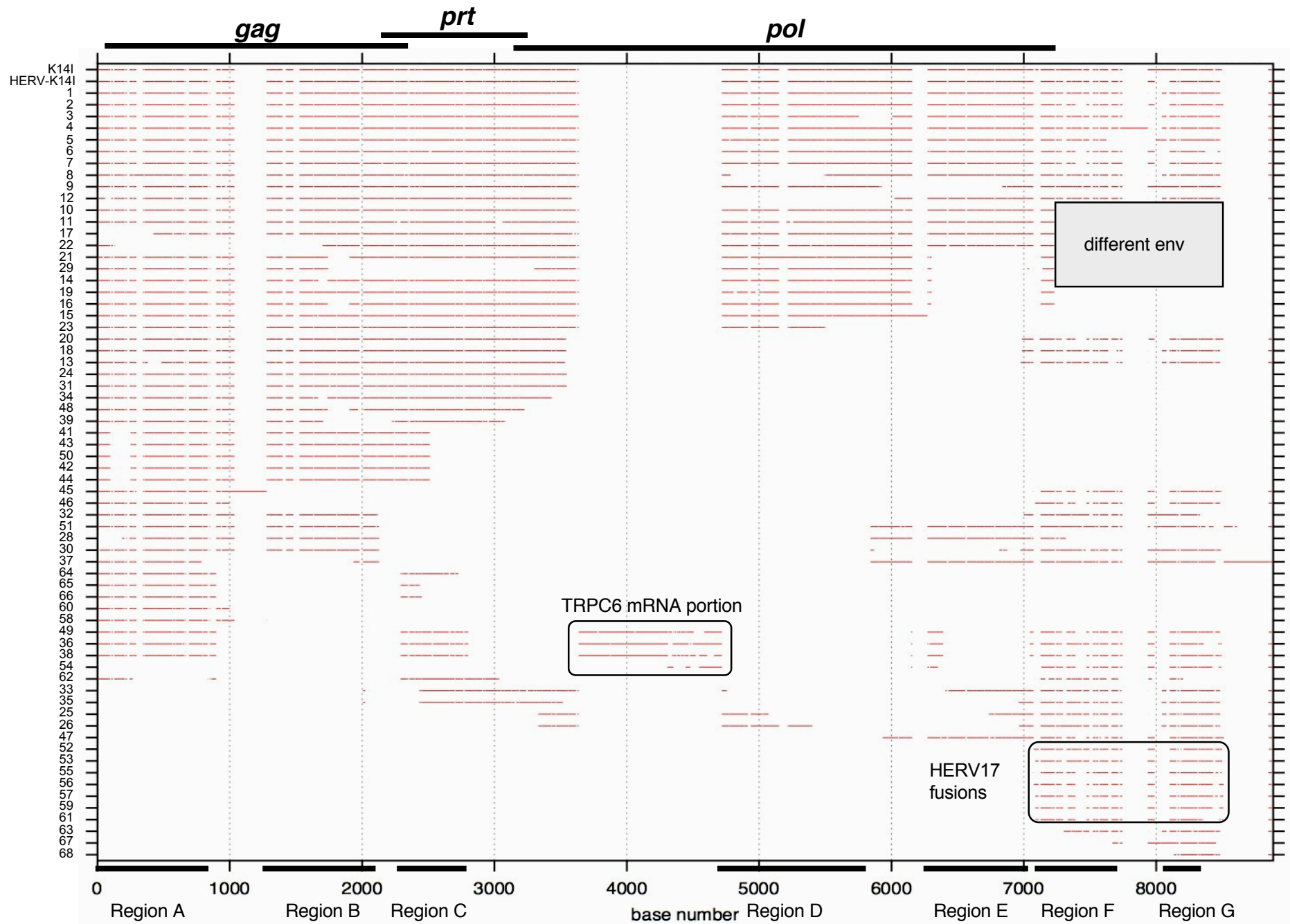
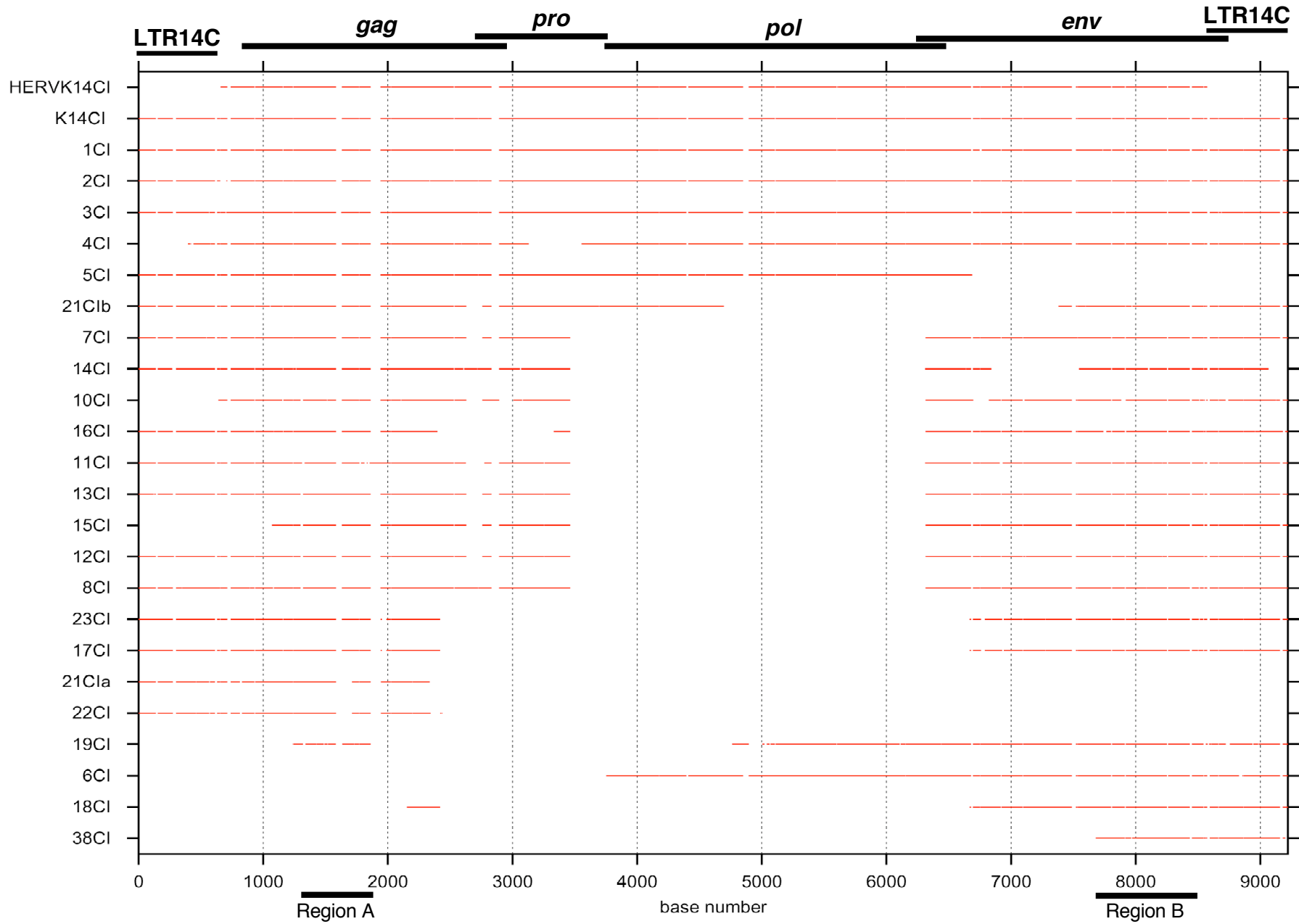


