Supplemental Material

Supplementary Table 1. Human T cell donor information.

Subject	Donated Material	Sex	Ethnicity	CCR5 genotype
1	Apheresis product	Female	Data not collected	WT/WT
2	Whole Blood	Male	Caucasian	WT/WT
3	Whole Blood	Male	Caucasian	WT/WT
4	Whole Blood	Male	South Asian	WT/WT
5	Whole Blood	Male	Caucasian	WT/WT
6	Whole Blood	Female	Caucasian	WT/WT
7	Whole Blood	Male	East Asian	WT/WT

Supplementary Table 2. Oligonucleotides used in this study.

Primer	Sequence (5'-->3')*	Description	Coordinates **	Split* *	Cell ***
CCR5-F	TGGTGGCTGTGTTTGCGTCTC	Forward primer for detecting CCR5 d32 mutation			
CCR5-R	AGCGGCAGGACCAGCCCCAAG	Reverse primer for detecting CCR5 d32 mutation			
RTPrime	CTCCACACTAACACTTGTCTCTCCG	RT primer for making cDNA from HIV mRNA transcripts	9309:9333		HOS, T
keo003	TTAGAAACATCAGAGGGCTGTAG	Forward primer to check for presence of full length (9kb) cDNA	939:961		HOS, T
keo004	TGGGATAGGTGGATTATTTGTC	Reverse primer to check for presence of full length (9kb) cDNA	1538:1559		HOS, T
F1.1.1	ACAGCGACGAAGACCTCC	Forward primer to determine ratio of spliced to incompletely spliced forms	5987:6004		HOS, T
R1.1.1	TGCCAAGAATCCGTTCAC	Reverse primer to determine ratio of spliced to incompletely spliced forms	8458:8475		HOS, T
F1.1.2	AAGTAGTGTGTGCCCGTCTG	Forward primer to determine ratio of spliced to incompletely spliced forms	550:569		HOS, T
R1.1.2	ATGTCCTGCTTGATATTCACAAC	Reverse primer to determine ratio of spliced to incompletely spliced forms	5434:5456		HOS, T
F1.2	CGTCTGTTGTGTGACTCTGG	Forward primer to determien relative use of exons 2-5	564:583		HOS, T
R1.2	AGGTCTTCGTCGCTGTCTC	Reverse primer to determine relative use of exons 2-5	5984:6002		HOS, T
F1.3	CAAGTAGTGTGTGCCCGTC	Forward primer to determine ratios of incompletely spliced (~4kb) transcripts	549:567		HOS, T
R1.3	GTGGTGGTTGCTTCTCTCC	Reverse primer to determine ratios of incompletely spliced (~4kb) transcripts	6353:6371		HOS, T
F1.4	GATCTCTCGACGCAGGACT	Forward primer to determine ratios of completely spliced (~2kb) transcripts	679:697		HOS, T
R1.4	GGGAGGAGGGTTGCTTT	Reverse primer to determine ratios of completely spliced (~2kb) transcripts	6038:6043; 8368:8378	s	HOS, T
F1.5	TTTGGAAAGGACCAGCAA	Forward primer to determine ratios of transcripts including exon 2	4929:4946		HOS, T
R1.5	CTGTCTTGAGGAGGTCTTCG	Reverse primer to determine ratios of transcripts including exon 2	5994:6013		HOS, T
F2.1.1	CCGTCTGTTGTGTGACTCTG	Forward primer to determine frequency of splicing from D1 to A1a	563:582		HOS, T
R2.1.1	TTGCTGCCATTGTCTGTATG	Reverse primer to determine frequency of splicing from D1 to A1a	4568:4587		HOS, T
F2.1.2	GCGACTGCAGGCAGATG	Forward primer to characterize splicing downstream when D1 is spliced to A1a	736:742; 4542:4551	s	HOS, T
R2.1.2'	AGGTCTTCGTCGCTGTCTC	Reverse primer to characterize splicing downstream when D1 is spliced to A1a	5984:6002		HOS
R2.1.2	TGACTGTCTTGAGGAGGTCTTC	Reverse primer to characterize splicing downstream when D1 is spliced to A1a	5995:6016		T
F2.1.3.1	GGACTCGGCTTGCTGAAG	Forward primer to characterize splicing upstream when D1a is spliced to A1	693:710		HOS, T
R2.1.3.1	ATCTCTGCTGTCCCTGTCC	Reverse primer to characterize splicing upstream when D1a is spliced to A1	4715:4720; 4912:4924	s	HOS, T
F2.1.3.2	GACCGGAAAGCGAAAGAG	Forward primer to characterize splicing upstream when D1a is spliced to A3	650:667		HOS, T
R2.1.3.2	CGACACCCAATTCTGTCC	Reverse primer to characterize splicing upstream when D1a is spliced to A3	4715:4720; 5776:5787	s	HOS, T
F2.2.1	AGACAGCGACGAAGACCTC	Forward primer to determine frequency of splicing from D4 to A5	5985:6003		HOS, T

Primer	Sequence (5'-->3')*	Description	Coordinates **	Split* *	Cell ***
R2.2.1	GCTTGTGGTGATATAGAAAGAGC	Reverse primer to determine frequency of splicing from D4 to A5	6695:6717		HOS, T
F2.2.3	CACAGCAAGAGGCGAGG	Forward primer to determine upstream splicing events when D5 is spliced to A7	715:731		HOS, T
R2.2.3	TGGGAGGAGGGTCTTATTTC	Reverse primer to determine upstream splicing events when D5 is spliced to A7	6722:6729; 8368:8379	s	HOS, T
F2.3.1	AGGACTCGGCTTGCTGAA	Forward primer to determine upstream splicing events when D4 is spliced to A7a	692:709		HOS, T
R2.3.1	CCCTGCCTAACTTGCTTTG	Reverse primer to determine upstream splicing events when D4 is spliced to A7a	6037:6043; 8334:8345	s	HOS, T
F2.3.2	CAGGACTCGGCTTGCTG	Forward primer to determine upstream splicing events when D4-A7b splice occurs	691:707		T
R2.3.2	GAATATCCCTGCTGCTTTGA	Reverse primer to determine upstream splicing events when D4-A7b splice occurs	6036:6043; 8340:8351	s	T
F2.4.1	CAGCGACGAAGACCTCC	Forward primer to determine relative use of splice acceptors in A7 and A8 clusters	5988:6004		HOS, T
R2.4.1	TGGTTAGCAGGGTGTAGCA	Reverse primer to determine relative use of splice acceptors in A7 and A8 custers	9273:9291		HOS, T
F2.4.2	ATCTCTCGACGCAGGACTC	Forward primer to determine upstream splicing events when D4 is spliced to A8	680:698		HOS, T
R2.4.2	CCTGGTGTGTAGTTTGCTTTG	Reverse primer to determine upstream splicing events when D4 is spliced to A8	6037:6043; 9177:9190	s	HOS, T
F3.1	AATCCCCAAAGTCAAGGAGTAG	Forward primer for unbiased analysis of splicing, first third of genome	4658:4679		HOS, T
R3.1	AGGAGGTCTTCGTCGCTG	Reverse primer for unbiased analysis of splicing, first third of genome	5988:6005		HOS, T
F3.2	GGGAATGATGGAGAAAGGA	Forward primer for unbiased analysis of splicing, second third of genome	6663:6681		HOS, T
R3.2	GAGGAAGAGGCACAGGTTC	Reverse primer for unbiased analysis of splicing, second third of genome	8496:8514		HOS, T
F3.3	CCGAAGGAACAGAAGAAGAAG	Forward primer for unbiased analysis of splicing, last third of genome	8402:8422		HOS, T
R3.3	CTTGTAGCACCATCCAAAGG	Reverse primer for unbiased analysis of splicing, last third of genome	9201:9220		HOS, T
F4.1	GTGTGCCCGTCTGTTGTG	Forward primer for full transcript structures	557:574		T
R4.1	CCATCCAAAGGTCAGTGG	Reverse primer for full transcript structures	9194:9211		T
keo056	CAGGGACCGGAAAGCGAAAGA	Forward primer to clone full length transcripts	646:666		HOS
keo057	GGAAAGTCCCCAGCGGAAAGT	Reverse primer to clone full length transcripts	9429:9449		HOS
keo058	GACCGGAAAGCGAAAGAGAAACC	Forward primer to confirm presence of ~1kb transcripts	650:672		HOS, T
keo059	GTCCCTTGTAGATAGCTCAGTGTCA	Reverse primer to confirm presence of ~1kb transcripts	9406:9430		HOS, T
keo070	AGACTCATCAAGTTTCTCTATCAAAG CAATAGTCCTCCCAGCCTCGA	forward quikchange primer for pIRES2-AcGFP1-Tat1 to make 1ex tat	6016:6043; 8368:8386		
keo071	TCGAGGCTGGGAGGACTATTGCTTT GATAGAGAAACTTGATGAGTCT	reverse quikchange primer for pIRES2-AcGFP1-Tat1 to make 1ex tat	6016:6043; 8368:8386		

*For amplification primers used on samples from HOS cells, each primer had a fivemer of 5'-GGTAT-3' appended to the 5' end to control for any biases in blunt end ligation step of PacBio SMRTBell hairpin loop attachment.

**Coordinates refer to HIV-$1_{89.6}$ reference sequence. Primers that span a known conserved or cryptic exon-exon junction are designated as split ("s") and the coordinates reflect the removed intron.

***Most primers were used on HIV RNA/cDNA from both HOS-CD4-CCR5 cells (HOS) and CD4-positive T cells (T). Primers that were used to amplify template from only one cell type are indicated.

Supplementary Table 3. HIV-$1_{89.6}$ splice donors and acceptors.

