

**Supplemental Material for:**

**Protein coding potential of transposable elements in vertebrate genomes.**

Evgeny M Zdobnov<sup>1\*</sup>, Monica Campillos<sup>1\*</sup>, Eoghan D Harrington<sup>1</sup>, David Torrents<sup>1</sup>  
and Peer Bork<sup>1,2</sup>

<sup>1</sup>EMBL, Meyerhofstr.1, 69117 Heidelberg, Germany

<sup>2</sup>MDC Berlin-Buch, Robert-Roessle-Str. 10, Germany

Address correspondence to:

Peer Bork

EMBL Heidelberg

Meyerhofstrasse 1

D-69117 Heidelberg, Germany

e-mail: [bork@embl.de](mailto:bork@embl.de)

phone: +49 6221-387-526

fax: +49 6221-387-517

## Supplementary tables:

**Supp Table 1.** The list of 85 RETRA gene characteristic domains. For each domain, the InterPro description as well as Pfam and Interpro accession numbers are shown.

Pfam	InterPro	Description
PF00075	IPR002156	RNase H
PF00077	IPR001995	Peptidase A2A, retrovirus
PF00078	IPR000477	RNA-directed DNA polymerase (Reverse transcriptase)
PF00424	IPR000625	Anti-repression trans-activator protein, REV protein
PF00429	IPR002050	ENV polyprotein (coat polyprotein)
PF00469	IPR001558	HIV negative factor Nef
PF00516	IPR000777	Envelope glycoprotein GP120
PF00517	IPR000328	Envelope Polyprotein GP41
PF00522	IPR000012	Retroviral VpR/VpX protein
PF00539	IPR001831	Immunodeficiency virus transactivating regulatory protein (Tat)
PF00540	IPR000071	Retroviral matrix protein p17
PF00552	IPR001037	Retroviral integrase, C-terminal
PF00558	IPR008187	Vpu protein
PF00559	IPR000475	Retroviral Vif (Viral infectivity) protein
PF00589	IPR002104	Phage integrase
PF00607	IPR000721	Retroviral nucleocapsid protein Gag
PF00665	IPR001584	Integrase, catalytic domain
PF00692	IPR008180	DeoxyUTP pyrophosphatase
PF00872	IPR001207	Transposase,
PF00906	IPR002006	Hepatitis core antigen
PF00971	IPR001361	Gp90 coat polyprotein of equine infectious anemia virus (EIAV)
PF01021	IPR001042	Peptidase A11B, Ty1 A and B
PF01045	IPR001027	Gp45 coat polyprotein of equine infectious anemia virus (EIAV)
PF01054	IPR001213	Mouse mammary tumor virus superantigen
PF01140	IPR000840	Matrix protein (MA)
PF01141	IPR002079	Gag polyprotein, inner coat protein p12
PF01359	IPR001888	Transposase, type 1
PF01385	IPR001959	Transposase (probable), IS891/IS1136/IS1341
PF01498	IPR002492	Transposase, Tc1/Tc3
PF01526	IPR002513	Transposase Tn3
PF01527	IPR002514	Transposase IS3/IS911
PF01548	IPR002525	Transposase, IS111A/IS1328/IS1533
PF01609	IPR002559	Transposase, IS4
PF01610	IPR002560	Transposase, IS204/IS1001/IS1096/IS1165
PF01710	IPR002622	Transposase, Synechocystis PCC 6803
PF01797	IPR002686	Transposase IS200-like
PF02022	IPR003308	Integrase, N-terminal zinc-binding
PF02093	IPR003036	Core shell protein Gag P30
PF02228	IPR003139	Major core protein p19
PF02281	IPR003201	Transposase Tn5
PF02316	IPR003314	MuA-transposase/repressor protein CI, DNA-binding
PF02337	IPR003322	Retroviral GAG p10 protein
PF02371	IPR003346	Transposase IS116/IS110/IS902
PF02813	IPR004028	Retroviral M domain

