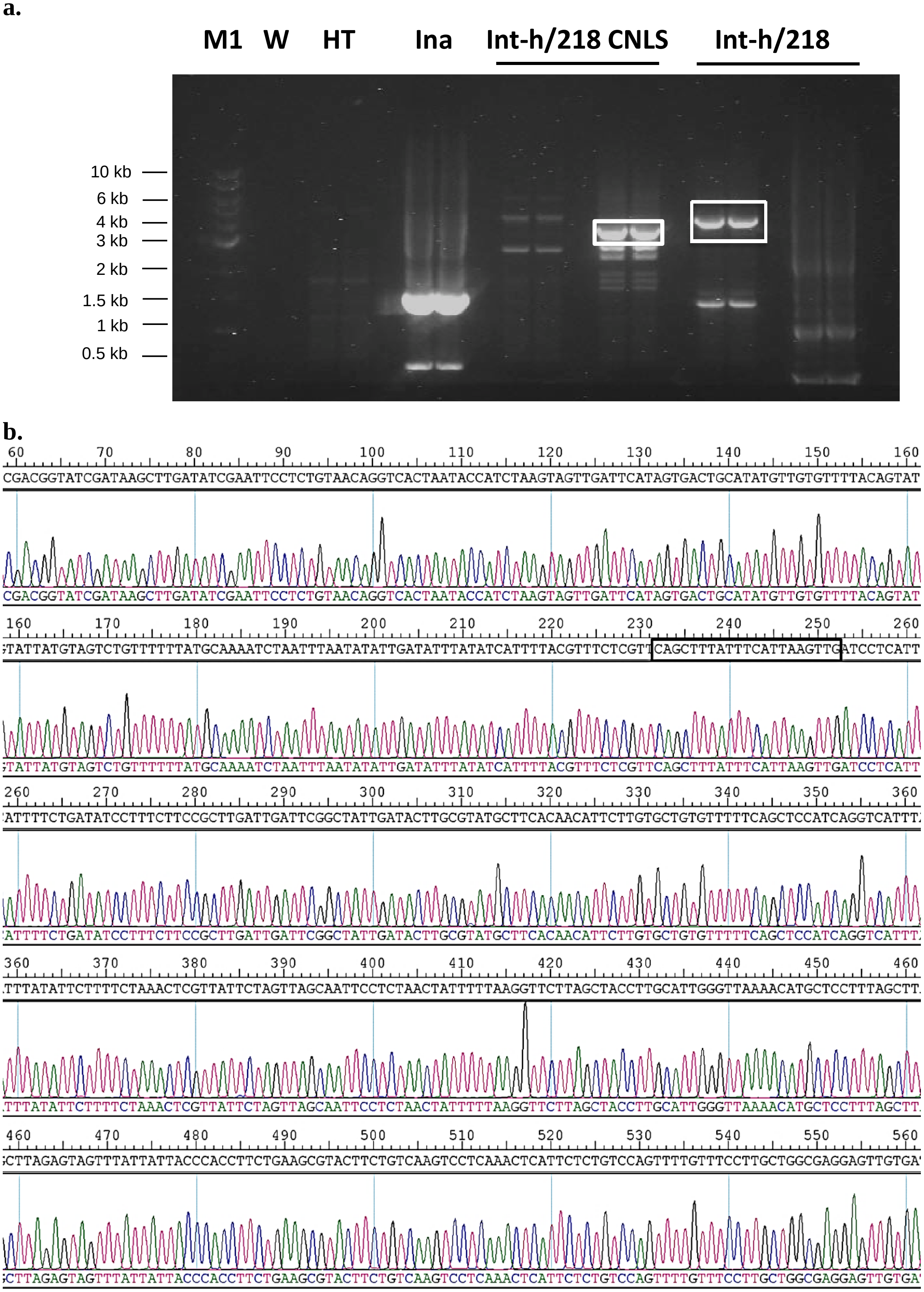
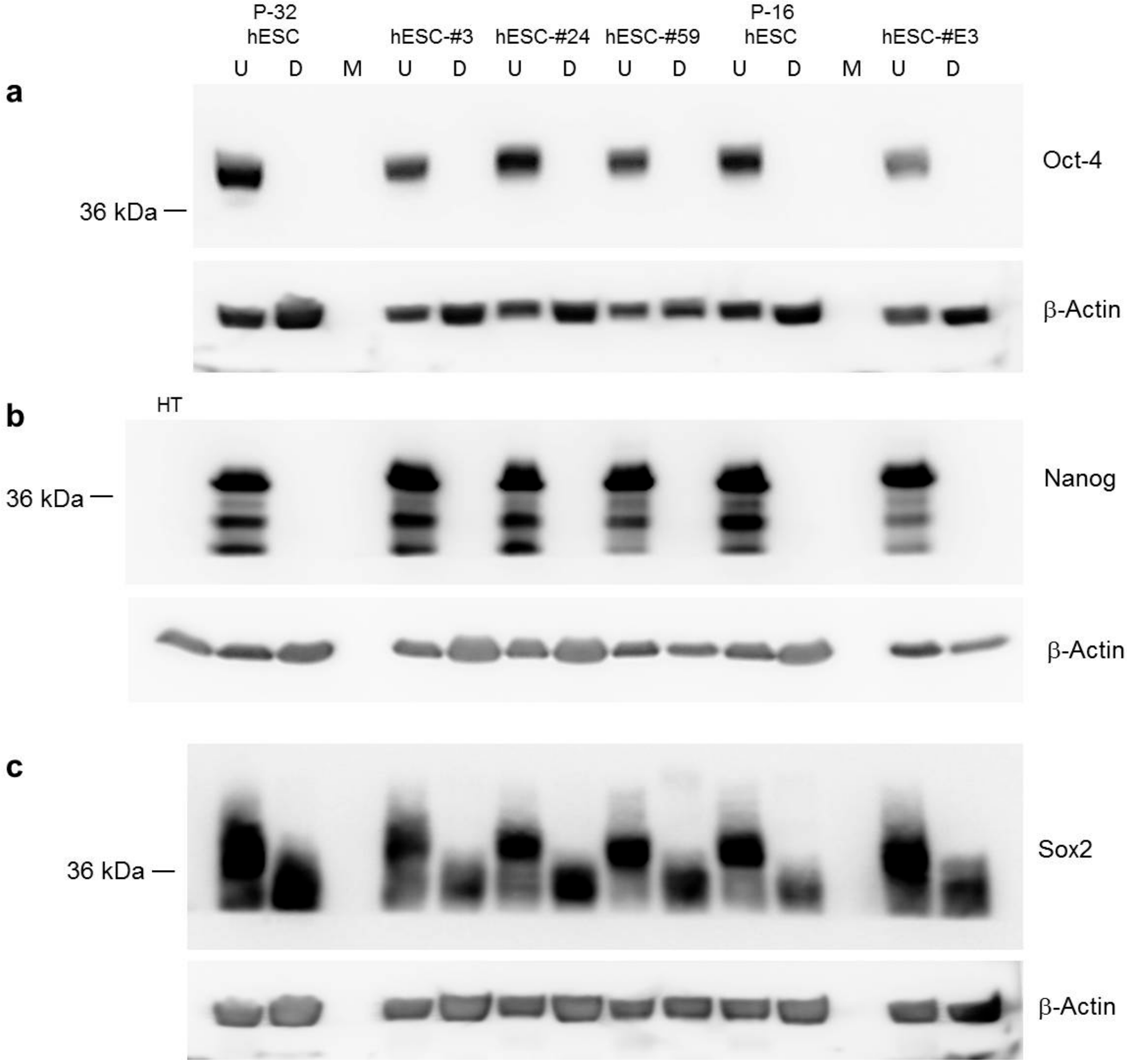
Supplementary Figure S1. Inverse PCR for identification of endogenous targets in HT1080.



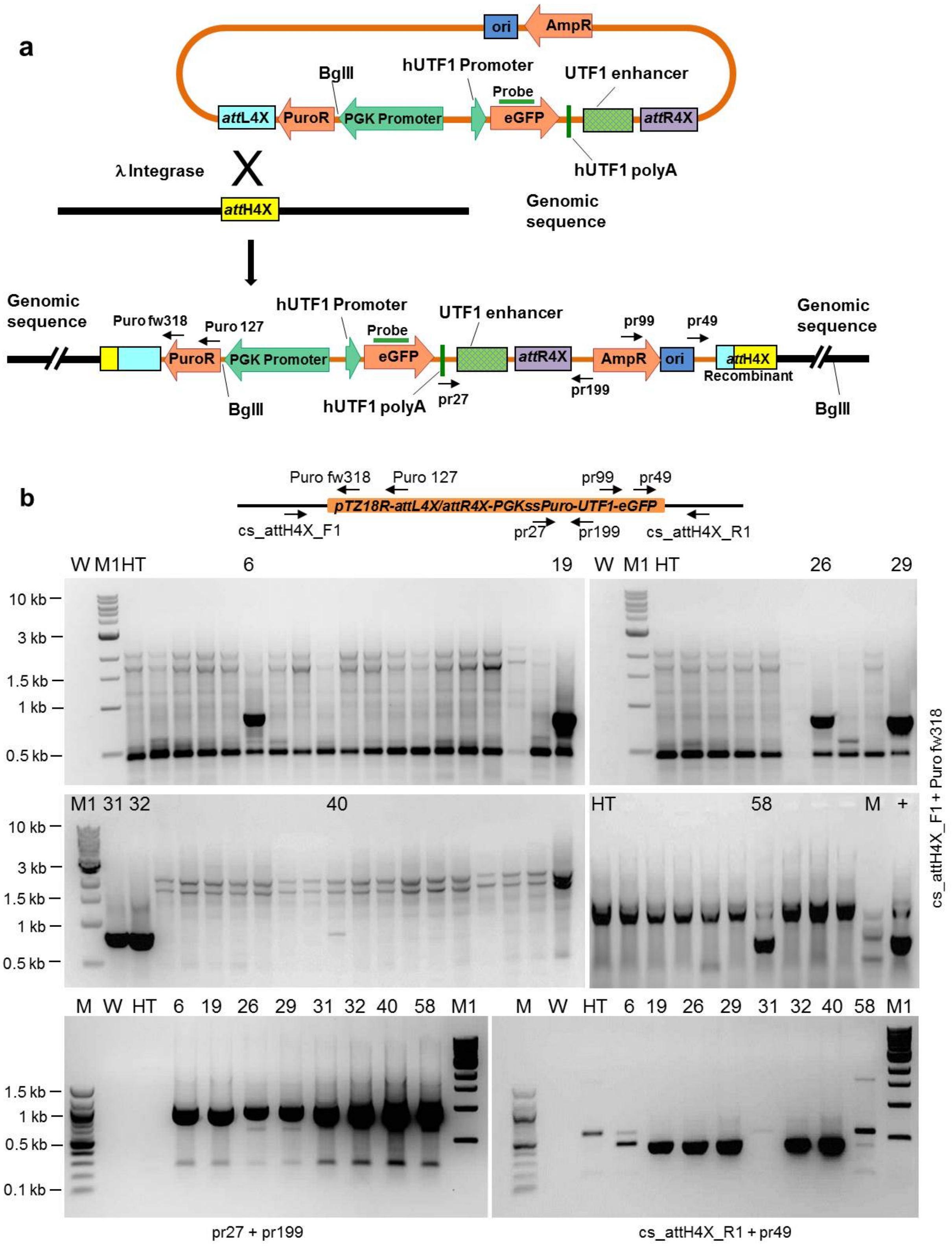
Supplementary Figure S1. Inverse PCR for identification of endogenous targets in HT1080. (a) Inverse PCR and corresponding nested PCRs were performed using genomic DNA from pooled puromycin-resistant HT1080 colonies obtained through transfections as described in Materials and Methods. PCR amplified products that were larger than 3kb (fragments marked in white boxes) were extracted and sequenced. W, no DNA template control; HT, negative control (genomic DNA template from parental HT1080 cells); Ina, genomic DNA template from puromycin resistant colonies obtained through co-transfection of *pattP4X-PGKssPuro* and *pCMVssIna*, the latter expressed inactive Int as negative control; Int-h/218CNLS, genomic DNA template from puromycin resistant colonies obtained through co-transfection of *pattP4X-PGKssPuro* and *pCMVssInt-h/218CNLS*; Int-h/218, genomic DNA template from puromycin resistant colonies obtained through co-transfection of *pattP4X-PGKssPuro* and *pCMVssInt-h/218*; M1, 1kb DNA ladder. **(b)** Sequencing analysis of the PCR product obtained in the inverse PCR with genomic DNA template from puromycin resistant colonies obtained through co-transfection of *pattP4X-PGKssPuro* and *pCMVssInt-h/218*. *att*R site (P-O-H’) resulting from recombination of *attP*4X with the endogenous sequence is marked in the box. Sequencing analysis of the PCR product obtained in the inverse PCR with genomic DNA template from puromycin resistant colonies obtained through co-transfection of *pattP4X-PGKssPuro* and *pCMVssInt-h/218*CNLS revealed a random integration of the targeting vector into chromosome 6 (sequence not shown).

Supplementary Figure S2. Western Blot analysis for pluripotency markers Oct-4, Nanog and Sox2 in the targeted hESC clones and differentiated derivatives.

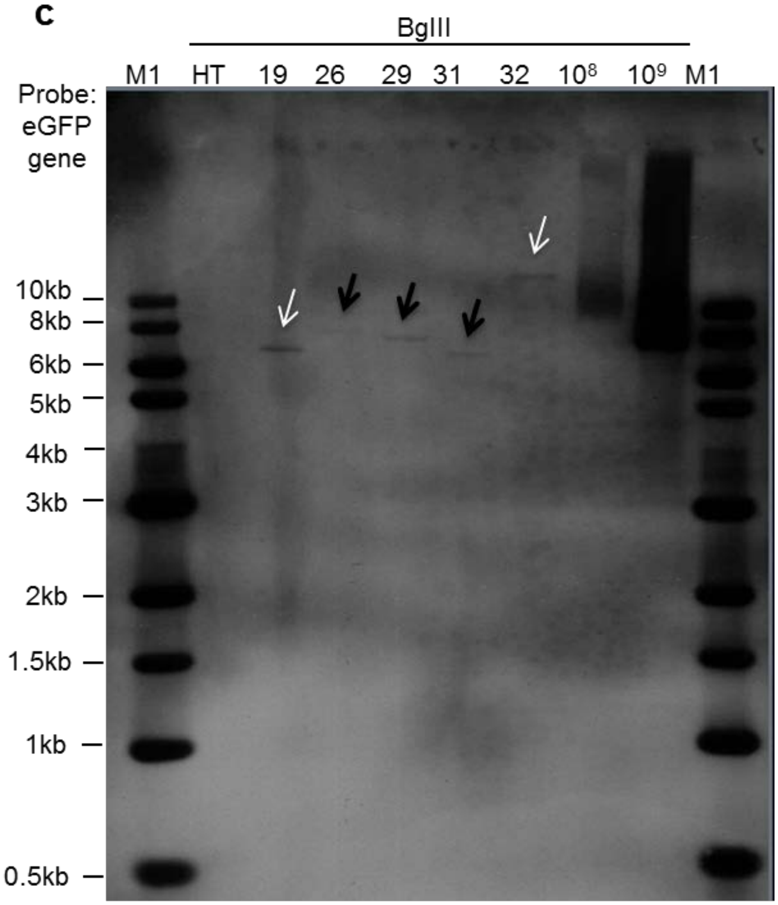


Supplementary Figure S2. Western Blot analysis for pluripotency markers Oct-4, Nanog and Sox2 in the targeted hESC clones and differentiated derivatives. Cell lysates from undifferentiated parental hESCs, the targeted clones (hESC#3, hESC#24, hESC#59 and hESC#E3) and their respective RA-induced differentiated derivatives were subjected to Western blot analysis and probed with antibodies against Oct-4, Nanog and Sox2 (top panels of a, b and c, respectively). Oct-4, Nanog and Sox2 expression was detected only in hESC clones. β-actin protein was used as loading control in each case (bottom panels of a, b and c). U, undifferentiated; D, differentiated; HT, cell lysate of HT1080 as control for differentiated cells; M, Marker lane; P-16, passage 16; P-32, passage 32.

