Genomic HEXploring allows landscaping of novel potential splicing regulatory elements

Steffen Erkelenz, Stephan Theiss, Marianne Otte, Marek Widera, Jan Otto Peter and Heiner Schaal

Supplemental Material

Supplemental Figure 1. As a complement to the quantification of spliced mRNA levels shown in Fig. 7, both spliced and unspliced messages are shown for all four mutations in HIV-1 exon 4 as key experiments (upper right panel). The remainder of Suppl. Fig. 1 corresponds to Fig. 7 and is shown here for completeness.

Supplemental Figure 2. Exonic splicing motif difference (ESMD) between mutant and reference sequence, derived from the well-established algorithms ESEfinder, RESCUE-ESE, FAS-ESS-hex3, PESX and PESR, is highly correlated with HEXplorer score difference in all 29 mutations of HIV-1 exons 1, 2, 2b, 3 and 4 also shown in Fig. 8 ($r = 0.93$).

Supplemental Figure 3. Both (A) exon-intron hexamer score Z_{EI} and (B) weak-strong hexamer score Z_{WS} quantitatively correlate with ESRseq score ($r = 0.72$) for 1,182 ESE and 1,090 ESS hexamers identified as exonic splicing regulatory elements by Ke et al. (2011).

Supplemental Figure 4. ESRseq score difference (mutant – reference) (Ke et al. (2011)) is highly correlated with HEXplorer score difference in all 29 mutations of HIV-1 exons 1, 2, 2b, 3 and 4 also shown in Fig. 8 ($r = 0.93$).

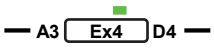
Supplemental Table 1. Primers used for site-directed mutagenesis for LTR ex2 ex3 splicing reporters.

Supplemental Table 2. Primers used for site-directed mutagenesis for SV-*env* splicing reporters.

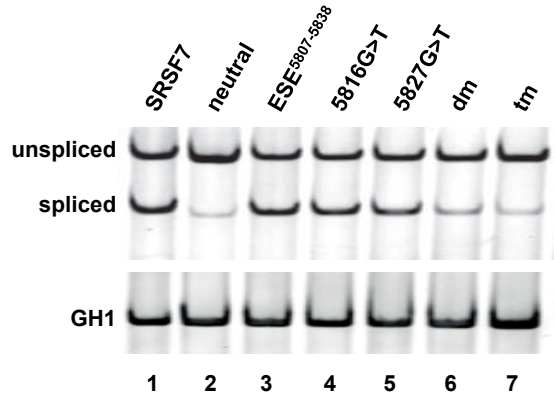
Supplemental Table 3. Primers used for semi-quantitative RT-PCR analyses.

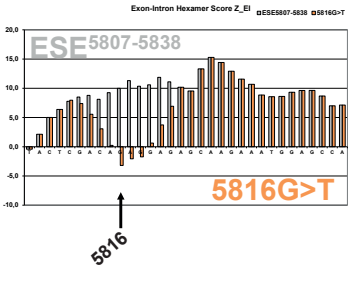
Reference:

Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J. and Chasin, L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. Genome Res., 21, 1360-1374.

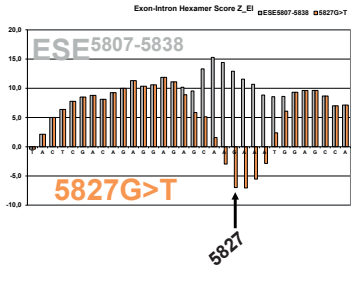


ESE$^{5807-5838}$	TACTCGACAGAGGAGAGCAAGAAATGGAGCCA
5816G>T	---------T----------
5827G>T	----------T----------
5816G>T; 5827G>T (dm)	---------T----------T----------
5816G>T; 5821G>T; 5827G>T (tm)	---------T----T-----T----------

