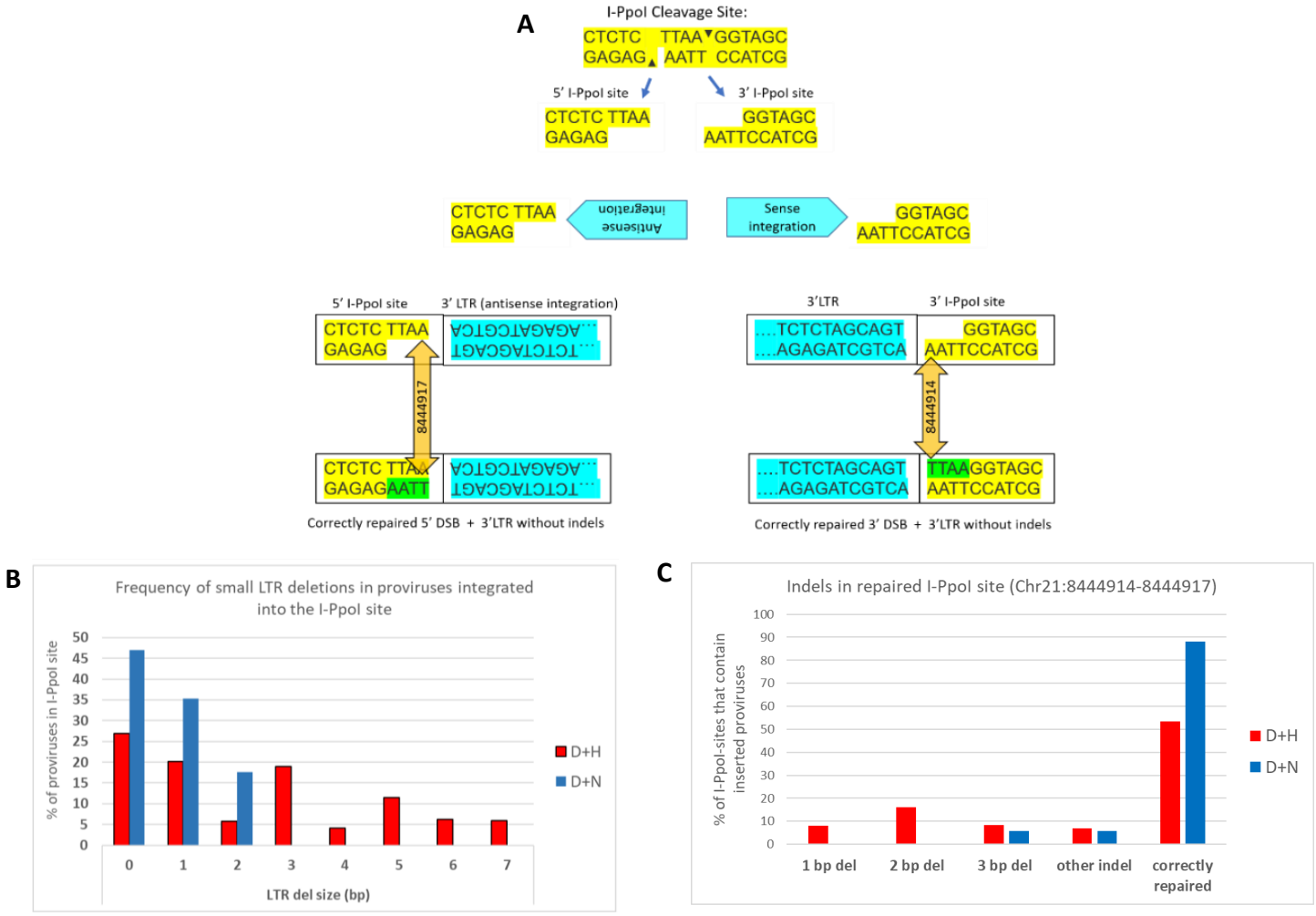

Figure S4: Characterization of indels in proviruses inserted into the cleaved I-PpoI site on chromosome 21 (Chr21:8444914-8444917) A) An illustration of provirus integration into the cleaved I-PpoI site with the genomic coordinates of the cleaved and repaired I-PpoI site nucleotides shown (orange arrows). The sequence of the unprocessed 3'LTR without deletions is highlighted with turquoise. **B)** Frequency of small deletions in the 3' LTRs of the inserted proviruses, where processing of the 3'GT dinucleotide by IN is not expected. **C)** Characterization of the I-PpoI site after vector insertion into the site. After error-free repair of the 5' or 3' I-PpoI sites the genomic sequence after the provirus matches exactly the nucleotides highlighted in yellow and green at the bottom of A.

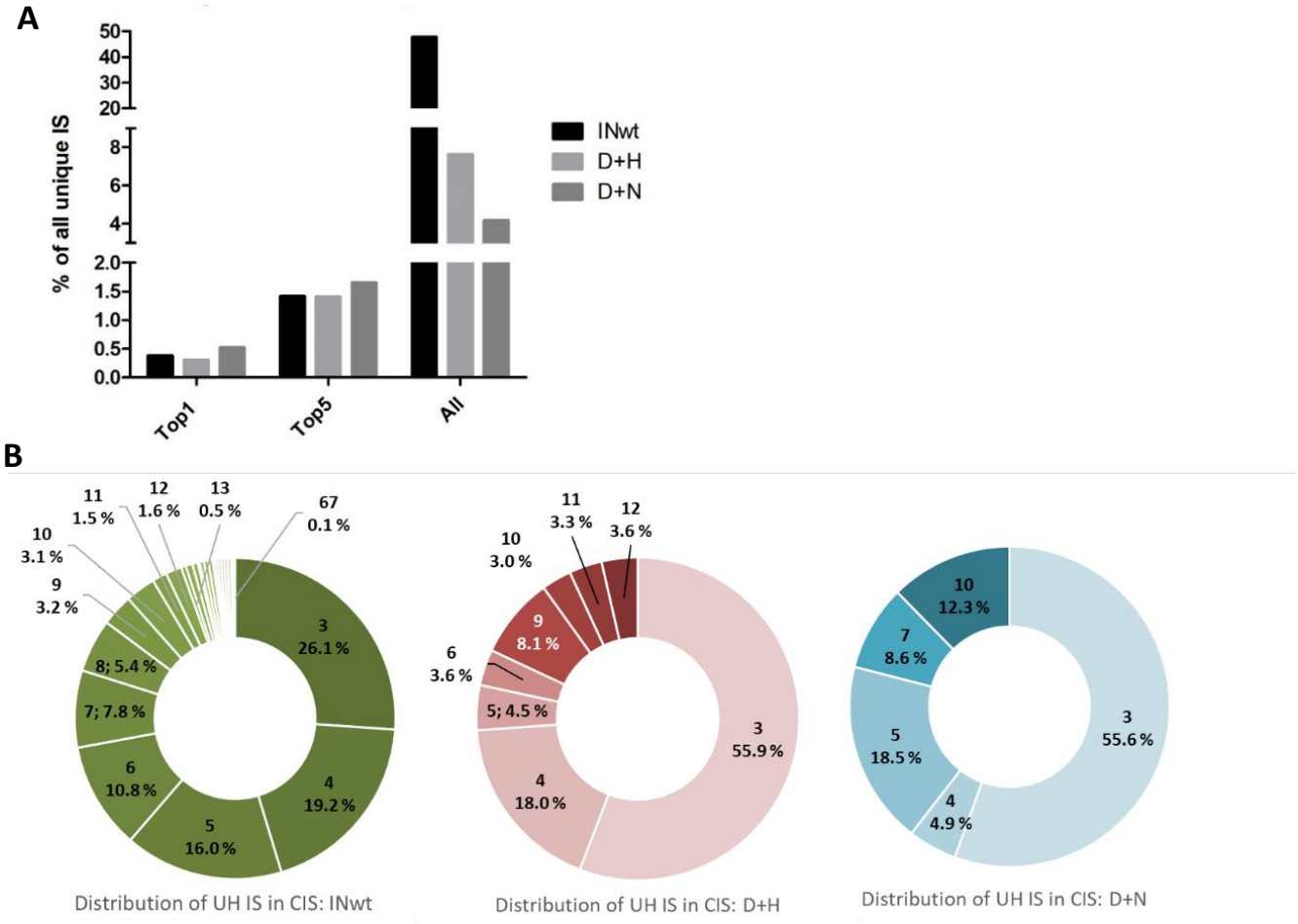

Figure S5: Characterization of integration site localization into common insertion sites (CIS). A) The proportion of CIS-associated IS of all unique IS in the strongest identified hotspot (Top1), in the five most targeted hotspots (Top5) and from all unique IS (All). **B)** Distribution of unique IS into CIS of different orders. The number shown above the percentage value is the CIS order (strength) that equals the number of IS within a CIS. For INwt LVs, CIS of the orders 14 to 66 are not shown for clarity. The percentages denote the fraction of all IS within CIS of the specified order of all CIS-associated IS.