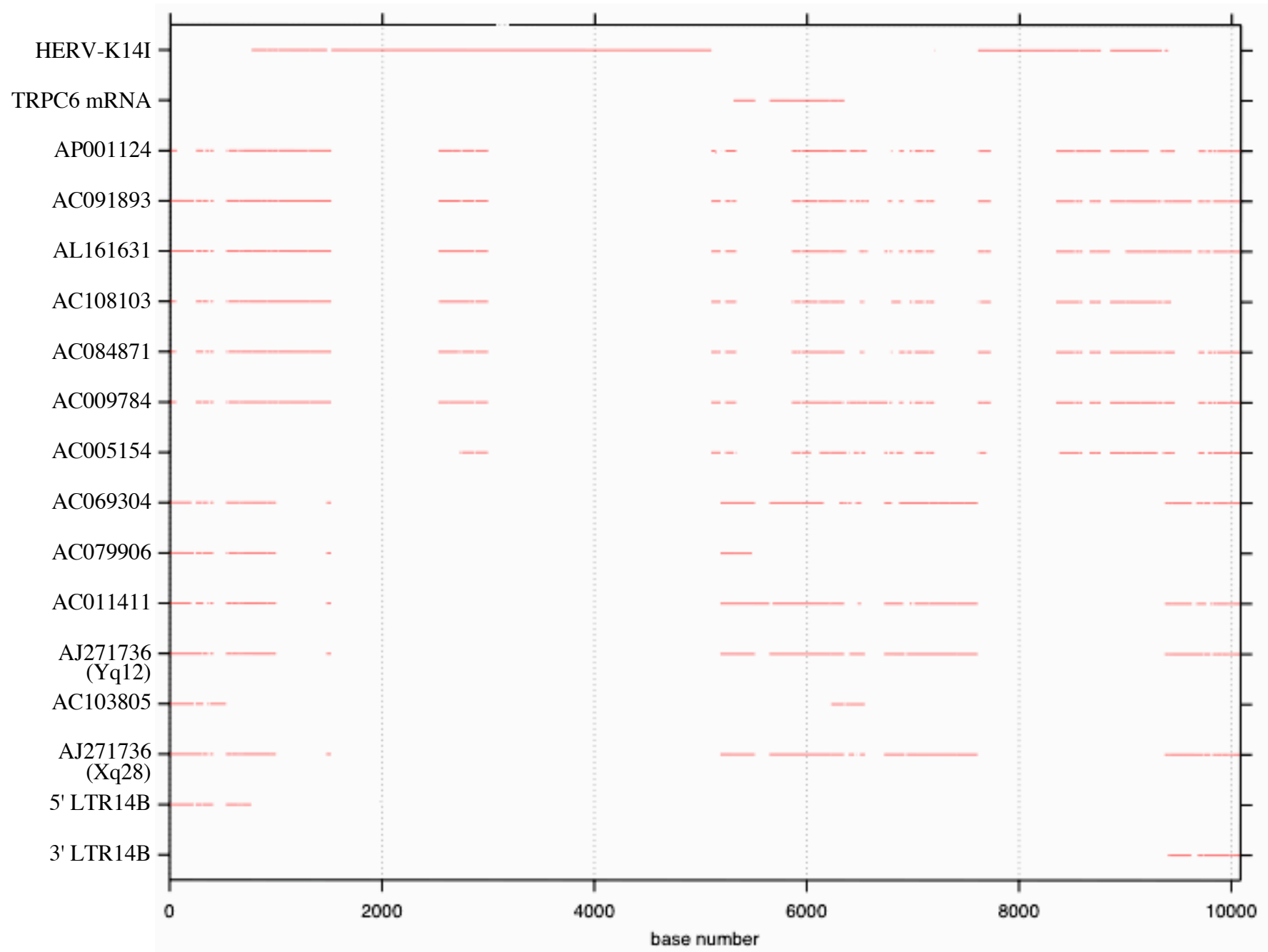
Supplementary Table B

1	2	3	4	5	6	7		8	9	10	11	12
PV no.	SCORE	START	END	IDENTITY	STRAND	POSITION	ACCESSION NO	5'LTR14C	3'LTR14C	LTR similarity	Repeats in PV	PV in Repeats
1CI	7203	0	7417	98.6%	-	16p12.3	AC130456.2	+	+	91		L2
2CI	6321	0	7417	93.1%	-	15q21.1	AC051619.7	+	+	92	AluSg/x	AluSg/x
3CI	6160	7	7392	92.3%	+	2p12	AC079896.6	+	+	91		THE1B-int
4CI	5871	7	7417	92.2%	+	13q31.1	AL445588.12	+	+	85		MER4(C)-int
5CI	4715	0	5654	92.1%	-	4q33	AC069306.6	+			LTR12	
6CI	3895	2833	7417	92.7%	+	8q24.3	AF235103.4				LTR5B, HERVK	
7CI	3735	15	7417	91.2%	+	6p12.3	AL391538.11	+	+	89	AluY	
8CI	3575	8	7411	89.5%	-	Yq11.223	AC010086.4	+	+	83	AluY	
9Cla, b	3573	0	7410	89.5%	+	Yq11.221	/	+	+	83		MER117, MER20
10CI	3456	7	7401	91.5%	+	1q44	AL627095.8/AL390728.37		+		AluY	
11CI	3352	28	7414	88.8%	-	Yp11.2	AC009952.4	+	+	86	AluYc	
12CI	3348	18	7414	88.7%	+	Yq11.223	AC021107.3	+	+	84	AluYc	
13CI	3231	18	7266	88.6%	-	Yq11.223	AC024236.5/AC007322.4	+	+	86	AluYc	
14CI	3215	0	7414	90.7%	+	19q13.2	AC092071.2/AC011541.7	+	+	89		
15CI	3040	375	7292	88.5%	+	Yp11.2	AC007967.3		+			
16CI	2887	16	7390	89.0%	-	Yp11.2	AC016749.4	+	+	84		
17CI = 9C1b	2404	26	7410	88.3%	-	Yq11.221	AC022486.4/AC007379.2	+	+	80		
18CI	1726	1330	7417	92.0%	+	8q21.12	AC068700.5/AC100854.2		+			
19CI	1537	3839	7008	82.3%	-	7q36.1	AC069304.7					
21Cla	1299	0	7023	88.8%	+	Yq11.221	AC007742.4	+	+	83		ALR/Alpha, MER117, MER20
21C1b = 9C1a							AC022486.4/AC007379.2					
22CI	962	0	1461	86.1%	-	Yq11.221	AC024183.4	+				ALR/Alpha
23CI = 9C1b	698	26	1056	86.9%	+	Yq11.221	AC022486.4/AC007379.2	+	+	80		
38CI	209	7137	7414	88.4%	+	Yq11.221	AC007379.2					

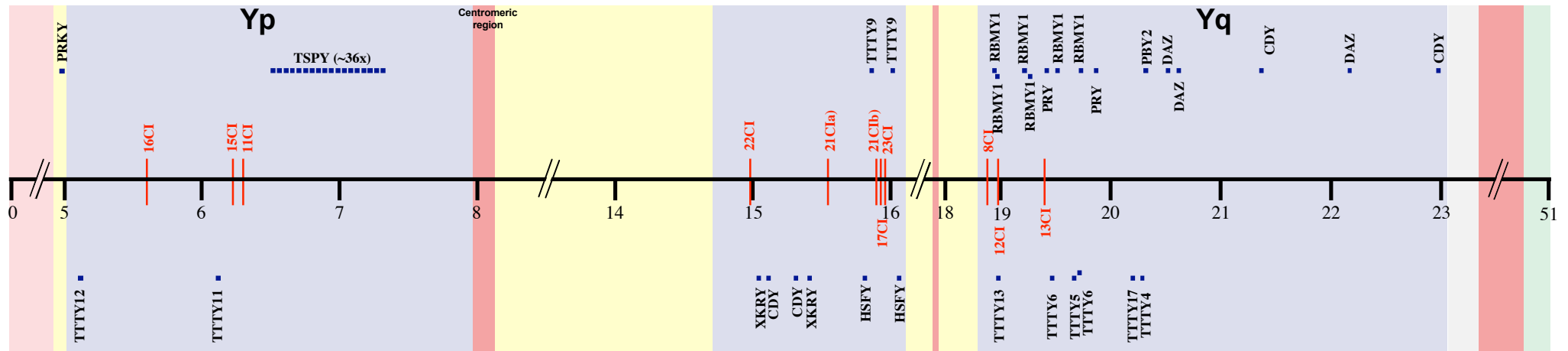


Flockerzi et al., Suppl. Fig. A





Flockerzi et al., Suppl. Fig. C



Suppl. Fig. D: Schematic localization of HERV-K14CI loci in intrachromosomally duplicated regions of the human Y chromosome. The figure is adapted from a recently published map of the human Y chromosome (Skaletsky et al. 2003) (ref. 38). Ampliconic regions, consisting of intrachromosomal duplications, are shown in blue. Heterochromatin, X-degenerate and pseudoautosomal regions are in red, yellow and green, respectively. The locations of several Y chromosome-located genes is shown. HERV-K14CI loci (red lines) were mapped according to their relative distances from those genes. Numbers refer to the ones given in Suppl. Table B. Only relevant regions of the Y chromosome are shown. The scale is in Mb.