Supplementary Figure S3. Targeting the endogenous *att*H4X sequence with *att*L4X/*att*R4X in HT1080.



c



Supplementary Figure S3. Targeting the endogenous *att*H4X sequence with *att*L4X/*att*R4X in HT1080. (a) Schematic diagram of *pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP* targeting vector after integration into *att*H4X. Relative binding positions of primers (Puro fw318, Puro rev127, pr27, pr199, pr99 and pr49) and the probe for eGFP are indicated. **(b)** Screening for *att*L4X/*att*R4X x *att*H4X recombination events. Eight clones positive for *att*L4X x *att*H4X recombination events at L1 loci were identified in an initial screen involving semi-nested PCR with primers cs_*att*H4X_F1 and Puro rev127 using templates from a primary PCR (primers cs_*att*H4X_F1 and Puro fw318). PCR-amplifications (~850 bp) were detected for clones (6, 19, 26, 29, 31, 32, 40, 58; top panel). PCR was performed to verify the presence of unrecombined *att*R4X in all 8 clones, using primers pr27 and pr199. PCR products (~1000bp) were detected in all (Bottom left panel). Recombinant *att*H4X junction (generated from *att*L4X x *att*H4X) recombination was screened by semi-nested PCR with primers cs_*att*H4X_R1 and pr49 using samples from a primary PCR (primers cs_*att*H4X_R1 and pr99 where). PCR products of expected size (~400 bp) were detected for clones 19, 26, 29, 32, 40 (Bottom right panel). W, no DNA template control; HT, negative control (genomic DNA from parental HT1080 cells); 1 kb DNA ladder. PCR products were verified by sequencing analysis. **(c)** Southern blot analysis. Genomic DNA purified from the five targeted HT1080 clones and parental HT1080 cells were subjected to digestion with *Bgl*II. PCR-derived Digoxigenin-labelled probe to the eGFP gene was employed. Lanes: 1kb DNA ladder; HT1080, parental DNA; 19, 26, 29, 31, 32 genomic DNA from targeted HT1080 clones;

10^8, 10^9 copies of linearized targeting vector as positive control. White arrow heads indicate fragments of the expected size and black arrow heads indicate extra or unexpected fragments in the targeted clones.

Supplementary Table 1. List of Primers (5’-3’)

Int fwd PstI	CCAACTGCAGCTCGAGGCCACCATGGGAAGAAGGCGAAG
Int_rev_XbaI	GAGCTCTAGATCATTATTTGATTTCA
CNLS_xbal_int	GAGCTCTAGATTATCATACCTTTCTCTTCTTCTTAGGGCCGCCTCCGGATTTGATTTCAATTTTG
KpnI_UTF1_fwd	GATTGGTACCTTAGGACCCGGCGGGCGGGG
SDM-Int-Y342A-F	CGGTCACCATGGCATCACAGGCTCGTGATGACAGAGGCAGGGA
SDM-Int-Y342A-R	TCCCTGCCTCTGTCATCACGAGCCTGTGATGCCATGGTGACCG
attP(4X)_sdm (For)	CGTTTCTCGTTCAGCTTTATTTCATTAAGTTGGCATTATAAAAAAGC
attP(4X)_sdm (Rev)	GCTTTTTTATAATGCCAACTTAATGAAATAAAGCTGAACGAGAAACG
INT-HIS-XbaI Rev	TATCCGAGATCTCTAGATTAGTGGTGATGGTGATGATGTTTGATTTCAATTTTGTCCCA
NLS-HIS-XbaI Rev	TATCCGAGATCTCTAGATTAGTGGTGATGGTGATGATGTACCTTTCTCTTCTTCTTAGG
EF1Fw_EcoRI	ATCGCCGAATTCGGCTCCGGTGCCCGTCAGTGGGCAG
EF1Rev-NheI	CGAATTGCTAGCTCACGACACCTGAAATGGAAGAA
EF1aRe-NsiI	GTGTATATGCATCCACCACACTGGACTAGTGGATC
EcoRV_EF_fwd	TCGGGATATCGGCTCCGGTGCCCGTCAGTG
ClaI_bgh_bpa_rev	GTGTATCGATGCCGCTACAGGGCGCGTGGGGAT
pMRNA-Inth218 Rev	TATCCGAGATCTGCTAGCTCATTATTTGATTTCAATTTTG
pMRNA-Inth218 Fw	ATCGCCGAATTCCTCGAGGCCACCATGGGAAGAAGG
C3Int-CNLS-BglII Rev	TATCCGAGATCTGCTAGCTTATCATACCTTTCTCTTCTTC
pMRNA-Seq Rev	TATTCTTCCTACTCAGGCTTTATTC
pMRNA-Seq For	TTGCGTTTGAGACGGGCGACAGATC
ClaI_UTF1 enhancer_rev	GTGTATCGATGCCACCACACCCGGCTAATTTTTG
attL(BOP’)Fwd (KpnI)	AAGTGGTACCTGAAGCCTGCTTTATTTCATTAAGTTGG
attL(BOP’)Rev(NotI)	GTGAGCGGCCGCCCGGGAGGGAGTGGGACAAAATTGAAATC
attR (POB’) Fwd (ClaI)	GTCTATCGATTCTGTTACAGGTCACTAATACCAT
attR (POB’) Rev (HindIII)	CTTGAAGCTTGTTCGCTCAAGTTAATGAAATAAAGCTGA
attP_fwd	GATTCATAGTGACTGCATATGTTGTG
attP_rev	GACTGATAGTGACCTGTTCGTTGC
Puro rev24	CACCGTGGGCTTGTACTCGGTC
Puro rev303	GGAACCGCTCAACTCGGCCATGC
Puro fw509	CTACGAGCGGCTCGGCTTCACC
Puro fw318	CAACAGATGGAAGGCCTCCTGG
Puro 127	GAGCTGCAAGAACTCTTCCTCAC
Pgk_rev	CAGGTGAATATCAAATCCTCCTCG
cs_*att*H4X_F1	GAGTGTTTTCCAACTTGGTTCCATT
cs_attH4X_R1	AAAACACAGCACGAGAACTTCGTGA

cs_*att*H4X_F2	CCTGTCTTGCTAGGTTGGGAAGT
cs_*att*H4X_R2	TAGAGGAATTGCTAACTAGAATAACCA
pr21	CACAGGAAACAGCTATGACCATG
pr27	GAACGTCTGGGGTGTGATCCT
pr49	CCAATACGCAAACCGCCTCTC
pr99	CTGAATGAAGCCATACCAAACGAC
pr199	CGTTTACAATTTCCCATTCGCCATTC
3-GF1	CATGTGAGATGGGTCTTCCAAATACAG
3-GF2	GGTATTAAAGTCTCCCACCGTTATTGT
3-GR1	CACACAGAAACCCCATCCATAGGTCAC
3-GR2	TTTTATGCTCACAGTGTAAACAAAGCC
24-GF1	AAATGTAGTCAGGCTTGGTGATTCAC
24-GF2	TACTGGAAAGAAGCAAAGAGGTATC
24-GR1	GAACTTCTAGCAATTACCAGCACAC
24-GR2	GTTTTCCCCTCAACTGTAAACAAAG
59-GF1	GTGGTTATTTTGCCCATTAGTTGATGC
59-GF2	CTTAAATAGAGCACACGGATGGGAC
59-GR1	CCTCATATGGTAGAACTCCAGCTGG
59-GR2	CCACTTTTCCATTGGTTTTATGCTTGG
E3-GF1	TTCAGGAGTTGTTGTAAGGCAGGC
E3-GR1	CTCTGGGCAGGGCATCTCTGAAAG
E3-GF2	CCTTTCTCTCTGGCTGCCCTTAAC
E3-GR2	GCACCAACAGCAAAGACCAAAGG
K3-GF2	GCGCCCACTCTCTTCTGACTTGTAG
K3-GR2	GACGTCCACACAGAAACCCCACTG
K3-GF1	CTTCAGGAGCTCTTGTAAGGCAGGC
K3-GR1	CCAGGGAAACAGGGTCTAGAGTGG
A3-GF2	GGCATGGGTGTTAGGGCCATTATG
A3-GR2	CGTCTAAAGGATCACAACTCCTTGCC
A3-GF1	AGTGGGAAATCTGGGGTCTCACTG
A3-GR1	GGAAGGAACGGGCAGCAATCTTTG
EGFP fwd	CCACATGAAGCAGCACGACTTC
EGFP rev	GGGTGTTCTGCTGGTAGTGGTC

Supplementary Table 2. *att*H sequence hits obtained using CA/TGCTTTNNNNNNNTAAG/CTTG as query sequence in bioinformatics search along with their chromosomal location and occurrence in the Human genome.

	***att*H sequences in Human genome identified using bioinformatics approach**	**Chromosome number, nucleotide co-ordinates. (GRCh38/hg38 Assembly)**	**Occurrence per genome**
1	CAACTTAATGAAATAAAGCAG (*att*H7X)	1; 106854666 - 106854686 1; 157347000 - 157347020 3; 137376026 - 137376046 5; 64475470 - 64475490 12; 33493985 - 33494005 12; 88438383 - 88438403 12; 102493126 - 102493146	7
2	CAGCTTTTTCAATTTAACTTG	6; 31334317 - 31334337 6_GL000256v2_alt; 2636515-2636535 6_GL000255v2_alt; 2589974-2589994	3
3	CAGCTTTCATAAAATAAGTTG	4; 143909914 - 143909934 14; 50663569 - 50663589	2
4	CAACTTAGCAGACAAAAGCAG	2; 27626505 – 27626525 X; 112382682 -112382702	2
5	CAAGTTATTCTTCAAAAGCAG	1; 19236066 - 19236086	1
6	CAACTTAAAAGGAAAAAGCTG	1; 82574082 - 82574102	1
7	CAACTTATAACTGTAAAGCAG	1; 97122439 - 97122459	1
8	CAACTTATTGAATGAAAGCAG	1; 106471437 - 106471457	1
9	CAACTTAAGACTTGAAAGCTG	1; 110011423 -110011443	1
10	CTGCTTTGGGAAAATAACTTG	1; 116915861- 116915881	1
11	CTGCTTTTCTTTCCTAACTTG	1; 163425240 - 163425260	1
12	CAGCTTTAAAAACCTAACTTG	1; 173189485 - 173189505	1
13	CAGCTTTCTGCATATAAGTTG	1; 182796593 182796613	1
14	CAGCTTTTCTATGTTAACTTG	1; 183243106 -183243126	1
15	CAACTTACGGTACAAAAGCTG	1; 183343809 - 183343829	1
16	CAACTTACTGATGTAAAGCAG	1; 208844994 - 208845014	1
17	CAGCTTTTGTTCCTTAACTTG	1; 217312051 - 217312071	1
18	CTGCTTTTTCCTTTTAAGTTG	1; 244735375 - 244735395	1
19	CAACTTATCCTCTTAAAGCAG	2; 4855275 - 4855295	1
20	CAAGTTAGGAAACTAAAGCAG	2; 6601353 - 6601373	1
21	CAGCTTTTTGCCCCTAACTTG	2; 23652921 - 23652941	1
22	CTGCTTTTGTCTGCTAAGTTG	2; 23652921 - 23652941	1
23	CAACTTATAGCCACAAAGCAG	2; 35463859 - 35463879	1
24	CAGCTTTCTATACATAAGTTG	2; 187498529 -187498549	1
25	CTGCTTTCATTGGTTAACTTG	2; 203262530 - 203262550	1

26	CAACTTAGGAAAGAAAAGCTG	2; 211166036 -211166056	1
27	CAAGTTATTCTTAGAAAGCTG	2; 235648033 - 235648053	1
28	CAAGTTAAAGGACAAAAGCTG	2; 241393667 - 241393687	1
29	CAGCTTTTATCCATTAAGTTG	3; 17021726 - 17021746	1
30	CAACTTAGTGCGTAAAAGCTG	3; 34949152 - 34949172	1
31	CAACTTAAAGTTATAAAGCAG	3; 49747167 - 49747187	1
32	CTGCTTTATTAACTTAAGTTG	3; 65371692 - 65371712	1
33	CTGCTTTTCCATATTAAGTTG	3; 84840603 - 84840623	1
34	CTGCTTTGATTTTATAACTTG	3; 95975581 - 95975601	1
35	CAGCTTTAAATAATTAACTTG	3; 105465612 - 105465632	1
36	CAAGTTACAATCTAAAAGCTG	3; 112543794 - 112543814	1
37	CTGCTTTCCCCAGCTAACTTG	3; 125195186 -125195206	1
38	CAACTTAAAATTGAAAAGCAG	3; 163098187 - 163098207	1
39	CAAGTTATTATATTAAAGCAG	3; 180355103 - 180355123	1
40	CAACTTACTGTTTCAAAGCTG	3; 196391805 - 196391825	1
41	CAAGTTATAATCAGAAAGCTG	4; 9810661 - 9810681	1
42	CTGCTTTTTTTTTCTAAGTTG	4; 14139515 - 14139535	1
43	CAGCTTTCAGATATTAAGTTG	4; 27810251 - 27810271	1
44	CAAGTTACAATGTAAAAGCTG	4; 33816397 - 33816417	1
45	CAGCTTTGTGATAATAACTTG	4; 57395339 - 57395359	1
46	CAGCTTTCTTTGTTTAAGTTG	4; 96050717 - 96050737	1
47	CAACTTAGTGAATAAAAGCAG	4; 96575810 - 96575830	1
48	CAACTTAAATTTCAAAAGCAG	4; 107399965 -107399985	1
49	CAACTTACACCTCCAAAGCTG	4; 119228746 -119228766	1
50	CAAGTTATTTAAGCAAAGCAG	4; 143440508 - 143440528	1
51	CAGCTTTTATGAGCTAAGTTG	4; 172008324 -172008344	1
52	CAAGTTATAATAGTAAAGCTG	4; 178746053 - 178746073	1
53	CTGCTTTAAATCTGTAACTTG	4; 184924992 - 184925012	1
54	CTGCTTTAAAGATGTAACTTG	5; 2196198 - 2196218	1
55	CTGCTTTCTGATTTTAAGTTG	5; 6210040 - 6210060	1
56	CAGCTTTGGTCCTATAACTTG	5; 17384798 - 17384818	1
57	CAGCTTTAAAATTCTAACTTG	5; 40747039 - 40747059	1
58	CTGCTTTAAGATTTTAACTTG	5; 74779791 - 74779811	1
59	CAAGTTAAGACAGAAAAGCTG	5; 83929507 - 83929527	1
60	CAGCTTTAATTTCCTAACTTG	5; 83977220 - 83977240	1
61	CTGCTTTGTTTCATTAAGTTG	5; 106998974- 106998994	1
62	CAACTTACACAGCCAAAGCAG	5; 107019917- 107019937	1
63	CAAGTTATGCTTATAAAGCTG	5; 131651814 - 131651834	1
64	CAAGTTAAAGGGCAAAAGCAG	6; 25652052 - 25652072	1
65	CTGCTTTAAAATTTTAAGTTG	6; 35836451 - 35836471	1
66	CAAGTTAAATATTAAAAGCTG	6; 57254511 - 57254531	1
67	CAAGTTAAATGATAAAAGCTG	6; 78368464 - 78368484	1
68	CTGCTTTAAAATATTAACTTG	6; 91002384 - 91002404	1