PF02892	IPR003656	BED finger
PF02914	IPR004189	Bacteriophage Mu transposase
PF02920	IPR004191	Tn916 integrase, N-terminal DNA binding
PF02959	IPR004120	HTLV Tat
PF02992	IPR004242	En/Spm-like transposon proteins
PF02994	IPR004244	L1 transposable element
PF02998	IPR004247	Lentiviral Tat protein
PF03004	IPR004252	Putative plant transposon protein
PF03017	IPR004264	TNP1/EN/SPM-like transposon protein
PF03050	IPR004291	Transposase IS66
PF03056	IPR004295	Env gp36 protein, HERV/MMTV type
PF03078	IPR004312	Arabidopsis retrotransposon ORF-1 protein
PF03108	IPR004332	Plant MuDR transposase
PF03184	IPR004875	CENP-B protein
PF03221	IPR004906	Tc5 transposase
PF03274	IPR004956	Foamy virus BEL 1/2 protein
PF03276	IPR004957	Spumavirus gag protein
PF03400	IPR005063	IS1 transposase
PF03408	IPR005070	Foamy virus envelope protein
PF03539	IPR001641	Peptidase A9, spumavirus
PF03708	IPR005166	Avian retrovirus envelope protein, gp85
PF03732	IPR005162	Retrotransposon gag protein
PF03811	IPR003220	Insertion element protein
PF04094	IPR007228	Protein of unknown function DUF390
PF04160	IPR007289	Orf-X
PF04218	IPR006695	CENP-B, N-terminal DNA-binding
PF04195	IPR007321	Putative gypsy type transposon
PF04236	IPR007350	Tc5 transposase, C-terminal
PF04693	IPR006783	Archaeal putative transposase ISC1217
PF04740	IPR006829	Bacillus transposase protein
PF04754	IPR006842	Putative transposase, YhgA-like
PF04937	IPR007021	Protein of unknown function DUF659
PF04986	IPR007069	Putative transposase
PF05380	IPR008042	Retrotransposon, Pao
PF05457	IPR008877	Sulfolobus transposase
PF05598	IPR008490	Sulfolobus solfataricus protein of unknown function DUF772
PF05599	IPR008596	Deltaretrovirus Tax
PF05699	IPR008906	HAT dimerisation
PF05851	IPR008668	Lentivirus virion infectivity factor
PF05858	IPR008411	Bovine immunodeficiency virus surface envelope
PF07282	IPR010095	Transposase, IS605 OrfB

**Supp Table 2.** URLs of sequence data sets used in this work as of Oct 2003.

<b>Data URL</b>	<b>Number of transcripts</b>
<b>Human 33</b>	
<u>Ensembl</u>	
ftp.ensembl.org/pub/current_human/data/fasta/pep/Homo_sapiens.NCBI33.pep.fa.gz	32035
ftp.ensembl.org/pub/current_human/data/fasta/pep/Homo_sapiens.NCBI33.pep_genscan.fa.gz	65771
ftp.ensembl.org/pub/current_human/data/fasta/dna/Homo_sapiens.NCBI33.contig.fa.gz	
<u>NCBI</u>	
ftp.ncbi.nih.gov/genomes/H_sapiens/protein/protein.fa.gz	37490
ftp.ncbi.nih.gov/genomes/H_sapiens/protein/GS_prot.fsa.gz	53234
<u>TWINS</u>	
genes.cse.wustl.edu/predictions/human/NCBI33/against_mouse_5-01-03/NCBI33_5-01-03.tgz	25434
<b>Rat</b>	
<u>Ensembl</u>	
ftp.ensembl.org/pub/current_rat/data/fasta/pep/Rattus_norvegicus.RGSC2.pep.fa.gz	28904
ftp.ensembl.org/pub/current_rat/data/fasta/pep/Rattus_norvegicus.RGSC2.pep_genscan.fa.gz	79083
ftp.ensembl.org/pub/current_rat/data/fasta/dna/Rattus_norvegicus.RGSC2.contig.fa.gz	
<u>NCBI</u>	
ftp.ncbi.nih.gov/genomes/R_norvegicus/protein/Gnomon_prot.fsa.gz	42537
ftp.ncbi.nih.gov/genomes/R_norvegicus/protein/protein.fa.gz	21205
<u>TWINS</u>	
http://genes.cse.wustl.edu/predictions/rat/Jun03/rat_Jun03_07-11-03.tgz	25499
<b>Mouse</b>	
<u>Ensembl</u>	
ftp.ensembl.org/pub/current_mouse/data/fasta/pep/Mus_musculus.NCBIM30.pep.fa.gz	32911
ftp.ensembl.org/pub/current_mouse/data/fasta/pep/Mus_musculus.NCBIM30.pep_genscan.fa.gz	94075
ftp.ensembl.org/pub/current_mouse/data/fasta/dna/Mus_musculus.NCBIM30.contig.fa.gz	
<u>NCBI</u>	
ftp.ncbi.nih.gov/genomes/M_musculus/MGSCv3_Release3/protein/GS_prot.fsa.gz	52392
ftp.ncbi.nih.gov/genomes/M_musculus/MGSCv3_Release3/protein/protein.fa.gz	36861
<u>TWINS</u>	
http://genes.cse.wustl.edu/predictions/mouse/5-01-03/toUCSC/MGSCv3_30_v_NCBI31_5-1-03.tar	24282
<b>Fugu</b>	
<u>Ensembl</u>	
ftp.ensembl.org/pub/current_fugu/data/fasta/pep/Fugu_rubripes.FUGU2.pep.fa.gz	37245
ftp.ensembl.org/pub/current_fugu/data/fasta/pep/Fugu_rubripes.FUGU2.pep_genscan.fa.gz	29625
<u>DOE JOINT GENOME INSTITUTE</u>	
ftp.jgi-psf.org/pub/JGI_data/Fugu/fugu_v3_prot.fasta.Z	37439
ftp.jgi-psf.org/pub/JGI_data/Fugu/genscan_proteome.fasta	38633