3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
AAGAAAAATCAGGAAATGGAGAATGTTAAACAGATTAAAGGCTGTTAATGCTGTAATTCACCCATGGGTCACACTGCAACCAGGGCTGCCCTCCCCAA
K K K S G K W R M L T D L R A V N A V I Q P M G A L Q P G L P S P>

3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
CAATGATCCAAAATACCTGGCCCTCTCATAGTGTAGATCTAAAGGATGCTTTTACCATTCCCTTAGCTGCCCAAGATTATGAAAAATTTGCTTTTAC
T M I P K Y W P L I V I D L K D C F F T I P L A A Q D Y E K F A F T>

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
TGTTCCGCATATAAAAGAACAGCAGGACAGATACCATTTGAAAGTACTACCAAGGCATGTTAAATAGCCCGACTATTTGTTCAAATTTATGTC
V P A I N N K E A P A D R Y H W K V L P Q G M L N S P T I C Q T Y V>

3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
GGGAAAGCTTAAAGCCAGTTAGAGAACAGTTTAAAAAATGTTATATTACATTTACATGGATGATATTTTATGTGACAGCTGAAACATAGGGAAGATAA
G K A I K P V R E Q F K K C Y I I H Y M D D I L C A A E T R E E L>

3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
TGTTATGCTACAAACAGTTAGAAAAGCTGTTAAATGCAGCAGGGTAAATATAGCCCCGATAAAAATCCAAACTTCTACTCCCTTTCAATATTTAGGAAT
M L C Y K Q L E K A V N A A G L I I A P D K I Q T S T P F Q Y L G M>

3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
GAAGGTAGAACAAGTGTCTATTAGCCCTCAAAGGTTCAAATTCGAAGAGATAATTTAAAACTTTAAATGATTTTCAAAAATTTATTAGGAGACATTAA
K V E Q S A I K P Q K V Q I R R D N L K T L N D F Q K L L G D I N>

4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
TGGATTCATCAACTTTAGGCATTCCTACCTATGCTATGTCTCACCTTTTCTACTTTACGAGGTGATTTAACTTAAACAGTAATGCTCCCTGTCCA
W I H P T L G I P T Y A M S H L F S T L R G D S N L N S K C S L S>

4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
AAGAAGCATTTGGAGAACTCAATTAATTGAAGAAAAATTCACCAAGCACAAGTGAACAAATTAACCCCTATGACGCCATTACAGTTTGTAGTTTTC
K E A L E E L Q L I E E K I Q Q A Q V K R I N P M Q P L Q F L V F P>

4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
TACTAACATTACACAGGAGTTTGTTCACAGGATGATCTGGTGTGGTGGCTTTTCTACCTCACAAATACAACAAAACAGCTCCTCTGACTTA
T K H S P T G V I V Q Q D D L V E W L F L P H N T T K T L T L Y L>

4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
GATCAAAATGCTGTGCTAGTAGGACAAGCAGCTGCGCACAAAAGTTAATGGATATGATCCAAATCAGATTATAGTTCCATTAAACAAAACAA
D Q I A V L V G Q A R L R T T K L M G Y D P N Q I I V P L T K Q Q>

4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
TTCAACAAGCTTATATTATCCAGGAATGGCAAGTTAATTTGGCAGGTTTTGGTGGCTTTTGTGATAATCATTAATCTTAACTAAAATTTCCAAAT
I Q Q A Y I N S Q E W Q V N L A G F V G I L D N H Y P K S K I F Q F>

4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
TCTAAATTAACATCTGGATTTGCTTCTATTACTCAAAAAGCCCTATTGAAGGGGCATTACTGTTTTACTGATGGATCTAGTAATGAAAAAGCC
L K L T S W I L P S I T Q K A P I E G A I T V F T D G S S N G K A>

4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
TCATTTGACAGGACCTCAACAACAAAGTTTTTCAAACGACTTTGCTTCTGCTCAAAAGGGCTGAACCTATGGCTGTGATAACAGTGTAAAAACTTTTAAAC
S F A G P Q Q V F Q T D F A S A Q R A E L M A V I T V L K T F K>

4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
AGCCATGAACATTTCTGATTCAGCCTATGTAGTGAAGCCACACAAAATTTGAATGTGCTTAAATTCAAAATGCTGACTGTGAACAACTTAATCT
Q P V N I V S A Y V V O A Q N I E C A L I O N V T D E Q L N L>

4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
TTTATTTTCACTTTTACAGCAAGCAGTACAACAAGCCATTTCCCTTTTCTATATCACTCATATGAGAGCACATACTAACCCTCCCTGGCCCTTTAACTAAA
L F H S L Q Q A V Q Q R H S P F Y I T H M R A H T N L P G P L T K>

4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
CTTAATCAAAGGCCGATGCTTGTCTGACGCTTTGCTGATGCACAAACATTTCCATTCTTTAACCCCTTAAATGCTGACGGCTTAGAAAAAGAT
L N Q R A D A L V S A A F A D A Q T F H S L T H L N A A G L R K R>

5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
ATGGTCTATCATGGAACAAGCTAAAGAAATTTGTCACAACACTGTTCTGCGCTGCCAAGTCTGCATCTGCCACATCAAGGAACAGGAGTTAACCCCTAGAGG
Y G L S W K Q A K E I V Q H C S A C Q V L H L P H Q G T G V N P R G>

5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
TTTATCTCAAATTCATCTGGCAGATGGATGTAACACATATTCCTGCTTTTGGAAAATTTGCTCCTTTGTTTCATGTTTTCAGTAGATACCTATTACATTT
L S P N S I W Q M D V T H I P A F G K L S F V H V S V D T Y S H F>

5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
ATCTGGGACATCTCAACAGGGGAAGCTACAGCTCATGTTAAAGACATCTTTTATCTTGTCTTTTCAGTTATGGGAAATCCCAAAAATCAAAACTG
I W A T C T Q T G E A T A H V K R H L L S C I P S E V M G I P E K I K T>

5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
ATAACGGCCAGGATACGTAGTAAGCCATGGCTACATTTTTTCAACAATGGAATATTTACCATACTACGGGTATTTCCATATAACTCACAAGGACAAGC
D N G P G Y C S K A M A T F F Q Q W N I T H T T G I P Y N S Q G Q A>

5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
AATAGTGAAGAAGAGCTAATCGTACTTTAAAACTCAAATAACAAGCAAAAGGAGGACCCAGGAATATAAAACACATATGCAATTTGCAATTTAGCT
I V E R A N R T L K T Q I Q K Q K G G D Q E Y K T P H M Q L H L A>

5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
TTATTAACATTAATTTTAAATTTTACAAAAAGTACAACCCATGACTGACGCTGAACAACCTGACAGGACAAAAGGAAAATAAAAAGGCTGGACAAG
L L T L N F L N L Q K D Q P M T A A E Q H L T G Q K E N K K A G Q>

5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
ATATATGTTGGAGGATGCACATACAAGAGCTGGGAAAAGGAAAGACAATATATGGGAAAGAGGATTTGCTTGTCTCTCCAGGTGACAATCAGGT
D I W W R D A H T K S W E K G K T I I W G R G F A C V S P G D N Q V>

5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
GCCTGTGTTGGTGGCCACCAAACATCTGAAGATCTATCATGAGCCACAGACTAGTGGACCCACCTGTACAGTGCAAAATGAAGGTTTAAAGGATGCTT
P V W V P T K H L K I Y H E P Q H L V D P P V Q C K L K V *>