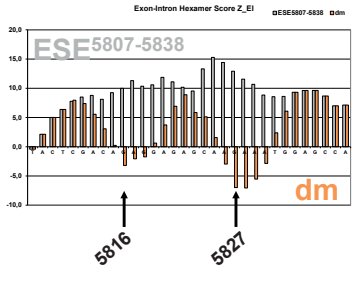




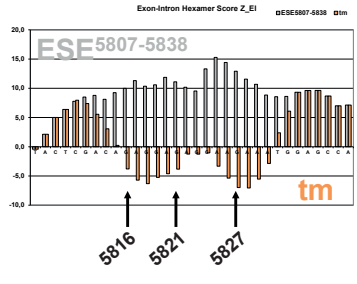
ΔHZ_{EI} = -79.10



ΔHZ_{EI} = -119.40

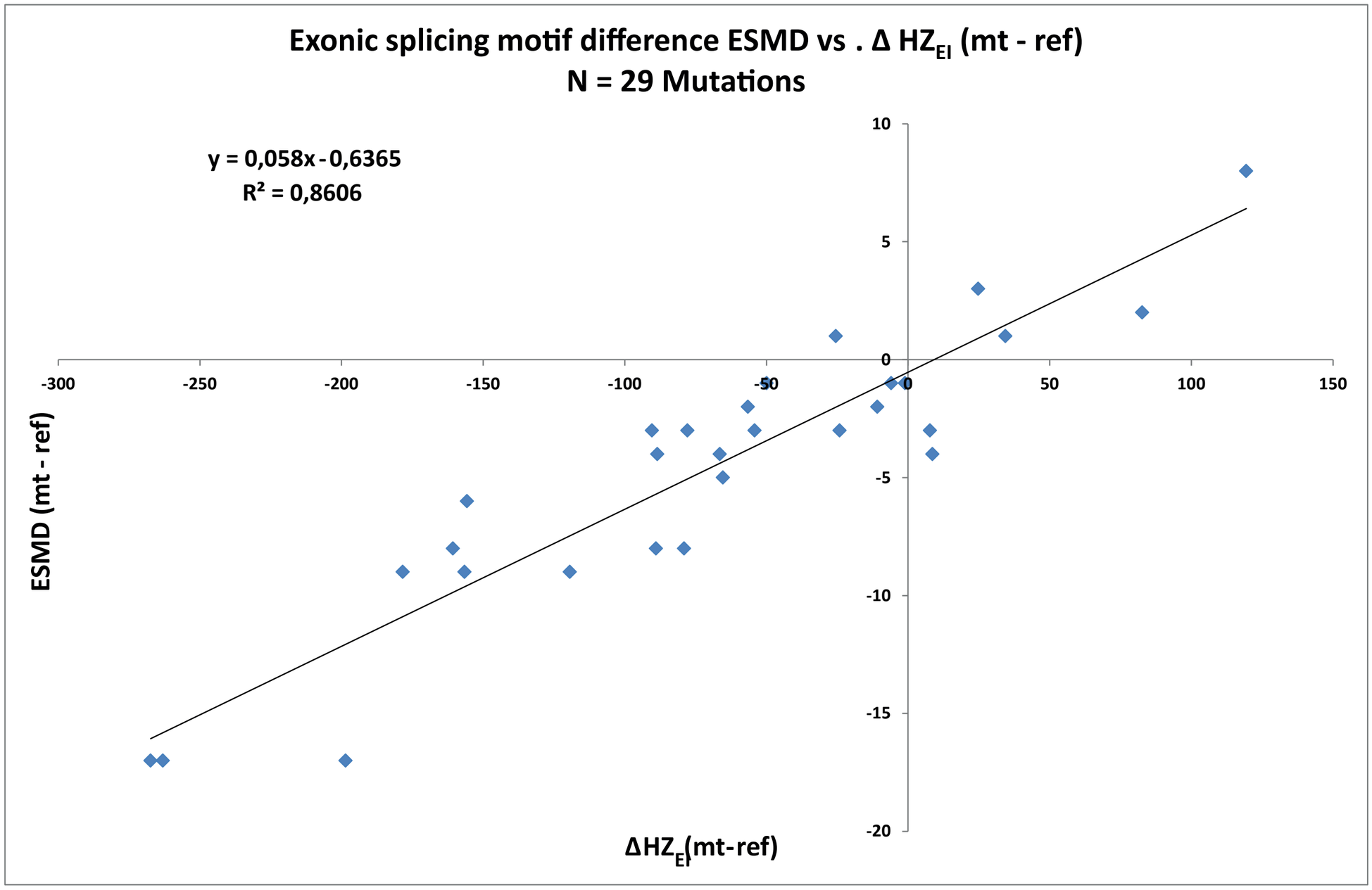


ΔHZ_{EI} = -198.60



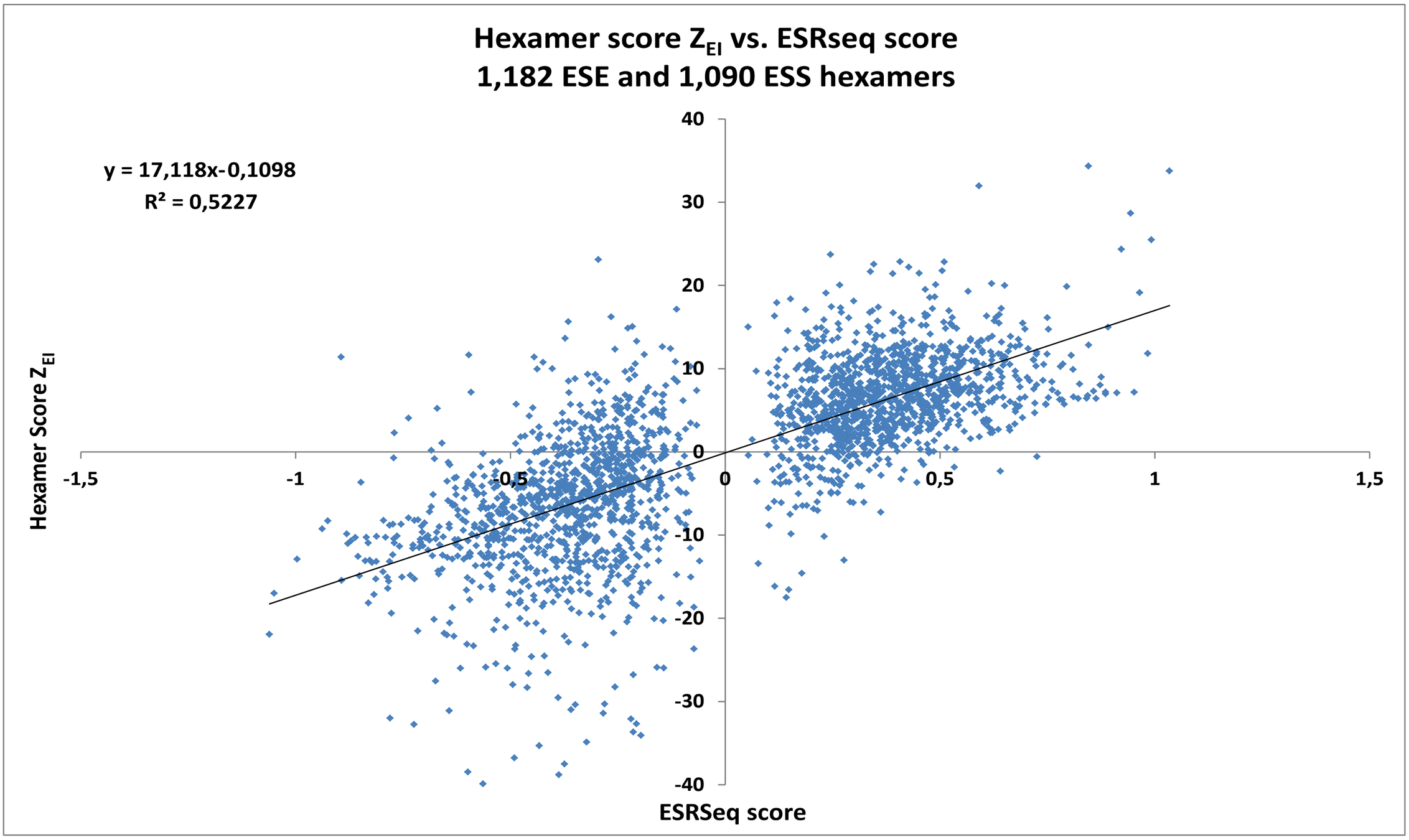
ΔHZ_{EI} = -263.09

Supplemental Figure 1

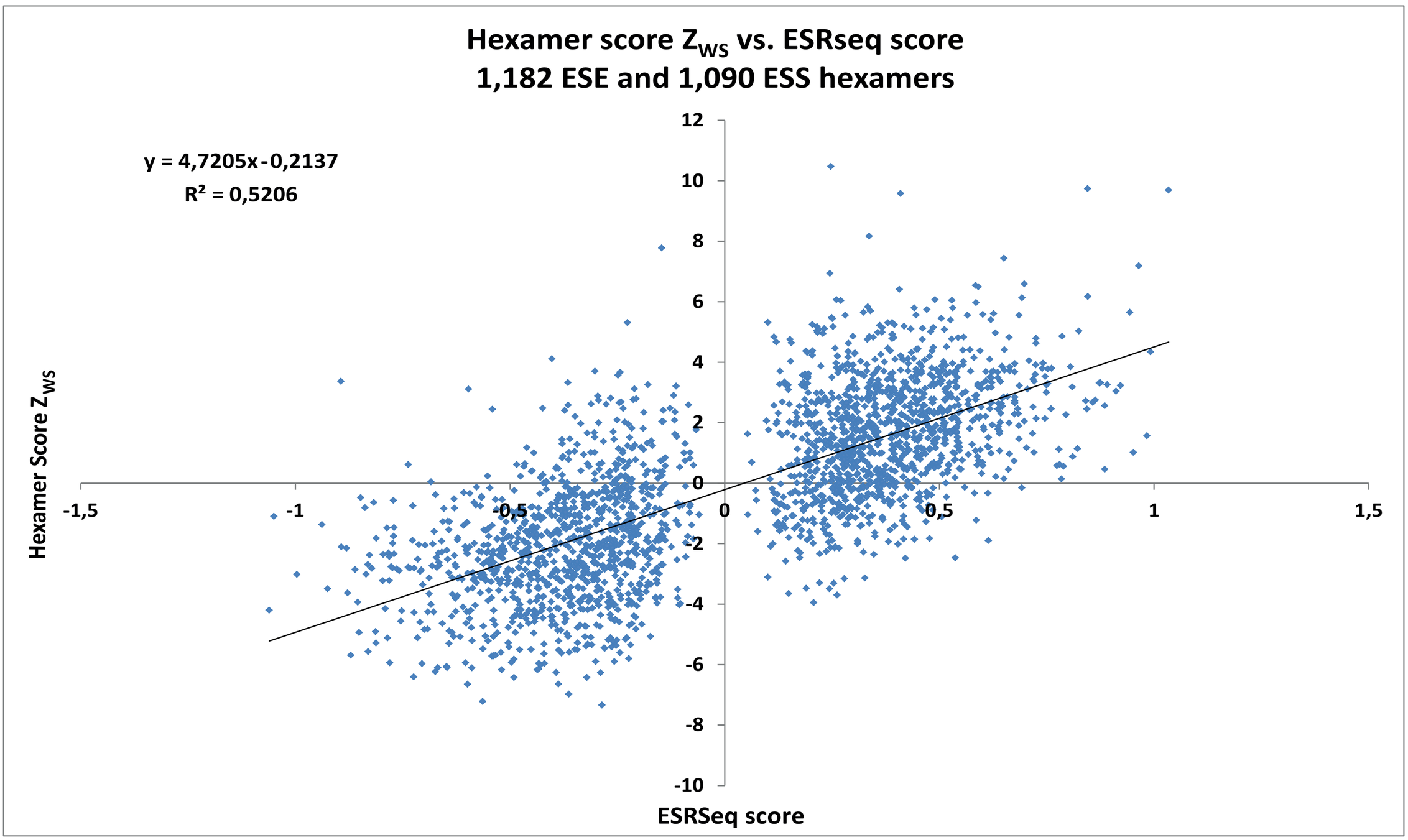


Supplemental Figure 2

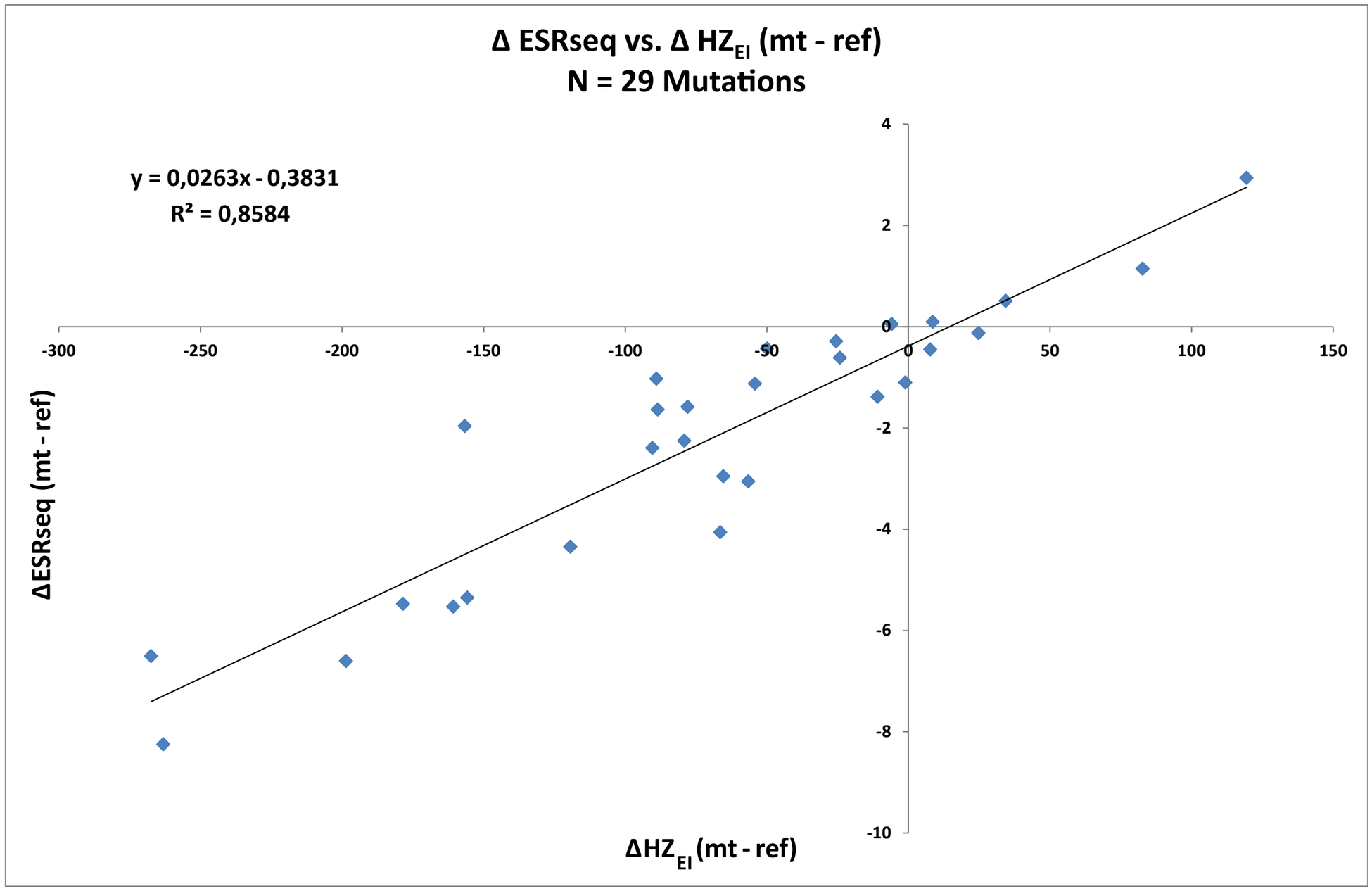
A



B



Supplemental Figure 3



Supplemental Figure 4

Genomic HEXploring allows landscaping of novel potential splicing regulatory elements