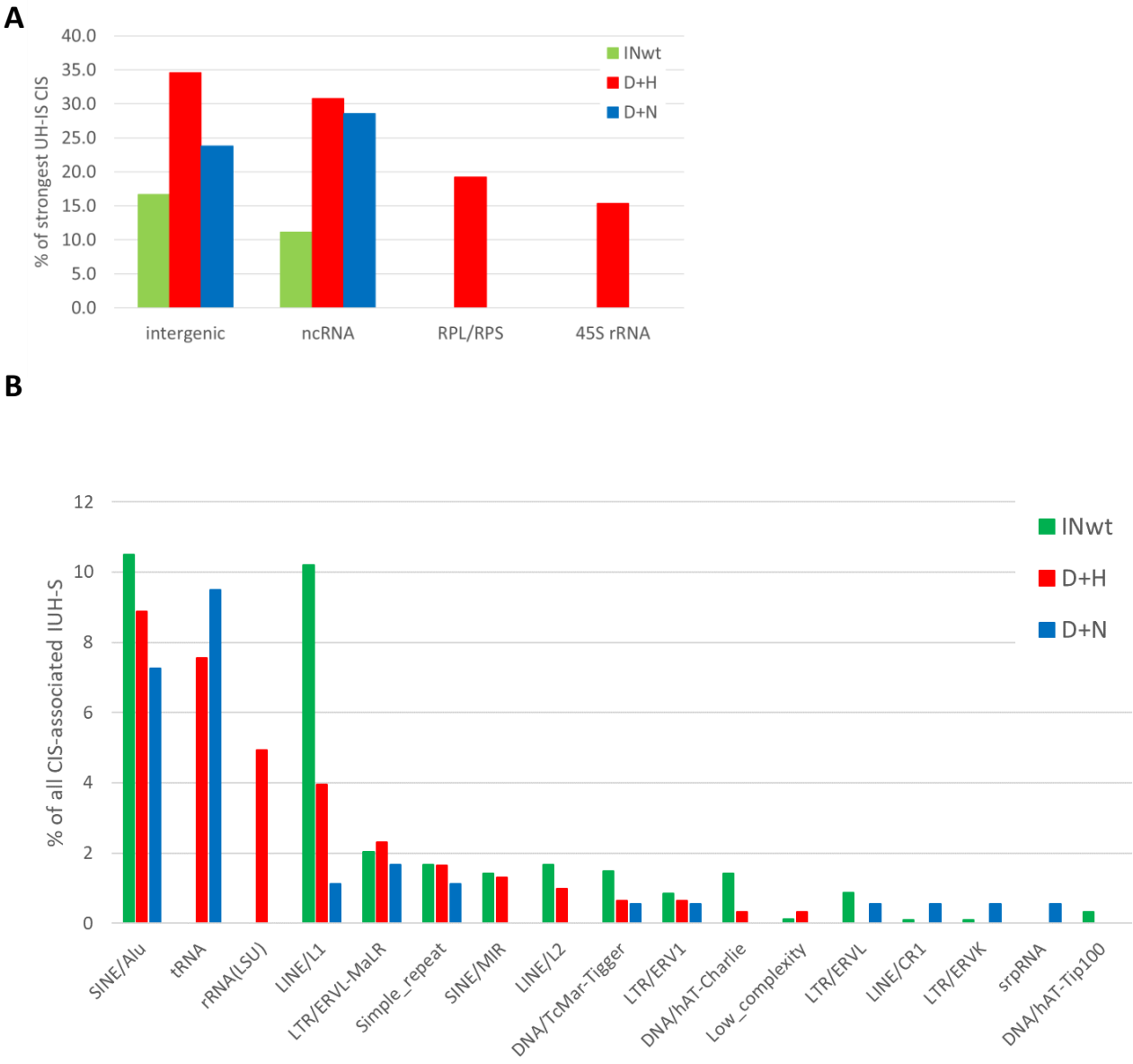

Figure S6: **Characterization of the differences in preferential integration target site selection between LVs containing the D+H fusion protein, the D+N fusion protein or INwt. A)** The proportion of selected features present in the ~15 most targeted integration hotspots among the unique IS (UH-IS) (Table 1) of the different LVs. **B)** The proportion of different repeat types present in CIS-associated unique IS reads. The numbers of CIS-contained IS are: 8450 for LV INwt; 333 for LV D+H and 81 for LV D+N. CIS: common integration site; LSU: ribosome large subunit -contained rRNA (28S rRNA gene).

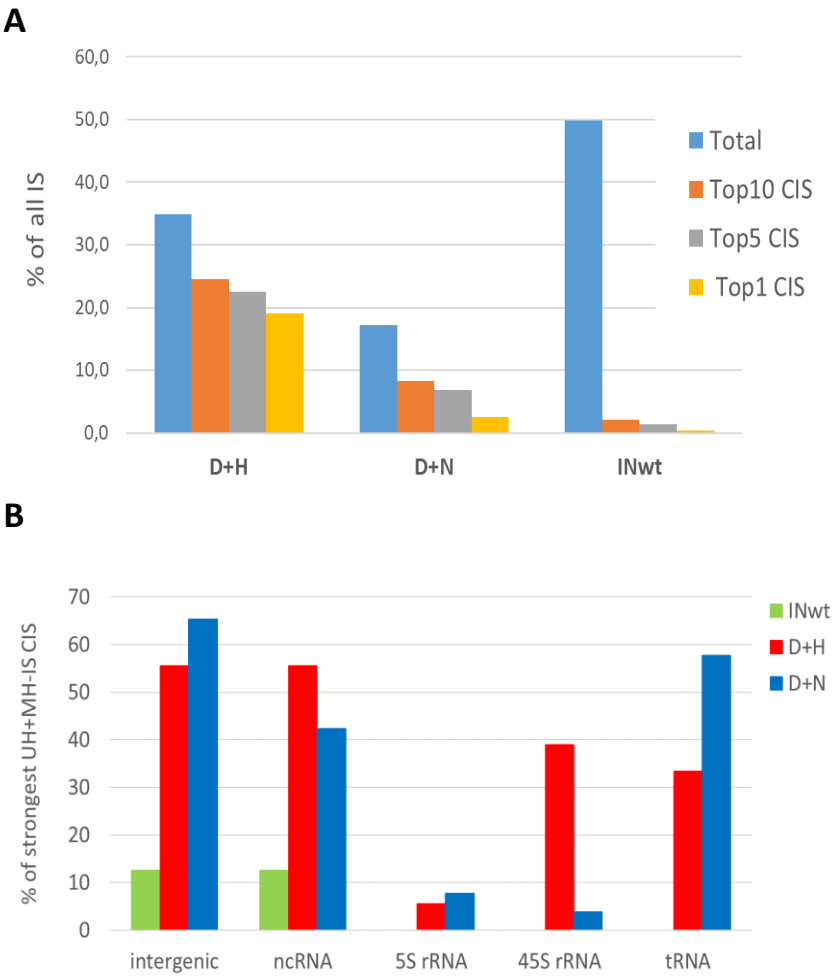

Figure S7: Characterization of the tendency of different LVs to form integration hotspots when both unique and multiple hit integration sites (UH and MH IS, respectively) are considered, and the differences between LVs in targeting specific features for integration. **A)** The proportion of CIS-associated IS of all IS in the strongest identified hotspot (Top1 CIS), in the ten, five or first most targeted hotspots (Top10 CIS, Top5 CIS and Top1 CIS, respectively) and from all IS (Total). **B)** The proportion of selected features present in the ~15 most targeted integration hotspots analyzed from the complete IS data of the different LVs (Table 2). The numbers of CIS-associated IS are 2506 for LV D+H; 498 for LV D+N and 10367 for LV INwt.

