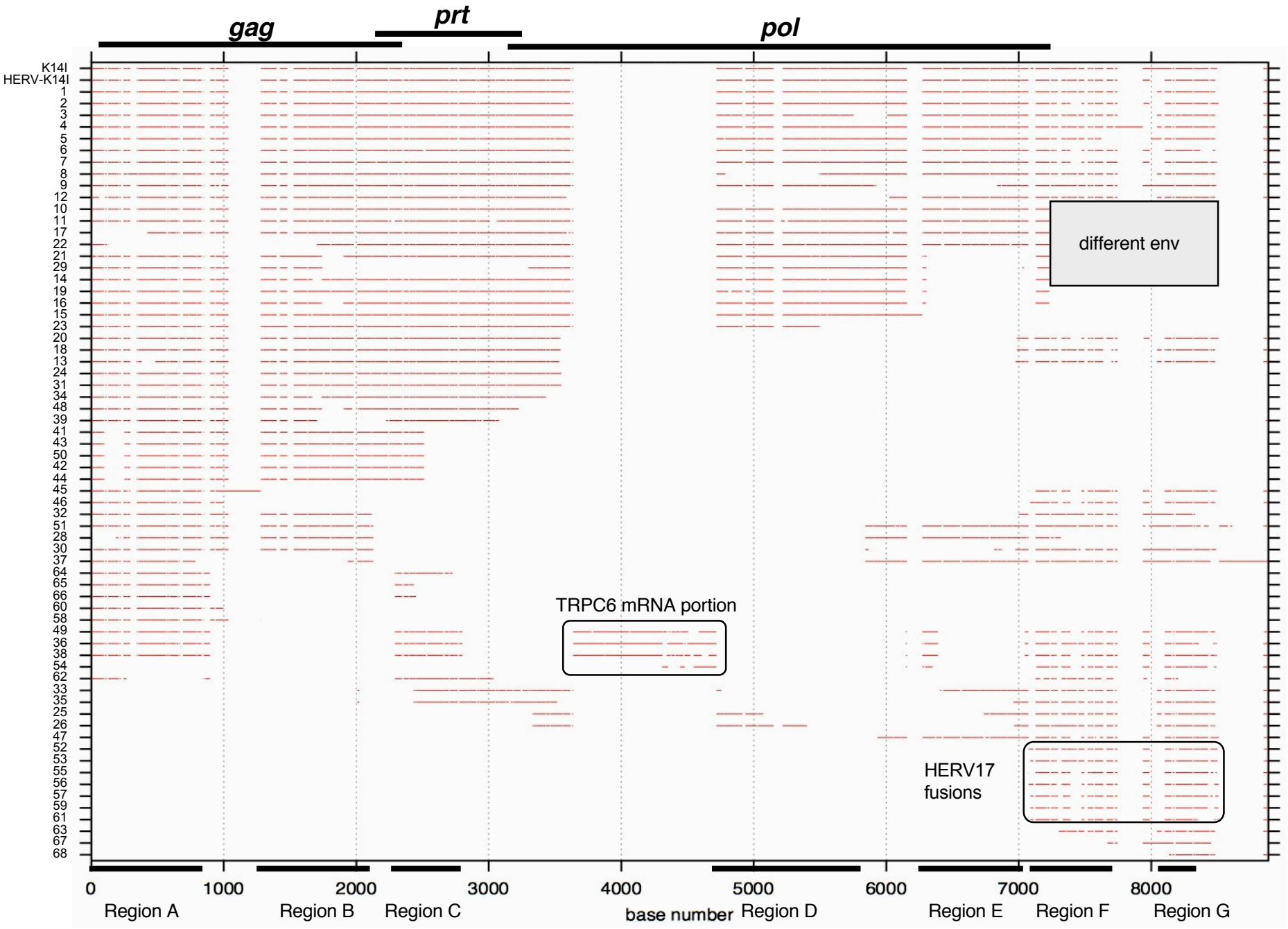
Supplementary Table A

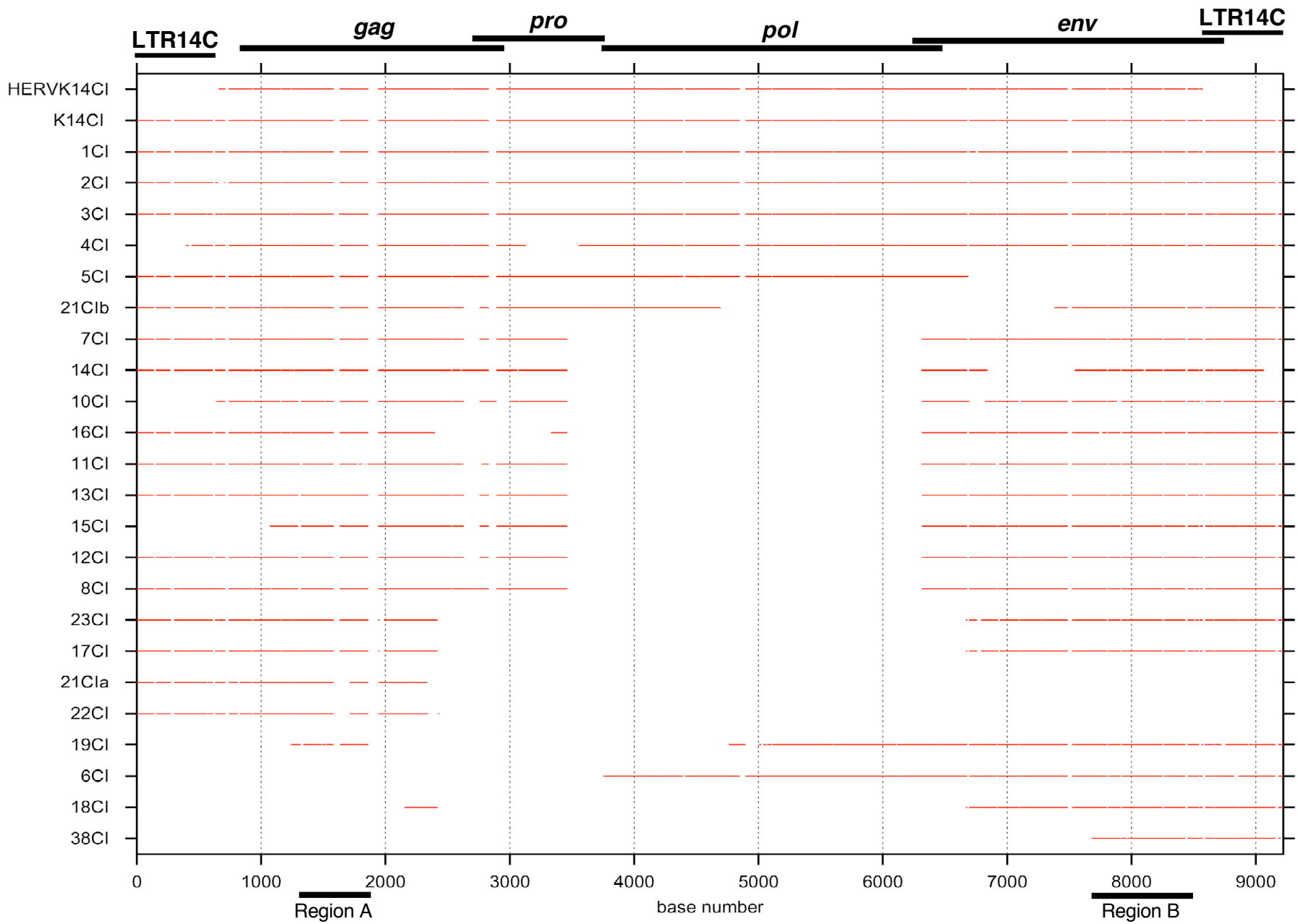
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
PV no	SCORE	START	END	IDENTITY	STRAND	POSITION	ACCESSION NO	5´LTR14A	3´LTR14A	5´LTR14B	3´LTR14B	LTR14Bv	LTR similarity	putat. env	HERV-W fusion	TRPC6 mRNA	Repeats in PV	PV in Repeats
1	5364	0	5945	95.6%	–	Xq22.1	AL109963.4/Z70280.1	+	+				91				Mer41A	AluSg
2	5166	0	5945	94.4%	–	2p11.2	AC096770.3			+	+		94				Mer 41C	Mer41A
3	4901	0	5869	94.4%	+	7q34	AC008155.10	+	+				93				AluY	
4	4895	435	5945	95.2%	–	Xq24	AL391358.14/AL391474.8	+	+				92				Mer41C, LTR12, LTR12C	L1MDa
5	4839	0	5682	94.7%	+	4q25	AC003100.1			+	+		86				Mer41C	L1
6	4747	0	5530	94.0%	–	Xq21.31	AL022154.2			+	+		90					L1M1
7	4637	0	5421	93.6%	–	4q22.1	AC108516.5/AC112250.4			+	+		92					
8	4485	177	5945	94.3%	–	17q25.1	AC087651.19	+	+				90				AluSp, 3x AluY, LTR12	
9	4091	10	5572	93.1%	+	15q24.2	AC027243.20		+	+	2x		91				AluY, Mer 41C	
10	3989	76	5142	89.8%	+	16p11.2	AC106730.2			+	+	+	85	+				
11	3886	0	5142	89.7%	–	11q13.4	AP003717.3			+	+	+	96	+			AluY	
12	3792	378	5945	94.3%	–	1p31.3	AL162739.24	+	+				89				Mer41C	
13	3323	0	5942	93.9%	–	22q12.2	AC003072.1	+	+				86				Mer41C, L2	
14	3274	80	4312	90.3%	+	1q32.1	AC119673.3/AL713965.10			+				+				L1PA5
15	3271	209	4281	90.9%	+	12q13.11	AC089987.26			+	+		89					
16	3253	0	4312	90.0%	+	14q32.32	AL161669.5			+								
17	3241	320	4554	90.4%	+	19q13.43	AC010326.7							+			2x AluY, AluYc	
18	3217	0	5945	92.5%	–	5p14.2	AC118456.2	+	+				89				Mer41C, LTR12	L1P