Steffen Erkelenz[1], Stephan Theiss[2], Marianne Otte[3], Marek Widera[1], Jan Otto Peter[1] and Heiner Schaal[1,*]

Supplemental Tables

Supplemental Table 1. LTR ex2 ex3 splicing reporters

Cloned construct	Primer	Sequence
forward		
LTR ex2 ex3 ESSV $^{-}$	#2949	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTG AAT ATC AAG CAG GAC ATA AC 3’
LTR ex2 ex3 ESSV^{-} ESE -24G>T	#2951	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTT AAT ATC AAG CAG GAC ATA ACA AGG 3’
LTR ex2 ex3 ESSV^{-} ESE -24G>A	#2986	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTA AAT ATC AAG CAG GAC ATA ACA AGG 3’
LTR ex2 ex3 ESSV^{-} ESE -17A>T	#2950	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTG AAT ATC TAG CAG GAC ATA ACA AGG TAG G 3’
LTR ex2 ex3 ESSV^{-} ESE -17A>C	#2987	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTG AAT ATC CAG CAG GAC ATA ACA AGG TAG G 3’
LTR ex2 ex3 ESSV^{-} ESE -17A>G	#3016	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTT AAT ATC CAG

		CAG GAC ATA ACA AGG TAG G 3’
LTR ex2 ex3 ESSV⁻ ESE -15G>T,-17A>T	#3017	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTG AAT ATC TAT CAG GAC ATA ACA AGG TAG GAT C 3’
LTR ex2 ex3 ESSV⁻ ESE -24G>T,-17A>T	#2985	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTT AAT ATC TAG CAG GAC ATA ACA AGG TAG G 3’
LTR ex2 ex3 ESSV⁻ ESE -24G>T,-17A>C	#3015	5’ TTT CAG ACT CTG CTA TAA GAA AGG CCT TAT TCT GAC ACA TAG TTC TCC CTC TGT GTT AAT ATC CAG CAG GAC ATA ACA AGG TAG G 3’
LTR ex2 ex3 ESSV⁻ ESE -15G>T	#3202	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA TAT CAA TCA GGA CAT AAC AAG GTA GGA T
LTR ex2 ex3 ESSV⁻ ESE -15G>C	#3203	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA TAT CAA CCA GGA CAT AAC AAG GTA GGA T 3’
LTR ex2 ex3 ESSV⁻ ESE -26G>T	#3204	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TTT GAA TAT CAA GCA GGA CAT AAC AAG 3’
LTR ex2 ex3 ESSV⁻ ESE -26G>C	#3205	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TCT GAA TAT CAA GCA GGA CAT AAC AAG 3’
LTR ex2 ex3 ESSV⁻ ESE -21T>G	#3206	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA GAT CAA