69	CTGCTTTGAGTCAGTAAGTTG	6; 108690628 - 108690648	1
70	CAAGTTAATAAATAAAAGCAG	6; 130502719 - 130502739	1
71	CAGCTTTTTGAAAGTAACTTG	6; 134899030 - 134899050	1
72	CAAGTTATTTAAAAAAAGCTG	6; 152991109 - 152991129	1
73	CAGCTTTGAAGGCATAAGTTG	6; 158673743 - 158673763	1
74	CAGCTTTAGGATAGTAACTTG	7; 2377245 - 2377265	1
75	CTGCTTTACATCATTAACTTG	7; 11318471 - 11318491	1
76	CAAGTTACACACTTAAAGCTG	7; 17361521 - 17361541	1
77	CAAGTTAGTTTTTGAAAGCTG	7; 49809083 - 49809103	1
78	CTGCTTTTCCTTGATAACTTG	7; 53800631 - 53800651	1
79	CAGCTTTTAAACTGTAACTTG	7; 64579860 - 64579880	1
80	CAGCTTTCACTCATTAAGTTG	7; 83597543 - 83597563	1
81	CTGCTTTGAATGTATAAGTTG	7; 94663282 - 94663302	1
82	CAAGTTAAACAGCAAAAGCTG	7; 97256143 - 97256163	1
83	CAAGTTACATGGTGAAAGCAG	7; 115330936 - 115330956	1
84	CAAGTTATTATCTGAAAGCTG	7; 120883773 - 120883793	1
85	CTGCTTTCATCAGTTAAGTTG	7; 139627071 - 139627091	1
86	CAGCTTTACAGTAATAACTTG	7; 157361587 - 157361607	1
87	CAACTTAAGAGCTTAAAGCAG	8; 3722953 - 3722973	1
88	CAAGTTAGGATGGGAAAGCTG	8; 58610454 - 58610474	1
89	CAAGTTACCCTTTAAAAGCAG	8; 71334382 - 71334402	1
90	CTGCTTTGCCAAACTAAGTTG	8; 90449166 - 90449186	1
91	CAAGTTAACCAGAGAAAGCAG	8; 100827768 - 100827788	1
92	CTGCTTTAAGACTATAAGTTG	8; 127114987 - 127115007	1
93	CAACTTAGCTCCTCAAAGCAG	8; 128124895 - 128124915	1
94	CAGCTTTTATTTCATAACTTG	9; 12155661 - 12155681	1
95	CAAGTTAATTATTCAAAGCAG	9; 26695059 - 26695079	1
96	CTGCTTTGATCAATTAACTTG	9; 109146961 - 109146981	1
97	CTGCTTTTAAGGTATAAGTTG	9; 113422032 - 113422052	1
98	CAACTTACACTTATAAAGCAG	9; 121160218 - 121160238	1
99	CTGCTTTTAGTTTTTAAGTTG	9; 129892186 - 129892206	1
100	CAGCTTTTTGGTAGTAAGTTG	10; 6851438 - 6851458	1
101	CAAGTTAATCCATCAAAGCTG	10; 72501178 - 72501198	1
102	CTGCTTTTCTTCTGTAACTTG	10; 119438226 - 119438246	1
103	CAACTTAGCACAACAAAGCTG	10; 128454453 - 128454473	1
104	CAGCTTTTAGGGCATAACTTG	11; 13552326 - 13552346	1
105	CTGCTTTAAACTTTTAAGTTG	11; 16752336 - 16752356	1
106	CAAGTTAAATGTAAAAAGCAG	11; 72812396 - 72812416	1
107	CAAGTTATTAAAATAAAGCAG	11; 74576294 - 74576314	1
108	CAGCTTTTTTGAATTAAGTTG	11; 103756334 - 103756354	1
109	CAGCTTTAATGTGCTAACTTG	11; 107797655 - 107797675	1
110	CTGCTTTCCAAGGGTAACTTG	11; 116912580 - 116912600	1
111	CAAGTTATCCAAAAAAAGCAG	11; 123287626 - 123287646	1

112	CAAGTTAAAATATGAAAGCAG	12; 12845396 - 12845416	1
113	CAAGTTAAAAAAGAAAAGCAG	12; 29161214 - 29161234	1
114	CAACTTATGACTTAAAAGCAG	12; 47114357 - 47114377	1
115	CAACTTAGACAGAGAAAGCAG	12; 52677529 - 52677549	1
116	CAAGTTAAAACCAGAAAGCTG	12; 84928830 - 84928850	1
117	CAAGTTAGAAAGCTAAAGCAG	12; 129048549 - 129048569	1
118	CAAGTTAAGTTTGAAAAGCAG	13; 53365028 - 53365048	1
119	CTGCTTTTGTCTCTTAAGTTG	13; 78663499 - 78663519	1
120	CTGCTTTAGTGAGTTAACTTG	13; 82818196 - 82818216	1
121	CAACTTATTCAATGAAAGCAG	14; 42949179 - 42949199	1
122	CAACTTACATCAGAAAAGCAG	15; 40032198 - 40032218	1
123	CAGCTTTTGTGTCTTAACTTG	15; 53332459 - 53332479	1
124	CAAGTTACAATTAAAAAGCTG	15; 101712807 - 101712827	1
125	CAGCTTTGTCAATGTAAGTTG	17; 43557851 - 43557871	1
126	CTGCTTTTTATGTTTAACTTG	18; 52621151 - 52621171	1
127	CTGCTTTCTGTAATTAAGTTG	18; 53222226 - 53222246	1
128	CAAGTTAAGATTTTAAAGCAG	19; 7801117 - 7801137	1
129	CAAGTTAAACAAGAAAAGCTG	20; 5390775 - 5390795	1
130	CTGCTTTTATTTAATAACTTG	20; 9367358 - 9367378	1
131	CTGCTTTTAAATACTAAGTTG	20; 19923109 - 19923129	1
132	CTGCTTTCAAACACTAACTTG	21; 35612903 - 35612923	1
133	CTGCTTTTTAATTATAAGTTG	22; 27899206 - 27899226	1
134	CTGCTTTGTTCTCTTAACTTG	X; 15961638 - 15961658	1
135	CTGCTTTCATTAAATAAGTTG	X; 26371828 - 26371848	1
136	CAGCTTTAATGGTCTAACTTG	X; 43113758 - 43113778	1
137	CTGCTTTGGATTCTTAAGTTG	X; 74743279 - 74743299	1
138	CAGCTTTGAACTTGTAACTTG	X; 91736877 - 91736897	1
139	CTGCTTTTTTTTTATAAGTTG	X; 111195171- 111195191	1
140	CTGCTTTTCATAAGTAAGTTG	X; 124065934 124065954	1

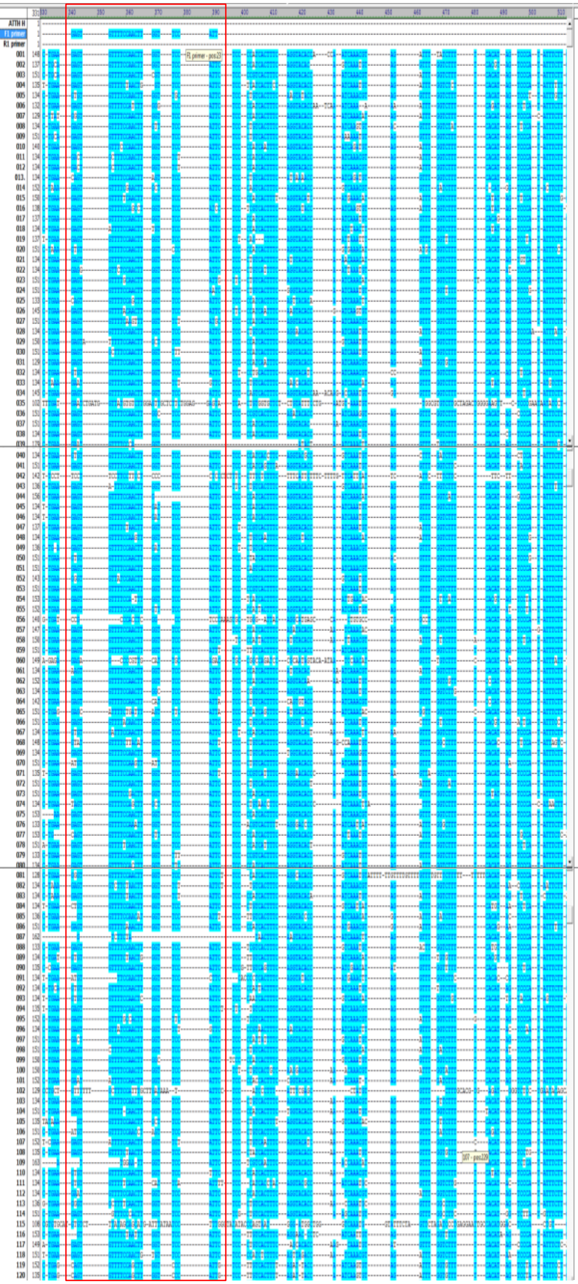
Note: The underlined sequences have been found in the updated version of Human genome assembly (GRCh38/hg38)

Supplementary Table 3. A list of *att*H7X sequences (underlined) along with their flanking sequences, *LINE-1* family and location in the Human genome.

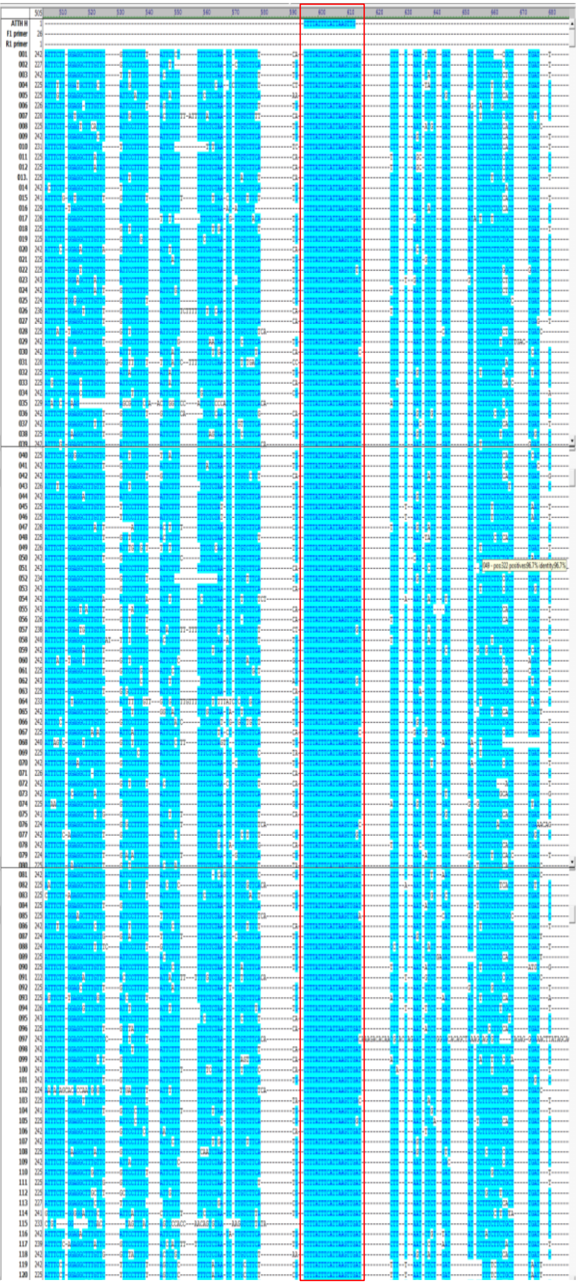
attH7X chromosomal locus	Flanking sequences	L1 Family	Location; Flanking genes
chr1:106854666-106854686 Reverse strand	TCAATAGCTGAATCAATCAAGAAGAAAAAAGGATATCAGAGATTGAAGAT**CAACTTAATGAAATAAAGCAG**GAAGACAAGATTAGAGAAAAAAGAAAAGGAATGAAAAAAACTTCCAAGAT	L1PA7	Intergenic; Gene ENSG00000237460 Location Chromosome 1: 106,818,239-106,838,167 Gene type Known lincRNA Gene ENSG00000198890 protein arginine methyltransferase 6 (PRMT6) Location Chromosome 1: 107,056,679-107,059,294 Gene type Known protein coding
chr5:64475470-64475490 Reverse strand	TCAATAGCTGAATCGATCGAGCAGAAGAAATGATATCAGAGATTGGAGAT**CAACTTAATGAAATAAAGCAG**GAAGACAAGATTACAGAAAAAAGAACGAAAAGTAATGAACAAAGCCTCCA	L1P3	Gene ENSG00000236068 Location Chromosome 5: 64,404,512-64,486,615 Gene type Known unitary pseudogene
chr12:33493985-33494005 Reverse strand	TCAACAGCCAAATCAATCAAGCGGAAGAAAGGATATCAGAGATTGAAGAT**CAACTTAATGAAATAAAGCAG**GAAAACAATATTAGATAAAAAAGAATGAAAAGGAACAAACAAAGCCTCTA	L1PA7	Intergenic; Gene ENSG00000110975 synaptotagmin X Location Chromosome 12: 33,374,238-33,439,819 Gene type Known protein coding Gene ENSG00000212475 RNA, U6 small nuclear 400, pseudogene Location Chromosome 12: 33,576,095-33,576,200 Gene type Known snRNA
chr12:88438383-88438403 Reverse strand	TCAATAGCTGAATCAATCAAACAGAAGAAAGGATATCAAAGATTGAAGAT**CAACTTAATGAAATAAAGCAG**GAAGACAAGATTAGAGAAAAAAGAATGAAAAGGAATGAACAAAGCCTCCA	L1PA7	Intergenic; Gene ENSG00000199245 Y RNA Location Chromosome 12: 88,430,442-88,430,550 Gene type Novel misc RNA Gene ENSG00000049130 KIT ligand (KITLG) Location Chromosome 12: 88,492,793-88,580,851 Gene type Known protein coding
chr1:157347000-157347020 Forward strand	TGAAGGCTTTGTTCATTCCTTTTCATTCTTTTTTCTCTAATCTTGTCTTC**CTGCTTTATTTCATTAAGTTG**ATCTTCAATCTCTGATATCCTTTCTTCTGCTTCATCAATTCGGCTACTGA	L1PA7	Intergenic; Gene ENSG00000237189 Location Chromosome 1: 157,287,703-157,288,053 Gene type Known lincRNA Gene ENSG00000143297 Fc receptor-like 5 (FCRL 5) Location Chromosome 1: 157,513,377-157,552,520 Gene type Known protein coding
chr3:137376026-137376046 Forward strand	TTCTTGAAGTCTTTGTTTCTTTTCATTCTTTTTTCTCTACTCTTGTCTGC**CTGCTTTATTTCATTAAGTTG**ATCTTCAATATTTGGTATTCTTTCTTCTGCTTGATCAATTCAGCTGTTTA	L1PA7	Intergenic; Gene ENSG00000174564 interleukin 20 receptor beta (IL20RB) Location Chromosome 3: 136,946,230-137,011,085 Gene type Known protein coding Gene ENSG00000201325 RNA, 5S ribosomal pseudogene 142 (RNA5SP142) Location Chromosome 3: 137,518,134-137,518,255 Gene type Known rRNA
chr12:102493126-102493146 Forward strand	TGGAGGCTTTGTTCATTCCTTTTTATTCTTTTTTCTCTAATCTTGTCTTC**CTGCTTTATTTCATTAAGTTG**ATCTTCAATCTCTGATATCCATTCTTCCACTTGATCGATTTGACTATTGA	L1PA7	Intergenic; Gene ENSG00000017427 (IGF 1) insulin-like growth factor 1 (somatomedin C) Location Chromosome 12: 102,395,867-102,480,645 Gene type Known protein coding Gene ENSG00000257254 Location Chromosome 12: 102,513,251-102,513,655 Gene type Known processed pseudogene

Supplementary Table 4. Alignment of 935 LINE 1 elements containing 18bp attH4X sequence flanked by 300bp upstream and downstream sequences (Vector NTI , AlignX) to design the primer pairs (cs_ attH4X _F1 & cs_ attH4X _R1) and (cs_ attH4X _F2 & cs_ attH4X _R2, not shown) in the consensus flanking region of attH4X. The respective subfamilies of 935 LINE1 elements that are shown in the right column correspond to the serial numbers of the sequences on the left side of the alignment.