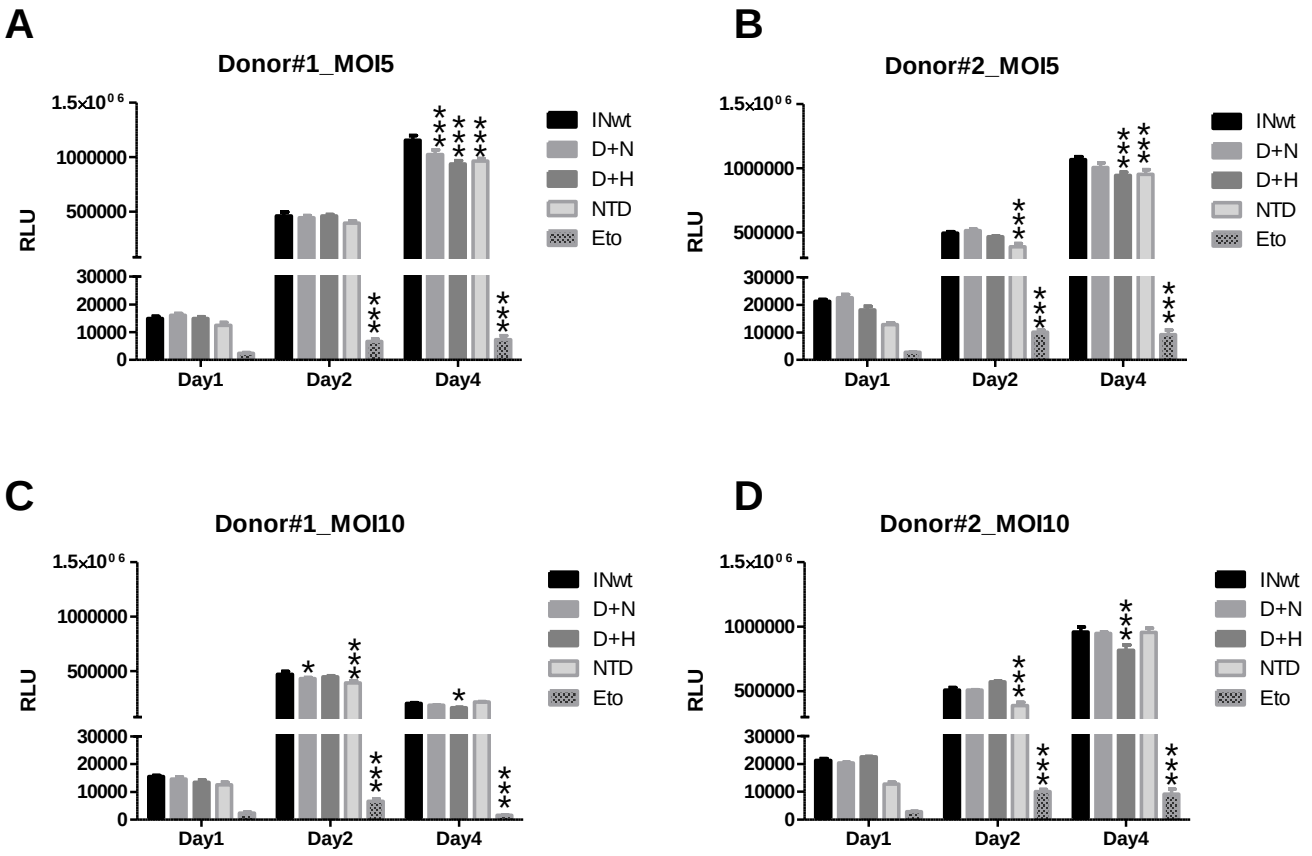

Figure S8: Viability of primary human T cells transduced with different LVs at days one to four post transduction. A) and B): viability of T cells extracted from donors 1 (A) and 2 (B) transduced with 5k vp/cell. C) and D): viability of T cells extracted from donors 1 (C) and 2 (D) transduced with 10k vp/cell. All statistical comparisons were done by comparing other groups to the INwt control. *:$p<0.05$; ***:$p<0.001$; RLU: relative light unit; Eto: etoposide; NTD: non-transduced cells. The RLU values represent means (with SEM) of measurements from triplicate wells.