19	3209	178	4312	89.5%	+	10q22.1	AC073370.7			+							LTR13, AluY	
20	3199	0	5942	92.8%	+	6q14.1	AL450327.12			2x	+		88					
21	3049	179	4312	89.5%	–	1q42.12	AC099066.3			+	2x	+	91	+			4x AluY, HERVKC4, L1M4, SVA	
22	2733	1166	4845	88.1%	+	10q26.3	AL731769.4			+		+		+				
23	2684	27	3621	88.7%	–	19p13.11	AC078899.1			+							L1PA4	
24	2437	0	2862	93.2%	+	10p15.1	AL391427.9	+										
25	2341	1827	5945	94.2%	–	2q23.3	AC007394.4		+								Mer41A	
26	2316	1827	5934	93.4%	–	11q11	AC022882.5		+								Mer41A	
27	2266	2204	5142	89.0%	–	Yp11.2	/				+							
28	2220	188	5230	93.0%	+	15q22.2	AC126323.6										2x AluY	
29	2201	132	4306	90.5%	+	3q12.1	AC079065.8			+		+		+			AluY	
30	2096	59	5884	92.5%	+	10q22.3	AL356095.11		+								Mer41C	
31	2087	0	5678	92.5%	+	15q21.2	AL391427.9			+	+		89				Mer41C	
32	1978	36	5804	91.9%	–	19q13.33	AC008750.9			+							Mer41C-int, AluY	
33	1975	1827	5678	92.8%	+	Yp11.2	AC002992.1		+								2x LTR12, Mer41A	
34	1848	338	2777	89.9%	+	1q32.1	AC119673.3			+								
35	1737	1827	5945	93.9%	–	6q23.2	AL078590.28		+								Mer41A	
36	1715	0	5945	91.3%	+	11q23.3	AP001124.4			+	2x		91			+		L1MEd
37	1669	3975	5945	94.0%	+	4p13	AC098590.3	+	+				89				Mer41C	L1
38	1667	0	5945	90.7%	–	4q35.1	AC084871.4			2x	+		88			+	Merr41B, HERVK3	
39	1581	0	2434	90.8%	+	2q12.2	AC009505.3			+	+		92				HERVK14C	
40	1433	0	1896	90.2%	–	4p15.33	/	+										
41	1433	0	1896	90.2%	+	4_random	AL354822.8	+										
42	1424	0	1896	90.1%	–	17_random	AL354822.8/AC145212.1	+										