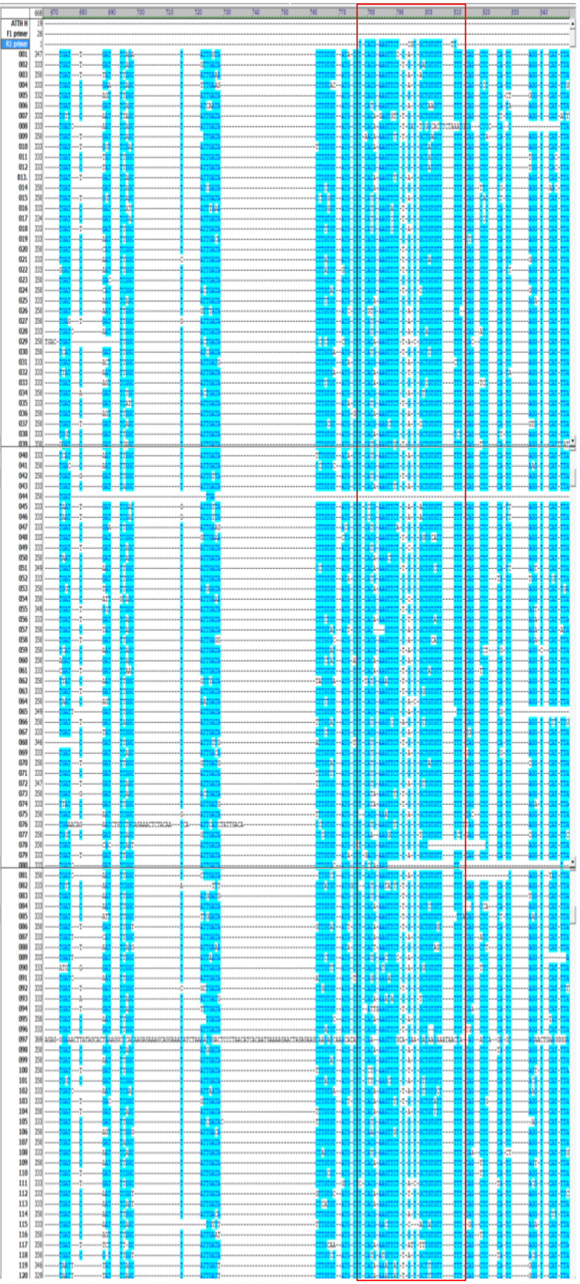
cs_attH4X_F1 primer
alignment
GAGTGTTTTCCAACTTGGTTCCATT



attH4X
alignment
CTTTATTTCATTAAGTTG



cs_attH4X_R1 primer
alignment
TCACGAAGTTCTCGTGCTGTGTTTT



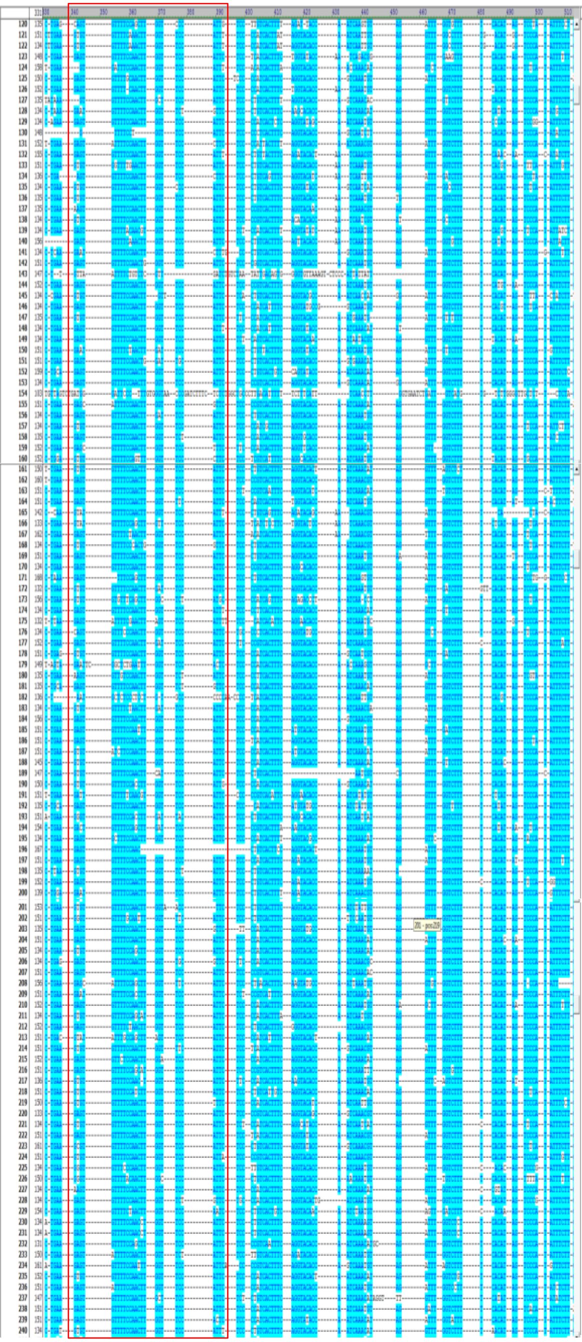
1-120 seq
LINE-1
subfamilies

L1P3
L1P2
L1PA3
L1P2
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA5
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA8
L1P2
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1P2
L1P3
L1PA7
L1P3
L1PA7
L1PA6
L1PA7
L1PA7
L1PA6
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1P3
L1PA7
L1PA7
L1P3
L1P2
L1PA7
L1PA7
L1P3

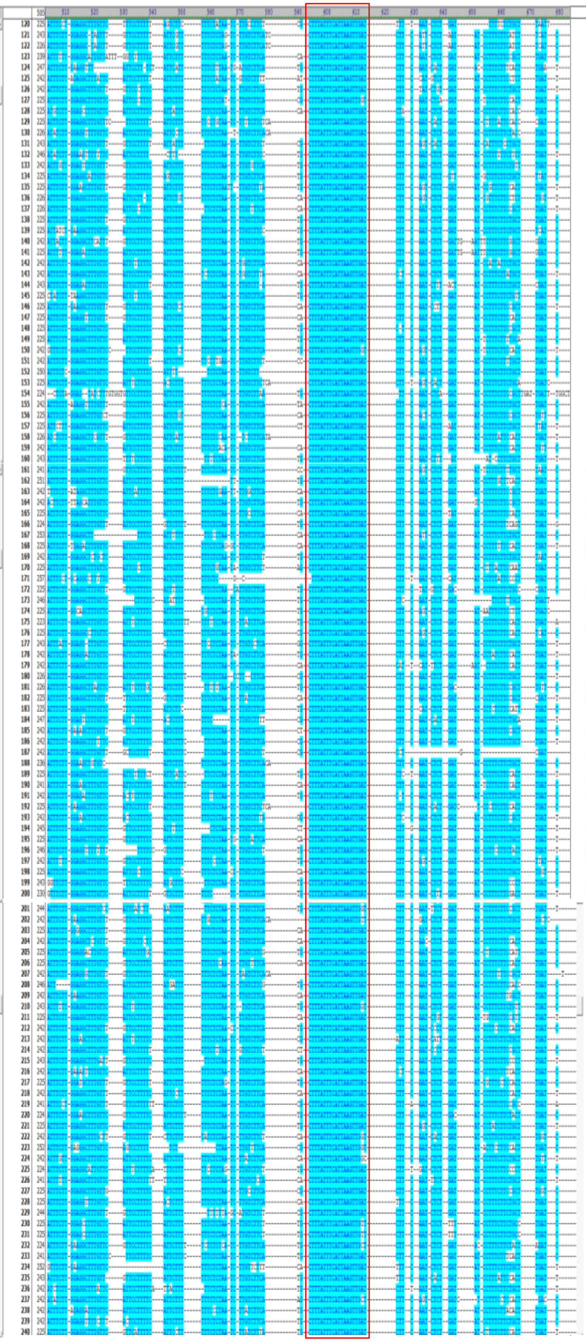
L1P2
L1P3
L1PA7
L1PA8
L1PA7
L1P2
L1PA7
L1P3
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA5
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA8A
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1P3
L1P2
L1P2
L1P2
L1PA7
L1P2
L1P3
L1PA7
L1P3
L1P3
L1PA7
L1P2
L1P2
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7

L1PA3/L1PA7
L1PA8
L1P3
L1PA7
L1P2
L1P2

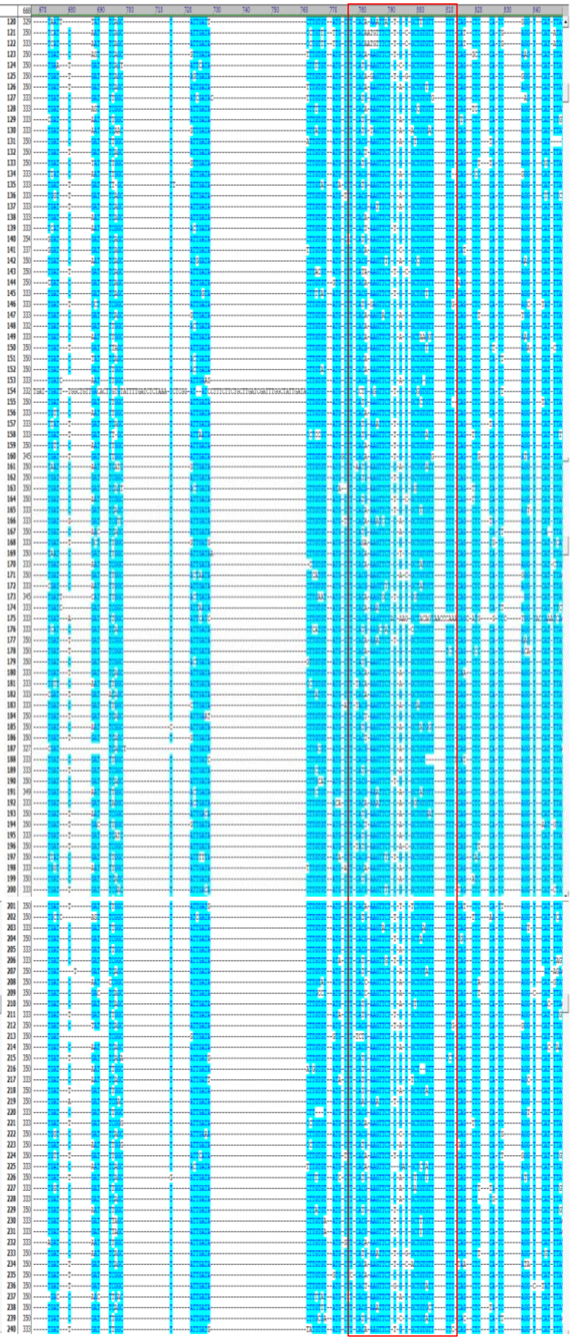
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment