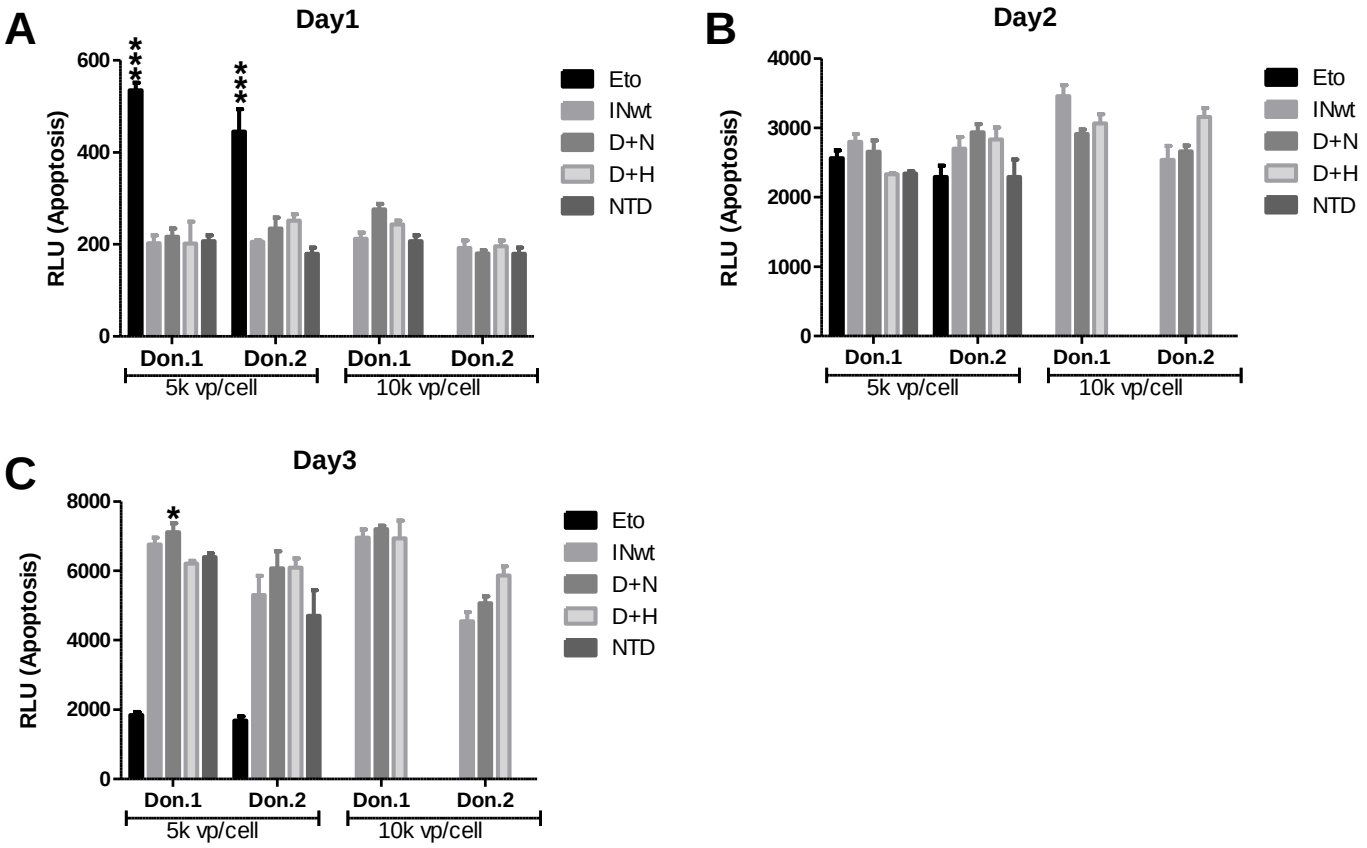

Figure S9: Measurement of apoptosis in primary human T cells transduced with different LVs. Apoptotic cells were detected at days one (A), two (B) and three (C) post transduction from primary T cells transduced with 5k and 10k vp/cell (shown below the X-axis). All vector-transduced cells were compared to the non-transduced control (NTD) analyzing each time point separately. *:$p<0.05$; ***:$p<0.001$; RLU: relative light unit; Eto: etoposide; NTD: non-transduced cells; vp: vector particle; Don: donor. The RLU values represent means (with SEM) of measurements from triplicate wells.

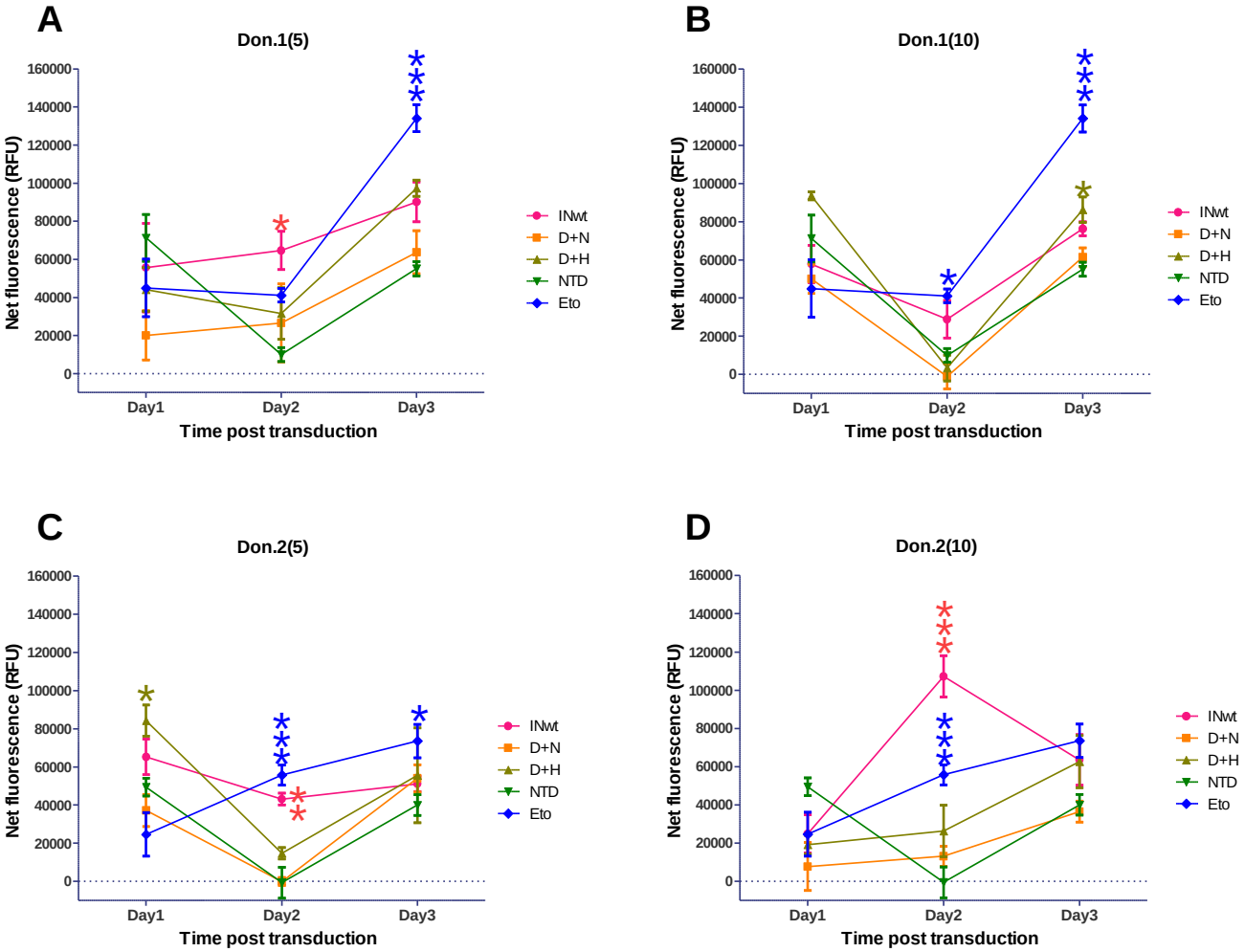

Figure S10: Measurement of necrosis at days one to three post transduction in primary human T cells transduced with different LVs. A: Detection of necrotic cells in donor 1-derived T cells transduced with 5k vector particles (vp) per cell. B: Detection of necrotic cells in donor 1-derived T cells transduced with 10k vp/cell. C: Detection of necrotic cells in donor 2-derived T cells transduced with 5k vp/cell. D: Detection of necrotic cells in donor 2-derived T cells transduced with 10k vp/cell. All vector groups were compared to the non-transduced (NTD) cell control. *:p<0.05; **:p<0.01; ***:p<0.001; RFU: relative fluorescence unit; Eto: etoposide; NTD: non-transduced cells; Don., donor. The values shown represent means (with SEM) of net fluorescence from triplicate wells per vector group.

Table S1: Full 15 bp I-PpoI recognition and cleavage sites in the human genome (Dec. 2013 GRCh38/hg38).

i: intron e: exon. LSU: large subunit of the ribosome that contains the 28S rRNA.

Chr	Locus	Gene	i/e	strand	Repeat	Nearest gene
1	chr1:57987624-57987610	DAB1	i1	minus	na	
1	chr1:237603118-237603132	RYR2	i35	plus	LSU rRNA	
2	chr2:132279870-132279856	intergenic		minus	LSU rRNA	ANKRD30BL
3	chr3:56330051-56330037	ERC2	i2	minus	na	
7	chr7:69062508-69062522	intergenic		plus	LSU rRNA	LOC100507468
8	chr8:69690276-69690262	SLCO5A1	i6	minus	LSU rRNA	
11	chr11:77886539-77886553	INTS4/AAMDC	i21/i4,5,6	plus	LSU rRNA	
20	chr20:30512873-30512859	intergenic		minus	LSU rRNA	MLLT10P1 (/5.8S rRNA/ FRG1BP)
21	**chr21:8217639-8217653**	**RNA28SN2**	**e1**	**plus**	**LSU rRNA**	
21	**chr21:8400677-8400691**	**RNA28SN3**	**e1**	**plus**	**LSU rRNA**	
21	**chr21:8444909-8444923**	**RNA28SN1**	**e1**	**plus**	**LSU rRNA**	
X	chrx:109054229-109054243	intergenic		plus	LSU rRNA	MIR6087

Table S2: Details of the starting material used for integration site sequencing.