43	1410	0	1882	89.9%	+	1p11.2	AL354822.8/AC145212.1	+										
44	1369	0	1831	89.3%	+	1p11.2	AL354822.8/AC145212.1	+										
45	1359	24	5945	92.9%	+	3q26.31	AC108667.9			+	+		87				AluY, HERV17	
46	1349	0	5945	91.8%	–	11p15.1	AC103974.6			2x	2x		93					L1PA15
47	1294	4080	5684	93.4%	–	Xq26.3	AL590618.18				+							
48	1136	0	1652	88.1%	+	19p13.2	AC008758.7			+								
49	1116	0	5945	91.8%	–	5p14.1	AC091893.3			+	+		90			+	HERVK3	L1P
50	1040	208	1515	90.2%	+	1p11.2	AL354822.8/AC145212.1	+										
51	1016	3978	5270	90.9%	+	Yq11.21	AC006998.3			2x	2x		83				AluY, Mer41-int	
52	747	5037	5945	93.0%	+	13q31.3	AL353757.9				2x				+			
53	730	5037	5943	92.8%	–	13q21.31	AL356253.7				+				+			
54	701	5048	5945	92.5%	+	7p14.3	AC005154.1				+					+		
55	689	5078	5945	92.4%	–	5p13.3	AC008768.8				+				+		Mer41A	
56	684	5037	5935	92.8%	+	6q27	AL121956.21				2x				+		Mer41C	
57	679	5037	5943	91.3%	–	9p21.1	AL513132.6				2x				+		Mer41C	
58	667	27	817	92.3%	+	16p12.3	AC141273.2			+								
59	662	5037	5945	91.9%	+	12p12.3	AC073504.18				2x				+			
60	649	0	783	91.3%	–	19p12	AC073534.4			+								
61	634	5037	5945	92.8%	–	6q27	AL022069.1				2x				+			
62	607	1704	2449	91.1%	–	19p13.11	AC010329.3	+										
63	597	5205	5942	93.3%	–	2p13.3	AC105054.6		+								Mer41C	
64	596	0	689	94.0%	+	5p14.2	AC108103.2			+								
65	593	23	692	94.0%	–	7q33	AC009784.2	+										
66	581	0	699	91.3%	–	9q31.1	AL161631.20	+										
67	364	5457	5884	93.2%	+	6_random	AL662873.3										Mer41A-int, Mer41A	
68	297	5609	5945	95.2%	+	22q11.23	AL022324.1		+								Mer41C	