	ID	HIV-$1_{89.6}$ Coordinate*	Type	Usage Proportion**	Sequence***
Splice Donors	D1b	725	Novel	0.0102	ag|GCgagg
	D1	742	Conserved	0.9389	tg|GTgagt
	D1c	746	Cryptic	0.047	ga|GTacgc
	D1a	4720	Cryptic	0.0001	ag|GTaaga
	D2	4961	Conserved	0.0587	ag|GTgaag
	D3	5462	Conserved	0.0686	ag|GTagga
	D4	6043	Conserved	0.824	ca|GTaagt
	D5	6729	Cryptic	0.0004	ag|GTaaag
	D6	8422	Novel	0.0011	ag|GTggag
Splice Acceptors	A1a	4542	Cryptic	0.0001	tctcctaaaattAG|c
	A1	4912	Conserved	0.0577	cgggtttattacAG|g
	A2	5389	Conserved	0.2031	tgattgtttttcAG|a
	A3	5776	Conserved	0.0396	tttattcatttcAG|a
	A4c	5935	Conserved	0.1212	ctttcattgccaAG|c
	A4a	5953	Conserved	0.0638	tttcataacaaaAG|g
	A4b	5959	Conserved	0.089	aacaaaaggcttAG|g
	A5	5975	Conserved	0.6139	atctcctatggcAG|g
	A5a	5979	Novel	0.0964	cctatggcaggaAG|a
	A5b	5982	Novel	0.0077	atggcaggaagaAG|c
	A6	6602	Cryptic	0	ccactctgtgttAC|t
	A6a	6654	Novel	0.0003	actaatcccactAG|t
	A7a	8334	Cryptic	0.0016	tctatagtaaatAG|a
	A7b	8340	Cryptic	0.0114	gtaaatagagttAG|g
	A7c	8344	Novel	0.0023	atagagttaggcAG|g
	A7d	8362	Novel	0.0146	attcaccattatCG|t
	A7	8368	Conserved	0.8738	cattatcgtttcAG|a
	A7e	8381	Novel	0.02	gaccctcctcccAG|c
	A7f	8485	Novel	0.0007	tggcacttttctGG|g
	A8d	9128	Novel	0.0137	tccttgatctgtGG|g
	A8a	9144	Cryptic	0.0371	ctaccacacacaAG|g
	A8c	9156	Novel	0.4571	aggcttcttcccAG|a
	A8	9164	Cryptic	0.0293	tcccagattggcAG|a
	A8b	9177	Cryptic	0.0048	gaactacacaccAG|g
	A8e	9183	Novel	0.0036	cacaccagggccAG|g
	A8f	9221	Novel	0.013	gatggtgctacaAG|c

* Coordinate indicates last base of exon for splice donors, first base of exon for acceptors

**Usage proportion is calculated as the fraction of sequence reads spanning the indicated splice site that contain a splice event involving that site.

***Sequence in 89.6 strain of HIV. Exon/intron boundaries are indicated by "|". +1 and +2 positions within the introns following splice donors (positions of consensus GT) are shown in capitals. Similarly -2 and -1 positions within the introns preceding acceptors (positions of consensus AG) are shown in capitals. Consensus for human splice donors is AG|GTRAGT and for splice acceptors, YYTTYYYYYYNCAG|G (1).

Supplementary Table 4. HIV-$1_{89.6}$ splice junction counts in sequences from T cells

		Splice Donors								
		D1b*	**D1**	D1c	D1a	**D2**	**D3**	**D4**	D5	D6*
Splice Acceptors	**A1**	151	12738	113	4	0	0	0	0	0
	A1a	0	4	0	0	0	0	0	0	0
	A2	407	43287	138	3	2367	0	0	0	0
	A3	109	7241	23	6	1761	456	0	0	0
	A4a	150	10696	52	0	228	354	0	0	0
	A4b	257	15442	66	0	777	873	0	0	0
	A4c	255	20605	64	7	588	1254	0	0	0
	A5	1058	104005	356	2	7166	9209	0	0	0
	A5a*	210	7883	88	5	336	425	0	0	0
	A5b*	193	1301	36	1	71	80	0	0	0
	A6a*	0	1	0	0	0	0	33	0	0
	A7a	0	0	0	0	0	0	166	6	0
	A7b	0	2	0	0	0	0	1216	0	0
	A7c*	0	1	0	0	0	0	252	1	0
	A7d*	3	4	1	0	0	0	295	0	0
	A7	38	695	6	0	19	13	105508	38	0
	A7e*	16	62	5	0	2	3	1763	0	0
	A7f*	1	2	1	0	0	0	11	0	11
	A8d*	0	9	1	0	0	0	159	0	3
	A8a	1	33	0	0	1	0	411	2	3
	A8c*	13	287	4	0	6	5	5142	1	49
	A8	5	21	1	0	0	0	492	0	6
	A8b	3	11	0	0	0	0	62	0	0
	A8e*	0	9	0	0	2	0	45	0	1
	A8f*	1	2	0	0	0	0	92	0	0

Splice acceptors and splice donors are labeled as identified in Supplmentary Table 3. Widely conserved major splice acceptors and donors are in bold. Asterisks indicate novel splice and donors identified in this work. Displayed counts are sequence reads with gaps in the alignment to the HIV-$1_{89.6}$ consensus (introns) beginning at the indicated splice acceptor and ending at the indicated donor. For this analysis all filtered sequence reads from T cells were pooled.

Supplementary Table 5. Ratios of HIV-$1_{89.6}$ transcripts in T cells and HOS cells.

Class	Primer Pair*	Transcript number**	splicing	percent of class				<5 reads****
				T Cells (48hpi)	HOS (18hpi)	HOS (24hpi)	HOS (48hpi)	
partially spliced (4kb)	1.3	Vif 2	D1^A1	0.21%	0.00%	0.00%	0.00%	
	1.3	Vpr 3	D1^A2	0.46%	0.00%	0.58%	0.53%	
	1.3	Vpr 4	D1^A1|D2^A2	0.00%	0.00%	0.00%	0.00%	0 reads
	1.3	Tat 5	D1^A3	1.92%	5.41%	6.43%	19.25%	
	1.3	Tat 6	D1^A1|D2^A3	0.62%	2.70%	0.00%	0.00%	
	1.3	Tat 7	D1^A2|D3^A3	0.14%	0.00%	0.00%	0.00%	
	1.3	Tat 8	D1^A1|D2^A2|D3^A3	0.04%	0.00%	0.00%	0.00%	<5 reads
	1.3	Env/Vpu 4	D1^A4c	9.56%	2.70%	3.51%	4.28%	
	1.3	Env/Vpu 8	D1^A1|D2^A4c	0.27%	0.00%	0.00%	0.00%	
	1.3	Env/Vpu 12	D1^A2|D3^A4c	0.36%	0.00%	0.00%	0.00%	
	1.3	Env/Vpu 16	D1^A1|D2^A2|D3^A4c	0.01%	0.00%	0.00%	0.00%	<5 reads
	1.3	Env/Vpu 3	D1^A4a	5.60%	5.41%	4.09%	3.74%	
	1.3	Env/Vpu 7	D1^A1|D2^A4a	0.09%	0.00%	0.00%	0.00%	
	1.3	Env/Vpu 11	D1^A2|D3^A4a	0.17%	0.00%	0.00%	0.00%	
	1.3	Env/Vpu 15	D1^A1|D2^A2|D3^A4a	0.00%	0.00%	0.00%	0.00%	0 reads
	1.3	Env/Vpu 2	D1^A4b	8.72%	8.11%	8.19%	4.81%	
	1.3	Env/Vpu 6	D1^A1|D2^A4b	0.58%	0.00%	0.00%	0.00%	
	1.3	Env/Vpu 10	D1^A2|D3^A4b	0.31%	0.00%	0.00%	0.53%	
	1.3	Env/Vpu 14	D1^A1|D2^A2|D3^A4b	0.03%	0.00%	0.00%	0.00%	<5 reads
	1.3	Env/Vpu 1	D1^A5	63.64%	70.27%	74.27%	65.24%	
	1.3	Env/Vpu 5	D1^A1|D2^A5	2.66%	2.70%	1.75%	0.53%	
	1.3	Env/Vpu 9	D1^A2|D3^A5	4.50%	2.70%	1.17%	1.07%	
	1.3	Env/Vpu 13	D1^A1|D2^A2|D3^A5	0.11%	0.00%	0.00%	0.00%	
completely spliced (2kb)	1.4	Vpr 1	D1^A2|D4^A7	0.14%	0.00%	0.13%	0.08%	
	1.4	Vpr 2	D1^A1|D2^A2|D4^A7	0.01%	0.00%	0.00%	0.00%	
	1.4	Tat 1	D1^A3|D4^A7	2.69%	5.00%	5.45%	9.19%	
	1.4	Tat 2	D1^A1|D2^A3|D4^A7	0.67%	0.43%	0.85%	0.31%	
	1.4	Tat 3	D1^A2|D3^A3|D4^A7	0.14%	0.65%	0.09%	0.00%	
	1.4	Tat 4	D1^A1|D2^A2|D3^A3|D4^A7	0.01%	0.00%	0.00%	0.00%	
	1.4	Rev 3	D1^A4c|D4^A7	12.76%	1.30%	2.64%	3.59%	
	1.4	Rev 6	D1^A1|D2^A4c|D4^A7	0.28%	0.43%	0.00%	0.00%	
	1.4	Rev 9	D1^A2|D3^A4c|D4^A7	0.54%	0.00%	0.09%	0.03%	
	1.4	Rev 12	D1^A1|D2^A2|D3^A4c|D4^A7	0.03%	0.00%	0.00%	0.00%	
	1.4	Rev 2	D1^A4a|D4^A7	7.18%	4.35%	3.83%	2.88%	
	1.4	Rev 5	D1^A1|D2^A4a|D4^A7	0.14%	0.00%	0.04%	0.00%	
	1.4	Rev 8	D1^A2|D3^A4a|D4^A7	0.32%	0.43%	0.13%	0.00%	
	1.4	Rev 11	D1^A1|D2^A2|D3^A4a|D4^A7	0.01%	0.00%	0.00%	0.00%	
	1.4	Rev 1	D1^A4b|D4^A7	9.38%	8.91%	6.64%	4.10%	
	1.4	Rev 4	D1^A1|D2^A4b|D4^A7	0.56%	0.00%	0.09%	0.00%	
	1.4	Rev 7	D1^A2|D3^A4b|D4^A7	0.62%	0.00%	0.09%	0.00%	
	1.4	Rev 10	D1^A1|D2^A2|D3^A4b|D4^A7	0.07%	0.00%	0.04%	0.00%	
	1.4 and 4.1	Nef 2	D1^A5|D4^A7	46.86%	72.39%	71.96%	78.29%	
	1.4	Nef 3	D1^A1|D2^A5|D4^A7	4.69%	3.26%	4.43%	1.07%	
	1.4	Nef 4	D1^A2|D3^A5|D4^A7	4.96%	2.39%	3.11%	0.42%	
	1.4	Nef 5	D1^A1|D2^A2|D3^A5|D4^A7	0.34%	0.43%	0.43%	0.03%	
	4.1	Novel (Nef 9)	D1^A5a|D4^A7	3.57%				
	4.1	Novel (Nef 11)	D1^A2|D3^A5a|D4^A7	0.28%				
	4.1	Nef 1	D1-A7	2.01%				
	4.1	Novel	D1^A5|D4^A7b	1.33%				
	4.1	Novel	D1^A5|D4^A7e	0.44%				