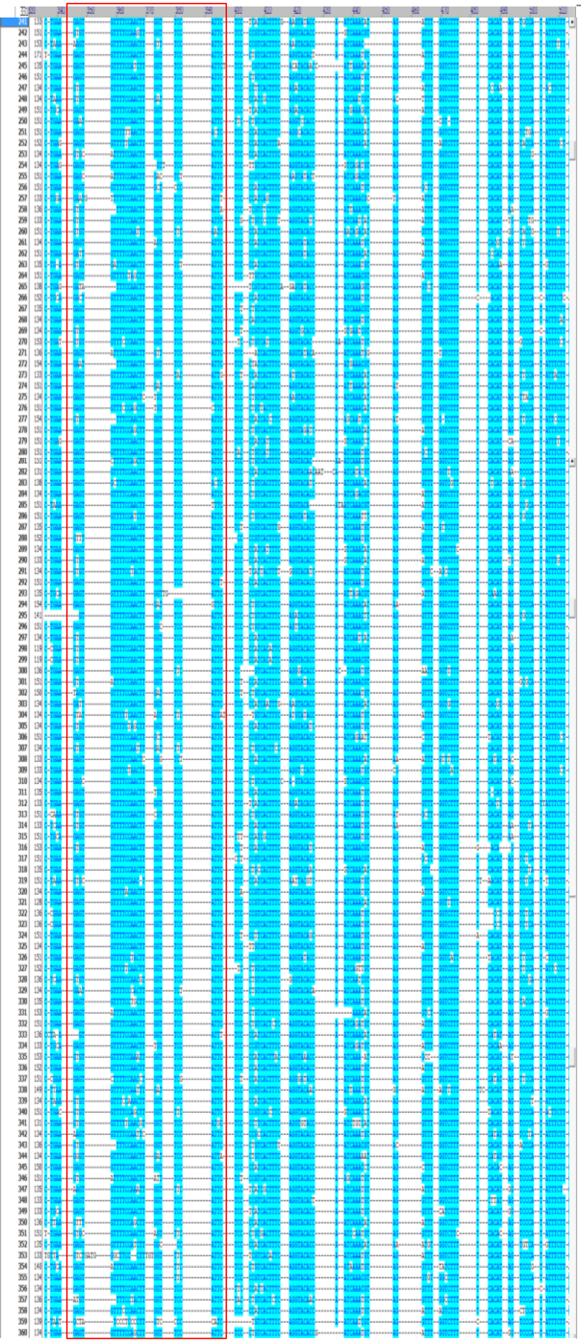


120-240 seq *LINE-1* subfamilies

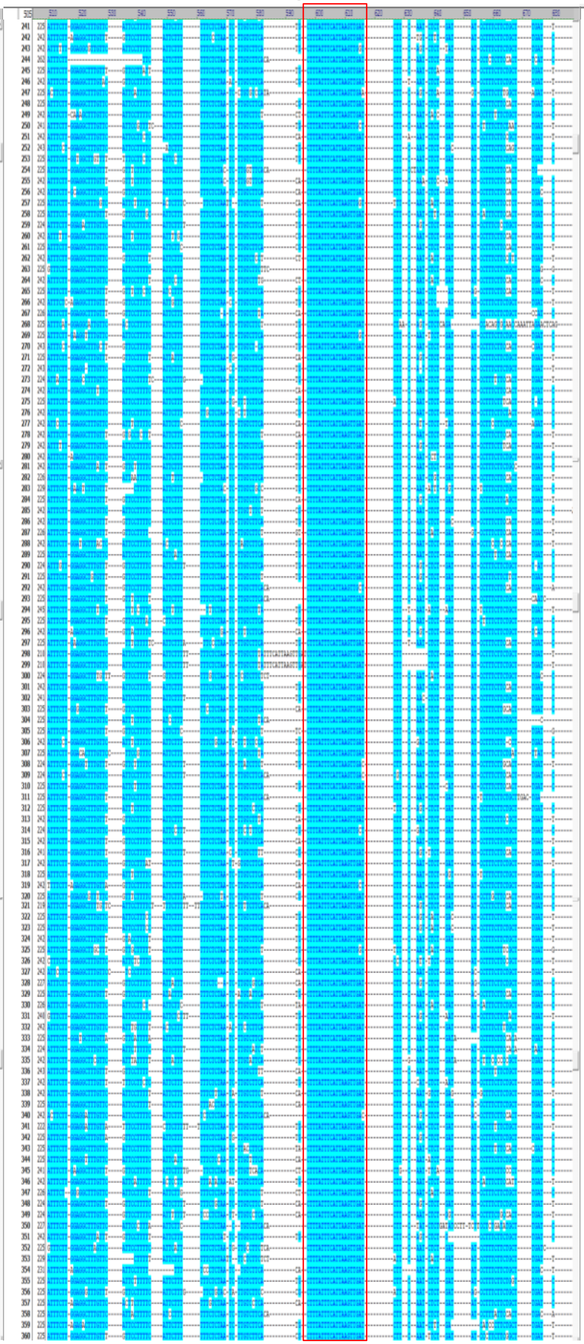
L1P2
L1PA8
L1PA8
L1PA7
L1PA7
L1P2
L1P2
L1P2
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA8
L1PA8
L1PA7
L1PA7
L1PA8
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1P3
L1PA4
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3/L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA6
L1PA6
L1PA7
L1PA7
L1PA7
L1PA7
L1PA4
L1PA7
L1PA5
L1PA8
L1PA8
L1PA7
L1P2
L1P2
L1P3
L1PA7
L1P3
L1PA7

L1PA7
L1PA8
L1PA9
L1PA10
L1PA11
L1PA12
L1PA13
L1PA14
L1PA15
L1PA16
L1PA17
L1PA18
L1PA19
L1PA20
L1PA21
L1PA22
L1PA23
L1PA24
L1PA25
L1PA26
L1PA27
L1PA28
L1PA29
L1PA30
L1PA31
L1PA32
L1PA33
L1PA34
L1PA35
L1PA36
L1PA37
L1PA38
L1PA39
L1PA40
L1PA41
L1PA42
L1PA43
L1PA44
L1PA45
L1PA46
L1PA47
L1PA48
L1PA49
L1PA50
L1PA51
L1PA52
L1PA53
L1PA54
L1PA55
L1PA56
L1PA57
L1PA58
L1PA59
L1PA60
L1PA61
L1PA62
L1PA63
L1PA64
L1PA65
L1PA66

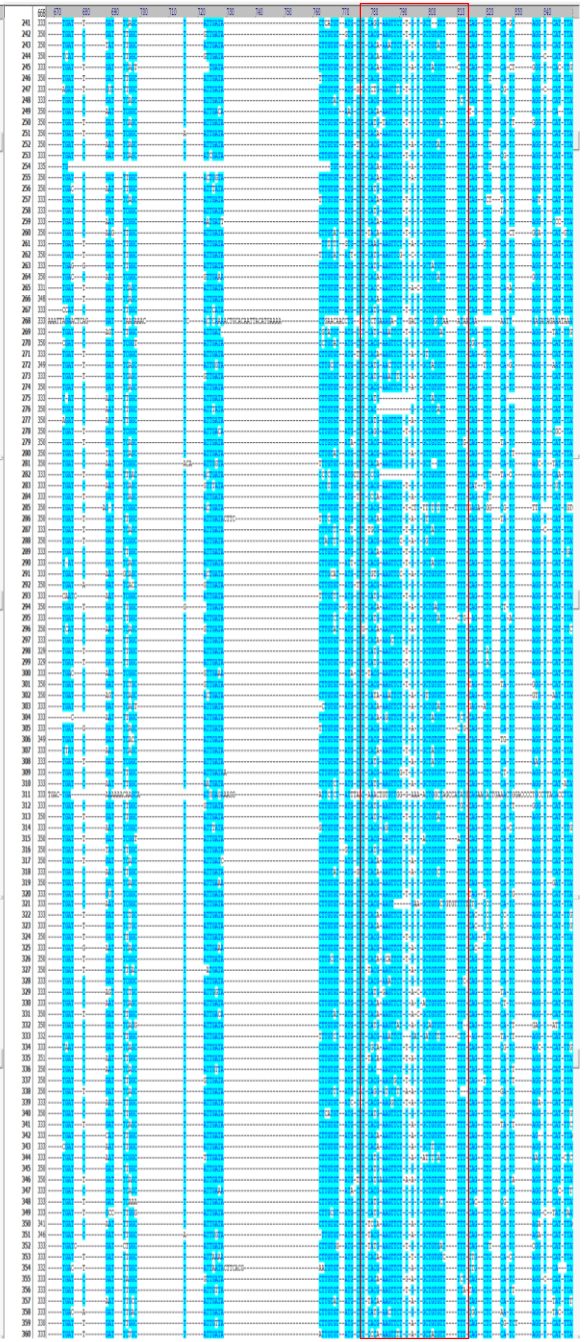
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment

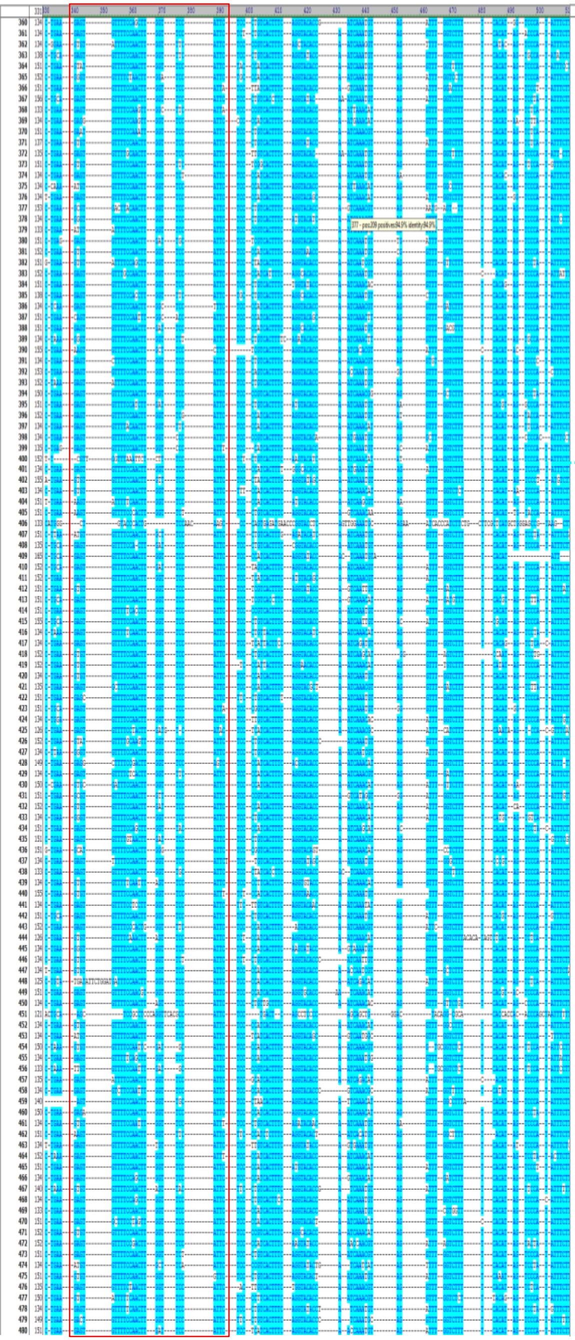


240-360 seq LINE1 subfamilies

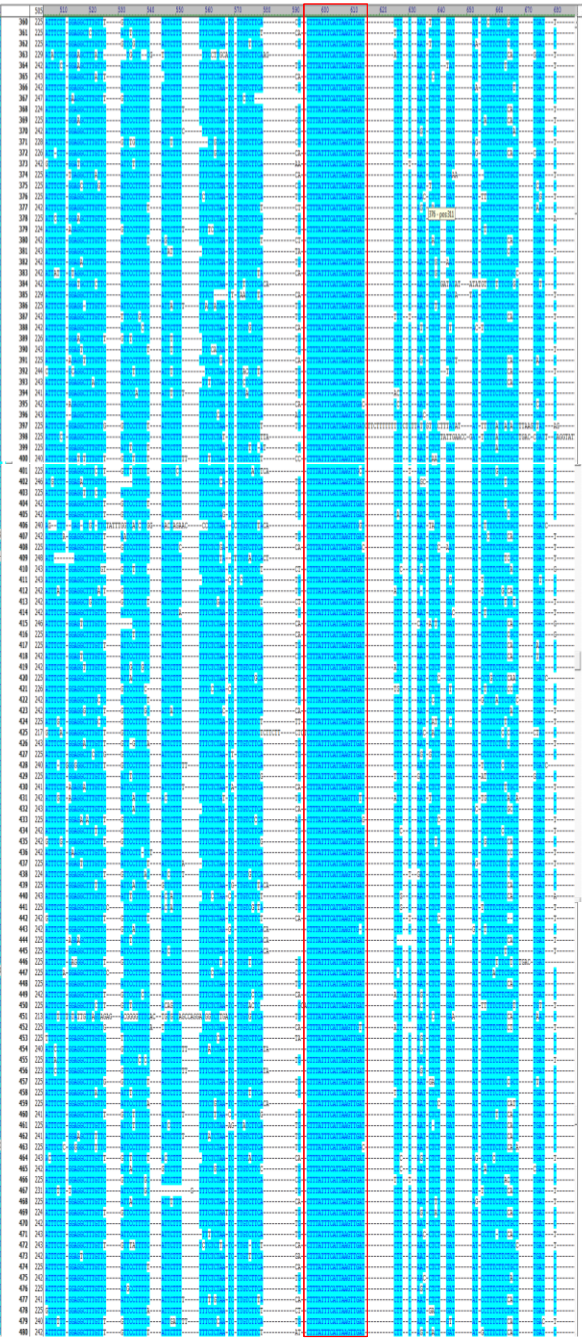
L1PA7
L1P3
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA8
L1PA7
L1P2
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7

L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA4
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1P3
L1P3
L1PA7
L1PA8
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA6
L1PA7
L1P3
L1PA7
L1P2
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1P3
L1PA8
L1PA7
L1PA7
L1P3
L1PA7

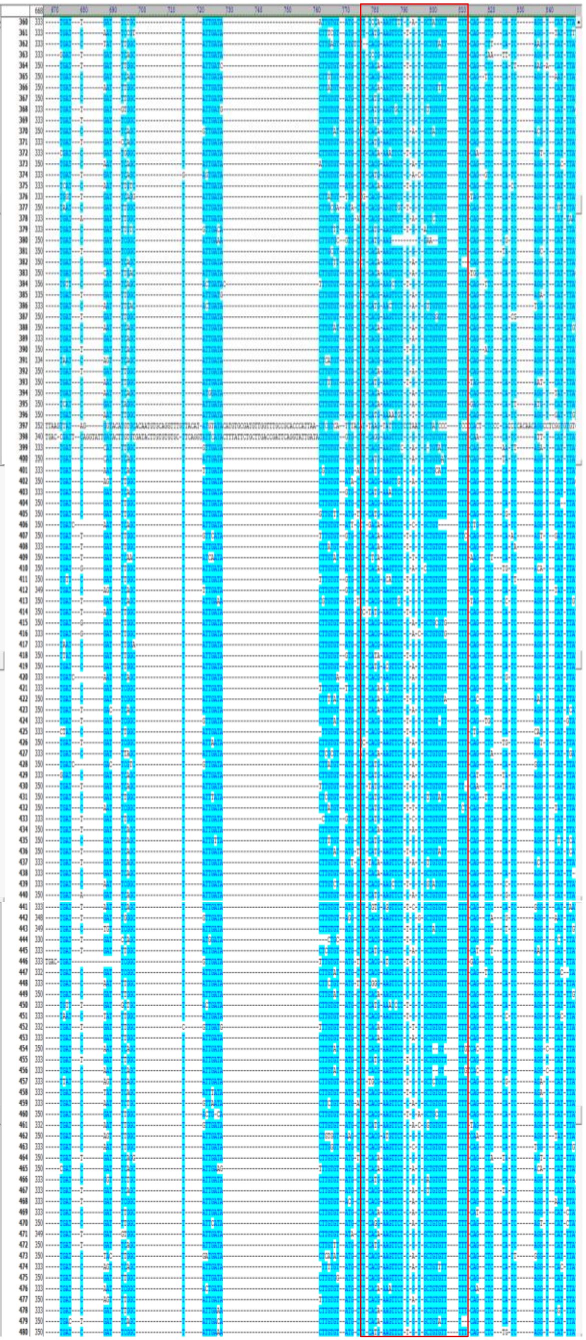
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment

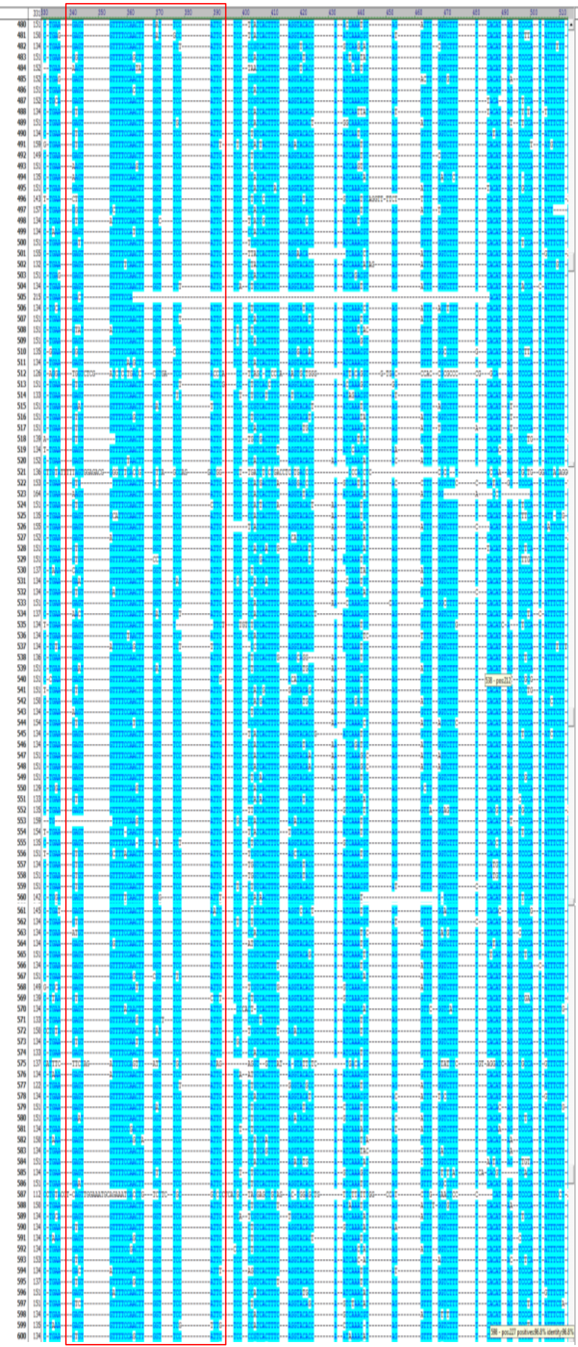


360-480 seq *LINE-1* subfamilies

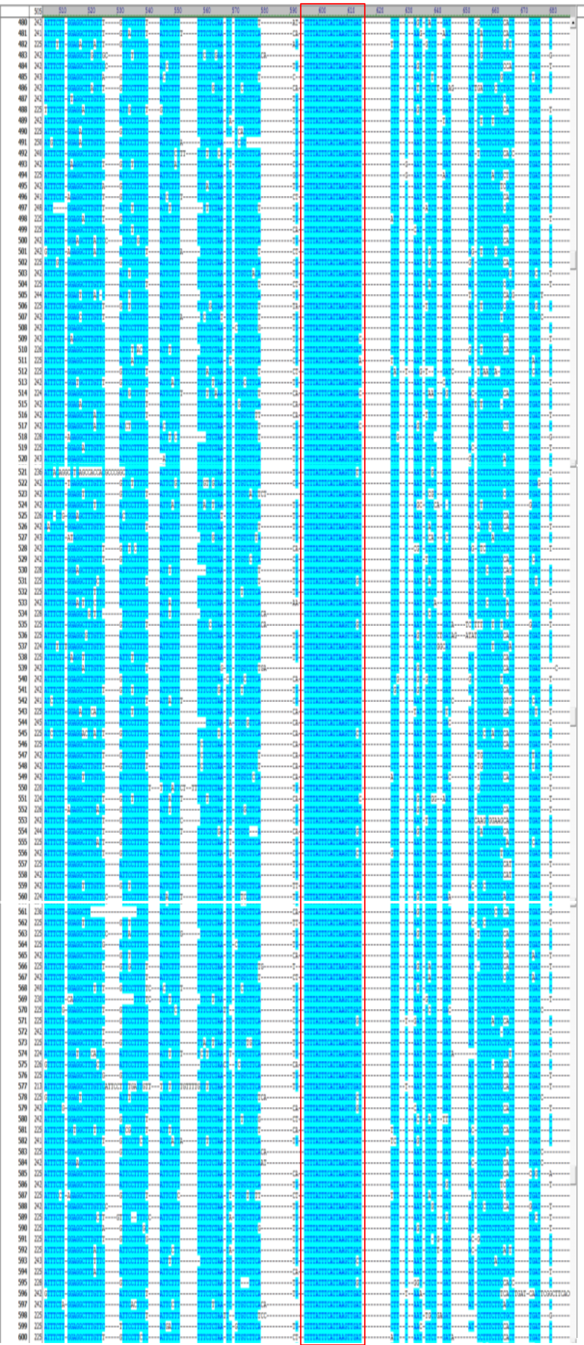
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA6
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1P2
L1P3
L1PA7
L1PA8
L1PA16
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P3
L1P3
L1PA7
L1PA7
L1PA3
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA8
L1P2
L1PA7

L1P2
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1PA8
L1PA7
L1PA7
L1P3
L1ME4a
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA6
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA8
L1P3
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA6
L1P1
L1PA7

cs_attH4X_F1 primer
alignment



attH4X
alignment



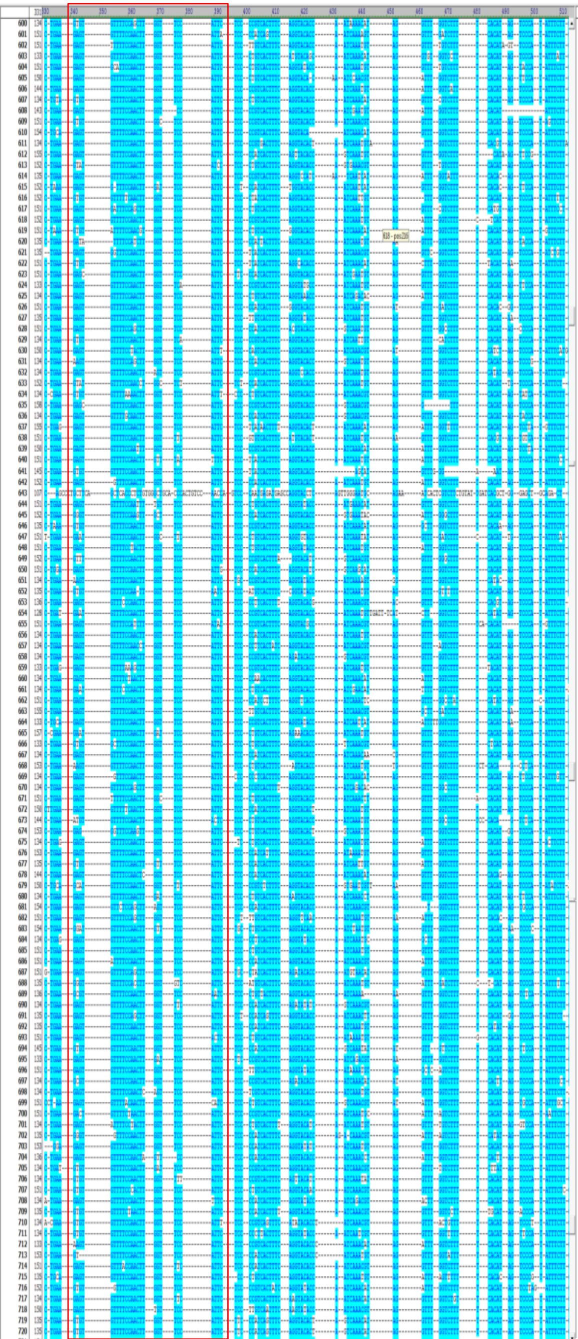
cs_attH4X_R1 primer
alignment

480-600 seq
LINE-1
subfamilies

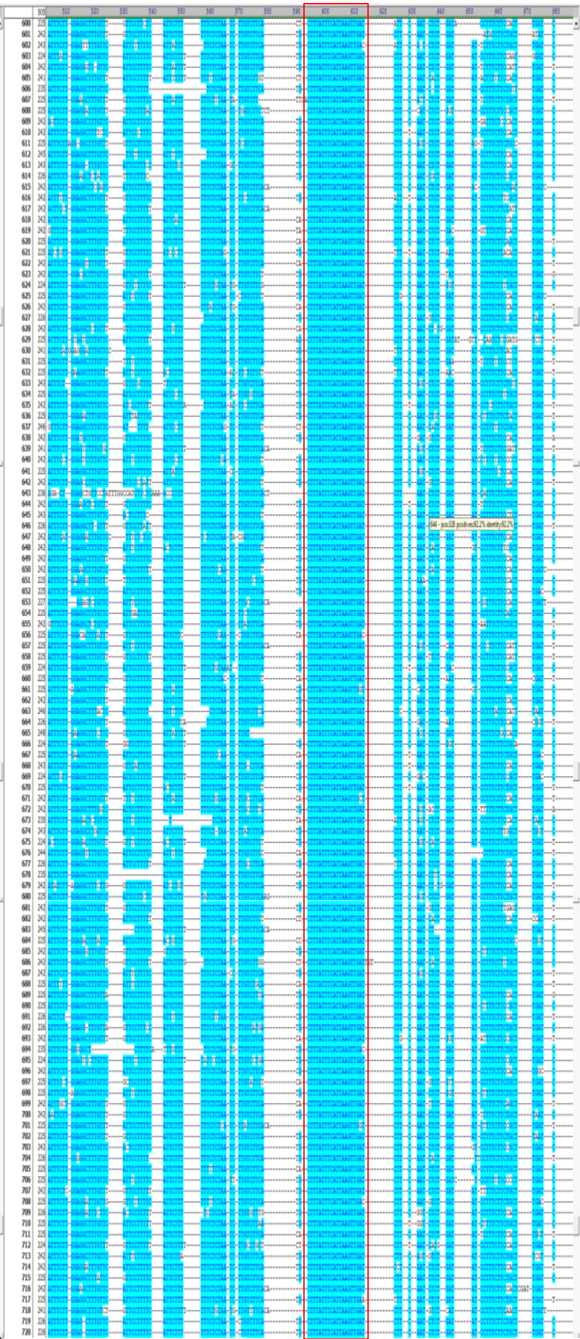
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P2
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA6
L1P3
L1PA7
L1PA7
L1PA7
L1P3
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA3
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA4
L1PA7
L1P3
L1P3
L1PA7
L1PA7

L1PA7
L1P3
L1P3
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA6
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7

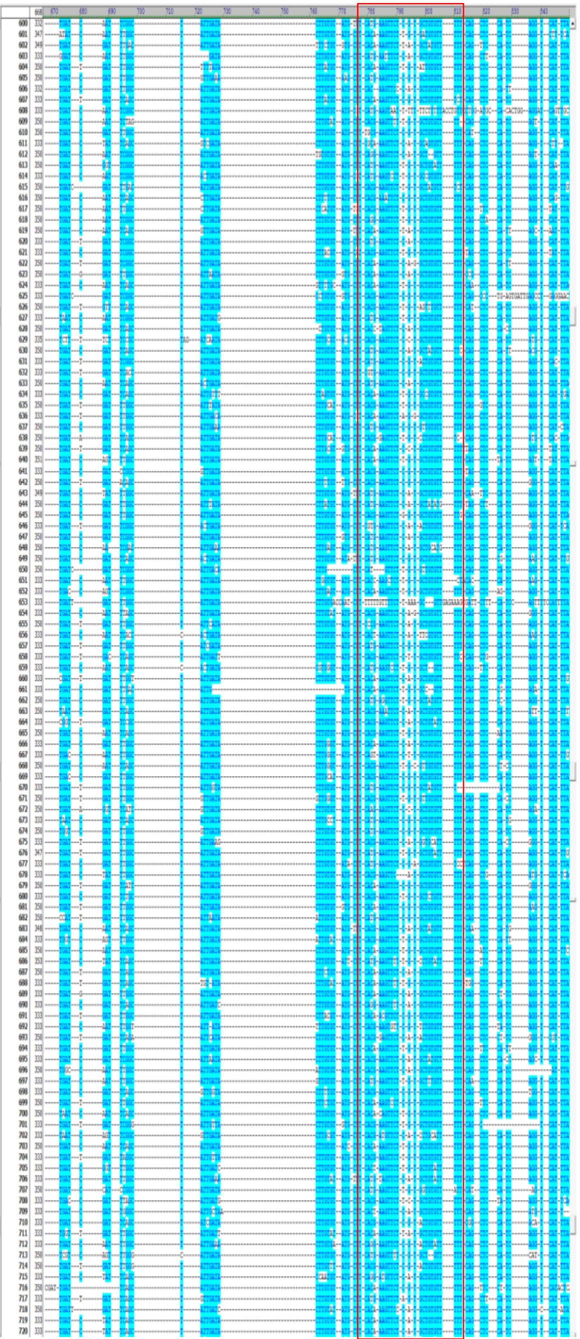
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment

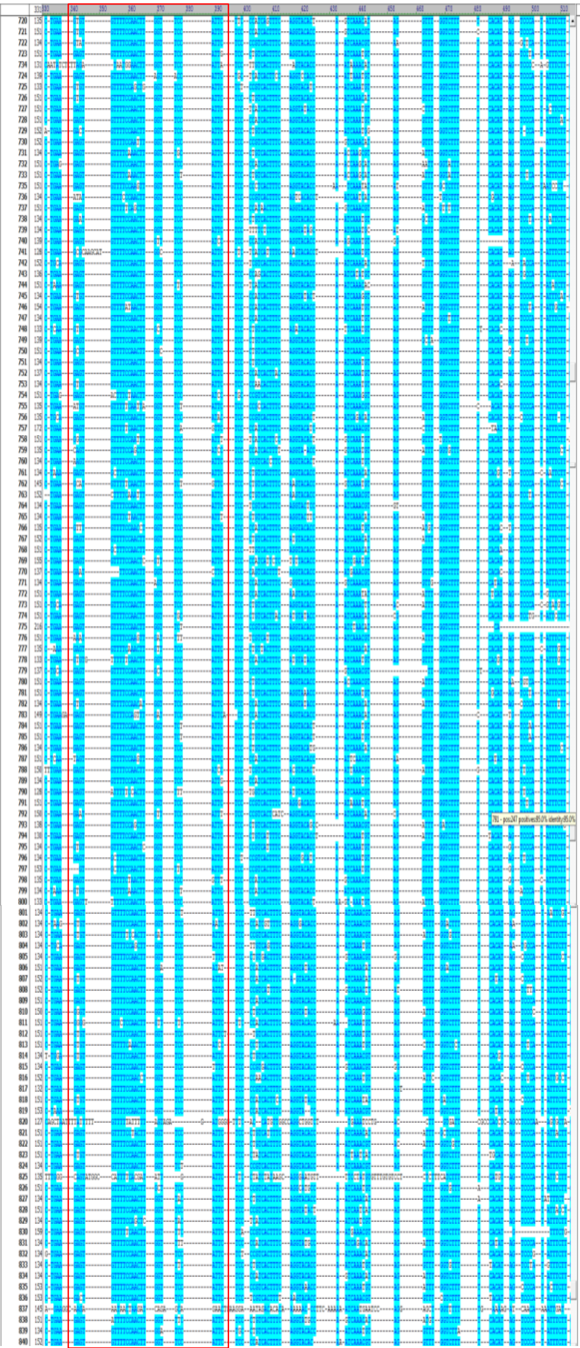


600-720 seq *LINE-1* subfamilies

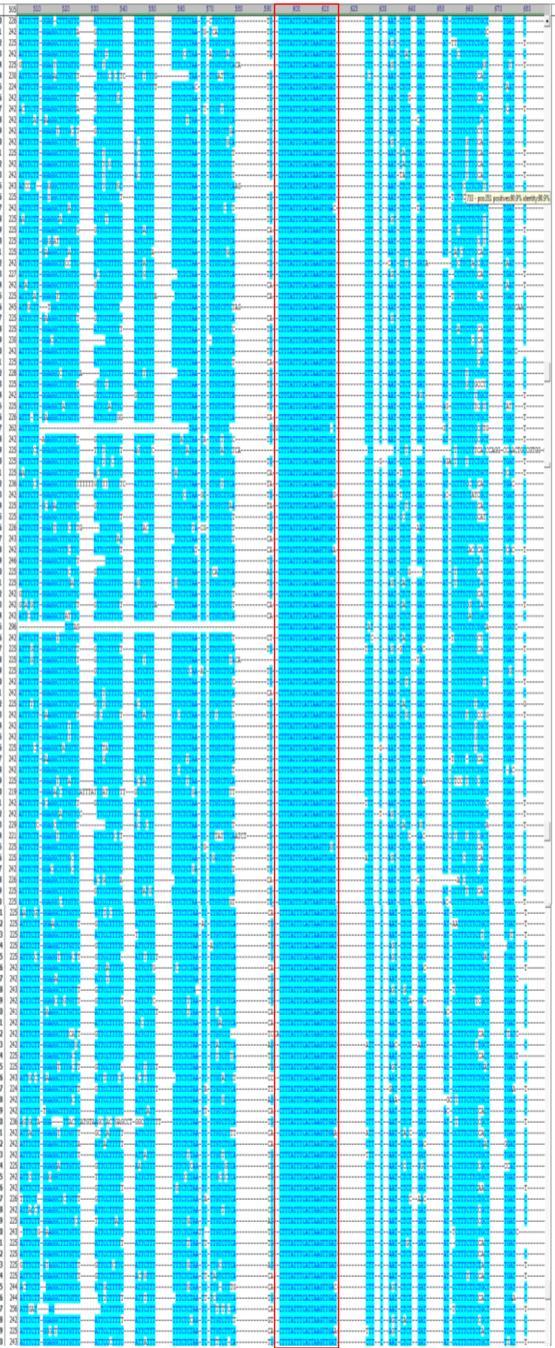
L1PA7
L1P3
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1P2
L1PA8
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA12
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1P3
L1PA7
L1P2
L1PA5
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7