LV	Integrase content in LV	MOI	% EGFP+ cells*	Sampling time point (days post td.)	ng of gDNA in MuA rxn	Linker #	MID #
D+H	IN_{D64V}+IN-I-$PpoI_{H78A}$	4	93,05	2	242,34	4	13
				2	198	6	14
D+N	IN_{D64V}+ IN-I-$PpoI_{N119A}$	4	97,02	3	129,78	7	15
				3	247,02	8	16
INwt	INwt	1	82,93	3	238,74	9	17
				3	163,98	10	18

LV: lentivirus vector; IN: integrase; td: transduction; gDNA: genomic DNA; rxn: reaction; MID: molecular identifier; MOI: multiplicity of infection.*: EGFP expression measured with flow cytometry from triplicate wells at the day of gDNA extraction.

Table S3: Localization of integration sites with respect oncogenes and their upstream and downstream regions.

	UH IS in or near oncogenes (% of all IS)	Intragenic IS		Intergenic IS			Med. oncogene Length. Kb
		IS within oncogenes (% of all IS)	Median dist. to TSS (kb)	% of interg. IS upstream of Oncog.	Median dist. (kb)to TSS of upstream IS	Median dist. (kb) to TSS of downstream IS	
INwt	15,8	13,2	55,4	57,2	52,0	40,6	123,0
D+H	10,4	8,1	54,2	53,7	37,2	39,8	115,2
D+N	12,0	9,5	40,1	49,3	34,1	55,3	99,5

UH IS: unique hit integration site. TSS: transcription start site. dist: distance. Med., median.

Table S4: Results of the ddPCR measurements of targeted integration within a 235 bp window around the 28S rRNA gene -contained I-PpoI site in MRC-5 cells.

		Copy number per cell							Targeted integrations of all integrated vector forms			
		Targeted 28S rRNA integration			All vector genomes	Episomal vector genomes	Production plasmid	Integrated vector copies	Targeted integration			
LV	Experiment (replicate)	28S int. Forw.	28S int. Rev.	28S int. Total	NHEJ	1-LTR	pLV	NHEJ-1-LTR-pLV	% of integrated	Average/ replicate	Average/ LV	LV
IN_{D64V}	1	0.000	0.00	0.00	0.26	0.20	0.01	0.05	**0.0 %**	**0.0 %**	**0.0 %**	IN_{D64V}
		0.000	0.00	0.00	0.10	0.08	0.00	0.02	**0.0 %**			
		0.000	0.00	0.00	0.14	0.12	0.01	0.01	**0.0 %**			
	2	0.000	0.00	0.00	0.66	0.46	0.54	-0.35	**0.0 %**	**0.0 %**		
		0.000	0.00	0.00	0.47	0.24	0.02	0.21	**0.1 %**			
		0.000	0.00	0.00	0.75	0.12	0.07	0.57	**0.0 %**			
INwt	1	0.017	0.01	0.03	13.67	0.88	0.06	12.73	**0.2 %**	**0.1 %**	**0.1 %**	INwt
		0.010	0.00	0.01	13.34	0.92	0.09	12.33	**0.1 %**			
		0.006	0.01	0.01	14.29	0.78	0.09	13.42	**0.1 %**			
	2	0.006	0.02	0.02	39.18	7.43	0.46	31.29	**0.1 %**	**0.1 %**		
		0.014	0.01	0.02	54.60	12.02	0.68	41.90	**0.1 %**			
		0.023	0.03	0.05	59.92	17.48	0.40	42.04	**0.1 %**			
D+N	1	0.001	0.00	0.00	1.73	1.01	0.13	0.59	**0.1 %**	**0.2 %**	**0.2 %**	D+N
		0.000	0.00	0.00	2.26	1.47	0.02	0.77	**0.0 %**			
		0.001	0.00	0.00	1.76	1.05	0.03	0.68	**0.6 %**			
	2	0.000	0.00	0.00	1.08	0.21	0.00	0.87	**0.0 %**	**0.1 %**		
		0.000	0.00	0.00	0.90	0.26	0.03	0.61	**0.2 %**			
		0.000	0.00	0.00	1.39	0.49	0.04	0.86	**0.0 %**			
D+H	1	0.047	0.05	0.10	3.41	3.19	0.05	0.17	**58.8 %**	**25.3 %**	**20.9 %**	D+H
		0.036	0.03	0.06	5.38	3.78	0.04	1.56	**4.1 %**			
		0.042	0.05	0.09	2.58	1.87	0.02	0.69	**13.0 %**			
	2	0.268	0.24	0.51	11.88	9.72	0.14	2.02	**25.1 %**	**16.4 %**		
		0.269	0.23	0.50	11.85	8.37	0.21	3.28	**15.1 %**			
		0.312	0.25	0.56	10.32	3.91	0.16	6.26	**9.0 %**			

NHEJ: Non-homologous end joining; int.: integration; 28S: 28S rRNA gene; Forw.: forward; Rev.:Reverse; LTR: long terminal repeat

Table S5: ddPCR-based measurement of targeted integration within a 235 bp window around the 28S rRNA gene -contained I-PpoI site in selected and unselected hTERT-RPE1 cells.

		Copy number per cell				Targeted integrations of all integrated vector forms		
		All vector genomes	28S integrations	Episomal vector genomes	Integrated vector copies	Targeted integration		
	Sample	WPRE	28S int. Forw.	1-LTR	WPRE-1-LTR	% of integrated	Average targeting %	Actual targeting %*
Unselected (d13 p.td)	D+H	0.52	0.01	0.20	0.32	4.3 %	**4.4 %**	**8.8 %**
		0.43	0.01	0.16	0.26	3.8 %		
		0.50	0.01	0.20	0.30	4.4 %		
		0.56	0.02	0.19	0.37	5.1 %		
	IN_{D64V}	0.21	0.00	0.13	0.08	0.0 %	**0.0 %**	**0.0 %**
		0.21	0.00	0.07	0.13	0.0 %		
		0.21	0.00	0.08	0.13	0.0 %		
		0.22	0.00	0.08	0.14	0.0 %		
Selected (d15 p.td)	D+H (replicate 1)	2.96	0.07	1.11	1.85	3.6 %	**4.2 %**	**8.4 %**
		3.12	0.08	0.79	2.33	3.4 %		
		2.56	0.08	0.88	1.68	5.0 %		
		2.78	0.07	1.22	1.56	4.7 %		
	D+H (replicate 2)	3.53	0.06	1.50	2.03	3.1 %	**3.3 %**	**6.6 %**
		3.51	0.07	1.46	2.06	3.5 %		
		3.42	0.07	1.40	2.02	3.7 %		
		4.00	0.07	1.59	2.41	2.8 %		
	IN_{D64V}	1.68	0.00	0.35	1.32	0.1 %	**0.0 %**	**0.1 %**
		1.67	0.00	0.31	1.36	0.0 %		
		1.64	0.00	0.22	1.41	0.0 %		
		1.71	0.00	0.32	1.39	0.0 %		

* Expected integration targeting efficiency (according to MRC-5 experiments) if 28S-targeted integration was also studied in reverse orientation; d: day; p.td: post transduction; LV: lentivirus vector

Table S6: Comparison of lentivirus vector common integration site genes in human mouse hematochimeras with the common integration sites identified in this study using the unique integration sites of LV INwt.

CIS-associated gene[23]	Alias gene name	Gene present in a CIS of LV INwt?	CIS order of LV INwt
PACS		yes	44
RAB40C		yes	33
HLA		yes (HLA-E)	27
NPLOC4		yes	25
SPDYC		yes	17
SAPS2	PPP6R2	yes	15
ZGPAT		yes	12
FBXL11	KDM2A	yes	9
ANKFY1		yes	8
RPA1		yes	8
SMYD4		yes	8
QRICH1		yes	8
USP48		yes	7
FCHSD2		yes	7
SMARCC1		yes	7
NSD1		yes	6
CENTD2	ARAP1	yes	4
FRYL		yes	4
CARD8		yes	4
EIF2C3	AGO3	yes	3
PSCD1	CYTH1	yes	3
NF1		yes	3
ABCA3		no	na
CBL		no	na
CDC27		no	na
HORMAD2		no	na
SP1		no	na
TAPBP		no	na
VAV1		no	na
WDR82		no	na
GPATCH8		no	na

LV: lentivirus vector; CIS: common integration site; na: not applicable.