Class	Primer Pair*	Transcript number**	splicing	percent of class				<5 reads****
				T Cells (48hpi)	HOS (18hpi)	HOS (24hpi)	HOS (48hpi)	
completely spliced (1kb)	4.1	Novel	D1^A5|D4^A8d	0.80%				
	4.1	Novel	D1^A8d	0.32%				
	4.1	Novel	D1^A4c|D4^A8a	0.78%				
	4.1	Novel	D1^A4b|D4^A8a	0.96%				
	4.1	Novel	D1^A5|D4^A8a	2.09%				
	4.1	Novel	D1^A1|D2^A5|D4^A8a	0.30%				
	4.1	Novel	D1^A5a|D4^A8a	0.15%				
	4.1		D1^A8a	1.90%				
	4.1	Novel (Tat^8c 1)	D1^A3|D4^A8c	0.64%				
	4.1	Novel (Ref 3)	D1-A4c|D4-A8c	8.18%				
	4.1	Novel (Ref 6)	D1-A1|D2-A4c|D4-A8c	0.38%				
	4.1	Novel (Ref 9)	D1-A2|D3-A4c|D4-A8c	0.36%				
	4.1	Novel (Ref 2)	D1-A4a|D4-A8c	3.72%				
	4.1	Novel (Ref 1)	D1-A4b|D4-A8c	6.81%				
	4.1	Novel (Ref 4)	D1-A1|D2-A4b|D4-A8c	0.59%				
	4.1	Novel (Ref 7)	D1-A2|D3-A4b|D4-A8c	0.37%				
	4.1	Novel	D1^A5|D4^A8c	41.41%				
	4.1	Novel	D1^A1|D2^A5|D4^A8c	2.41%				
	4.1	Novel	D1^A2|D3^A5|D4^A8c	2.53%				
	4.1	Novel	D1^A5|D4^A7|D6^A8c	0.37%				
	4.1	Novel	D1^A5a|D4^A8c	3.75%				
	4.1	Novel	D1^A8c	19.05%				
	4.1	Novel	D1^A1|D2^A8c	0.09%				
	4.1	Novel	D1^A2|D3^A8c	0.15%				
	4.1	Novel	D1^A4c|D4^A8	0.36%				
	4.1	Novel	D1^A4b|D4^A8	0.10%				
	4.1		D1^A5|D4^A8	0.71%				
	4.1		D1^A8	0.55%				
	4.1		D1^A8b	0.17%				

*Reads used in these calculations came from HIV-1 sequence reads that had been amplified with the indicated primer pairs (F1.3/R1.3, F1.4/R1.4, or F4.1/R4.1). Nef 2 was used to normalize reads from primer pair 4.1 for comparison to ~2kb transcripts. Primer pair 4.1 was not used to amplify messages produced in HOS-CD4-CCR5, therefore transcripts identified with this primer pair could not be compared in HOS cells.

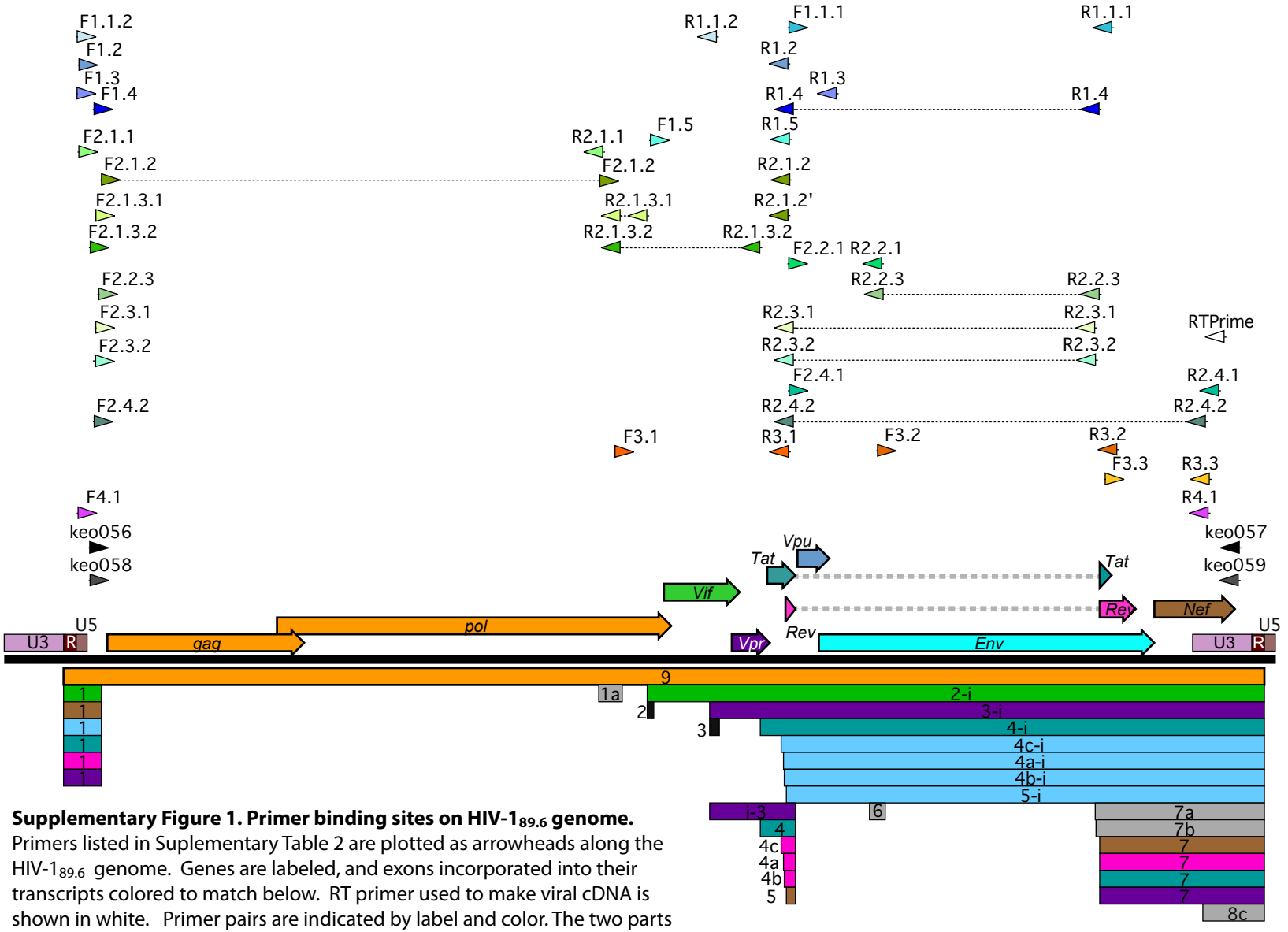
**Naming of transcripts is according to convention established in (2). Novel Tat^8c and Ref encoding transcripts were named similarly. New Nef transcripts were named by extension of the convention. Previously detected unnamed transcripts are left blank.

*** "|" indicates continuous exon. "^" indicates removed intron. Site labels are as in Table 5-3.

****Transcripts detected in less than 5 sequence reads in T cells

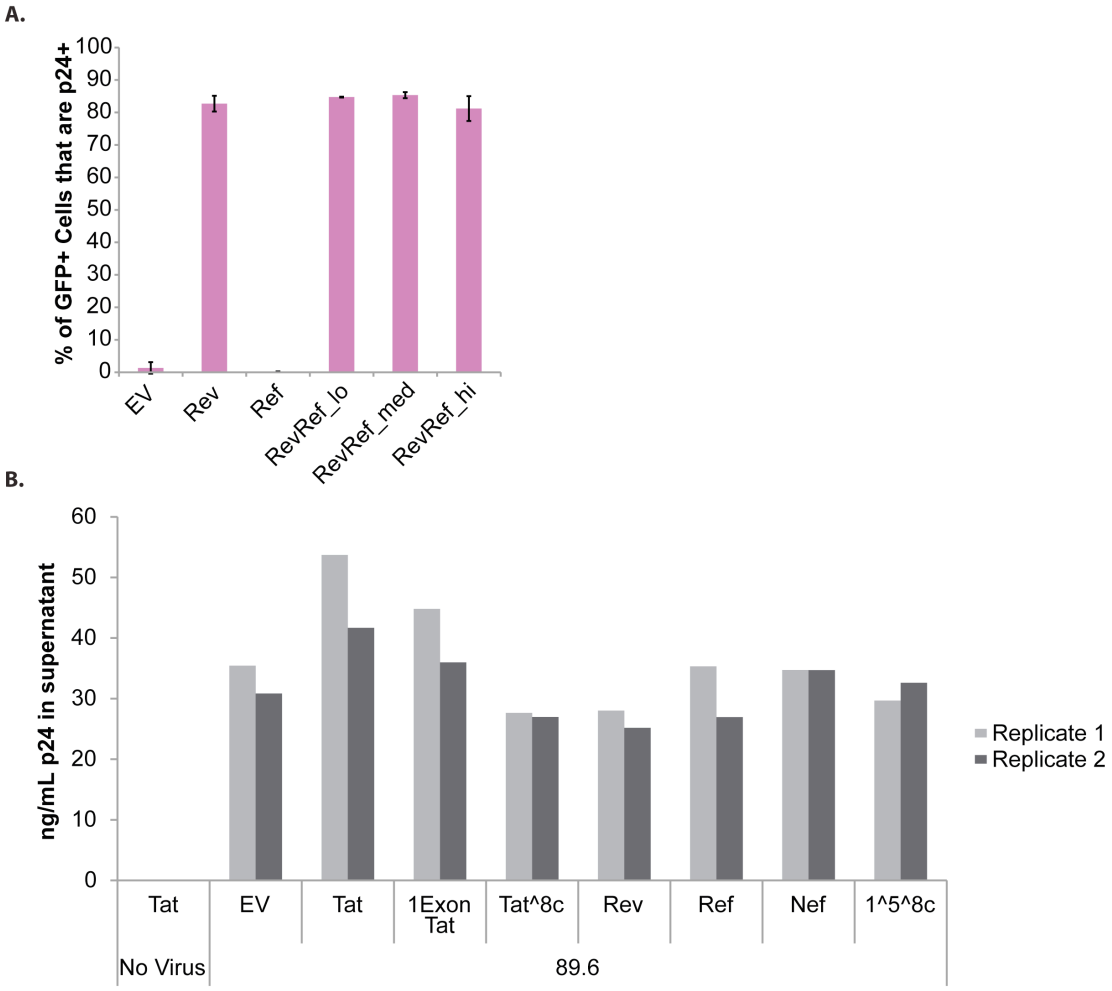
Supplementary Table 6. ORFs and amino acid sequences of Tat^8c and Ref.