**Supp Table 3.** Similarity, ESTs, best-reciprocal hit (BRH) and synteny analysis of human Ensembl genes with RETRA domains. For each gene we show its Ensembl gene (GENE\_ID) and protein (PROT\_ID) accession numbers, preceded by supporting analysis of: a) the closest region in the genome using BLAT search, showing the percent identity (PID) and the length coverage (COV) of this match, and b) the number of ESTs that match unambiguously (UNAM) and ambiguously (AMB). The BRH and synteny for the pairs of human-mouse (H/M), human-rat (H/R) and mouse-rat (M/R) genomes were indicated with (+) when identified using the automatic procedure, marked with (\*) when the assignment was done manually, and with (~) where the RETRA domain is not present in rat or mouse genomes.

Similarity		ESTs		GENE_ID	PROT_ID	BRH			Synteny		
PID	COV	UNAM	AMB			H/M	H/R	M/R	H/M	H/R	M/R
IPR001995 (PF00077) Peptidase A2A, retrovirus											
95	100	0	0	ENSG00000112307	ENSP00000230051						
96	80	13	8	ENSG00000179858	ENSP00000329107						
99.2	100	0	27	ENSG00000184379	ENSP00000332276	+	+	+			
IPR001037 (PF00552) Retroviral integrase, C-terminal											
96.1	100	77	27	ENSG00000185161	ENSP00000331756	+					
IPR000721 (PF00607) Retroviral nucleocapsid protein Gag											
90.8	99	7	9	ENSG00000121514	ENSP00000240857		+				
88.9	100	1	0	ENSG00000176556	ENSP00000316720						
99.2	100	0	27	ENSG00000184379	ENSP00000332276	+	+	+			
IPR001888 (PF01359) Transposase, type 1											
89.8	50	72	0	ENSG00000170364	ENSP00000305060	~	~	~	~	~	~
72.7	53	0	0	ENSG00000183366	ENSP00000329769						
IPR003322 (PF02337) Retroviral GAG p10 protein											
90.8	99	7	9	ENSG00000121514	ENSP00000240857		+				
96.1	100	0	0	ENSG00000171778	ENSP00000306385						
99.2	100	0	9	ENSG00000176416	ENSP00000314601						
99.2	100	0	27	ENSG00000184379	ENSP00000332276	+	+	+			
IPR004244 (PF02994) L1 transposable element											
95	99	0	0	ENSG00000133848	ENSP00000256230						
99.6	100	0	145	ENSG00000145152	ENSP00000273744						
90.2	55	65	0	ENSG00000162606	ENSP00000294615				*	*	*
99.6	100	0	154	ENSG00000175442	ENSP00000307837						
99.9	100	0	188	ENSG00000173238	ENSP00000307952						
99.6	100	0	212	ENSG00000173458	ENSP00000307967						
99.5	100	0	191	ENSG00000173286	ENSP00000307994						
99.8	100	0	134	ENSG00000172918	ENSP00000308002						
100	100	0	191	ENSG00000172980	ENSP00000308040						
99.9	100	0	156	ENSG00000175524	ENSP00000308059						
99.5	100	0	107	ENSG00000174531	ENSP00000308072						
99.2	100	0	164	ENSG00000173785	ENSP00000308102						
99.6	100	0	97	ENSG00000174735	ENSP00000308135						
88	49	0	0	ENSG00000174656	ENSP00000308177						
99.8	100	0	132	ENSG00000173140	ENSP00000308257						
100	100	0	191	ENSG00000175525	ENSP00000308277						
99.4	100	0	72	ENSG00000174416	ENSP00000308278						
99.5	100	0	105	ENSG00000174895	ENSP00000308287						