Table S7: HIV-1 recurrent integration genes and LV INwt UH-CIS that are within a 100 kb distance from one another.

Table S8: DdPCR results of targeted integration detection in primary T cells at day two post transduction.

	Copy number per cell						Targeted	
Day 2 p. td.	**All vector genomes**	**Targeted 28S integration**		**Production plasmid**	**Episomal vector genomes**	**Integrated vector genomes**	**Targeted integration**	
Sample	**NHEJ**	**28Sint**	**Rev-28Sint**	**pLV**	**1-LTR**	**NHEJ-pLV-1-LTR**	**% targeted**	**Average**
Donor 1 NTD-1	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	0.0 %
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 1 NTD-2	0.01	0.00	0.00	0.00	0.00	0.01	0.0 %	
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 1 INwt, replicate 1, 5K VP/cell	28.31	0.01	0.01	0.47	4.32	23.52	0.1 %	0.0 %
	28.90	0.00	0.00	0.44	3.49	24.97	0.0 %	
Donor 1 INwt, replicate 2, 5K VP/cell	35.43	0.01	0.00	0.65	4.69	30.09	0.0 %	
	32.73	0.01	0.00	0.63	5.61	26.49	0.0 %	
Donor 1 INwt, replicate 1, 10K VP/cell	31.74	0.01	0.00	0.57	4.84	26.33	0.0 %	0.0 %
	35.85	0.00	0.00	0.59	4.53	30.73	0.0 %	
Donor 1 INwt, replicate 2, 10K VP/cell	32.83	0.01	0.00	0.47	4.19	28.17	0.0 %	
	33.50	0.01	0.00	0.45	4.47	28.58	0.0 %	
Donor 1 D+H, replicate 1, 5K VP/cell	17.89	0.02	0.03	0.51	7.53	9.85	0.5 %	0.4 %
	19.69	0.03	0.02	0.48	6.64	12.56	0.4 %	
Donor 1 D+H, replicate 2, 5K VP/cell	17.46	0.02	0.02	0.53	5.68	11.25	0.3 %	
	16.15	0.02	0.03	0.52	6.84	8.79	0.6 %	
Donor 1 D+H, replicate 1, 10K VP/cell	29.27	0.05	0.05	0.68	11.56	17.03	0.6 %	0.6 %
	31.36	0.04	0.07	0.73	11.19	19.43	0.6 %	
Donor 1 D+H, replicate 2, 10K VP/cell	28.12	0.05	0.05	0.63	11.06	16.44	0.6 %	
	27.72	0.04	0.05	0.64	12.51	14.57	0.6 %	
Donor 2 NTD-1	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	0.0 %
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 2 NTD-2	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 2 INwt, replicate 1, 5K VP/cell	15.26	0.00	0.00	0.29	1.51	13.46	0.0 %	0.0 %
	12.39	0.00	0.00	0.23	1.52	10.63	0.0 %	
Donor 2 INwt, replicate 2, 5K VP/cell	16.53	0.00	0.00	0.31	2.31	13.90	0.0 %	
	17.13	0.00	0.00	0.32	2.32	14.50	0.0 %	
Donor 2 INwt, replicate 1, 10K VP/cell	19.06	0.00	0.00	0.43	2.60	16.03	0.0 %	0.0 %
	17.37	0.00	0.00	0.37	1.76	15.24	0.0 %	
Donor 2 INwt, replicate 2, 10K VP/cell	18.79	0.00	0.00	0.28	2.00	16.50	0.0 %	
	17.70	0.00	0.00	0.35	1.82	15.53	0.0 %	
Donor 2 D+H, replicate 1, 5K VP/cell	10.10	0.01	0.01	0.29	3.19	6.62	0.4 %	0.3 %
	9.99	0.01	0.02	0.33	2.80	6.86	0.4 %	
Donor 2 D+H, replicate 2, 5K VP/cell	9.38	0.00	0.01	0.34	2.81	6.23	0.2 %	
	10.52	0.01	0.01	0.24	2.91	7.37	0.3 %	
Donor 2 D+H, replicate 1, 10K VP/cell	14.44	0.02	0.01	0.38	5.05	9.01	0.3 %	0.3 %
	15.44	0.02	0.01	0.50	4.04	10.90	0.3 %	
Donor 2 D+H, replicate 2, 10K VP/cell	26.00	0.02	0.02	1.16	6.92	17.92	0.2 %	
	21.97	0.02	0.01	1.12	7.05	13.81	0.2 %	

p. td: post transduction; vp: vector particle; NTD: non-transduced cells

Table S9: DdPCR results of targeted integration detection in primary T cells at day ten post transduction.

	Copy number per cell						Targeted	
Day 10 p.td.	All vector genomes	Targeted 28S integration		Production plasmid	Episomal vector genomes	Integrated vector genomes	Targeted integration	
Sample	NHEJ	28Sint	Rev-28Sint	pLV	1-LTR	NHEJ-pLV-1-LTR	% targeted	Average
Donor 1 NTD-1	0.01	0.00	0.00	0.00	0.00	0.01	0.0 %	0.0 %
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 1 NTD-2	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 1 INwt, replicate 1, 5K VP/cell	6.10	0.00	0.00	0.04	0.19	5.87	0.1 %	0.1 %
	6.51	0.00	0.00	0.02	0.26	6.23	0.1 %	
Donor 1 INwt, replicate 2, 5K VP/cell	6.75	0.00	0.00	0.02	0.24	6.49	0.1 %	
	6.51	0.00	0.00	0.04	0.26	6.21	0.1 %	
Donor 1 INwt, replicate 1, 10K VP/cell	8.50	0.00	0.00	0.03	0.34	8.13	0.0 %	0.0 %
	8.71	0.00	0.00	0.03	0.31	8.37	0.1 %	
Donor 1 INwt, replicate 2, 10K VP/cell	8.37	0.00	0.00	0.04	0.37	7.95	0.1 %	
	7.77	0.00	0.00	0.05	0.24	7.49	0.1 %	
Donor 1 D+H, replicate 1, 5K VP/cell	0.35	0.00	0.00	0.01	0.12	0.22	1.3 %	4.4 %
	0.33	0.01	0.01	0.01	0.09	0.24	4.7 %	
Donor 1 D+H, replicate 2, 5K VP/cell	0.29	0.01	0.01	0.01	0.10	0.19	7.8 %	
	0.27	0.00	0.01	0.00	0.08	0.19	3.6 %	
Donor 1 D+H, replicate 1, 10K VP/cell	0.47	0.00	0.01	0.01	0.15	0.30	4.3 %	4.8 %
	0.49	0.01	0.01	0.00	0.15	0.33	4.4 %	
Donor 1 D+H, replicate 2, 10K VP/cell	0.53	0.01	0.01	0.00	0.14	0.39	5.6 %	
	0.51	0.01	0.01	0.01	0.16	0.35	5.1 %	
Donor 2 NTD-1	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	0.0 %
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 2 NTD-2	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
	0.00	0.00	0.00	0.00	0.00	0.00	0.0 %	
Donor 2 INwt, replicate 1, 5K VP/cell	2.60	0.00	0.00	0.00	0.12	2.48	0.1 %	0.1 %
	2.40	0.00	0.00	0.00	0.08	2.32	0.0 %	
Donor 2 INwt, replicate 2, 5K VP/cell	2.37	0.00	0.00	0.00	0.12	2.25	0.2 %	
	2.41	0.00	0.00	0.02	0.10	2.29	0.0 %	
Donor 2 INwt, replicate 1, 10K VP/cell	3.21	0.00	0.00	0.03	0.14	3.04	0.0 %	0.1 %
	3.03	0.00	0.00	0.04	0.07	2.91	0.0 %	
Donor 2 INwt, replicate 2, 10K VP/cell	3.26	0.00	0.00	0.01	0.12	3.13	0.1 %	
	3.64	0.00	0.01	0.06	0.09	3.49	0.2 %	
Donor 2 D+H, replicate 1, 5K VP/cell	0.14	0.00	0.00	0.00	0.03	0.12	5.0 %	2.7 %
	0.26	0.00	0.00	0.00	0.07	0.19	0.7 %	
Donor 2 D+H, replicate 2, 5K VP/cell	0.13	0.00	0.00	0.01	0.07	0.05	0.0 %	
	0.10	0.00	0.00	0.00	0.04	0.06	5.2 %	
Donor 2 D+H, replicate 1, 10K VP/cell	0.15	0.00	0.00	0.01	0.03	0.12	5.7 %	3.9 %
	0.11	0.00	0.00	0.00	0.01	0.10	0.0 %	
Donor 2 D+H, replicate 2, 10K VP/cell	0.34	0.00	0.01	0.02	0.12	0.20	3.7 %	
	0.29	0.01	0.00	0.01	0.14	0.14	6.0 %	

p. td: post transduction; vp: vector particle; NTD: non-transduced cells

Table S10: Quantification of the DJ region and the 18S rRNA gene copies in the genomes of primary T cells at day two post transduction.