Trans-cript	Exons	cDNA ORF and Amino Acid sequence
Tat^8c	1-4-8c	ATG GAG CCA GTA AAT CCT AGC CTA GAG CCC TGG AAG CAT CCA GGA AGT CAG CCT M E P V N P S L E P W K H P G S Q P AAA ACT GCT TGT ACC AAT TGC TAT TGC AAA AAA TGT TGC TTT CAT TGC CAA GCT K T A C T N C Y C K K C C F H C Q A TGT TTC ATA ACA AAA GGC TTA GGC ATC TCC TAT GGC AGG AAG AAG CGG AGA CAG C F I T K G L G I S Y G R K K R R Q CGA CGA AGA CCT CCT CAA GAC AGT CAG ACT CAT CAA GTT TCT CTA TCA AAG CA|A R R R P P Q D S Q T H Q V S L S K Q TTG GCA GAA CTA CAC ACC AGG GCC AGG AAT CAG ATA TCC ACT GAC CTT TGG ATG L A E L H T R A R N Q I S T D L W M GTG CTA CAA GCT AGT ACC AGT TGA V L Q A S T S *
Ref	1-4c/a/b 8c	ATG GCA GGA AGA AGC GGA GAC AGC GAC GAA GAC CTC CTC AAG ACA GTC AGA CTC M A G R S G D S D E D L L K T V R L ATC AAG TTT CTC TAT CAA AGC A|AT TGG CAG AAC TAC ACA CCA GGG CCA GGA ATC I K F L Y Q S N W Q N Y T P G P G I AGA TAT CCA CTG ACC TTT GGA TGG TGC TAC AAG CTA GTA CCA GTT GAG CCA GAT R Y P L T F G W C Y K L V P V E P D GAA GGA GAG AAC AAC AGA GAG GAC AAC AGC TTG CTA CAC CCT GCT AAC CAG CAT E G E N N R E D N S L L H P A N Q H GGA GTA GAA GAC TCG GAG AGA CAA GTG TTA GTG TGG AGG TTT GAC AGC CGC CTA G V E D S E R Q V L V W R F D S R L GCA TTC CAT CAC GTG GCC CGA GAG CTG CAT CCG GAG TAC TTC AAG AAC TGA A F H H V A R E L H P E Y F K N *



Supplementary Figure 1. Primer binding sites on HIV-$1_{89.6}$ genome. Primers listed in Suplementary Table 2 are plotted as arrowheads along the HIV-$1_{89.6}$ genome. Genes are labeled, and exons incorporated into their transcripts colored to match below. RT primer used to make viral cDNA is shown in white. Primer pairs are indicated by label and color. The two parts of "split" primers that span exon junctions are joined by dashed lines.

Supplementary Figure 2.



S**upplementary Figure 2. Overexpression of cloned novel D4^A8c cDNAs. A.** HIV-$1_{89.6}$ products were assayed for Rev activity by co-transfection of 293T cells with cDNA expression plasmids and a reporter plasmid expressing Rev-dependent Gag/Gag-Pol. Intracellular Gag expression was assayed by FACS, gating on GFP positive cells (to eliminated untransfected or poorly expressing cells). Rev and Ref were expressed alone or mixed (constant Rev, increasing Ref). Apart from rev, no cDNAs appeared to encode Rev activity. **B.** To determine whether transcripts with splicing to A8c affected infection efficiency, HOS-CD4-CCR5 cells were transfected with each expression construct and then infected with HIV-$1_{89.6}$. Gag (p24) released into the supernatant was measured at 72hpi. Only Tat (1 and 2 exon) appeared to stimulate p24 production. Results were consistent with FACS analysis of intracellular Gag production (not shown).

Supplementary Methods

Cell culture and viral infections. HIV-$1_{89.6}$ was generated by the University of Pennsylvania Center for Aids Research. 293T cells were transfected with plasmid containing the viral clone, and harvested virus was passaged in SupT1 cells once. Virus was titered by p24 antigen content.

Human osteosarcoma cells expressing CD4 and CCR5 (HOS-CD4-CCR5) were grown and passaged in DMEM with GlutaMAX (Invitrogen) supplemented with 10% FBS (Sigma-Aldrich), and 1 ug/mL Puromycin (Calbiochem). Low passage (<20) cells were plated in 10 cm plates and grown overnight to 80% confluency. HIV-$1_{89.6}$ equivalent to 2.25 micrograms of p24 was added to each plate in a total of 3mL D10. At 4hpi, cells were washed twice in PBS and media was replaced with 10mL fresh D10 with puromycin and 21.5uM Nelfinavir to prevent second round infection. At 18, 24, and 48hpi, cells were washed in PBS, trypsinized, and pelleted.

Primary $CD4^+$ T cells were isolated by the University of Pennsylvania Center for AIDS research Immunology core from apheresis product (Subject 1) or 40 mL whole blood (Subjects 2-7) using the RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technologies). All donors were healthy. Cells were genotyped as below.

T Cells were stimulated for 3 days at 0.5 x10^6 cells per milliliter in R10 media (RPMI 1640 with GlutaMAX (Invitrogen) supplemented with 10% FBS (Sigma-Aldrich) with 100 units U/mL recombinant IL2 (Novartis) + 5ug/mL PHAL (Sigma-Aldrich). Cells from each donor were infected in triplicate. For each infection, three million cells were plated in each of 2 wells of a six well plate in 3 mL R10 media containing ~270 ng p24 of HIV-$1_{89.6}$ per 10^6 cells. Cells were infected by spinoculation for 2 hr at 1200 x g

and 37°C. Plates were incubated at 37°C for an additional 2 hours, and then cells were pooled and volume increased to 12 mL R10 media with 100 U/mL IL2 and 100 pg/mL Enfuvirtide (T-20) to restrict infection to a single round. We have observed that ~5-6% of cells become infected at 48hpi using this protocol (data not shown). At 48hpi, cells were washed twice in PBS and pelleted for RNA preparation.

The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HOS-CD4-CCR5 from Dr. Nathaniel Landau(1,2), Nelfinavir, and T-20, Fusion Inhibitor from AIDS, NIAID.

Genotyping of the CCR5 allele. Genomic DNA was isolated from T cells (subject 1) or from buccal swab (subjects 2-7) using DNeasy Blood and Tissue Kit (Qiagen). Approximately 35 ng of gDNA and each of three control plasmid mixes containing CCR5 alleles (WT, Δ32, or both) were amplified in 25 uL with 200 pmols each of primers CCR5-F and CCR5-R (Table 5-1) using Go Taq Green Master Mix (Promega). Amplification was performed as follows: 5 min at 95°; 35 cycles of 30s at 95°C, 30s at 60°C, and 45s at 72°C; and finally 2 min at 72°C. Amplification products were resolved by 2% agarose gel electrophoresis and stained with ethidium bromide.

Single molecule amplification performed by RainDance Technologies. Two μL of each cDNA sample was added to the Template Mix consisting of 4.7 μL 10× High-Fidelity Buffer (Invitrogen), 1.26 μL of $MgSO_4$ (Invitrogen), 1.44 μL 10 mM dNTP (New England Biolabs), 3.6 μL Betaine (Sigma), 3.6 μL of RDT Droplet Stabilizer (RainDance Technologies), 1.8 μL dimethyl sulfoxide (Sigma) and 0.7 μL 5 units/μL of

Platinum High-Fidelity Taq (Invitrogen). The samples were brought to a final volume of 25 µL with Nuclease Free Water, Teknova (Fisher).

PCR droplets were generated on the RDT 1000 (RainDance Technologies) using the manufacturer's recommended protocol. To process a single sample, the user placed onto the RDT 1000 a single tube containing 25 µL of Template Mix, the custom primer droplet library (RainDance Technologies) and a disposable microfluidic chip (RainDance Technologies). The custom primer droplet library consists of a collection of unique droplets in which each droplet contains a unique primer pair that targets a specific HIV sequence. Within each droplet the concentration of each primer was 5.2 µM to allow for uniform amplicification of the different target sequences. The custom primer libraries for this study contained 18 (HOS-CD4-CCR5 cells) or 20 (primary T cells) PCR primer pairs designed to amplify different HIV RNA isoforms. The RDT 1000 paired a single template droplet with a single primer droplet, and merged the two droplets into a single PCR droplet using an electric field. All of the resulting PCR droplets were automatically dispensed as an emulsion into a single PCR tube and transferred to a standard thermal cycler for PCR amplification. Each cDNA sample generated more than 1,000,000 PCR droplets.

Following processing of samples on the RDT 1000 with the custom HIV primer library the samples were thermal cycled on a Bio-Rad PTC-225 thermal cycler with the following profile: 94°C for 2 minutes; 55 cycles at 94°C for 30 seconds, 54°C for 30 seconds and 68°C for 6 minutes; followed by 68°C for 10 minutes and hold at 4°C.

Following PCR amplification the emulsion of PCR droplets were broken to release the amplification products, 50 uL of RDT 1000 Droplet Destabilizer (RainDance

Technologies) was added to the emulsion, the sample vortexed for 15 seconds and then spun in a microcentrifuge at 12,000xg for 10 minutes. The oil from below the aqueous phase was removed and the sample was purified over a MinElute column (Qiagen, 28004) following the manufacturer's recommended protocol. The sample was eluted from the column with 11 μL of the Qiagen Elution Buffer. The amplicon yields for each sample were determined by running 1 μL of the eluted samples on an Agilent Bioanalyzer DNA 12,000 chip.

Cloning transcripts with A8c splicing. cDNAs of transcripts with exon structures 1-4-8c (*tat^8c)*, 1-4b-8c (*ref*), and 1-5-8c, as well as *tat* (2 exon), *rev* and *nef* cDNAs were cloned into the TOPO vector as described in the methods. After sequence verification, inserts were excised by restriction digest using EcoRI (New England Biolabs) and ligated into the expression vector pIRES2-AcGFP1 (Clonetech). This vector contains a CMV promoter, and encodes and IRES and GFP following the inserted gene for production of a bicistronic message. Correct ligation was confirmed by Sanger sequencing. Plasmid encoding 1-exon Tat was created from the *tat* plasmid using the Quikchange II XL Site directed mutagenesis kit (Stratagene) to introduce the appropriate stop codon after the first exon with the primers keo070 and keo071 (Supplementary Table 2).