5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
TTTTGCTATACTGTTGCACAAGAAGGATAAGCCTCGATTGCTTTCTGTCGCTTCTGTTAATCAGAAGGGGCTGTTTCTCATTATCAGTGGTAAGTT
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
TTACCCCATGGTAATTAACAAAGAGGCAGAAGCTGATTTACAATGCTTCAGCAATGGCAGGCTCCCGGCTACAGCCACAAAAGTTTTTGTCTCTGTT
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
TCAGTAGATTTACTAACCTGGGGTGGGGTATGCTTGTGTTTTCAGGAGATGAACAACCGTGGATGACCCCTCAAGATGTGACAGCACCATGGAACA
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
GGAGACTGGAGGACCCATGGATCCCAACCATGGACCGGGTCCCCAGTATGAGCCATGAGCCAGTTGAATCTGAATCGGAAGATGGAACGAGGACCGA
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
CCAGAGTCACTGCTGACATCAACCCCATTAACATGGGACAGATCAAGAAAACCCACAGGAAGCTGAGAACTGCTGGAGCACAAGTTTTTACCATT
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
GCTGGGATTCAGAGTACAATAGATGCTTAACGGACCAATGCTTTCTGACTCAGCTTCTCTACCTCTGAATAACAAGAGACCCCTAATAGTTAGGACGAA
6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
TATCATCGCCCCATTTCAGCATGAAGAAGTTACAGAAGACGGACCTTATCTTCTGCAACCCCTAGGATTAAGGGTCTCTTTGAAAACGGGAAAGGGG
6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
AGATATGTTGGGAAGCATTCAAACAGAGGACCTCCATTTTGAATAAGGGCTAAGAAAATGAAGCTGGATCACCAACCGGCAATTAAGGGCTGCACAGCC
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700
TGCAATTCCTTCTCAATTAATAAGAGGCCACCTTATGCTAGTAATATGATAGCTAGTAATAATGATGTTGCTTCTTTTACAAAAAGAGAGGGG PPT

6710 LTR 6720 6730 6740 6750 6760 6770 6780 6790 6800
GGCATGGGAGAAAAGCTGAGTGTGGGAAAAGCTGAGGAGGGCTTGCATGCTGACATAATGAAAAGAGTCTTGAACATGTCCGGGTTCCAGGGTCT
6810 6820 6830 6840 6850 6860 6870 6880 6890 6900
AAAACCCCTCGTGGCCTTTGGAAACCAAGCTCTGTGCTAAAGGGTGAAGGCTACCCTGACGCCACATAATCTAAGCCAGGGCATAAAAACCCCTCGT
6910 6920 6930 6940 6950 6960 6970 6980 6990 7000
GCTTGGATAGAATCCAGGGCTCGTGGCCCTGGAATGTGCTAGACTGTGCTGCTTCTGCTCCTTCTCCAGGATCGATTGTATCTTTAGTTAAAA
7010 7020 7030 7040 7050 7060 7070 7080 7090 7100
GAACCTGCTCTCATTATCTCAAGTAGCAGAACATGTTCATATGCTCAAAGGAAATGCTAAAACCATCACAGCTGTAGATCATCGCTTAAATGCAACTT
7110 7120 7130 7140 7150 7160 7170 7180 7190 7200
GCCCTTCGACCCACATTTCTACCACCTGTTCTTTTGTGATCACAATAAATAGCTGGGCTTCCAGAGCTCGGGGCCCTTTGAGCCTCATACTT
7210 7220 7230 7240 7250 7260 7270 7280 7290 7300
AGCGTGGCCCCCTGACCCACTTTCTCTCAAATGCTTTTTCTATTTCTTTGACTCCCGCGGACTTCGTCACCCAGACCTGGTGGTGGTCTG
7310
ATCACCCCAACA

pol

LTR14B

RT

RNaseH

Integrase

HERV-K14I, 3' portion, variant with putative *env* gene

5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
 TTATTAACATTAATTTTTTAAATTTACAAAAAGATCAACCCATGACTGCAGCTGAACAACACCTGACAGGACAAAAGAAAATAAAAGGGTGGACAAG
 L L T L N F L N L Q K D Q P M T A A E Q H L T G Q K E N K K A G Q>

5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
 ATATATGGTGGAGGATGCACATACAAGAGCTGGGAAAAGGAAAGACAATATATGGGGAAGAGGATTTGCTTGTCTCTCCAGGTGACAATCAGGT
 D I W W R D A H T K S W E K G K T I I W G R G F A C V S P G D N Q V>

5710 5720 5730 5740 5750 # 5760 5770 5780 5790 5800
 GCCTGTGGGTGCCCAACATCTGAAGATCTATCATGAGCCACAGCAGGAAGAGGACTCTGGGAAGAGCCAGAACTCCCGATACGAGTGTGGC
 P V W V P T K H L K I Y H E P Q Q E E R T L G R A R T P D T S D G>
 * R S I M S H S R K R G L W E E P E L P I R V M A>

5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
 ACAAATGAACATCTCAGAGACAAGGAAAAGCCGAGAATATCATCAGGCAGATCCTCCAACATGGGGACAAATCAAGAACTGGGCACAGATGGCAGAG
 T N E H L R D K G K D R E Y S S G R S S N M G T N Q E T G T D G R>
 Q M N I S E T K E K T E N T H Q A D P P T W G Q I K K L A Q M A E>

5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
 GACAACCTGAGAGCACAACAACCAAAAACAACTAGTAACCTAATGGTGGCCATGCTGGCAGTACTCACTGTGGCAGTAAGCCCTCCCTACTATAGGAG
 G Q P E S T E Q T K N N *
 D N L R A Q N K P K T T S N L M V A M L A V L T V A V S L P T I G>

6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
 CAACTCAAATTTACTTTATTTGGGCATATGTCCTTCTCTTAAATTAGGTCTGTGAGTTGGATGGACCCCTGTTATTGAGGTGTACACCAATGACAG
 A T Q N F T Y W A Y V P F P P L I R S V S W M D P V I E V Y T N D S>

6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
 TACCTGGATGCCTGAGCCATAGATAACAGAGGGCAATGCATCCTAATGAGGAAGGGATGAAAATGAATATATCAATAGGATATAAATATCTCCCAATA
 T W M P E P I D N R G P M H P N E E G M K M N I S I G Y K Y P P I>

6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
 TGCCTGGGACCTGCTGGATGCTTACAAATGGCACACAAGTCTGGTGGCAGTAATCCCTGGAAAAATAAATCACAGGCTGCTTACATATGATTT
 C L G P A A G C L Q I G T Q V W L A V I P G K N K S Q A A L H M I>

6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
 CAGGCAAAAGTTAAAAATAAACCTTCCAGCCCTAAAACCTCAGCAGTTCTATCCAAAACAGCTCAAATGTAACAAAATAGAGTTTGGTTGAAGGTGT
 S G Q S L K Y N H S S P K T Q Q F Y P N K L K C K Q N R V W F E G V>

6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
 GGATGCTTGAATTTGGGAAGATTTGATAGCAAGTAAGGCTGAGGTGCTACAAAATAATTCCTATGGAATCGTCATTGATTTGGTCCCTAAGGGGATTTTT
 D A W N W E D C I A S K A E V L Q N N S Y G I V I D W S P K G I F>

6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
 AAAAAAATGCACCAGAAGGCCTTCTGTAGAACAATAATAAACCGATGGAATCGTTGGCAACAGAATCATAACAATATATTCAGACAGAAGCTG
 K K N C T R R P S C R T N I I N R W N R W Q Q N H T Q Y I Q T E A>

6610 6620 6630 6640 6650 6660 6670 6680 6690 6700
 ATGTCCTATCATATGGAGTCTGTGGCATTGTCACCCCTAGTCCAAAATGATATCTCCTGCCATAGGACAGGAACATTCAGAATTATGGAATTAGC
 D V P I I W S P G G I V T P S P K M I S P A I G Q E H S E L W K L A>

6710 6720 6730 6740 6750 6760 6770 6780 6790 6800
 TATGGCTCAAAGTTCAATCAAATTTGGGAGGTAATAATAAGTATAAAGGGGAAGGTAAGATAAATATGTTCTTTCTTTCTTAAACAGG
 M A Q S S I K I W E G K Y N K Y K G E G G K D K Y V L S F L S N R>