Sample	CN DJ	Average DJ	CN 18SrRNA	Average 18SrRNA
Donor 1 NTD-1	17.31	17.63	594.53	501.93
	17.13		369.54	
Donor 1 NTD-2	18.63		514.97	
	17.46		528.68	
Donor 1 INwt, replicate 1, 10K VP/cell	18.51	17.85	676.39	682.76
	17.40		754.53	
Donor 1 INwt, replicate 2, 10K VP/cell	18.14		456.37	
	17.36		843.77	
Donor 1 D+H, replicate 1, 10K VP/cell	18.98	16.54	535.81	701.19
	14.71		490.89	
Donor 1 D+H, replicate 2, 10K VP/cell	16.49		1 040.95	
	16.00		737.11	
Donor 2 NTD-1	13.23	13.14	484.96	478.25
	12.21		451.65	
Donor 2 NTD-2	15.05		614.73	
	12.10		361.67	
Donor 2 INwt, replicate 1, 10K VP/cell	13.89	14.33	690.84	538.84
	13.62		740.72	
Donor 2 INwt, replicate 2, 10K VP/cell	14.15		370.03	
	15.66		353.77	
Donor 2 D+H, replicate 1, 10K VP/cell	12.98	13.85	607.67	659.56
	12.90		649.42	
Donor 2 D+H, replicate 2, 10K VP/cell	14.48		862.98	
	15.05		518.17	

CN: copy number; DJ: distal junction; vp: vector particle; NTD: non-transduced cells.

Table S11: RT-ddPCR measurements of provirus transcripts originating from the 28S rRNA gene locus in primary T cells at day two post transduction.

	Gene expression ratio to control assay				Gene expression ratio comparison
Day 2 p. td.	**Total provirus expression**		**Provirus transcripts from the 28S rRNA locus (sense-orientation integration)**		**28S rRNA locus transcripts of total provirus transcripts**
Sample	**Ratio (WPRE/IPO8)**	**Average (WPRE/IPO8)**	**Ratio (28Sint/IPO8)**	**Average (28Sint/IPO8)**	**28Sint-ratio/ WPRE-ratio**
Donor 1 NTD-1	0.00	**0.00**	0.00	**0.00**	**0.0 %**
	0.00		0.00		
Donor 1 NTD-2	0.00		0.00		
	0.00		0.00		
Donor 1 INwt, replicate 1, 5K VP/cell	19.51	**25.36**	0.00	**0.00**	**0.0 %**
	2 030.04		0.00		
Donor 1 INwt, replicate 2, 5K VP/cell	28.66		0.00		
	27.93		0.00		
Donor 1 INwt, replicate 1, 10K VP/cell	*1 875.80*	**22.70**	0.00	**0.00**	**0.0 %**
	1 893.12		0.00		
Donor 1 INwt, replicate 2, 10K VP/cell	*2 116.56*		0.00		
	22.70		0.00		
Donor 1 D+H, replicate 1, 5K VP/cell	1.81	**2.64**	0.01	**0.00**	**0.1 %**
	1.82		0.00		
Donor 1 D+H, replicate 2, 5K VP/cell	3.15		0.00		
	3.77		0.00		
Donor 1 D+H, replicate 1, 10K VP/cell	3.19	**3.04**	0.03	**0.03**	**0.8 %**
	3.21		0.03		
Donor 1 D+H, replicate 2, 10K VP/cell	2.79		0.02		
	2.96		0.02		
Donor 2 NTD-1	0.01	**0.00**	0.00	**0.00**	**0.0 %**
	0.00		0.00		
Donor 2 NTD-2	0.00		0.00		
	0.00		0.00		
Donor 2 INwt, replicate 1, 5K VP/cell	14.07	**12.71**	0.00	**0.00**	**0.0 %**
	14.72		0.00		
Donor 2 INwt, replicate 2, 5K VP/cell	11.03		0.00		
	11.01		0.00		
Donor 2 INwt, replicate 1, 10K VP/cell	16.67	**14.54**	0.00	**0.00**	**0.0 %**
	14.44		0.00		
Donor 2 INwt, replicate 2, 10K VP/cell	11.21		0.00		
	15.83		0.00		
Donor 2 D+H, replicate 1, 5K VP/cell	1.45	**1.25**	0.01	**0.01**	**0.4 %**
	1.42		0.01		
Donor 2 D+H, replicate 2, 5K VP/cell	1.02		0.00		
	1.10		0.00		
Donor 2 D+H, replicate 1, 10K VP/cell	1.47	**2.39**	0.01	**0.01**	**0.3 %**
	1.43		0.01		
Donor 2 D+H, replicate 2, 10K VP/cell	3.35		0.01		
	3.30		0.01		

p. td: post transduction; vp: vector particle; NTD: non-transduced cells

value too high to reliably quantitate; not included in average

Table S12: RT-ddPCR measurements of provirus transcripts originating from the 28S rRNA gene locus in primary T cells at day ten post transduction.

	Gene expression ratio to control assay				Gene expression ratio comparison
Day 10 p.td.	**Total provirus expression**		**Provirus transcripts from the 28S rRNA locus (sense-orientation integration)**		**28S rRNA locus transcripts of total provirus transcripts**
Sample	**Ratio (WPRE/IPO8)**	**Average (WPRE/IPO8)**	**Ratio (28Sint/IPO8)**	**Average (28Sint/IPO8)**	**28Sint-ratio/ WPRE-ratio**
Donor 1 NTD-1	0.00	**0.00**	0.00	**0.000**	**0.0 %**
	0.00		0.00		
Donor 1 NTD-2	0.00		0.00		
	0.00		0.00		
Donor 1 INwt, replicate 1, 5K VP/cell	11.60	**12.70**	0.00	**0.000**	**0.0 %**
	11.01		0.00		
Donor 1 INwt, replicate 2, 5K VP/cell	12.54		0.00		
	13.97		0.00		
Donor 1 INwt, replicate 1, 10K VP/cell	31.27	**23.67**	0.00	**0.000**	**0.0 %**
	29.89		0.00		
Donor 1 INwt, replicate 2, 10K VP/cell	16.24		0.00		
	17.31		0.00		
Donor 1 D+H, replicate 1, 5K VP/cell	0.63	**0.63**	0.01	**0.013**	**2.0 %**
	0.61		0.01		
Donor 1 D+H, replicate 2, 5K VP/cell	0.65		0.01		
	0.64		0.01		
Donor 1 D+H, replicate 1, 10K VP/cell	1.67	**1.41**	0.00	**0.007**	**0.5 %**
	1.65		0.00		
Donor 1 D+H, replicate 2, 10K VP/cell	1.19		0.01		
	1.14		0.01		
Donor 2 NTD-1	0.01	**0.00**	0.00	**0.000**	**0.0 %**
	0.00		0.00		
Donor 2 NTD-2	0.00		0.00		
	0.00		0.00		
Donor 2 INwt, replicate 1, 5K VP/cell	4.58	**6.88**	0.00	**0.000**	**0.0 %**
	4.61		0.00		
Donor 2 INwt, replicate 2, 5K VP/cell	8.84		0.00		
	9.46		0.00		
Donor 2 INwt, replicate 1, 10K VP/cell	9.25	**7.58**	0.00	**0.000**	**0.0 %**
	9.01		0.00		
Donor 2 INwt, replicate 2, 10K VP/cell	6.16		0.00		
	5.91		0.00		
Donor 2 D+H, replicate 1, 5K VP/cell	0.65	**0.72**	0.00	**0.001**	**0.2 %**
	0.59		0.00		
Donor 2 D+H, replicate 2, 5K VP/cell	0.75		0.00		
	0.90		0.00		
Donor 2 D+H, replicate 1, 10K VP/cell	0.00	**3.83**	0.04	**0.009**	**0.2 %**
	0.18		0.00		
Donor 2 D+H, replicate 2, 10K VP/cell	7.41		0.00		
	7.74		0.00		