Tat activity assay. TZM-bl cells (kind gift of Dr. Robert W. Doms) were grown in D10 medium (DMEM supplemented with 10% FBS (Sigma-Aldrich)) in a 96 well plate to 80% confluency. These cells stably express a Tat-inducible luciferase. Cells were transiently transfected by lipofection using the Lipofectamine 2000 reagent (Invitrogen)

according to the manufacturer's instructions. Transfections were done in quadruplicate for each expression construct and the empty pIRES2-AcGFP1 vector with a total of 0.2 ug of DNA per well. Mock transfections were also performed. Expression of different plasmid constructs was confirmed to be equivalent by fluorescence microscopy, counting GFP-expressing cells (data not shown). Luciferase production was measured as emitted luminescence using the Bright-Glo luciferase assay system (Promega).

Rev activity assay. In order to determine whether the products expressed from our cDNAs had Rev activity, we co-transfected (as above) HEK-293T cells with each of our plasmids and pCMVGagPol-RRE-R, a reporter plasmid from which Gag and Pol are expressed in a Rev-dependent manner (kind gift of David Rekosh)(3). Cells were plated in 12 well plates and transfected with a total of 2.1 ug plasmid DNA per well: 0.7 ug pCMVGagPol-RRE-R plus 0.7 ug of experimental plasmid or a mix of 0.7 ug of the *rev* plasmid with 0.08 ug (Rev low), 0.23 ug (Ref medium) or 0.7 ug (Ref high) of the plasmid containing the *ref* cDNA insert. Empty vector (pIRES2-AcGFP) was added to bring the total DNA to 2.1 ug. Reporter *gag* expression was measured by flow cytometry and by ELISA of p24 (Capsid) in supernatants (below).

HIV Replication assays. In order to determine whether expression of any of the cloned cDNAs affected HIV-$1_{89.6}$ infection efficiency, we transiently transfected each of the expression plasmids into HOS-CD4-CCR5 cells by lipofection essentially as above in 24 well plates with 0.8 ug plasmid per well. 24 hours post transfection, cells were infected with HIV-$1_{89.6}$ equivalent to 125 ng p24 in 300 uL D10 medium per well. Spinoculation

was performed at 1200g for 2hr at 37°C. Cells were then incubated at 37°C for an additional 4 hours, after which they were washed with PBS and fresh media was applied. Cells were analyzed for intracellular Gag expression by flow cytometry and for release of p24 into supernatant by ELISA by the University of Pennsylvania Center for AIDS Research.

Flow cytometry. 1-2x10^6 cells were stained per co-transfection for flow cytometry. All incubations were at room temperature. Cells were first washed in PBS and then twice in FACS wash buffer (PBS, 2.5% FBS, 2mM EDTA). Cells were fixed and permeabilized with CytoFix/CytoPerm (BD) for 20 minutes and washed with Perm-Wash Buffer (BD) before staining with anti-HIV-Gag-PE (Beckman Coulter) for 60 min. Finally cells were washed in FACS wash buffer and resuspended in 3% PFA. Samples were run on a LSRII (BD) and analyzed with FlowJo 8.8.6 (Treestar). Cells were gated as follows: lymphocytes (SSC-A by FSC-A), then singlets (FSC-A by FSC-H), then GFP+ cells (FSC-A by GFP), then by Gag expression (FSC-A by Gag).

Proteosome inhibition and over-expression for Western Blotting. HEK-293T cells were transfected as above with expression constructs in 6 well plates, 4 ug plasmid DNA per well. At 40 hours post transfection, 3 uL of DMSO or 10mM MG132 in DMSO was added per well.

Supplementary Report 1. Additional Novel HIV-189.6 transcripts.

Several novel splice donors and acceptors, or novel combinations of known exons were observed in this study. Notable examples omitted from the main text are discussed below.

Splicing at D2 does not require splicing at A1.

The first major conserved splice acceptor in HIV-1 is A1 and, together with donor D2, it defines the boundaries of exon 1. In all previous examples of splicing from D2, it has been in the context of a fully spliced exon 2, and it has been shown that splicing at D2 helps drive splicing from D1 to A1 via exon definition (6); however, our data showed for the first time that splicing from D2 to several downstream acceptors can occur without upstream splicing to A1. We could not distinguish whether these transcripts contained an intron upstream of D2 or whether the D2 splice junctions were the first splice events in these transcripts. The latter splicing pattern is particularly interesting as it would create spliced transcripts containing the ψ packaging signal (normally disrupted by splicing at D1), although it is likely repressed (7). Packaging of such subgenomic RNAs would poison HIV replication.

A Tat-env-env fusion.

Another transcript of interest contains the first coding exon of *tat* (exon 3) spliced to a novel small exon within *env* and in frame with *env*, with an additional splicing event to the final major HIV-1 exon (exon 7) again in frame with *env*. This would encode a Tat-

Env-Env fusion protein containing the first 71 residues of Tat, a novel histidine, and two stretches of Env: residues 145-169 followed by the N terminal residues 716-853. This fusion is reminiscent of $p28^{tev}$ (Tev) first described by Benko and colleagues; however, the 5' boundary of the small central exon in Tev is an acceptor (A6) that is found in HXB2 but not conserved (8,9). The acceptor (dubbed A6a) used in our novel proposed Tat-Env-Env message lies downstream of A6 and is also not well conserved (although it is found sporadically in most clades of HIV-1). Both Tev and the Tat-Env-Env transcript require splice donor D5, which is conserved in 89.6 and several clade B viruses, but not in other HIV-1 clades.

Clusters of splice sites near conserved Donor 1, Acceptor 5, and Acceptor 7

The majority of novel splice sites were identified in clusters near previously observed sites. Two infrequently used splice donors were detected near D1. Use of these in place of D1 would modify the 5' untranslated leader on all spliced mRNAs without affecting any known HIV-1 ORFs. One was novel (D1b) and use of the other, 4 nucleotides downstream of D1 (D1c), had only been observed in viruses in which D1 was mutated (2). Similarly, two novel acceptors downstream of exon 5 are predicted to shorten exon 5 but should not affect the *nef, env,* or *vpu* ORFs encoded by mRNAs that use exon 5. Novel acceptors near A7 were not redundant. Two (A7c and A7d, Supplementary Table 3) would add novel amino acids between the protein domains encoded by the first and second coding exons of *tat* and *rev*. Splicing at A7e is predicted to produce Tat-Rev and Rev-Env fusions, while use of A7f would create Tat-Env and Rev-Tat fusions.

These redundant splice sites may be evidence that *cis*-acting splice enhancers (SEs) in HIV-1 may be promiscuous, with the ability to attract splicing factors to novel splice donors and acceptors that evolve within close proximity to the conserved sites which the SEs normally benefit. For example, a strong exonic splice enhancer known as the GAR is located just downstream of A5 (10,11). Redundancies at vital sites may be evolutionarily beneficial for the virus, protecting important spliced transcripts. Abbink and Berkhout have reported that, in viruses where the use of the major donor in the upstream leader, D1, is compromised, mutation of a GC within the dimerization initiation sequence (DIS) stem to GT can rescue splicing and replication of HIV-1 (12). This GC is just 7 nucleotides upstream of the GC that defines D1b, also found within the DIS. We speculate that both upstream GCs can act as "backup" donors for D1.

Novel splice acceptors near Acceptor 8.

Several new acceptors were detected downstream of A7, near cryptic acceptor A8 within the *nef* ORF and 3' LTR. As discussed in the text splicing of ~2kb transcripts to acceptors A8a, b and c creates transcripts that encode the amino-terminal region of Tat with appended novel amino acids. Transcripts using A8d, and A8e in place of A7 are predicted to do the same. As with A8a, b, and c, splicing at A8e is also predicted to create Rev-Nef transcripts. By contrast, transcripts in which exons 4a, 4b, and 4c are spliced to A8d would encode the first 26 amino acids of Rev fused to 35 novel amino acids. Splicing to A8f in place of A7 would produce transcripts encoding 1-exon Tat and first 25 amino acids of Rev fused to 4 novel amino acids. The wealth of splice acceptors in this region of HIV-$1_{89.6}$ suggests that splicing factors are efficiently attracted to the 3'

end of the transcribed RNAs. Sequences around putative acceptors A8e and A8f are fairly well conserved within HIV-1s (>60%), again suggesting functionality of splicing in this region.

A novel splice donor in Exon 7.

Finally, we detected a putative splice donor within exon 7, upstream of A8 which we have termed D6. Due to our experimental design, we could only detect splice junctions involving this donor and the cluster of A8 acceptors. However, it is conceivable that, in RNA transcripts that read through the poly-adenylation site into adjacent host genomic DNA, this splice donor could be paired with host splice acceptors to create novel chimeric viral-host transcripts as has been observed with lentiviral gene-therapy vectors (13).

Supplementary Report 2. Generalized linear models of changes in use of mutually exclusive HIV-1 splice acceptors

Reads splicing from D1 to one of five mutually exclusive acceptors, D3, D4c, D4a, D4b, D5, and D5a, in three primers, 1.2, 1.3 and 1.4, were collected. Since these data are based on counts, we modeled them as Poisson distributed with an extra variance term allowing for additional variance using a quasi-Poisson generalized linear model with log link. We accounted for differences in sequencing effort by including the total number of D1 to mutually exclusive acceptors reads in each primer-sample as an offset. Differences in the read counts a) over time,b) between human donor and c) cell type were analyzed separately. A term was included for each acceptor and its interaction with the variable of interest. The models included primer and replicate terms and their individual interactions with acceptor to account for any confounding factors.

HOS vs T Cells

R command: `glm(count~offset(log(total)) + acceptor:primer + acceptor:isHos + acceptor, data=mutEx[mutEx$time==48,], family='quasipoisson')`

Difference between HOS and T cells may be confounded by run differences between early sequencing and later sequencing. Verification by agarose gel (Figure 4b) suggest that these differences are likely biological.

```
                Df Deviance Resid. Df Resid. Dev        F    Pr(>F)
NULL                             395     138330
acceptor         5   133985       390       4345 9003.990 < 2.2e-16 ***
acceptor:primer 12      751       378       3594   21.032 < 2.2e-16 ***
acceptor:isHos   6     2466       372       1127  138.105 < 2.2e-16 ***
```

So after accounting for primer-acceptor bias, the difference between HOS and T cells is significant.