6810 6820 6830 6840 6850 6860 6870 6880 6890 6900
 ACCTTTTGGATTCAAAGTTGTGTCCAGCCACCTTTTATGTTAGCAATAGGAAATATTAATCTTGCATATAAATCCCATTTTATCACCTGCCACGAAATGTC
 T F W I Q S C V Q P P F M L A I G N I T L D I N T H F I T C H E C>

6910 6920 6930 6940 6950 6960 6970 6980 6990 7000
 ACTTGTACTTGCATTAACCTCAACTTTTGATAAAAATCAAACCTATCTTGTAAATAGAGCCAGAGAAGGAGTTGGATCCCTGTGTCTCTAAATAGACC
 H L F T C I N S T F D K N Q T I L L I R A R E G V W I P V S L N R P>

7010 7020 7030 7040 7050 7060 7070 7080 7090 7100
 ATGGGAGGCATCACCATTCAATTTATAAAGTGAATCCTTAAAAAATTTTATCCCATCAAAGAAATTTATAGTTGCTCTTATATTTGCTATAATT
 W E A S P S I H I I T E I L K K L L S H S K R F I V A L I F A I I>

7110 7120 7130 7140 7150 7160 7170 7180 7190 7200
 GGCCTCATTGCTGTACCCTACTGCTGAGTAGTGGTGTGGCTTTACACTCATCTGTGCAAACTGTTAAATTTGTTGATAAATGGCAAAAGAAATCTTA
 G L I A V T T T A A V A G V A L H S S V Q T V K F V D K W Q K N S>

7210 7220 7230 7240 7250 7260 7270 7280 7290 7300
 CGAAATTTGAACTCTCAGGTTCAAACAGATCAAAAAATAGTTAATCAAATTAATGATCTCCGTGAGACAGTAATTTGGATGGGGGATCGTATCATAAG
T K L W N S Q V Q T D Q K I V N O I N D L R Q T V I W M G D R I I S>

7310 7320 7330 7340 7350 7360 7370 7380 7390 7400
 TTTAGAAACTAGAATTCAAATGCAATGTGATTTGGAATACATCTGACTTTTGCATTACTCCTCATAGTTATAATGAAACAGAACCAATGGGAAAAAGTT
L E T R I Q M Q C D W N T S D F C I T P H S Y N E T E H Q W E K V>

7410 7420 7430 7440 7450 7460 7470 7480 7490 7500
 AAATGCCATTTAGAAGGTAAGAGGAAATCTCTCCCTTGATATTGTAACCTGAAAGAGCAGGTTTGGAGCCCTCAGGCTCACTTAACCTGCTCT
K C H L E G K E E N L S L D I V K L K E Q V F E A S Q A H L T L L>

7510 7520 7530 7540 7550 7560 7570 7580 7590 7600
 CTGGAATGATATTTAAGCAAGCAGCTGATGGGTTGTCTGCAATCAATCTCTTAAATGGAATTAAGACCATTGGAAACGCTACACCTGTAACCTTGT
S G T D I L S K A A D G L S A I N P L K W I K T I G N A T P V N L V>

7610 7620 7630 7640 7650 7660 7670 7680 7690 7700
 TCTAATAATTATGTCTTGTGCTGCTCTGTTTCAGTCTACAGATGCAGAGCCACCTCTGGAGAGAAAGCCACCGCGAGAACAAGCAATGATAGCTGTG
 L I I M C L C C L V S V Y R C R S H L W R E S H R R E Q A M I A V>

7710 7720 7730 **LTR** 7740 7750 7760 7770 7780 7790 7800
 GCAGTTTTCGCAACAAAAAGAGAGACATGTTGGTAGAAGAGCTGAGGCAGGACTGGCTTGTCTGTCTAATGTAATAAGAGTCTTGAAGATGTCCTGG
 A V L Q N K K E R H V G R R A E A G L A C L S *

7810 7820 7830 7840 7850 7860 7870 7880 7890 7900
 GTCCAGGGTCAAACCCCTCGTGGCTTTGGAAACCAAGCTCTGTGCCAAAGGTTGGAAGGCTGCCCTGCCGACCAACATCTAAGCCAGGGCATAA

pol

env

LTR14Bv PPT

GP36

LTR14Bv	7910	7920	7930	7940	7950	7960	7970	7980	7990	8000	
	AACCCCTTGTGGCTTGGATGGAACCCAGGGCTCAGGGCATAAAACCCCTTGTAGCCTCTGGAATGTGTCAGACTTGCTGGCTCCTTGCTTCTTGCTCTC										
	8010	8020	8030	8040	8050	8060	8070	8080	8090	8100	
	CCAGGCTGTAAACATAATCTATCTTAACTCAAGCAACAGAGCATATCTATATCAAGTAGCAGAGCATGTTCCATATGCATCAAAGGAAATGCTAAACC										
	8110	8120	8130	8140	8150	8160	8170	8180	8190	8200	
	ATCACAGCTACGCTTGATGCACCACTACCTTTCTACCCCCACATCCTCAGCCCTCACCTGTTTACCCCCACATCTACATGTCCCTCACCACTGCTTCTT										
	8210	8220	8230	8240	8250	8260	8270	8280	8290	8300	
	TGTTTGATCACCAATAAATAGTGTGGGCTCCAGAGCTCGGGGCCTTCACAGCCTCCATACTAGCGTTGGCCCCCTGGACCCACTTTATGCACTCTTAAC										
	8310	8320	8330	8340	8350	8360	8370				
	TTGTCCTTTCTCATTCTTTGACTCTGCCGGACTTCGTAGCCCCATGGCCTGGTGTGGGTCTGATCACCCAACA										

Suppl Fig. E: Annotated HERV-K14I consensus sequences generated in this study. The first sequence is the HERV-K14I having about 800 bp of *env*-unrelated sequence downstream from the *pol* gene. The proviral body is flanked by LTR14B sequences. The second part of the figure presents the 3' portion of a consensus sequence for HERV-K14I proviruses with a putative *env* gene downstream from *pol*. The 3' LTR is represented by LTR14Bv, that is associated with *env*-having HERV-K14I proviruses (see text). The start of the different *pol-env* gene region is indicated (#). Locations of proviral open reading frames for *gag*, *pro*, *pol* and *env* genes, primer binding site (PBS) and polypurine tract (PPT) are indicated on the left. Protein domains, as revealed by CD-search, are indicated on the right and are underlined in the protein sequences. The presented sequence (HERV-K14I without *env*) has been deposited in Rebase.

LTR14C

PBS

gag

pro

10 20 30 40 50 60 70 80 90 100
TGTGGGAAAGAGATTTCTGGGGTGCAGTTGAGTTGGTCTCCCTGTGTGAGACACCCATGGGAAGCCATGGGCGCCTCTGAGGAGAAAAGTCTCCTT

110 120 130 140 150 160 170 180 190 200
ATTGCTTTCATGCTTTTATGCCCGGAGAGCATAACCGCTCAGCGGATTTCCACAGGTTGCTCAGGGAGATAACACTCCCTTGAAGCAGTGGAGTATAATC

210 220 230 240 250 260 270 280 290 300
AAACATCTTGGTCTCTGAAACCCACTCCACCCGTTTCAGTCCCGATAAGTTAAAGATCTTAAGTAGTTTAGACACACGCTTTGCTCAAGGAAATT

310 320 330 340 350 360 370 380 390 400
CACAGAAACCGCCACTGCTACACATCTTATCGAATGACTCACGAGTTCTCCTTACATGATTAATCCCTTTCTCATCCCTTCTCCCTCCCATCTGCC

410 420 430 440 450 460 470 480 490 500
CTAAGAACAAGAGCTTGTAAACCAATAAATTGGTGGAGCCAAAGAGCTCTGGCCGCTGAGCAAGCTCCGATGCTCTGGTCCCTGGACCCGCTTTT

510 520 530 540 550 560 570 580 590 600
AAACGCTTATTCTGTCTCTTTCTAACTCCTTTGTCTCCGCCGACTCGGGGTACCCACTGGGTGGTGTGGGGCTGGTTTCCCAACATCTGGTCCCCAAC