Supplementary Table B

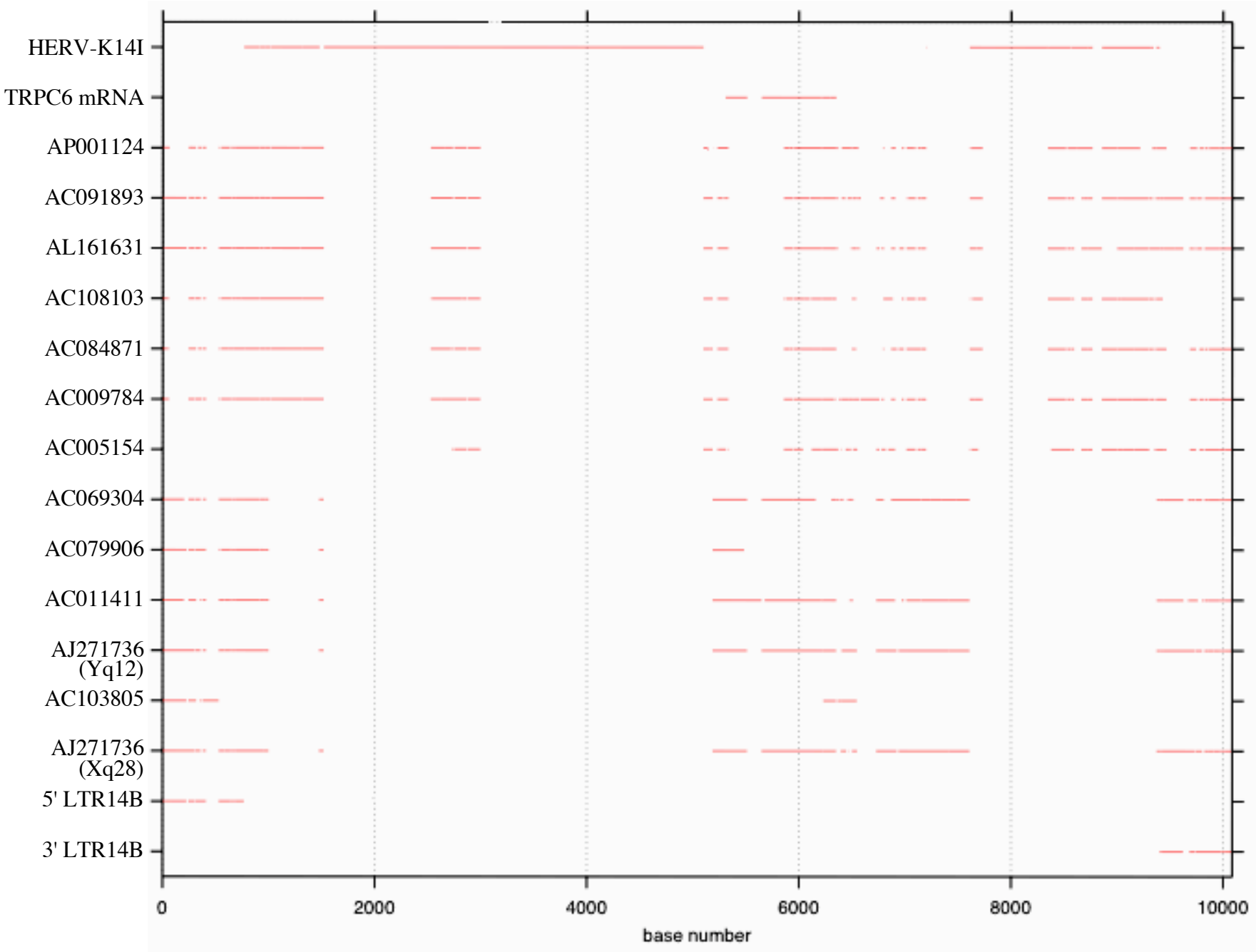
1	2	3	4	5	6	7		8	9	10	11	12
PV no.	**SCORE**	**START**	**END**	**IDENTITY**	**STRAND**	**POSITION**	**ACCESSION NO**	**5´LTR14C**	**3´LTR14C**	**LTR similarity**	**Repeats in PV**	**PV in Repeats**
1Cl	7203	0	7417	98.6%	-	16p12.3	AC130456.2	+	+	91		L2
2Cl	6321	0	7417	93.1%	-	15q21.1	AC051619.7	+	+	92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