The interesting terms in the model are:

```
                     Estimate Std. Error t value Pr(>|t|)
acceptorA3:isHosTRUE   1.471656   0.065862  22.345  < 2e-16 ***
acceptorA4a:isHosTRUE -0.944942   0.124619  -7.583 2.73e-13 ***
acceptorA4b:isHosTRUE -0.928536   0.105907  -8.767  < 2e-16 ***
acceptorA4c:isHosTRUE -1.227872   0.106658 -11.512  < 2e-16 ***
acceptorA5:isHosTRUE   0.090816   0.026075   3.483 0.000555 ***
acceptorA5a:isHosTRUE  0.630805   0.079395   7.945 2.33e-14 ***
```

So it appears A3 is up; A4c, A4a and A4b are down;A5 is up a little and A5a up in HOS.

HOS Over Time

R command: `glm(value~offset(log(total)) + acceptor + acceptor:primer + acceptor:time, data=mutEx[mutEx$isHos,], family='quasipoisson')`

Looking only within HOS, we see a significant linear effect of time:

```
                Df Deviance Resid. Df Resid. Dev         F    Pr(>F)
NULL                             53    17961.7
acceptor         5  17709.5        48      252.2 6698.1217 < 2.2e-16 ***
acceptor:primer 12     18.0        36      234.2    2.8346   0.01018 *
acceptor:time    6    217.8        30       16.4   68.6567 3.571e-16 ***
```

We are assuming that a particular acceptor will have the same change in all three primers here.

The interesting terms are:

```
                       Estimate  Std. Error t value Pr(>|t|)
acceptorA3:time       0.0247685  0.0017776  13.934 1.22e-14 ***
acceptorA4a:time     -0.0162083  0.0028115  -5.765 2.69e-06 ***
acceptorA4b:time     -0.0252632  0.0022712 -11.123 3.62e-12 ***
acceptorA4c:time      0.0158665  0.0030499   5.202 1.32e-05 ***
acceptorA5:time      -0.0019181  0.0006313  -3.038   0.0049 **
acceptorA5a:time      0.0049199  0.0019687   2.499   0.0182 *
```

So A3, A4c and A5a increase over time and A4a, A4b and A5 decrease over time. All of these coefficients are with a log link and linear and so multiplicative. That means that for example A3 will increase 2.5%/hour (exp(`.0247`)) or equivalently 81% (1.025^24) over 24 hours.

Between Human Comparison

R command: `glm(value~offset(log(total)) + acceptor + acceptor:run + acceptor:primer + acceptor:subject, data=mutEx[!mutEx$isHos,], family='quasipoisson')`

In humans, we added a term to account for any potential run bias between the three replicates. Subject refers to the seven human blood donors from which T cells were collected:

```
                 Df Deviance Resid. Df Resid. Dev          F    Pr(>F)
NULL                            377     128430
acceptor          5   126446       372       1985 19598.4128 < 2.2e-16 ***
acceptor:run     12      136       360       1849     8.7924 1.767e-14 ***
acceptor:primer  12      850       348        998    54.9144 < 2.2e-16 ***
acceptor:subject 36      597       312        401    12.8623 < 2.2e-16 ***
```

So after accounting for any run and primer bias, subject ID has a statistically significant effect on our observed counts. If we compare everything to subject 7, the interesting terms are:

```
                      Estimate Std. Error t value Pr(>|t|)
acceptorA3:subject6   -0.001399   0.072855  -0.019 0.984696
acceptorA4a:subject6  -0.112900   0.049440  -2.284 0.023068 *
acceptorA4b:subject6  -0.054327   0.040379  -1.345 0.179472
acceptorA4c:subject6   0.028292   0.033604   0.842 0.400475
acceptorA5:subject6    0.016825   0.016001   1.051 0.293848
acceptorA5a:subject6  -0.030853   0.060918  -0.506 0.612881
acceptorA3:subject5   -0.077666   0.074228  -1.046 0.296223
acceptorA4a:subject5  -0.114369   0.049816  -2.296 0.022348 *
acceptorA4b:subject5  -0.068366   0.040896  -1.672 0.095587 .
acceptorA4c:subject5  -0.085852   0.034749  -2.471 0.014023 *
acceptorA5:subject5    0.038879   0.016156   2.406 0.016690 *
acceptorA5a:subject5   0.078768   0.060384   1.304 0.193037
acceptorA3:subject4   -0.184906   0.095778  -1.931 0.054443 .
acceptorA4a:subject4   0.071861   0.057914   1.241 0.215602
acceptorA4b:subject4   0.126202   0.047136   2.677 0.007812 **
acceptorA4c:subject4  -0.100205   0.043028  -2.329 0.020507 *
acceptorA5:subject4   -0.001158   0.019691  -0.059 0.953142
acceptorA5a:subject4   0.023455   0.073526   0.319 0.749942
acceptorA3:subject3   -0.003511   0.086650  -0.041 0.967707
acceptorA4a:subject3   0.071069   0.055639   1.277 0.202439
acceptorA4b:subject3   0.006463   0.046993   0.138 0.890705
```

```
acceptorA4c:subject3  -0.063344   0.040758  -1.554 0.121162
acceptorA5:subject3    0.010519   0.018868   0.557 0.577593
acceptorA5a:subject3  -0.070948   0.072851  -0.974 0.330868
acceptorA3:subject2   -0.232935   0.091755  -2.539 0.011613 *
acceptorA4a:subject2   0.024054   0.056425   0.426 0.670183
acceptorA4b:subject2   0.110672   0.045347   2.441 0.015221 *
acceptorA4c:subject2   0.021760   0.039521   0.551 0.582303
acceptorA5:subject2   -0.003760   0.018691  -0.201 0.840706
acceptorA5a:subject2  -0.160746   0.073508  -2.187 0.029502 *
acceptorA3:subject1    0.095360   0.065563   1.454 0.146816
acceptorA4a:subject1   0.029316   0.044309   0.662 0.508696
acceptorA4b:subject1  -0.214372   0.038431  -5.578 5.28e-08 ***
acceptorA4c:subject1  -0.397412   0.033847 -11.742  < 2e-16 ***
acceptorA5:subject1    0.091438   0.014697   6.221 1.58e-09 ***
acceptorA5a:subject1   0.027466   0.055941   0.491 0.623788
```

So there were small but significant effects between subjects especially between subject 1 and subjects 2-7. Interestingly T cells were collected from apheresis product in subject 1 and from whole blood in subjects 2-7 although why this would affect later assays is unknown.

mutEx Data

primer	acceptor	subject	run	time	isHos	total
1.2	A3	1	1	48	FALSE	3691
1.2	A3	1	2	48	FALSE	2826
1.2	A3	1	3	48	FALSE	9560
1.2	A3	2	1	48	FALSE	2140
1.2	A3	2	2	48	FALSE	3980
1.2	A3	2	3	48	FALSE	1196
1.2	A3	3	1	48	FALSE	2553
1.2	A3	3	2	48	FALSE	1778
1.2	A3	3	3	48	FALSE	1407
1.2	A3	4	1	48	FALSE	2141
1.2	A3	4	2	48	FALSE	1607
1.2	A3	4	3	48	FALSE	1268
1.2	A3	5	1	48	FALSE	3430
1.2	A3	5	2	48	FALSE	1651
1.2	A3	5	3	48	FALSE	4767
1.2	A3	6	1	48	FALSE	4170
1.2	A3	6	2	48	FALSE	3211
1.2	A3	6	3	48	FALSE	3638
1.2	A3	7	1	48	FALSE	3338
1.2	A3	7	2	48	FALSE	1250
1.2	A3	7	3	48	FALSE	3574
1.2	A3	HOS	4	18	TRUE	319
1.2	A3	HOS	4	24	TRUE	1954
1.2	A3	HOS	4	48	TRUE	2763
1.2	A4a	1	1	48	FALSE	3691
1.2	A4a	1	2	48	FALSE	2826
1.2	A4a	1	3	48	FALSE	9560

1.2	A4a	2	1	48	FALSE	2140
1.2	A4a	2	2	48	FALSE	3980
1.2	A4a	2	3	48	FALSE	1196
1.2	A4a	3	1	48	FALSE	2553
1.2	A4a	3	2	48	FALSE	1778
1.2	A4a	3	3	48	FALSE	1407
1.2	A4a	4	1	48	FALSE	2141
1.2	A4a	4	2	48	FALSE	1607
1.2	A4a	4	3	48	FALSE	1268
1.2	A4a	5	1	48	FALSE	3430
1.2	A4a	5	2	48	FALSE	1651
1.2	A4a	5	3	48	FALSE	4767
1.2	A4a	6	1	48	FALSE	4170
1.2	A4a	6	2	48	FALSE	3211
1.2	A4a	6	3	48	FALSE	3638
1.2	A4a	7	1	48	FALSE	3338
1.2	A4a	7	2	48	FALSE	1250
1.2	A4a	7	3	48	FALSE	3574
1.2	A4a	HOS	4	18	TRUE	319
1.2	A4a	HOS	4	24	TRUE	1954
1.2	A4a	HOS	4	48	TRUE	2763
1.2	A4b	1	1	48	FALSE	3691
1.2	A4b	1	2	48	FALSE	2826
1.2	A4b	1	3	48	FALSE	9560
1.2	A4b	2	1	48	FALSE	2140
1.2	A4b	2	2	48	FALSE	3980
1.2	A4b	2	3	48	FALSE	1196
1.2	A4b	3	1	48	FALSE	2553
1.2	A4b	3	2	48	FALSE	1778
1.2	A4b	3	3	48	FALSE	1407
1.2	A4b	4	1	48	FALSE	2141
1.2	A4b	4	2	48	FALSE	1607
1.2	A4b	4	3	48	FALSE	1268
1.2	A4b	5	1	48	FALSE	3430
1.2	A4b	5	2	48	FALSE	1651
1.2	A4b	5	3	48	FALSE	4767
1.2	A4b	6	1	48	FALSE	4170
1.2	A4b	6	2	48	FALSE	3211
1.2	A4b	6	3	48	FALSE	3638
1.2	A4b	7	1	48	FALSE	3338
1.2	A4b	7	2	48	FALSE	1250
1.2	A4b	7	3	48	FALSE	3574
1.2	A4b	HOS	4	18	TRUE	319
1.2	A4b	HOS	4	24	TRUE	1954
1.2	A4b	HOS	4	48	TRUE	2763
1.2	A4c	1	1	48	FALSE	3691
1.2	A4c	1	2	48	FALSE	2826
1.2	A4c	1	3	48	FALSE	9560