610 620 630 640 650 660 670 680 690 700
CGCGGGCTCCCTATAATCTCTACAATAATCCAGTGAAGAAATGCCAGAGCTGGAAGAGTGGAGGACGACTGACAAAGACCGCCGAGTCTTTTCACT

710 720 730 740 750 760 770 780 790 800
TCAAGCTCTACAGGTAAGTAGGGCACTCAGAGAATTCAGGGTAACCTCAGGAAAATATGGGTGAGCTGAAAGTAAGTTTGCCTAAATACCTTAAGCCCTGG

810 820 830 840 850 860 870 880 890 900
TGCAGCAGTTATGTCACAGAGGGGTAATTTGAGTACCCAAAATCTCACATCTTTGTTCCATCTCATAGAAAAGTATTTCTCCTTGGTTCCTCCGAAAT

910 920 930 940 950 960 970 980 990 1000
TGGAAACATGAATGTAAGAAATGGGACAGGTCGGACTCAGACTTAAACGAGCACAACAGAGGGCCACGATATCCCTTCTCCACTTGGTCTGTGTGG

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
TCAGCAATTAACAGCAGTGGAGCCCTCCACACTGAGGAGGAGGAGGAGAAGTTTCAGGATGACATAGAAAAGTTTAATAATCAGGAGTCTGATGATC

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
AGCAAAGTGAACCATCACAGTCTAGTTTAAAAAGGGGAGAAATGGGAAGCTATATATGCTAACCTCCAAAACTTATGAAAGAACAGTTCCACCTAC

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
TGCAGAAACAGTGCACCTACTGCGCCTTAGGGGAAGTCCGGAATGGCCACCCACCTCAGCCTTATGAATTTTGGAAAYGGGAGCCTGAGACGGGG

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
CTTGCCACTCCCATTTGTCACGCCCCACCATTAATATGGCAAAGGGAAGCTTCAGGCTTGCACCAACAGCCAATACAGTGGAGGATGATCCAGGCT

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
GTCCACCTGTTAATATGTTAGAGGAATGCTGCAAGCCAGCCAAATACAAATATGGTGCAGGACAATCCAGGCTATCCATTTGCCAGGCACAGAAAT

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
GGGGATTTGGATGCTTGGCAGTTTCTGGTAATATTTCTCCAGCTGAGGAGCCAGAGAACATGCTCAAGCATGCTGGGAGCCATTTCCTTTTAAAAATA

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
TTAAAGACTTAAAGCAAGCAATTTGGACAATATGGGCCAAATTTCTCTTATGTTTCAATCTCTGTTTACAATCTGTGGCTTATAACCGGCATTAATACCTA

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
TGGATTTGGGAGTCATTTAGCCCGATCCACCTGTCCCTCTCAATTTCTCAATTTTAAACCTGGTGGACAGATGAAGCAACAAATCAGGCATGCAGAAA

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
TGCTCAAGCCCACTCCCATTAATATCACATCTGATCAATTTGGAATCGGACAGGATGGGGTACTGTAATCAACAGATGGTAAATGGGTGATGAG

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
GCTGTTGATCAGCTCAGAATATATGCTTAAAGCCCTGGGAAAAATTCATGACCTGGTACTACTTATCTCTTTTAACTCAGTTCCAGAGGGTCCAA

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
RGGAGCCTTATCCAGATTTTATCGCCATTTGCAAGACGCGGGCTCAAAGGCTATTTTGGATTCTCATGCCAGGAAAGTATGATCATTAGCTGCTTGCTTA

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
TGAAATGCTAATACAGAATGTCAGGCAGCAATTTAGACCTATTAAGGGAAGGAGCAGATCTAAATGAGGAAAAAACTTTAAGTGAATACATTAAGCCCTGT

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
GATGGCATTGGGGGGCATTATATAAGGCCAGCCTCCTTGTCTCAGCAATTTGGCTGGACTAAGGTAACAAAAACACACAGAGTGTCTCCCTGGATCTTCT

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
ATAAATGTTGGACAGATAGGACATACAAAAGAGAGTGTACAAAAGAGCCAAAAGGCAAAACTCAGGAGGAAAAAGCAGGGAACAGGTACCTTTCTCTAG

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
ATGTAATAAAGGAAAAACACTGGGCTAATCAATGCTATTCAAAGTTTGAATAACAGCGGACAGCCCTTCCCGGAAACGGACAGAGGGGGCCAGCCCCAGGCC

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
CTGATTCAAAATGGGGCATTCCCAATTCAGGATGGGACTCCCTGACTCCAAACGGAGTGTCCAGCCAGTCTATCCCTGTACAATGTACAGCAAT

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
GTCCCTCCACAGCTAAAAGTGGGGCAGTAGATTTATGCTGTACAAAAGCTGTATCCCTCCTTCTGGGAGCCCTCTAGGAAGGTCCCAATGGGAGTT

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
TACGCCCATTTGCCAAATGGCAGGCTGGGACTTATACTGGGAAGTTCAGCTTAAACTTAAAGGGAATTCAGTACATCTGGAGTGTGGACTCTGTGAT

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
GCCAGGAGAAATCAAATTTGTTATCTCTCCACTGTTCCCTGGAGTGTCTAAATCCAGGTGACAGAATAGCTCAACTGTTGCTTTTACCATATGTTAAGTT

2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
AGGAGAAAGCTCAGAAAAAGAACAGGAGATTTGGAAGCACAATTCAGCAGGCAAGGCTGCCTATTTGGGTAATCAAGTCTCTGACAATAGACCTATT

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
TGTATGGTCACTATTCAAGAAAAACAATTTGAGGGTCTGGTGCACAGGAGCAGATGTGTCGATCATAGCTCTTTATCAATGGCCAAAACACTGGCCCA