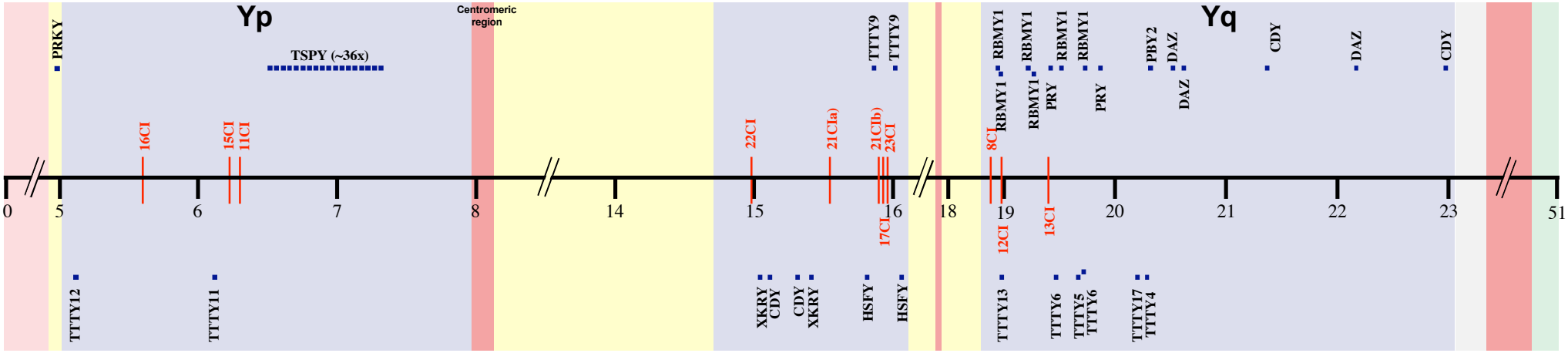
Flockerzi et al., Suppl. Fig. A



Flockerzi et al., Suppl. Figure B

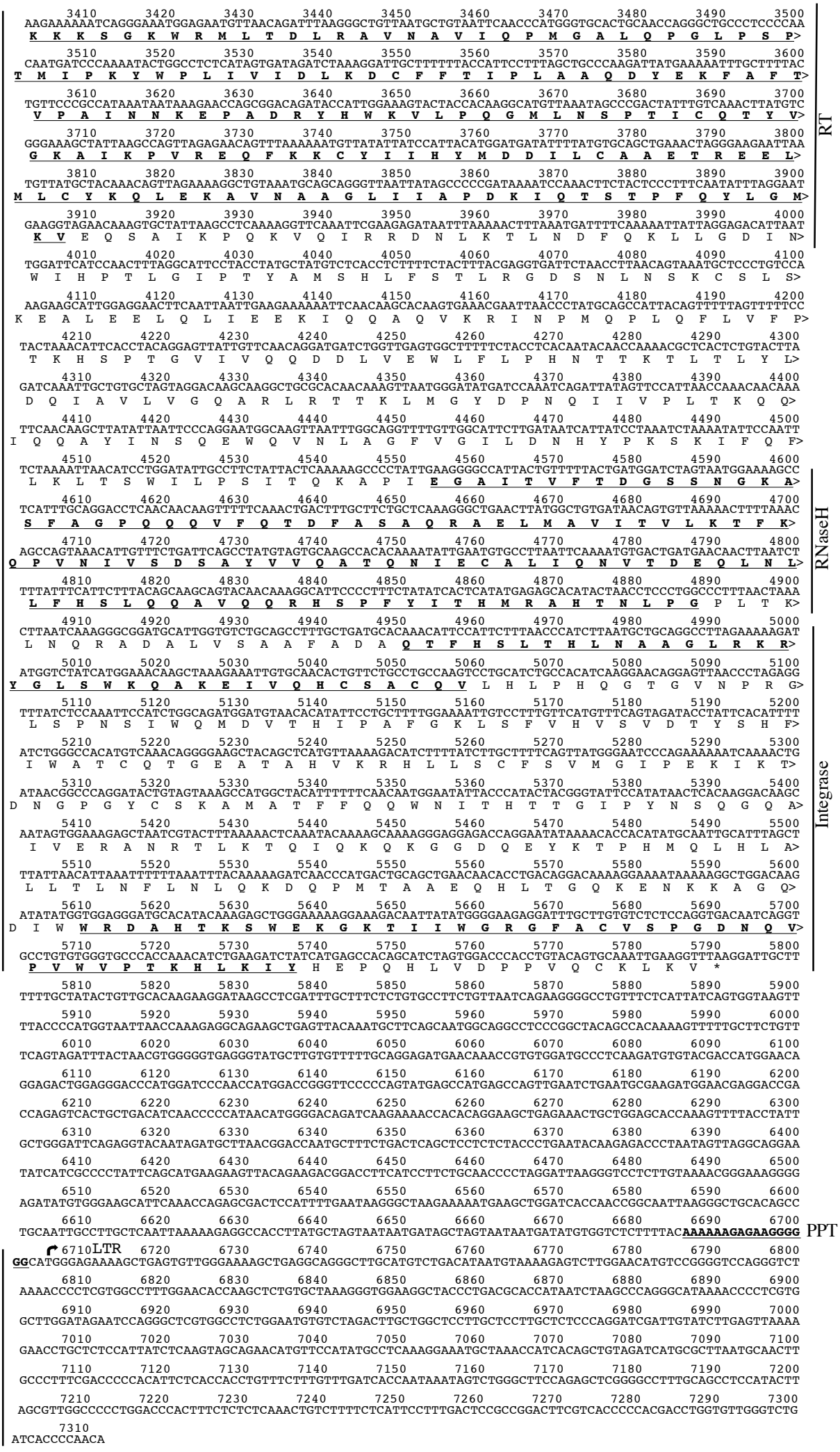


Flockerzi et al., Suppl. Fig. C

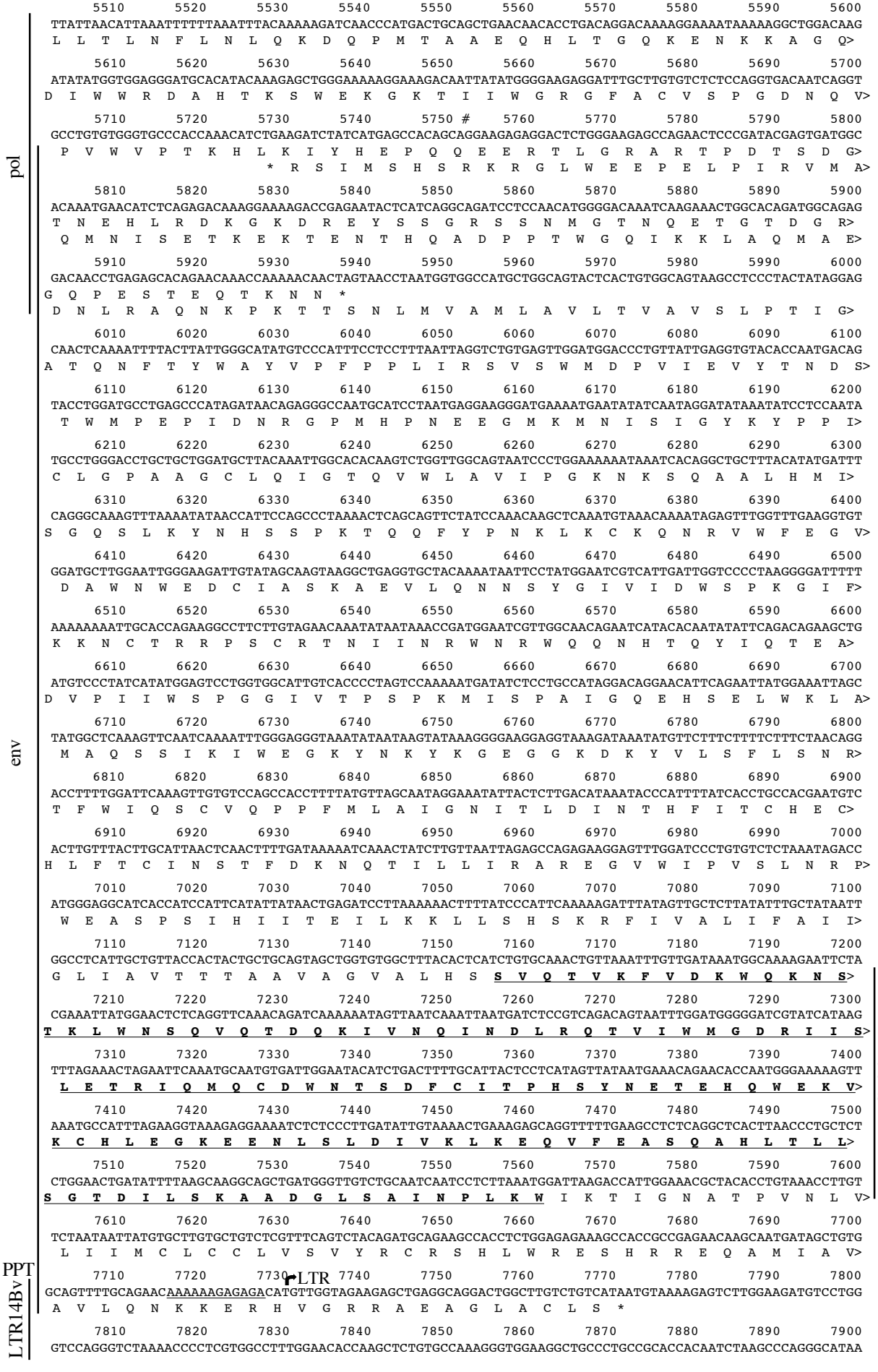


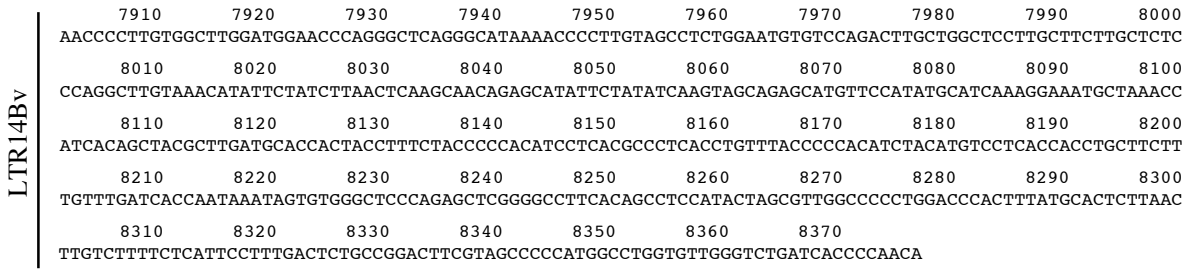
Suppl. Fig. D: Schematic localization of HERV-K14CI loci in intrachromosomally duplicated regions of the human Y chromosome. The figure is adapted from a recently published map of the human Y chromosome (Skaletsky et al. 2003) (ref. 38). Ampliconinc regions, consisting of intrachromosomal duplications, 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.