1.2	A4c	2	1	48	FALSE	2140
1.2	A4c	2	2	48	FALSE	3980
1.2	A4c	2	3	48	FALSE	1196
1.2	A4c	3	1	48	FALSE	2553
1.2	A4c	3	2	48	FALSE	1778
1.2	A4c	3	3	48	FALSE	1407
1.2	A4c	4	1	48	FALSE	2141
1.2	A4c	4	2	48	FALSE	1607
1.2	A4c	4	3	48	FALSE	1268
1.2	A4c	5	1	48	FALSE	3430
1.2	A4c	5	2	48	FALSE	1651
1.2	A4c	5	3	48	FALSE	4767
1.2	A4c	6	1	48	FALSE	4170
1.2	A4c	6	2	48	FALSE	3211
1.2	A4c	6	3	48	FALSE	3638
1.2	A4c	7	1	48	FALSE	3338
1.2	A4c	7	2	48	FALSE	1250
1.2	A4c	7	3	48	FALSE	3574
1.2	A4c	HOS	4	18	TRUE	319
1.2	A4c	HOS	4	24	TRUE	1954
1.2	A4c	HOS	4	48	TRUE	2763
1.2	A5	1	1	48	FALSE	3691
1.2	A5	1	2	48	FALSE	2826
1.2	A5	1	3	48	FALSE	9560
1.2	A5	2	1	48	FALSE	2140
1.2	A5	2	2	48	FALSE	3980
1.2	A5	2	3	48	FALSE	1196
1.2	A5	3	1	48	FALSE	2553
1.2	A5	3	2	48	FALSE	1778
1.2	A5	3	3	48	FALSE	1407
1.2	A5	4	1	48	FALSE	2141
1.2	A5	4	2	48	FALSE	1607
1.2	A5	4	3	48	FALSE	1268
1.2	A5	5	1	48	FALSE	3430
1.2	A5	5	2	48	FALSE	1651
1.2	A5	5	3	48	FALSE	4767
1.2	A5	6	1	48	FALSE	4170
1.2	A5	6	2	48	FALSE	3211
1.2	A5	6	3	48	FALSE	3638
1.2	A5	7	1	48	FALSE	3338
1.2	A5	7	2	48	FALSE	1250
1.2	A5	7	3	48	FALSE	3574
1.2	A5	HOS	4	18	TRUE	319
1.2	A5	HOS	4	24	TRUE	1954
1.2	A5	HOS	4	48	TRUE	2763
1.2	A9a	1	1	48	FALSE	3691
1.2	A9a	1	2	48	FALSE	2826
1.2	A9a	1	3	48	FALSE	9560

1.2	A9a	2	1	48	FALSE	2140
1.2	A9a	2	2	48	FALSE	3980
1.2	A9a	2	3	48	FALSE	1196
1.2	A9a	3	1	48	FALSE	2553
1.2	A9a	3	2	48	FALSE	1778
1.2	A9a	3	3	48	FALSE	1407
1.2	A9a	4	1	48	FALSE	2141
1.2	A9a	4	2	48	FALSE	1607
1.2	A9a	4	3	48	FALSE	1268
1.2	A9a	5	1	48	FALSE	3430
1.2	A9a	5	2	48	FALSE	1651
1.2	A9a	5	3	48	FALSE	4767
1.2	A9a	6	1	48	FALSE	4170
1.2	A9a	6	2	48	FALSE	3211
1.2	A9a	6	3	48	FALSE	3638
1.2	A9a	7	1	48	FALSE	3338
1.2	A9a	7	2	48	FALSE	1250
1.2	A9a	7	3	48	FALSE	3574
1.2	A9a	HOS	4	18	TRUE	319
1.2	A9a	HOS	4	24	TRUE	1954
1.2	A9a	HOS	4	48	TRUE	2763
1.4	A3	1	1	48	FALSE	2538
1.4	A3	1	2	48	FALSE	155
1.4	A3	1	3	48	FALSE	6483
1.4	A3	2	1	48	FALSE	1699
1.4	A3	2	2	48	FALSE	353
1.4	A3	2	3	48	FALSE	962
1.4	A3	3	1	48	FALSE	2231
1.4	A3	3	2	48	FALSE	149
1.4	A3	3	3	48	FALSE	1082
1.4	A3	4	1	48	FALSE	1869
1.4	A3	4	2	48	FALSE	90
1.4	A3	4	3	48	FALSE	991
1.4	A3	5	1	48	FALSE	2535
1.4	A3	5	2	48	FALSE	110
1.4	A3	5	3	48	FALSE	3851
1.4	A3	6	1	48	FALSE	3623
1.4	A3	6	2	48	FALSE	236
1.4	A3	6	3	48	FALSE	2244
1.4	A3	7	1	48	FALSE	3209
1.4	A3	7	2	48	FALSE	64
1.4	A3	7	3	48	FALSE	2834
1.4	A3	HOS	4	18	TRUE	452
1.4	A3	HOS	4	24	TRUE	2376
1.4	A3	HOS	4	48	TRUE	3987
1.4	A4a	1	1	48	FALSE	2538
1.4	A4a	1	2	48	FALSE	155
1.4	A4a	1	3	48	FALSE	6483

1.4	A4a	2	1	48	FALSE	1699
1.4	A4a	2	2	48	FALSE	353
1.4	A4a	2	3	48	FALSE	962
1.4	A4a	3	1	48	FALSE	2231
1.4	A4a	3	2	48	FALSE	149
1.4	A4a	3	3	48	FALSE	1082
1.4	A4a	4	1	48	FALSE	1869
1.4	A4a	4	2	48	FALSE	90
1.4	A4a	4	3	48	FALSE	991
1.4	A4a	5	1	48	FALSE	2535
1.4	A4a	5	2	48	FALSE	110
1.4	A4a	5	3	48	FALSE	3851
1.4	A4a	6	1	48	FALSE	3623
1.4	A4a	6	2	48	FALSE	236
1.4	A4a	6	3	48	FALSE	2244
1.4	A4a	7	1	48	FALSE	3209
1.4	A4a	7	2	48	FALSE	64
1.4	A4a	7	3	48	FALSE	2834
1.4	A4a	HOS	4	18	TRUE	452
1.4	A4a	HOS	4	24	TRUE	2376
1.4	A4a	HOS	4	48	TRUE	3987
1.4	A4b	1	1	48	FALSE	2538
1.4	A4b	1	2	48	FALSE	155
1.4	A4b	1	3	48	FALSE	6483
1.4	A4b	2	1	48	FALSE	1699
1.4	A4b	2	2	48	FALSE	353
1.4	A4b	2	3	48	FALSE	962
1.4	A4b	3	1	48	FALSE	2231
1.4	A4b	3	2	48	FALSE	149
1.4	A4b	3	3	48	FALSE	1082
1.4	A4b	4	1	48	FALSE	1869
1.4	A4b	4	2	48	FALSE	90
1.4	A4b	4	3	48	FALSE	991
1.4	A4b	5	1	48	FALSE	2535
1.4	A4b	5	2	48	FALSE	110
1.4	A4b	5	3	48	FALSE	3851
1.4	A4b	6	1	48	FALSE	3623
1.4	A4b	6	2	48	FALSE	236
1.4	A4b	6	3	48	FALSE	2244
1.4	A4b	7	1	48	FALSE	3209
1.4	A4b	7	2	48	FALSE	64
1.4	A4b	7	3	48	FALSE	2834
1.4	A4b	HOS	4	18	TRUE	452
1.4	A4b	HOS	4	24	TRUE	2376
1.4	A4b	HOS	4	48	TRUE	3987
1.4	A4c	1	1	48	FALSE	2538
1.4	A4c	1	2	48	FALSE	155
1.4	A4c	1	3	48	FALSE	6483

1.4	A4c	2	1	48	FALSE	1699
1.4	A4c	2	2	48	FALSE	353
1.4	A4c	2	3	48	FALSE	962
1.4	A4c	3	1	48	FALSE	2231
1.4	A4c	3	2	48	FALSE	149
1.4	A4c	3	3	48	FALSE	1082
1.4	A4c	4	1	48	FALSE	1869
1.4	A4c	4	2	48	FALSE	90
1.4	A4c	4	3	48	FALSE	991
1.4	A4c	5	1	48	FALSE	2535
1.4	A4c	5	2	48	FALSE	110
1.4	A4c	5	3	48	FALSE	3851
1.4	A4c	6	1	48	FALSE	3623
1.4	A4c	6	2	48	FALSE	236
1.4	A4c	6	3	48	FALSE	2244
1.4	A4c	7	1	48	FALSE	3209
1.4	A4c	7	2	48	FALSE	64
1.4	A4c	7	3	48	FALSE	2834
1.4	A4c	HOS	4	18	TRUE	452
1.4	A4c	HOS	4	24	TRUE	2376
1.4	A4c	HOS	4	48	TRUE	3987
1.4	A5	1	1	48	FALSE	2538
1.4	A5	1	2	48	FALSE	155
1.4	A5	1	3	48	FALSE	6483
1.4	A5	2	1	48	FALSE	1699
1.4	A5	2	2	48	FALSE	353
1.4	A5	2	3	48	FALSE	962
1.4	A5	3	1	48	FALSE	2231
1.4	A5	3	2	48	FALSE	149
1.4	A5	3	3	48	FALSE	1082
1.4	A5	4	1	48	FALSE	1869
1.4	A5	4	2	48	FALSE	90
1.4	A5	4	3	48	FALSE	991
1.4	A5	5	1	48	FALSE	2535
1.4	A5	5	2	48	FALSE	110
1.4	A5	5	3	48	FALSE	3851
1.4	A5	6	1	48	FALSE	3623
1.4	A5	6	2	48	FALSE	236
1.4	A5	6	3	48	FALSE	2244
1.4	A5	7	1	48	FALSE	3209
1.4	A5	7	2	48	FALSE	64
1.4	A5	7	3	48	FALSE	2834
1.4	A5	HOS	4	18	TRUE	452
1.4	A5	HOS	4	24	TRUE	2376
1.4	A5	HOS	4	48	TRUE	3987
1.4	A9a	1	1	48	FALSE	2538
1.4	A9a	1	2	48	FALSE	155
1.4	A9a	1	3	48	FALSE	6483