P10

p24

ZnF

dUTPase

Protease

3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
AACAAAAGGCCAGTGGGTCTTGTGGGGTCCAGAACTGCTTCAGAAGTTCCTTCAAGAGTACTTTTATCTTACCATGCTCTGGTCCAGAAGAACAGGAAGG
K Q K A P V G L V G V R T A S E V F Q S T F I L P C L G P E E Q E G >
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
CACATAAACTGTAAATATACCCATTCCCTGTTAACTTATGGGGGAGAGATCTGCTTCAGCAATGGGGCACAGAATAATTCATCCCTTCCCCGGTAT
T I K P V I I P I P V N L W G R D L L Q Q W G T E I S I P S P R Y >
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
GGTCAAGCTAGTCAAAAATAATGTCAAACATGGGCTATGTTCCAGAAAAGGCCTAGGAAAACAGGACGGGCATTTGAACCCATACAGGTTACTG
G Q A S Q K I M S N M G Y V P R K G L G K Q E T G I I E P I Q V T >
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
TAAAAATGACCGAAAAGTATAGTTACCATTTTTAGGGGTGGTCACTGTTGAGCCTCAAGTCTTCCCTTAAATGGAAGACCCAAAATCCTGTT
V K N D R K G L G Y H F *
* P K R I R L P F L G V V T V E P P S P I P L K W K T Q N P V >
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
TGGTCAAGCAGTGGCCACTTCTCAGAAAAATTTGGGGCCCTTACAGGAATAGTCAAGAGCAATTAACAAGGAACATTTGAGCCACATTTTCTC
W V K Q W P L S Q E K L G A L Q E L V K E Q L N K G N I E P T F S >
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
CATGAATTCCGCGAGTGTGTTAATAAAGAAAAATCCGGCAGATGGCCGATGCTAACCGACTTACGAGCGGTCAATGCAGTCATCCAGCAATGGGGGC
P W N S P V F V I K K K S G R W R M L T D L R A V N A V I Q P M G A >
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
CCTGCAGCAGGGCTTCATCCCCACCATGATTCCTAGAGACTGGCCATTAATAATATAGATTGAAGGATGCTTTTTAATATCCCTTAGCAGAG
L Q P G L P S P T M I P R D W P L I I I D L K D C F F N I P L A E >
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
TCTGATTTGAGAAATTTGCTTTTACTATTCCTGCTATGAACAACAAGGAACAGCAGCAGATATCATTGGAAAGTCCGACAGGTTATGTTGAATA
S D F E K F A F T I P A M N N K E P A A R Y H W K V L P Q G M L N >
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
GTCCTACTATTGTCAAACCTTTTGTGGGAAGGCTTCAACCTGTGAGAGATCAGTTCCAGATTGCTATATCATTATATATGACATATTTGTG
S P T I C Q T F V G K A I Q P V R D Q F P D S Y I I H Y M D D I L C >
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
TGCGCCGAAAATCGAGCAACTTATCCAGTGTATTTCATATTTACAGGAGGTGGTCCCAATGCTGGATTGCTCATAGCACCAGATAAAAATTCAAACG
A A E N R D Q L I Q C Y S Y L Q E V V A N A G L L A I A P D K I Q T >
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
GCCACTCTTTCCAAATATTTGGGAATGCAGGTTCCAGGAAAGGGCAATTAACCCCAAAAGGTTCAAATTCGAAAAGACTCTCGAAAACCTTTAAATGATT
A T P F Q Y L G M Q V Q E R A I K P Q K V Q I R K D S L K T L N D >
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
TTCAAAAATTTAGGGATATCAATTTGGATTTCGACCTACTTTGGGAATCCCTACCTATGCTATGCTAACTGTTCTTATTTAAGAGGAGACCCCTGC
F Q K L L G D I N W I R P T L G I P T Y A M S N L F S I L R G D P A >
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
TCTCAATAGTAAACGAGCACTGACTCTGAGGCTGACAAAAGAAATTAACAATGATTGAAGAAAAATACAACAGGCCAGGTTAATAGATAAGACTCAAGT
L N S K R E L T P E A D K E L Q M I E E K I Q Q A Q V N R I D S S >
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
TTACCATCAGTTCATTTGTTCCCTACTCTCCTACTCTACAGGGTATAGTTCAGAGTGGAGACTAGTTGAATGGTCTTTTTCGCTCACATA
L P L Q F I V F P T L H S P T G V I V Q S E D L V E W S F L P H N >
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
CTGTTAGAACACTCACAGTATACTTGGATCAGATGGCAATCTTGATTGGGCAAGCTACCTTAGAGTGTGTTAACTTTGTGGCTCAGATCCAGATAAGAT
T V R T L T V Y L D Q M A I L I G Q A H L R V V K L C G S D P D K I >
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
TATAGTTCGAATGAATAAAATCAGATTCGGCAGCCCTTTGTTAATTTGGTCAATTTGGCAAATAAATTTAGTGGCTTCATTTGAGTCTTTGACAATCAC
I V P M N K N Q I R Q A F V N L V N W Q I N L A G F I G V L D N H >
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
TATCCAAAACAATTTTTCAGTTTTAAAGCTAACAACATGGGCTTCCTCCAAAATACCATTGGTCTCCATTGGAAGGAGCAGTGTGTTTAA
Y P K N N F F Q F L K L T T W V L P K I T H C A P L E G A V T V F >
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
CAGATGGCTCTAGCCATGGAAAAGCAGCTTATGTGGGCACTAAAACAGAAATTTCAAACACTGACTTTCAAATCGCCACAGAGGGCTGAATACAGGCAGT
T D G S S H G K A A Y V G P K N R I I Q T D F Q S A Q R A E L Q A V >
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
TATAGCTGTGTTAGAAGCTTTAAGCAACCTGTAAATATTTGCTCTGCTTACAGCTTATGTTGTTCAAGCCACTCAATAGAACTGCACTCAATAAA
I A V L E D F K Q P V N I V S D S A Y V V Q A T Q Y I E T A L I K >
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
TATCTTTGAGGATGAACAACCTATCATCTTTTCTTTTACAAAAGCAGTGGCTGCTACATATTTTCTTTCTATATCATGCACATTCGAGCATATA
Y L V D E Q L Y Q L F S S L Q K A V R D H Y F P F Y I M H I R A Y >
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
CTAATCTCCCTGGGACCTTGTAAAGGCTAACGATCAAGCTGATTTACTAGTTTCCACTGTGCTTACTAATGCCCCAAGATTTTCACTCCCTAACCACTGT
T N L P G T V V R A N D Q A D L L V S T V L T N A Q D F H S L T H V >
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
TAATGCAGCAGGACTTAAACAAAAATATCAAAATTAATGAGACAGGCAAAAGACATTTGTGCAACATTCGCCCTCAGTCCAGTACCAACTGCCACAT
N A A G L K Q K Y Q I T W R Q A K D I V Q H C P Q C Q V L Q L P H >
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
GAAGGAGCTGGTGTAAACCCATGGGGATTAACCCCAATATGTTGTGGCAAATGGACGTAACCCATGTACCTTCACTGGGAAACTTTCATACGTCACATG
E G T G V N P W G L T P N M L W Q M D V T H V P S F G K L S Y V H >
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
TTACTATAGATACCTTTTCTATTTTGTGGCAACTTGCACAAACAGGTTGAGGACGCTGCTCATATTAAGACATTTACTTTCTCTGTTTTCGCTGCCAT
V T I D T F S H F V W A T C Q T G E A A A H I K R H L L S C F A A M >
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
GGGCATCCCAACAAAAGATTAACACAGCAATGGCCCAAGCTGATGTAATAACTTTTACAAAGCTTTCACAAAGCTTTCCTCAACAATGGCAGTACAGTACGGA
G I P Q K I K T D N G P G Y C T S K S L Q A F L Q Q W H I E H S T G >
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
ATACCCATAAATTTCTCAAGGCCAAGCCATTTGTTGAACAGGCTAATCGAAACCTTAAATCTCAATTAACAAAACAAAAGACAGAGGGGGAAACACAGAAAT
I P Y N S Q G Q A I V E Q A N R T L K S Q L Q K Q K T E G G T R E >
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
ACTCTACTCCCATATGAGCTACAATTTGGCTCTTATCACTTTAAATTTTTTGAATTTGCTTAGAGATCAGGTTACACAGCCAGCAGCAGCATTGAC
Y S T P H M Q L Q L A L I T L N F L N L S R D Q V T T A A E Q H L T >
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
AGGGCAAAAATAAAATCCCTCATGAAGGAAAACATGTTGTTGGAAGGACGTCAGAAAACCTGGGAAAAGGGCAAAAATCATACATGGGGTGGGGGG
G Q K I N P H E G K H V W W K D V S R T K T W E K G K I I T W G V R G >
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
TTTGGCTTGTATCTCAGCAGGAGAAATCAGCTTCTGGTGTGACCAAGACATCTTAAAGCTGTGCCATGAGCCAGAATCCAAGGAAGGAAAGAAAGA
L L V S H C Q E R I S F L S G Y P Q D I L S C A M S Q T N P R K R K R >
F A C I S P G E N Q L P V W V P T R H L K L C H E P E S K E E E K >

pro

pol

env

Protease

RT

RNaseH

Integrase

6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
 CCTCGGAACGTCCTGCACCCAGTTCATCAGATGAGTGAACATCTCTGTTGAGCAGATGGAAACCAGTAAACCTCACCAAGCACTCCACCGA
 P R N V P A P P V H Q M A Q M N I S V V E Q M E T S K T H Q A T P P >
 T S E R P C T P S S S D G S D E H L C *

6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
 CCTGGGGGAGATGAAGGACTAGCTCACATTGCAGAAGAGAACCTGAGTCTCAGAACAGCTGCCTGACCACCAAGTAACTAATGGTAGCTATGATGGT
 T W G Q M K R L A H I A E N L R S Q N K L L T T S N L M V A M M P >

6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
 GGTAATCTCCTGGTGGTGAAGTCTCCCCGAGTGGAGCAGATCAAAATACACTTATGGGCCTACATTCCATCCACCACCTGATAGCCCTGTTACA
 V I S L V V S L P V A E A D Q N Y T Y W A Y I P F P P L I R P V T >

6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
 TGGTTAGACCCCGAGTGGAGTTTATGTTAATGATAGTGTCTGGATGGCTGGACCAACAGATAACCGAGGTCTACTCATCCAGAGGAGGAAGGAATGT
 W L D P P V E V Y V N D S V W M P G P T D N R G P T H P E E E G M >