L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1P2
L1PA6
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA8
L1PA7
L1P3
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7

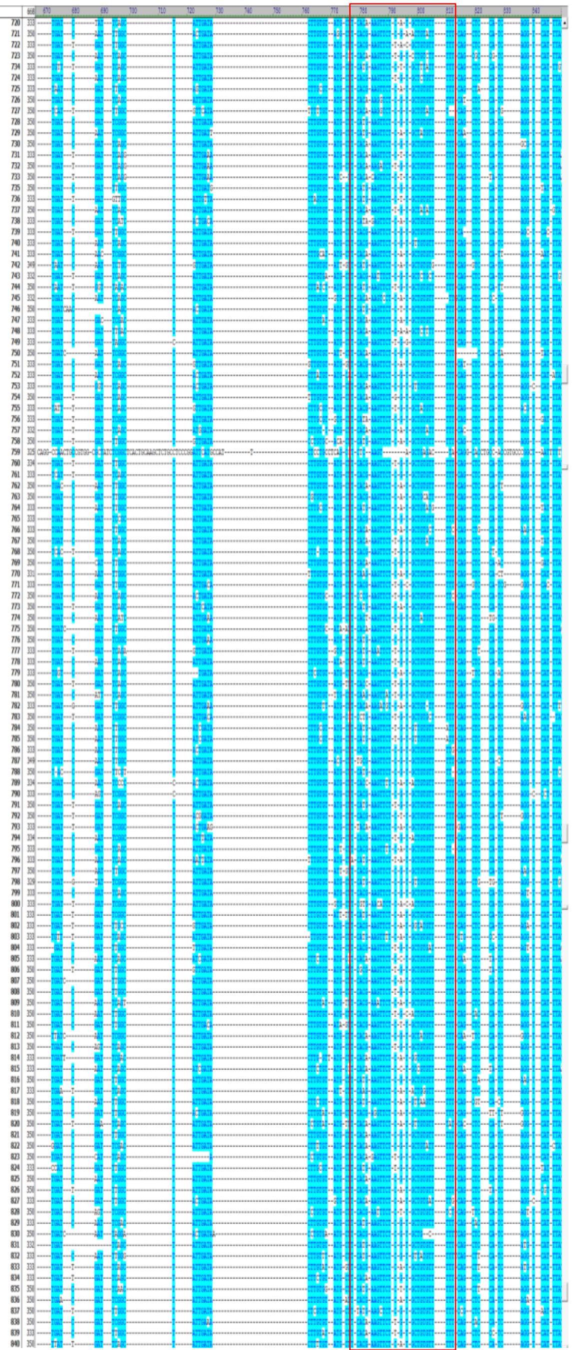
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment

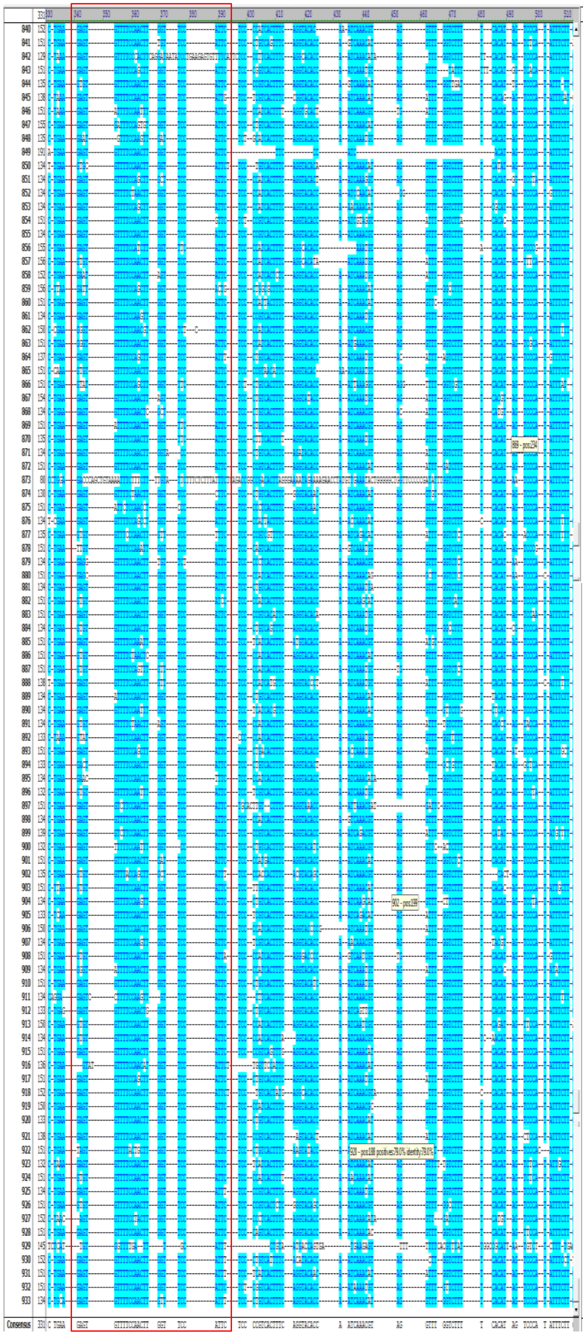


720-840 seq *LINE-1* subfamilies

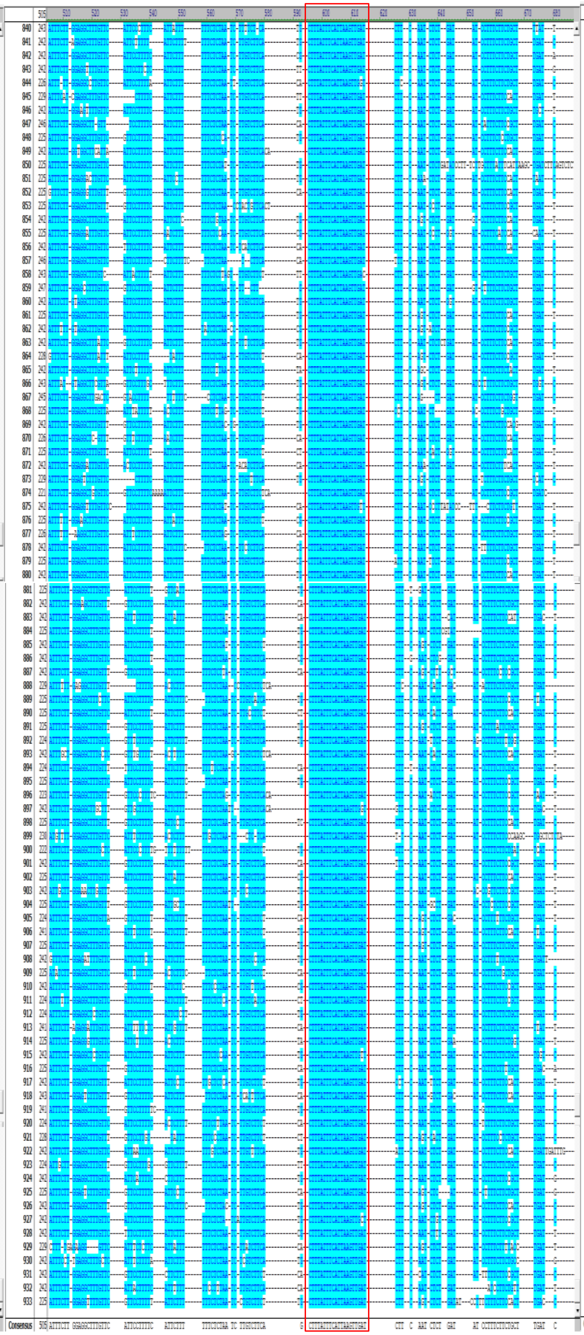
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1P2
L1P2
L1PA8
L1P3
L1PA7
L1PA7
L1P3
L1PA4
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA6
L1PA7
L1PA7
L1PA7
L1P2
L1P3
L1PA8A
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1MC2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7

L1PA7
L1PA7
L1P3
L1P3
L1P3
L1PA7
L1PA7
L1P3
L1P3
L1PA7
L1P3
L1P3
L1PA7
L1PB1
L1P2
L1PA7
L1PA3
L1PA7
L1PA3
L1P2
L1PA7
L1PA7
L1PA7
L1PA8A
L1P3
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7

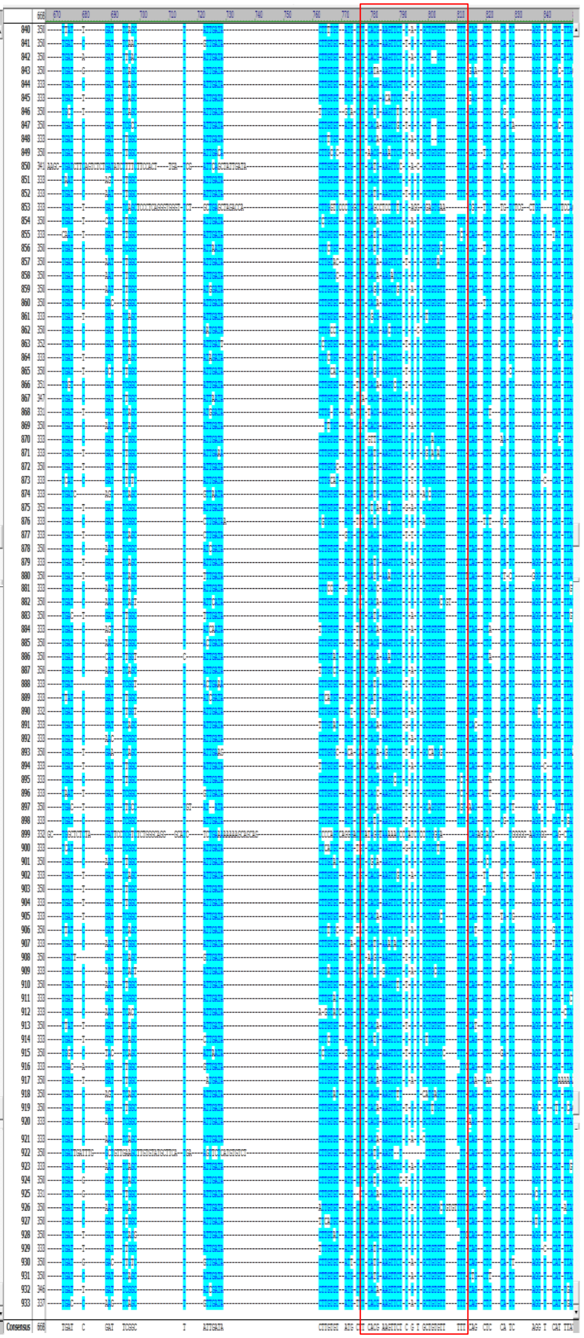
cs_attH4X_F1 primer alignment



attH4X alignment



cs_attH4X_R1 primer alignment



840-933 seq *LINE-1* subfamilies

L1PA7
L1P3
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA5
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1P3
L1P3
L1P3
L1PA7
L1PA7
L1PA7
L1PA8
L1PA7
L1PA7
L1PA7

L1P2
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1P2
L1P3
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1P3
L1PA7
L1P3
L1PA10
L1P3
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1PA7
L1P2
L1PA7
L1P3
L1PA7
L1PA7
L1P3
L1PA7
L1PA7
L1PA7
L1PA6
L1PA8
L1PA7

Supplementary Table 5. Information on *att*H4X sequences in the targeted clones from different cell lines.