LTR14B
PBS
LTR
gag
pro
pol
p10
p24
ZnF
dUTPase
Protease

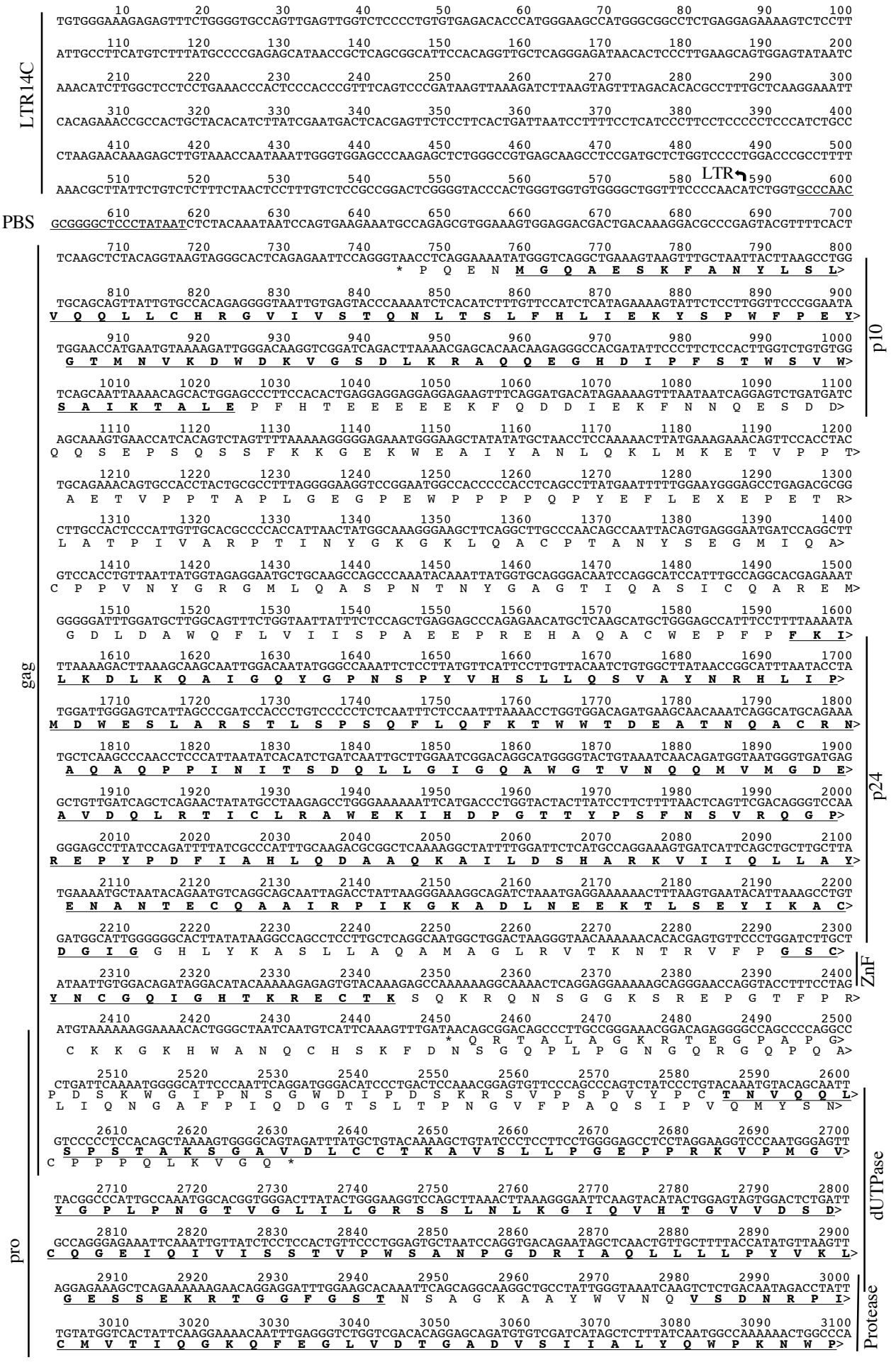

pol
RT
RNaseH
Integrase
LTR14B
PPT
LTR

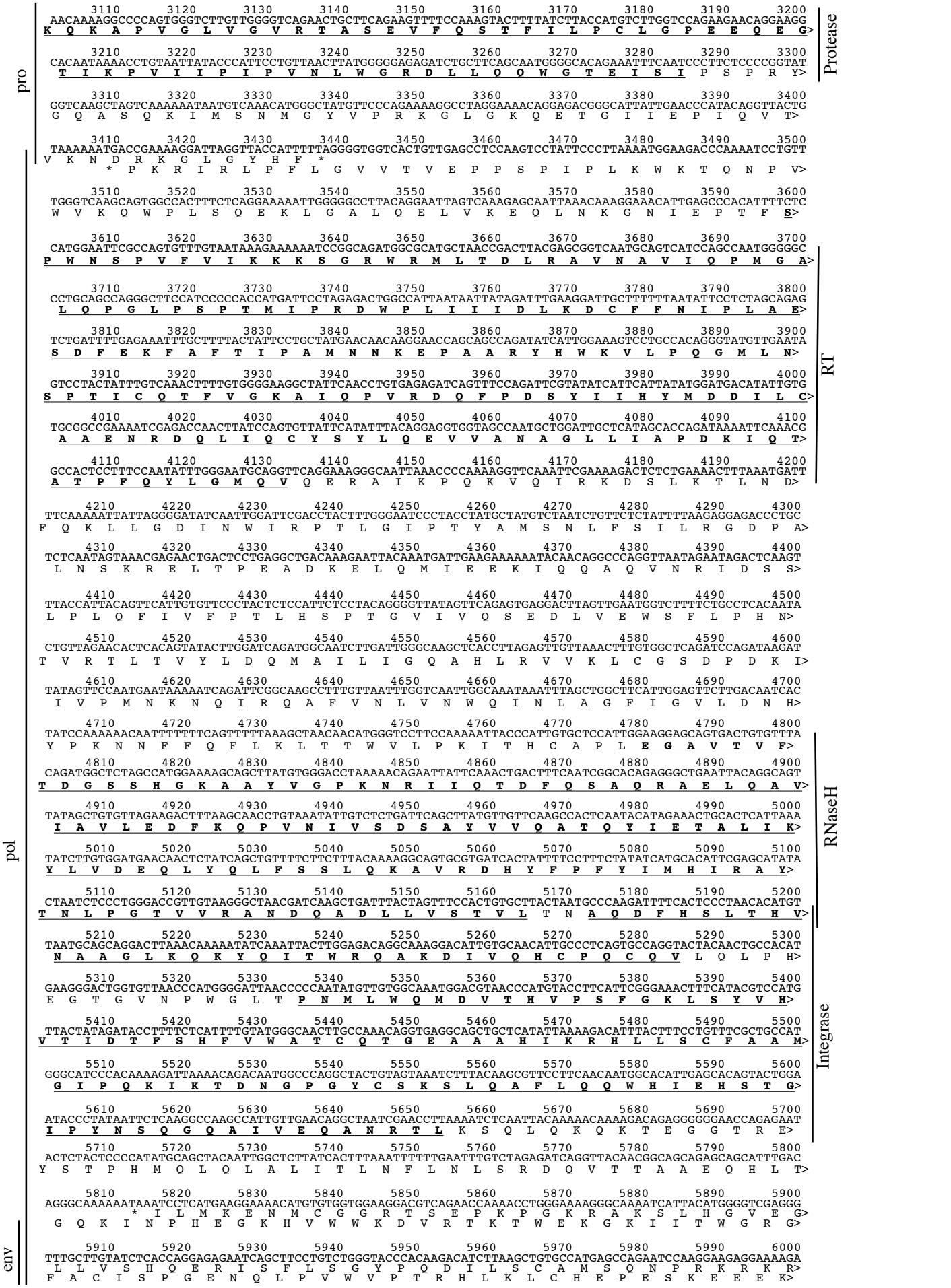
HERV-K14I, 3' portion, variant with putative *env* gene