6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
 TAATGAATGTTTCCATGGTATCGCTTTCCTCCCATCGCTGGGCCAGCAGCAGGATGTTTAAATATGATAAAACAAAGTTGGATGGTATGATGCC
 L M N V S I G Y R F P P I C L G P A C A G C L N Y D K Q S W M V Y V M >

6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
 TGCATAAATGGATCAAAAGCCTTATTCATGCAATCAGTGAAGAACAATTTCAATCTTTGGACACTATTAAATACCTTGAGCATGGCTATGTTATGACA
 A H N G S A V A I H A I S G R T F Q S L D T I K Y L E H G I Y V M T >

6610 6620 6630 6640 6650 6660 6670 6680 6690 6700
 CATCGCCAGATTAATAAATTTAAACCTAATAAGAAGCCCTGCCCTAGCAGGCCATAAATGGTCTGAAAAGCTAGAGGTGCTAACCTGGGAAGATTGTA
 H R Q I N K F K P N K K P C P R Q A T K W S E K L E V L T W E D C >

6710 6720 6730 6740 6750 6760 6770 6780 6790 6800
 TTGCAACAGCGCTGCTGACTGCAAAATAATTCCTATGGAATCATCTGATGGGCCCTTAGGGGACACTTTGCAGTAAATTTGACTGGACAGAGCAA
 I A N S A A V L Q N N S Y G I I I D W A P R G H F A V N C T G Q S K >

6810 6820 6830 6840 6850 6860 6870 6880 6890 6900
 AGATTGTAGAGACTCCTTTGCAATGACTACCCAGATAATGCACAAAATTTATAGAGAATGAAAACAAATTACCCTATTAAGTGGGAGGAGAAT
 D C R E T P F A N D Y P D N A P K L Y R R I E T N Y P I K W E E M >

6910 6920 6930 6940 6950 6960 6970 6980 6990 7000
 GGATGGCTCCTCCAAGCCCAAAATGATGATCCAATTATAAGTCCAGAACATCCAGAAATTTGTGGAATTAATGATGGCTCAAACCCCAATTCGGATT
 G M A P P S P K M I D P I I S P E H P E L W K L M M A Q T P I R I >

7010 7020 7030 7040 7050 7060 7070 7080 7090 7100
 GGAAGGAGAAATATAAACAGAGCCCATAGTAAAAAATTCGATTTGTTGAGCATTGACCTTAATCAGACGGTCCCATGACAGATGTTGTTAAACC
 W K G E Y K T E T H S K K L R F V A M T S N Q T V P Q S C V K P >

7110 7120 7130 7140 7150 7160 7170 7180 7190 7200
 TCCTTTATGTTGGCAGTGGGAAAAATTAATATCTACCTGACTCTCAAAACCATATGCTCAACTGTCTATCTTTTACCTGCATTAATTTACTCTTT
 P F M L A V L I N I L P D S Q T I S C L N C H L F T C I N S T >

7210 7220 7230 7240 7250 7260 7270 7280 7290 7300
 AATAAGGATAATAGCAATTTACTGGTTAGGGCCGAGAGGAGTTTGGATACCTGTTTCCCTCAATAGACCTTTGGGAGGCCTCTCCCTCCATACATATTA
 N K D N S I L L V R A R E G V W I P V S L N R P W E A Q T P S I H I >

7310 7320 7330 7340 7350 7360 7370 7380 7390 7400
 I T C T E V L K G I L N R S K R F I F T L I A V I M G L I A V T A T A >

7410 7420 7430 7440 7450 7460 7470 7480 7490 7500
 TGCTGTGGTGTCTTGGCACTCTTATCAAACTGCGGGCTTTGGATAGTTGGCAGAAAAATTTCTTAAAGCTTTGGAATTTCCCAAAGCCAAATA
 A A G V A L L H S I Q T A G F V D S W Q K N S S I K I >

7510 7520 7530 7540 7550 7560 7570 7580 7590 7600
 GATCAAAAATGGCAAATCAAATTAATGATCTCCGTCAAACAGTAATTTGGATGGGAGATCGGATTAAGAGCTTGGAGCATAGAATCAAATGCAATGTG
 D Q K L A N Q I T P S Y I W M G D R I M S L E H R I Q K C >

7610 7620 7630 7640 7650 7660 7670 7680 7690 7700
 ATTGGAATCTCTGATTTTGTATTACTCTAGCTCTTATAATGCCACTGAACACCCATGGGAGATGATTAGACATCACCTACAGGAAAGAAAGATAA
 D W N T S D I T P S Y I T E H H W E M I R H I Q K E A D N >

7710 7720 7730 7740 7750 7760 7770 7780 7790 7800
 TTTAACAATTAGATATGCTAACTGAAAAACAACCTTTTGGAGCATCTCAGGCCTCACTCACCTGTTGCCTGGAGCTGATATTTCTGGAGCCACT
 L T L D I A K L K K Q L F E A S Q A H L T L L P G A D I L A G A T >

7810 7820 7830 7840 7850 7860 7870 7880 7890 7900
 GATGGCCTTCTAATACCAATCTTAAAGTGGATTAAACCATGATGGATCAACAATTTGCAAAATTTATTTTGGTGTGTCTGTTTATGCTGTTTGT
 D G L S N T N P L K W I K T I G G S T I A N F I L V C V C L >

7910 7920 7930 7940 7950 7960 7970 7980 7990 8000
 TTTTAGTCTACAGATCGAGACGGCACCTTGGGAGAGAAGCCAGACACCGTGAACGAGCATTGATAGCAATGGCGGTATTAATAAAAAAATAATAGAA
 F L V Y R C R R H L G E A R H R E R A M I A M A V I N K K K L I E >

8010 8020 8030 8040 8050 8060 8070 8080 8090 8100
 GACAAAAAAGGGGACATGTGGGAAAGAGAGTTTCTGGGTGCCAGTTGAGTTGGTCTCCCTGTGTGAGACACCCATGGGAAGCCATGGGGCCCTCTGT
 T K K G D M W E R E F L G C Q L S W S P L C E T P M G S H G R P L >

8110 8120 8130 8140 8150 8160 8170 8180 8190 8200
 AGGAGAAAAGTCTCCTTATTCCTTATGCTCTTTATGCCCCGAGAGCATAAACCGCTCAGCGCATTCACAGGTTGCTCAGGGAGATAACACTCCCTTGA
 R R K V S L L P S C L Y A P R A *

8210 8220 8230 8240 8250 8260 8270 8280 8290 8300
 AGCAGTGGAGTATAATCAAACATCTTGGCTCCTCTGAAACCCACTCCACCCGTTTCAGTCCCGATAAGTTAAAGATCTTAAAGTATTAGACACACGC

8310 8320 8330 8340 8350 8360 8370 8380 8390 8400
 CTTTGCCTCAAGGAAATCACAGAAACCGCCACTGCTACACATCTTATGGAATGACTCACGAGTTCTCCTTCACTGATTAATCTTTCTCCTCATCCCTTCC

8410 8420 8430 8440 8450 8460 8470 8480 8490 8500
 TCCCTTCCCATCTGCCTAAGAACAAGAGCTTGTAAACCAATAAATTTGGTGGAGCCAAAGAGCTCTGGGCCGTGAGCAAGCCCTCCGATGCTCTGGTC

8510 8520 8530 8540 8550 8560 8570 8580 8590 8600
 CCCGACCCGCTTTTAAACGCTTATCTGTCTTTCTAACTCCTTTGTCTCCGCGGACTCGGGGTACCCACTGGGTGGTGGGGTGGTTTCCCC
 AACA

pol
 env
 PPT
 LTR14C
 LTR
 GP36

Suppl Fig. F: Annotated HERV-K14CI consensus sequence generated in this study. The proviral body is flanked by LTR14C sequences. Locations of proviral open reading frames for *gag*, *pro*, *pol* and *env* genes, primer binding site (PBS) and polypurine tract (PPT) are indicated on the left. Protein domains, as revealed by CD-search, are indicated on the right and are underlined in the protein sequences. The presented sequence has been deposited in Repbase.