1.4	A9a	2	1	48	FALSE	1699
1.4	A9a	2	2	48	FALSE	353
1.4	A9a	2	3	48	FALSE	962
1.4	A9a	3	1	48	FALSE	2231
1.4	A9a	3	2	48	FALSE	149
1.4	A9a	3	3	48	FALSE	1082
1.4	A9a	4	1	48	FALSE	1869
1.4	A9a	4	2	48	FALSE	90
1.4	A9a	4	3	48	FALSE	991
1.4	A9a	5	1	48	FALSE	2535
1.4	A9a	5	2	48	FALSE	110
1.4	A9a	5	3	48	FALSE	3851
1.4	A9a	6	1	48	FALSE	3623
1.4	A9a	6	2	48	FALSE	236
1.4	A9a	6	3	48	FALSE	2244
1.4	A9a	7	1	48	FALSE	3209
1.4	A9a	7	2	48	FALSE	64
1.4	A9a	7	3	48	FALSE	2834
1.4	A9a	HOS	4	18	TRUE	452
1.4	A9a	HOS	4	24	TRUE	2376
1.4	A9a	HOS	4	48	TRUE	3987
1.3	A3	1	1	48	FALSE	228
1.3	A3	1	2	48	FALSE	309
1.3	A3	1	3	48	FALSE	1010
1.3	A3	2	1	48	FALSE	90
1.3	A3	2	2	48	FALSE	170
1.3	A3	2	3	48	FALSE	103
1.3	A3	3	1	48	FALSE	204
1.3	A3	3	2	48	FALSE	142
1.3	A3	3	3	48	FALSE	135
1.3	A3	4	1	48	FALSE	233
1.3	A3	4	2	48	FALSE	125
1.3	A3	4	3	48	FALSE	180
1.3	A3	5	1	48	FALSE	123
1.3	A3	5	2	48	FALSE	95
1.3	A3	5	3	48	FALSE	303
1.3	A3	6	1	48	FALSE	322
1.3	A3	6	2	48	FALSE	249
1.3	A3	6	3	48	FALSE	281
1.3	A3	7	1	48	FALSE	245
1.3	A3	7	2	48	FALSE	129
1.3	A3	7	3	48	FALSE	317
1.3	A3	HOS	4	18	TRUE	37
1.3	A3	HOS	4	24	TRUE	173
1.3	A3	HOS	4	48	TRUE	194
1.3	A4a	1	1	48	FALSE	228
1.3	A4a	1	2	48	FALSE	309
1.3	A4a	1	3	48	FALSE	1010

1.3	A4a	2	1	48	FALSE	90
1.3	A4a	2	2	48	FALSE	170
1.3	A4a	2	3	48	FALSE	103
1.3	A4a	3	1	48	FALSE	204
1.3	A4a	3	2	48	FALSE	142
1.3	A4a	3	3	48	FALSE	135
1.3	A4a	4	1	48	FALSE	233
1.3	A4a	4	2	48	FALSE	125
1.3	A4a	4	3	48	FALSE	180
1.3	A4a	5	1	48	FALSE	123
1.3	A4a	5	2	48	FALSE	95
1.3	A4a	5	3	48	FALSE	303
1.3	A4a	6	1	48	FALSE	322
1.3	A4a	6	2	48	FALSE	249
1.3	A4a	6	3	48	FALSE	281
1.3	A4a	7	1	48	FALSE	245
1.3	A4a	7	2	48	FALSE	129
1.3	A4a	7	3	48	FALSE	317
1.3	A4a	HOS	4	18	TRUE	37
1.3	A4a	HOS	4	24	TRUE	173
1.3	A4a	HOS	4	48	TRUE	194
1.3	A4b	1	1	48	FALSE	228
1.3	A4b	1	2	48	FALSE	309
1.3	A4b	1	3	48	FALSE	1010
1.3	A4b	2	1	48	FALSE	90
1.3	A4b	2	2	48	FALSE	170
1.3	A4b	2	3	48	FALSE	103
1.3	A4b	3	1	48	FALSE	204
1.3	A4b	3	2	48	FALSE	142
1.3	A4b	3	3	48	FALSE	135
1.3	A4b	4	1	48	FALSE	233
1.3	A4b	4	2	48	FALSE	125
1.3	A4b	4	3	48	FALSE	180
1.3	A4b	5	1	48	FALSE	123
1.3	A4b	5	2	48	FALSE	95
1.3	A4b	5	3	48	FALSE	303
1.3	A4b	6	1	48	FALSE	322
1.3	A4b	6	2	48	FALSE	249
1.3	A4b	6	3	48	FALSE	281
1.3	A4b	7	1	48	FALSE	245
1.3	A4b	7	2	48	FALSE	129
1.3	A4b	7	3	48	FALSE	317
1.3	A4b	HOS	4	18	TRUE	37
1.3	A4b	HOS	4	24	TRUE	173
1.3	A4b	HOS	4	48	TRUE	194
1.3	A4c	1	1	48	FALSE	228
1.3	A4c	1	2	48	FALSE	309
1.3	A4c	1	3	48	FALSE	1010

1.3	A4c	2	1	48	FALSE	90
1.3	A4c	2	2	48	FALSE	170
1.3	A4c	2	3	48	FALSE	103
1.3	A4c	3	1	48	FALSE	204
1.3	A4c	3	2	48	FALSE	142
1.3	A4c	3	3	48	FALSE	135
1.3	A4c	4	1	48	FALSE	233
1.3	A4c	4	2	48	FALSE	125
1.3	A4c	4	3	48	FALSE	180
1.3	A4c	5	1	48	FALSE	123
1.3	A4c	5	2	48	FALSE	95
1.3	A4c	5	3	48	FALSE	303
1.3	A4c	6	1	48	FALSE	322
1.3	A4c	6	2	48	FALSE	249
1.3	A4c	6	3	48	FALSE	281
1.3	A4c	7	1	48	FALSE	245
1.3	A4c	7	2	48	FALSE	129
1.3	A4c	7	3	48	FALSE	317
1.3	A4c	HOS	4	18	TRUE	37
1.3	A4c	HOS	4	24	TRUE	173
1.3	A4c	HOS	4	48	TRUE	194
1.3	A5	1	1	48	FALSE	228
1.3	A5	1	2	48	FALSE	309
1.3	A5	1	3	48	FALSE	1010
1.3	A5	2	1	48	FALSE	90
1.3	A5	2	2	48	FALSE	170
1.3	A5	2	3	48	FALSE	103
1.3	A5	3	1	48	FALSE	204
1.3	A5	3	2	48	FALSE	142
1.3	A5	3	3	48	FALSE	135
1.3	A5	4	1	48	FALSE	233
1.3	A5	4	2	48	FALSE	125
1.3	A5	4	3	48	FALSE	180
1.3	A5	5	1	48	FALSE	123
1.3	A5	5	2	48	FALSE	95
1.3	A5	5	3	48	FALSE	303
1.3	A5	6	1	48	FALSE	322
1.3	A5	6	2	48	FALSE	249
1.3	A5	6	3	48	FALSE	281
1.3	A5	7	1	48	FALSE	245
1.3	A5	7	2	48	FALSE	129
1.3	A5	7	3	48	FALSE	317
1.3	A5	HOS	4	18	TRUE	37
1.3	A5	HOS	4	24	TRUE	173
1.3	A5	HOS	4	48	TRUE	194
1.3	A9a	1	1	48	FALSE	228
1.3	A9a	1	2	48	FALSE	309
1.3	A9a	1	3	48	FALSE	1010

1.3	A9a	2	1	48	FALSE	90
1.3	A9a	2	2	48	FALSE	170
1.3	A9a	2	3	48	FALSE	103
1.3	A9a	3	1	48	FALSE	204
1.3	A9a	3	2	48	FALSE	142
1.3	A9a	3	3	48	FALSE	135
1.3	A9a	4	1	48	FALSE	233
1.3	A9a	4	2	48	FALSE	125
1.3	A9a	4	3	48	FALSE	180
1.3	A9a	5	1	48	FALSE	123
1.3	A9a	5	2	48	FALSE	95
1.3	A9a	5	3	48	FALSE	303
1.3	A9a	6	1	48	FALSE	322
1.3	A9a	6	2	48	FALSE	249
1.3	A9a	6	3	48	FALSE	281
1.3	A9a	7	1	48	FALSE	245
1.3	A9a	7	2	48	FALSE	129
1.3	A9a	7	3	48	FALSE	317
1.3	A9a	HOS	4	18	TRUE	37
1.3	A9a	HOS	4	24	TRUE	173
1.3	A9a	HOS	4	48	TRUE	194

Supplemental References

1. Burset, M., Seledtsov, I.A. and Solovyev, V.V. (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic acids research*, 28, 4364-4375.
2. Purcell, D.F. and Martin, M.A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *Journal of virology*, 67, 6365-6378.
3. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M. *et al.* (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381, 661-666.
4. Landau, N.R. and Littman, D.R. (1992) Packaging system for rapid production of murine leukemia virus vectors with variable tropism. *Journal of virology*, 66, 5110-5113.
5. Srinivasakumar, N., Chazal, N., Helga-Maria, C., Prasad, S., Hammarskjold, M.L. and Rekosh, D. (1997) The effect of viral regulatory protein expression on gene delivery by human immunodeficiency virus type 1 vectors produced in stable packaging cell lines. *Journal of virology*, 71, 5841-5848.
6. Exline, C.M., Feng, Z. and Stoltzfus, C.M. (2008) Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. *Journal of virology*, 82, 3921-3931.
7. Bohne, J., Wodrich, H. and Krausslich, H.G. (2005) Splicing of human immunodeficiency virus RNA is position-dependent suggesting sequential removal of introns from the 5' end. *Nucleic acids research*, 33, 825-837.
8. Benko, D.M., Schwartz, S., Pavlakis, G.N. and Felber, B.K. (1990) A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. *Journal of virology*, 64, 2505-2518.
9. Salfeld, J., Gottlinger, H.G., Sia, R.A., Park, R.E., Sodroski, J.G. and Haseltine, W.A. (1990) A tripartite HIV-1 tat-env-rev fusion protein. *The EMBO journal*, 9, 965-970.
10. Asang, C., Hauber, I. and Schaal, H. (2008) Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic acids research*, 36, 1450-1463.
11. Caputi, M., Freund, M., Kammler, S., Asang, C. and Schaal, H. (2004) A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *Journal of virology*, 78, 6517-6526.
12. Abbink, T.E. and Berkhout, B. (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *Journal of virology*, 82, 3090-3098.
13. Almarza, D., Bussadori, G., Navarro, M., Mavilio, F., Larcher, F. and Murillas, R. (2011) Risk assessment in skin gene therapy: viral-cellular fusion transcripts generated by proviral transcriptional read-through in keratinocytes transduced with self-inactivating lentiviral vectors. *Gene therapy*, 18, 674-681.