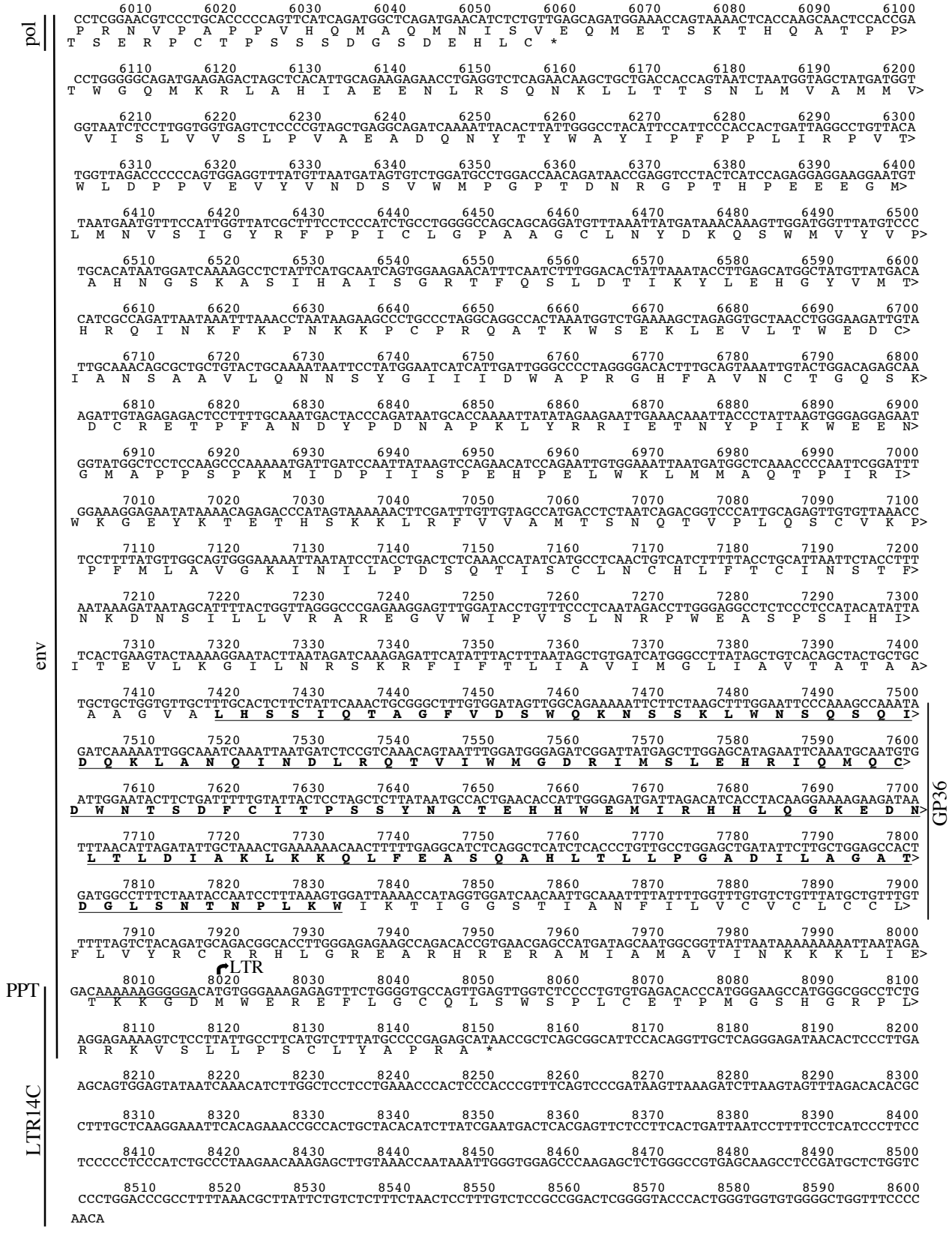




Suppl Fig. E: Annotated HERV-K14I consensus sequences generated in this study. The first sequence is the HERV-K14I having about 800 bp of *env*-unrelated sequence downstream from the *pol* gene. The proviral body is flanked by LTR14B sequences. The second part of the figure presents the 3' portion of 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 Repbase.


LTR14C
PBS
LTR
gag
pro
p10
p24
ZnF
dUTPase
Protease


pro
pol
env
Protease
RT
RNaseH
Integrase



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.