Targeted Clones (Targeting Vector)	Targeted *att*H4X sequence and resulting *att*L and *att*R sequences	Genomic location of targeted attH4X	Length of targeted *LINE-1* (bp)	*LINE-1* subfamily	Retro-transposition functional status*
HT1080 # 3 (pattP4X-PGKssPuro)	TTTTCTCTAATCTAGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTAGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr19:35188923-35188943 (-)	3188	L1P3	Truncated
HT1080 # 11 (pattP4X-PGKssPuro)	TTTTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr4:152517906–152517926	3623	L1PA7	Truncated
HT1080 # 18 (pattP4X-PGKssPuro)	TTTTCTCTGATCTTGTTTTCTCACTTTATTTCATTAATTTGATCTTCAGTCACTGATACCC TTTTCTCTGATCTTGTTTTCTCACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAATTTGATCTTCAGTCACTGATACCC	chr9:15026596 -15026616(-)	5511	L1PA6	Full length non intact
HT1080 # 19 (pattP4X-PGKssPuro)	TTTTTCTCTAATCTGTCTTCATGCTTTATTTCATTAAGTTGGTCTTCAATCTCTGATATCC TTTTTCTCTAATCTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGGTCTTCAATCTCTGATATCC	chr2:153753805-153753825(-)	6467	L1PA7	Full length non intact
HT1080 # 21 (pattP4X-PGKssPuro)	TTCTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTCTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr7:111345001-111345021	3455	L1PA7	Truncated
HeLa # H1-C7 (pattP4X-PGKssPuro)	TTTTCCCTAACCTTGTCTTCATGCTTTATTTCATTAAGTTGATTGTCAATCTCTGTTATCT TTTTCCCTAACCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATTGTCAATCTCTGTTATCT	chr4:90488464-90488484(-)	1592	L1P2	Truncated
HeLa # H1-C26 (pattP4X-PGKssPuro)	TTTTCTCTAATCTTGTCTTCGTGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCTCTAATCTTGTCTTCGTGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chr7:94450365-94450385	1581	L1PA7	Truncated
HeLa # H2-C1 & # H2-C12 (pattP4X-PGKssPuro)	TTTTCTCTAATCTTGTTTTCACGCTTTATTTCATTAAGTTGATCTTCAATTTCTGATAACC TTTTCTCTAATCTTGTTTTCACGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATTTCTGATAACC	Chr4:142034109-142034129	1070	L1PA7	Truncated
HeLa # H2-C15 (pattP4X-PGKssPuro)	TTTTCTCTAATATTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATATTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr1:55897841-55897861(-)	6568	L1PA7	Full length non intact
NEB-1 # N2-C3 (pattP4X-PGKssPuro)	TTTCTCTAATTCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAGTCTATGATATCC TTTCTCTAATTCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTATGATATCC	chr5:72277503-72277523	2323	L1PA7	Truncated
NEB-1 # N2-C11 (pattP4X-PGKssPuro)	TTTTCAGTAATCTTCTCTTCATGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCAGTAATCTTCTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chrX:31488017-31488037	6405	L1PA7	Full length non intact
HT1080_pC3#A (pattP4X-PGKssPuro) (C3mRNA Exp)	TTTTCTCTAATCTAGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTAGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr3:151604346-151604356(-)	2445	L1P3	Truncated
HT1080_pC3#B (pattP4X-PGKssPuro) (C3mRNA Exp)	TTTCCTCGAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCG TTTCCTCGAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCG	chr6:115967740-115967760(-)	4440	L1P3	Truncated
HT1080_mRNA C3#C (pattP4X-PGKssPuro) (C3mRNA Exp)	TTTTCTCTAATCTTGTCTTCTATCTTTATTTCATTAAGTTGATCTTCAGTCACTGATATGC TTTTCTCTAATCTTGTCTTCTATCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCACTGATATGC	chr6:110001570-110001590(-)	3346	L1PA7	Truncated
HT1080_mRNA C3#D (pattP4X-PGKssPuro) (C3mRNA Exp)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGCTCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGCTCTTCAATCTCTGATATCC	chr3:157783598-157783608(-)	522	L1P2	Truncated
HT1080_mRNA C3#E (pattP4X-PGKssPuro) (C3mRNA Exp)	TTTTCTCTAATCTTGTCTTCTCACTTTATTTCATTAAGTTGATCTTCCATCTCTGATATTT TTTTCTCTAATCTTGTCTTCTCACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCCATCTCTGATATTT	chr1:181239747-181239767	4244	L1PA7	Truncated
HT1080 _mRNA C3C#1 (pattP4X-PGKssPuro) (C3-CNLS mRNA)	TTTTCTCCAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGTTATCC TTTTCTCCAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGTTATCC	chr5:9938593-9938613(-)	1029	L1P1	Truncated
HT1080_pC3C#1 (pattP4X-PGKssPuro) pC3-CNLS	TGTTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TGTTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr3:114897186-114897206	2058	L1PA7	Truncated
HT1080 #89 (pattP4X-PGKssPuro-EF-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr7:17848142-17848162(-)	6461	L1PA7	Full length non intact
HT1080 #51 (pattP4X-PGKssPuro-EF-eGFP)	TTTTCTCTAATCTTGTCTGCATGCTTTATTTCATTAAGTTGATCTTCAATCTCCGATATCC TTTTCTCTAATCTTGTCTGCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCCGATATCC	chr4: 90010246-90010266	2189	L1PA7	Truncated
hESC#3 (pTZattP4X-UN-EF1a-eGFP)	TTTTCTCTAATCTTGTCTTCTCGCTTTATTTCATTAAGTTGGCATTATAAAAAAGCATTGC TTTTCTCTAATCTTGTCTTCTCGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGGCATTATAAAAAAGCATTGC	chrX: 43585650 -43585670	6259	L1PA7	Full length non intact

hESC#24 (pTZattP4X-UN-EF1□-eGFP)	TATTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATAGCC TATTCTCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATAGCC	chr5: 12672998 – 12673018	2152	L1P3	Truncated
hESC#59 (pTZattP4X-UN-EF1□-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr1: 84245320 - 84245340	4937	L1PA7	Truncated
hESC A3 clone (pTZattP4X-UN-EF1□-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAAACTCTGATATCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAAACTCTGATATCC	chrX:64787931-64787952	1916	L1PA7	Truncated
hESC E3 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGCCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTTATATCC TTTTCTCTAATCTTGCCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTTATATCC	chr1:213884581-213884602	6423	L1PA7	Full length non intact
hESC K4 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAAACTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTCGTATCC TTTTCTCTAAACTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTCGTATCC	chr13:46195229-46195250	6439	L1PA7	Full length non intact
hESC A20 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTTTGATAGCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTTTGATAGCC	chr5:72103592-72103613	6302	L1PA7	Full length non intact
hESC K1 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chr7:133653822-133653843	6485	L1PA7	Full length non intact
hESC L2 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TCTTCTCTAATCTTGTCTTCACGCTTTATTTCATTAAGTTGATCTTCAAACTCTGATATCC TCTTCTCTAATCTTGTCTTCACGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAAACTCTGATATCC	chr1:84910263-84910284	6472	L1PA7	Full length non intact
hESC B16 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATGTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCTCTAATGTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chr3:111164680-111164701	6423	L1PA7	Full length non intact
hESC K6 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chr7:133653822-133653843	6485	L1PA7	Full length non intact
hESC L1 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC TTTTCTCTAATCTTGGCTTCACGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAGTCTCTGATATCC	chr7:133653822-133653843	6485	L1PA7	Full length non intact
hESC M5 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAAACTTCTTTTCTCGCTTTATTTCATTAATTTGATCTTCAATCACTAATACGC TTTTCTCTAAACTTCTTTTCTCGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAATTTGATCTTCAATCACTAATACGC	chr1:76244691-76244712(-)	6151	L1PA4	Full length non intact
hESC M9 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGGTCTTCCATCTCTGATATCC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGGTCTTCCATCTCTGATATCC	chr7:80319732-80319753(-)	6371	L1PA7	Full length non intact
hESC A8 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCCAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGTTATCC TTTTCTCCAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGTTATCC	chr5:9938285-9938306(-)	1029	L1P1	Truncated
hESC A17 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAATCTCTGATATTC TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATTC	chr1:92719737-92719758(-)	5690	L1PA7	Truncated
hESC G4 clone (pattP4X-PGKssPuro-UN-EF-eGFP)	TTTTATCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTATCTAATCTTGTCTTCATGCTTTATTTCATTAAGTTG//CAGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chrX:38949476-38949497(-)	2162	L1PA7	Truncated
HT1080 -19 (pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP)	TTTTCTCTAATCTTGTCTTCGTGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCGTGCTTTATTTCATTAAGTTG//CTGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr2:150835444-150835464(-)	5080	L1PA7	Full length non intact
HT1080-26 (pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP)	TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTGATCTTCAGTCTCATATCTTT TTTTCTCTAATCTTGTCTTCACACTTTATTTCATTAAGTTG//CTGCTTTATTTCATTAAGTTGATCTTCAGTCTCATATCTTT	chr21:19146794-19146814	6451	L1PA7	Full length non intact
HT1080-29 (pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP)	TTTTCTCTAATCTTGTCTTCACGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC TTTTCTCTAATCTTGTCTTCACGCTTTATTTCATTAAGTTG//CTGCTTTATTTCATTAAGTTGATCTTCAATCTCTGATATCC	chr12:2050403-2050423	984	L1PA8A	Truncated
HT1080 -31 (pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP)	TTTCCTTGCATCTTGTCTTCAAGCTTTATTTCATTAAGTTGATATTCAATCTCTGATATCC TTTCCTTGCATCTTGTCTTCAAGCTTTATTTCATTAAGTTG//CTGCTTTATTTCATTAAGTTGATATTCAATCTCTGATATCC	chr4:142912896-142912916	5600	L1PA7	Full length non intact
HT1080-32 (pTZ18R-attL4X/attR4X-PGKssPuro-UTF1-EGFP)	TTTTCTCTAATCTTGTCTTCTCGCTTTATTTCATTAAGCTGATCTTCAATCTCTGTTATCC TTTTTTCTCTAATCTTGTCTTCTCGCTTTATTTCATTAAGTTG//CTGCTTTATTTCATTAAGCTGATCTTCAATCTCTGTTATCC	chr20:33750784-33750805(-)	2947	L1PA7	Truncated

*According to L1Base (Database for *LINE-1* elements)

Supplementary Table 6. Summary of Southern analysis results for the targeted clones in HT1080, HeLa, NEB-1 and hESC lines.

CLONES	***att*L junction**		***att*R junction**		**Copy Number**
	Expected Fragment size (bp)	**Observed Fragment size (kb)**	**Expected Fragment size (bp)**	**Observed Fragment size (kb)**	
	Restricted with *XbaI*		**Restricted with *PstI***		
HT1080 #3	4421	~4.4	4470	~4.4	2
HT1080 #11	3787	~3.7	4527	~4.5	2
HT1080 #18	3490	~3.5	4590	~4.6	1
HT1080 #19	1758	~1.7	4594	~4.6	1
HT1080 #21	3760	~5.1	4579	~4.7	1
HeLa#H1- C26	7084	~7	4376	~4.3	1
HeLa# H1- C7	4108	~4	4526	~4.5	Multiple
HeLa# H2-C1	8698	~8.7	13659	> 10	1
HeLa# H2-C12	8698	~3.1	13659	~4	1
HeLa# H2-C15	3618	~2.1	4596	~4.9	1
NEB-1#N2-C3	6020	~6	4645	~4.6	1
NEB-1#N2-C11	3493	~3.5	4500	~4.5	1
	Restricted with *HindIII*		**Restricted with *XbaI***		
hESC#3	8779	~8.6	6256	~6.2	1
hESC#24	4996	~4.7	6063	~6.1	1
hESC#59	4480	~4.3	15893	~16	1
	Restricted with *NdeI*		**Restricted with *XbaI***		
hESC-A3	10500	~8	6600	~10.5	1
hESC-E3	10500	~10.5	7400	~7.4	1
hESC-K3	10300	~5.8	6100	~5.7	1
			Restricted with *BglII*		
HT1080-19			6600	~6.6	1
HT1080-26			11000	~7.5	1
HT1080-29			15000	~7	1
HT1080-31			8800	~6.3	1
HT1080-32			17000	>15	1

Supplementary Table 7. Transcriptome analysis showing expression (log values) for upregulated genes (> log 2) and downregulated genes (< log2) along with their chromosomal location in the hESC clones (#3,#24, E3).

hESC #3 : Targeting at chrX: 43,585.650 – 43.585.670

Gene Name	Expression (Log value)	Chromosomal location
CSAG1	4.1465	chrX:151,903,228-151,909,518
MAGEA2B	3.1835	chrX:151,918,388-151,922,364
SP140L	3.583	chr2:231,191,894-231,268,445
MAGEA2	2.864	chrX:151,918,388-151,922,364
CSAG3	2.534	chrX:151,923,679-151,928,738
TRIM43B	2.4655	chr2:96,142,715-96,150,479
DMD	3.072	chrX:32,382,699-32,383,325
RPL10L	2.915	chr14:47,120,220-47,121,028
DNER	-2.322	chr2:230,312,032-230,451,615
ZIC2	-2.068	chr13:100,634,026-100,639,019
HIST1H2BB	-2.5265	chr6:26,184,023-26,184,404
LOC729506	-2.578	chr5:8,333,596-8,457,677
ZNF257	-3.5415	chr19:22,235,266-22,273,903
ZNF560	-3.887	chr19:9,577,031-9,609,279
ZNF492	-3.789	chr19:22,817,126-22,850,472
RWDD2B	-3.0945	chr21:30,378,080-30,391,685
XIST	-3.937	chrX:73,040,486-73,072,588
CLC	-4.4425	chr19:40,221,893-40,228,669
ZNF98	-4.966	chr19:22,573,899-22,605,148
ZNF676	-5.843	chr19:22,361,903-22,379,753
ZNF729	-6.006	chr19:22,469,252-22,499,978
HIST1H3C	-6.1145	chr6:26,045,639-26,046,097
ZNF208	-7.899	chr19:22,148,897-22,193,745

hESC #24: Targeting at chr5:12.672.998 -12.673.018

Gene Name	Expression (Log value)	Chromosomal location
ZNF560	-2.5195	chr19:9,577,031-9,609,279
XIST	-4.236	chrX:73,040,486-73,072,588

hESC E3: Targeting at chr1:213,884,581-213,884,602

Gene Name	Expression (Log value)	Chromosomal location
PIK3AP1	3.494	chr10:98,353,069-98,480,279
MAGEA2B	3.4025	chrX:151,918,388-151,922,364
MAGEA2	3.4285	chrX:151,918,388-151,922,364
CSAG1	3.3225	chrX:151,903,228-151,909,518
CSAG3	3.5585	chrX:151,923,679-151,928,738
PCDH10	3.2505	chr4:134,070,470-134,112,732
KHDC3L	3.659	chr6:74,072,400-74,073,898
NPPB	3.1175	chr1:11,917,521-11,918,992
SFTA1P	2.8615	chr10:10,826,402-10,836,877
S100A4	2.8155	chr1:153,516,095-153,518,282
PLAT	2.646	chr8:42,032,236-42,065,194
LOC339535	2.582	chr1:238,643,684-238,649,317
CDKN2B	2.613	chr9:22,002,902-22,009,312
SP140L	2.997	chr2:231,191,894-231,268,445
MMP9	2.503	chr20:44,637,547-44,645,200
HAPLN1	3.389	chr5:82,934,017-83,016,896
MYOF	2.942	chr10:95,066,186-95,242,074
COL5A1	3.0405	chr9:137,533,651-137,736,688
SYTL5	2.6955	chrX:37,892,787-37,988,073
TGFBI	2.9995	chr5:135,364,584-135,399,507
CAPN2	2.2885	chr1:223,900,119-223,963,720
CD46	2.212	chr1:207,925,383-207,968,861
CXCL14	2.2885	chr5:134,906,371-134,914,969
EBI3	2.209	chr19:4,229,540-4,237,524
TGM2	2.477	chr20:36,756,864-36,793,700
NDRG1	2.269	chr8:134,249,414-134,309,547
COL3A1	2.823	chr2:189,839,099-189,877,472
DPEP3	2.254	chr16:68,009,566-68,014,452
ACTA1	2.1155	chr1:229,566,993-229,569,843
FOXD1	2.5285	chr5:72,742,085-72,744,352
WNK4	2.5205	chr17:40,932,649-40,949,084
FRG1B	2.5545	chr20:29,624,035-29,653,908
ANGPT1	2.0485	chr8:108,261,710-108,510,254
ACTC1	2.2675	chr15:35,080,297-35,087,927
ZNF99	-2.3175	chr19:22,934,985-22,966,973
FAM107A	0.0745	chr3:58,549,845-58,563,491
COMTD1	-2.4025	chr10:76,993,729-76,995,770
LOC729506	-3.2075	chr5:8,333,596-8,457,677
CAT	-3.1955	chr11:34,460,472-34,493,607
ZNF560	-3.4215	chr19:9,577,031-9,609,279
ZNF257	-4.171	chr19:22,235,266-22,273,903
XIST	-4.566	chrX:73,040,486-73,072,588
ZNF98	-4.734	chr19:22,573,899-22,605,148
ZNF676	-5.019	chr19:22,361,903-22,379,753
ZNF729	-5.2215	chr19:22,469,252-22,499,978
ZNF208	-5.9805	chr19:22,148,897-22,193,745