		GCA GGA CAT AAC AAG GTA G 3’
LTR ex2 ex3 ESSV⁻ ESE -21T>A	#3207	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA AAT CAA GCA GGA CAT AAC AAG GTA G 3’
LTR ex2 ex3 ESSV⁻ ESE -9C>A	#3208	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA TAT CAA GCA GGA AAT AAC AAG GTA GGA TCT CTA CAA TAC 3’
LTR ex2 ex3 ESSV⁻ ESE -9C>G	#3209	5’ TTT TCA GAC TCT GCT ATA AGA AAG GCC TTA TTC TGA CAC ATA GTT CTC CCT CTG TGT GAA TAT CAA GCA GGA GAT AAC AAG GTA GGA TCT CTA CAA TAC 3’
LTR ex2 ex3 ESSV⁻ (pNEU)	#3297	5’ TTT TTC AGA ATC TGC TAT AAG AAA TAC CAT ATT AGG ACG TAT AGT ATC GCC ACG TTG TGA ATA TCA AGC AGG ACA TAA C 3’
LTR ex2 ex3 ESSV⁻ (-36A>C)	#4482	5’ TTT TCA GAA TCT GCT ATA AGA AAT ACC ATA TTA GGA CGT ATA GTT CGT CCT AGG TGT GAA TAT CAA GCA GGA CA 3‘
LTR ex2 ex3 ESSV⁻ (-29G>C)	#4481	5’ TTT TCA GAA TCT GCT ATA AGA AAT ACC ATA TTA GGA CGT ATA GTT AGT CCT ACG TGT GAA TAT CAA GCA GGA CAT AAC 3’
LTR ex2 ex3 ESSV⁻ (-29G>C; -36A>C; dm)	#4483	5‘ TTT TCA GAA TCT GCT ATA AGA AAT ACC ATA TTA GGA CGT ATA GTT CGT CCT ACG TGT GAA TAT CAA GCA GGA CAT AAC 3‘

reverse		
all	#2588	5’ CTT TAC GAT GCC ATT GGG A 3’

Supplemental Table 2. SV-*env* splicing reporters

Cloned construct	Primer	Sequence
SV exon 3 D4-env and SV exon 3 ESSV $^{-}$ D 4-env	#1906	5‘ GAT CGA ATT CAC TCT GCT ATA AGA A 3‘
	#1907	5‘ GAT CGA GCT CCT TGT TAT GTC CTG C 3‘
SV exon 3 (part I) D4-env	#1958	5‘ AAT TCA CTC TGC TAT AAG AAA GGC CTT ATG AGC T 3‘
	#1959	5‘ CAT AAG GCC TTT CTT ATA GCA GAG TG 3‘
SV exon 3 (part II) D4-env	#1960	5‘ AAT TCT TTA GGA CAC ATA GTT AGC CCT AGG GAG CT 3‘
	#1961	5‘ CCC TAG GGC TAA CTA TGT GTC CTA AAG 3‘
SV exon 3 (part II ESSV $^{-}$) D4-env	#1962	5‘ AAT TCT TTC TGA CAC ATA GTT CTC CCT CTG TGA GCT 3’
	#1963	5‘ CAC AGA GGG AGA ACT ATG TGT CAG AAA G 3‘
SV exon 3 (part III) D4-env	#1964	5‘ AAT TCT TGT GAA TAT CAA GCA GGA CAT AAC GAG CT 3‘
	#1965	5‘ CGT TAT GTC CTG CTT GAT ATT CAC AAG 3‘
SV ESE$^{705\text{-}29}$ D1-env/eGFP	#4473	5‘ AAT AAG AAT TCG CTG AAG CGC GCA CGG CAA GAG GCG GAG CTC CTG GTG AGT ACC TT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{705\text{-}29}$ 708G>T; 718C>G (dm) D1-env/eGFP	#4474	5‘ AAT AAG AAT TCG CTG TAG CGC GCA GGG CAA GAG GCG GAG CTC CTG GTG AGT ACC TT 3‘