p. td: post transduction; vp: vector particle; NTD: non-transduced cells

File S1: CIS analysis of all LVs.

File S2: Enriched GO terms in the CIS-associated genes of the analyzed LVs.

Supplemental Methods

Primer and linker sequences used for the extraction of LV integration sites.

Primer Name	Primer Sequence (5'-3')
ForwA + MID13 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATAGTAGTGAGACCCTTTTAGTCAGTGTGGAAAATC
ForwA + MID14 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGAGATACAGACCCTTTTAGTCAGTGTGGAAAATC
ForwA + MID15 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACGACGTAAGACCCTTTTAGTCAGTGTGGAAAATC
ForwA + MID16 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACGTACTAAGACCCTTTTAGTCAGTGTGGAAAATC
ForwA + MID17 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTCTAGTACAGACCCTTTTAGTCAGTGTGGAAAATC
ForwA + MID18 HIV LTR primer, PCR2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTACGTAGCAGACCCTTTTAGTCAGTGTGGAAAATC
Rev_P1b+ad4_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGACATTGCTTCTTCCCACTAGAG
Rev_P1b+ad6_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAACTTGCACTTCTGACCTAGCT
Rev_P1b+ad7_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAGACGAGTCAGTCCTACTAAAG
Rev_P1b+ad8_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAACGCGAGCCAGACTCCATATT
Rev_P1b+ad9_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATCGCTAGAGTACGGCCTTGAA
Rev_P1b+ad10_PCR2	CACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATATTGAGAGAGGGAAAGAGGC
L4 PCR1 primer	TTCAGGAGGTCACTTCGCACAT
L6 PCR1 primer	TAGACCGCTCAGAGGTCATACT
L7 PCR1 primer	CATCGTCGACACACGTGATGAC
L8 PCR1 primer	TATGCGGGACAGGTAATACGCG
L9 PCR1 primer	GGAATCTATGTAGCAGGTCGCT
L10 PCR1 primer	CGCTTTGAGCTATGAACCCTAT
MuL4 anneal	TTCAGGAGGTCACTTCGCACATTGCTTCTTCCCACTAGAGTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
MuL6 anneal	TAGACCGCTCAGAGGTCATACTTGCACTTCTGACCTAGCTTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
MuL7 anneal	CATCGTCGACACACGTGATGACGAGTCAGTCCTACTAAAGTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
MuL8 anneal	TATGCGGGACAGGTAATACGCGAGCCAGACTCCATATTTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
MuL9 anneal	GGAATCTATGTAGCAGGTCGCTAGAGTACGGCCTTGAATGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
MuL10 anneal	CGCTTTGAGCTATGAACCCTATTGAGAGAGGGAAAGAGGCTGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
Mu -- Donor	TCGGATGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACA/3AmMC7/
HIVLTR primer,PCR1	CTTAAGCCTCAATAAAGCTTGCCTTGAG

Details of the materials used in ddPCR.

Product	Bio-Rad Cat. No.	Manufacturing origin
Droplet generation oil for probes	1863005	USA
Droplet reader oil	1863004	USA
DG8™ Cartridges for QX200™/QX100™ Droplet Generator	1864008	Germany
DG8™ Gaskets for QX200™/QX100™ Droplet Generator	1863009	USA
ddPCR™ 96-Well Plates	12001925	USA
Piercable foil heat seal	1814040	UK
Supermix for probes (no dUTP)	1863025	USA

Primers and design of the ddPCR and RT-ddPCR assays used to estimate integration targeting near the I-PpoI site and transcription from the 28S rRNA gene locus.

Primer Name	Primer Sequence (5'-3')	Assay	Used in assay to detect
28Sint_FW	GCTCTCTGGCTAACTAGGGAA	28S int. Forw.	Transgene integration in the I-PpoI recognition site in the 28S rRNA gene (sense orientation); detection of transgene transcripts from the 28S rRNA gene locus with RT-ddPCR
28Sint_REV	GTTCATCCATTCATGCGCG		
28Sint_int	TGTGCCCGTCTGTTGTGTGACTCTGGT		
Rev-28Sint_FW	AGCAGTGGGTTCCCTAGTTA	28S int.Rev.	Transgene integration in the I-PpoI recognition site in the 28S rRNA gene (antisense orientation)
Rev-28Sint_REV	GTTCATCCATTCATGCGCG		
Rev-28Sint_int	CCAGAGAGCTCCCAGGCTCAGATCTGG		
1-LTR FW	GCTCGGTACCTTTAAGACCA	1-LTR	Episomal vector genomes
1-LTR REV	GTTTCCCTTTCGCTTTCAGG		
1-LTR int	AGTCAGTGTGGAAAATCTCTAGCAGTG		
NHEJ_fw	GGAAAATCTCTAGCAGTGGC	NHEJ	All vector genomes (MRC-5 and T cell integration targeting efficiency measurements)
NHEJ_rev	CCCGCTTAATACTGACGCT		
NHEJ_int	GCAAGAGGCGAGGGGCGGCG		
pLV-fw	GCCTTGAGTGCTTCAAGTAG	pLV	Production plasmid carry-over (transgene construct)
pLV-rev	CAAGTTCCTCTCACTCTCTG		
pLV-int	TGTGCCCGTCTGTTGTGTGACTCTGGT		
WPRE_FW	CACTGACAATTCCGTGGTGT	WPRE	All vector genomes (hTERT-RPE1 integration targeting efficiency measurements; RT-ddPCR)
WPRE_REV	CAGAATCCAGGTGGCAACA		
WPRE_int	ACGTCCTTTCCATGGCTGCTCGCCT		
DJgRNA3_FW	CATTTCCCAGCTTCCAGGAT	DJ	Quantification of possible deletions of the distal junction (DJ) sequences
DJgRNA3_REV	AGGAGCTTGGGATCTGTCTC		
DJgRNA3_int	TCGCAGGGCAACAGGGGCTGTGA		
18SrRNA_FW	CGCTACTACCGATTGGATGG	18S	Quantification of possible deletions of the 18S rRNA gene copies
18SrRNA_REV	CAAGTTCGACCGTCTTCTCA		
18SrRNA_int	AGGCCCTCGGATCGGCCCCG		

Reference gene assays: PrimePCR ddPCR Copy Number Assay:RPP30, Human (Bio-Rad Assay ID dHsaCP2500350)
PrimePCR ddPCR Expression Probe Assay:IPO8, Human (Bio-Rad Assay ID dHsaCPE5044719)
All assays from Bio-Rad (made in US); Dyes: 5' 6-FAM/HEX, quencher 3' Iowa Black FQ

PCR program used in ddPCR assays.

Program:		
95 °C	10:00	
94 °C	1:00	50 x
61 °C	2:00	
98 °C	10:00	
4 °C	hold	