	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{4932-62}$D1-env/eGFP	#4476	5‘ AAT AAG AAT TCT GGA AAG GAC CAG CAA AGC TCC TCT GGA AAG GAG CTC CTG GTG AGT ACC TT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{4932-62}$ 4942C>T; 4947A>T (dm) D1-env/eGFP	#4477	5‘ AAT AAG AAT TCT GGA AAG GAC TAG CAT AGC TCC TCT GGA AAG GAG CTC CTG GTG AGT ACC TT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5807-38}$ D1-env/eGFP	#4430	5‘ AAT AAG AAT TCT ACT CGA CAG AGG AGA GCA AGA AAT GGA GCC AGA GCT CCT GGT GAG TAC CTT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5807-38}$ 5816G>T D1-env/eGFP	#4431	5‘ AAT AAG AAT TCT ACT CGA CAT AGG AGA GCA AGA AAT GGA GCC AGA GCT CCT GGT GAG TAC CTT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5807-38}$ 5827G>T D1-env/eGFP	#4432	5‘ AAT AAG AAT TCT ACT CGA CAG AGG AGA GCA ATA AAT GGA GCC AGA GCT CCT GGT GAG TAC CTT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5807-38}$ 5816G>T; 5827G>T (dm) D1-env/eGFP	#4433	5‘ AAT AAG AAT TCT ACT CGA CAT AGG AGA GCA ATA AAT GGA GCC AGA GCT CCT GGT GAG TAC CTT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’

SV ESE$^{5807\text{-}38}$ 5816G>T; 5821G>T; 5827G>T (tm) D1-env/eGFP	#4434	5‘ AAT AAG AAT TCT ACT CGA CAT AGG ATA GCA ATA AAT GGA GCC AGA GCT CCT GGT GAG TAC CTT 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5005\text{-}32}$ D1-env/eGFP	#4965	5‘ GGG CCT AGG AAT TCT AGT GCC AAG AAG AAA AGC AAA GAT CAT GAG CTC CTG GTG AGT ACC TTA AGC TCT CCG 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5005\text{-}32}$ 5015A>T D4-env/eGFP	#4966	5‘ GGG CCT AGG AAT TCT AGT GCC AAG TAG AAA AGC AAA GAT CAT GAG CTC CTG GTG AGT ACC TTA AGC TCT CCG 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5005\text{-}32}$ 5025A>T D4-env/eGFP	#4967	5‘ GGG CCT AGG AAT TCT AGT GCC AAG AAG AAA AGC ATA GAT CAT GAG CTC CTG GTG AGT ACC TTA AGC TCT CCG 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
SV ESE$^{5005\text{-}32}$ 5015A>T; 5025A>T (dm) D4-env/eGFP	#4968	5‘ GGG CCT AGG AAT TCT AGT GCC AAG TAG AAA AGC ATA GAT CAT GAG CTC CTG GTG AGT ACC TTA AGC TCT CCG 3‘
	#640	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’

Supplemental Table 3. Primers used for semi-quantitative RT-PCR

Cloned construct	Primer	Sequence
Reporter mRNAs (LTR ex2 ex3)	#1544	5’ CTT GAA AGC GAA AGT AAA GC 3’
	#2588	5’ CTT TAC GAT GCC ATT GGG A 3’
	#3210	5’ TGA GGA GGC TTT TTT GGA GG 3’
	#3211	5’ TTC ACT AAT CGA ATG GAT CTG TC 3’
GH1 (transfection control)	#1224	5’ TCT TCC AGC CTC CCA TCA GCG TTT GG 3’
	#1225	5’ CAA CAG AAA TCC AAC CTA GAG CTG CT 3’
viral mRNA classes	#1544 (E1)	5’ CTT GAA AGC GAA AGT AAA GC 3’
	#3392 (E7)	5’ CGT CCC AGA TAA GTG CTA AGG 3’
	#640 (I4)	5’ CAA TAC TAC TTC TTG TGG GTT GG 3’
	#3632 (E4)	5’ TGG ATG CTT CCA GGG CTC 3’