99.5	100	0	71	ENSG00000172762	ENSP00000311619						
99.9	100	0	152	ENSG00000172773	ENSP00000311720						
99.9	100	0	179	ENSG00000175457	ENSP00000311721						
99.9	100	0	156	ENSG00000172946	ENSP00000311799						
99.8	100	0	207	ENSG00000173307	ENSP00000311870						
99.8	100	0	221	ENSG00000172544	ENSP00000311911						
100	100	0	182	ENSG00000172937	ENSP00000311924						
99.8	100	0	138	ENSG00000174763	ENSP00000311935						
99.4	100	0	147	ENSG00000175053	ENSP00000312040						
99.8	100	0	107	ENSG00000173550	ENSP00000312087		+				
99.7	100	0	234	ENSG00000174597	ENSP00000312149						
100	100	0	163	ENSG00000175113	ENSP00000312182						
100	100	0	208	ENSG00000174387	ENSP00000312245						
99.7	100	0	124	ENSG00000172743	ENSP00000312295						
99.7	100	0	173	ENSG00000174447	ENSP00000312429						
99.9	100	0	175	ENSG00000173096	ENSP00000312462						
99.9	100	0	212	ENSG00000175615	ENSP00000312467						
99.9	100	0	208	ENSG00000173018	ENSP00000312473						
99.9	100	0	185	ENSG00000175596	ENSP00000312480						
100	100	0	147	ENSG00000173912	ENSP00000312638						
99.9	100	0	111	ENSG00000174391	ENSP00000312653						
99.8	100	0	188	ENSG00000179917	ENSP00000312744						
99.8	100	0	143	ENSG00000176292	ENSP00000313692						
99.8	100	0	142	ENSG00000181011	ENSP00000313943						
99.3	100	0	80	ENSG00000177817	ENSP00000315082						
99.6	100	0	155	ENSG00000178203	ENSP00000315097						
99.8	100	0	106	ENSG00000175908	ENSP00000315537						
99.5	100	0	71	ENSG00000178277	ENSP00000315539						
99.2	100	0	123	ENSG00000178455	ENSP00000315660						
99.7	100	0	106	ENSG00000176210	ENSP00000315746						
99.8	100	0	220	ENSG00000176109	ENSP00000316485						
99.7	100	0	149	ENSG00000176610	ENSP00000318379						
77.5	22	1	0	ENSG00000179655	ENSP00000320780						
99.7	100	0	147	ENSG00000178677	ENSP00000320976						
99.4	100	0	99	ENSG00000177543	ENSP00000321280						
99.6	100	3	149	ENSG00000175770	ENSP00000322514						
99.5	100	0	97	ENSG00000178664	ENSP00000322561						
81.8	100	1	0	ENSG00000176802	ENSP00000323959						
100	100	0	150	ENSG00000176814	ENSP00000324049						
99.7	100	0	154	ENSG00000177751	ENSP00000324591						
99.7	100	0	119	ENSG00000179757	ENSP00000326461		+				
99.7	100	0	90	ENSG00000177996	ENSP00000326548						
99.6	100	0	112	ENSG00000182920	ENSP00000327744						
99.8	100	0	232	ENSG00000185396	ENSP00000327803						
99.9	100	0	208	ENSG00000183157	ENSP00000327926						
99.5	100	0	107	ENSG00000182105	ENSP00000328253						
99.5	100	0	145	ENSG00000184694	ENSP00000328530		+		+		
99.4	100	0	138	ENSG00000182506	ENSP00000328759						
99.1	100	0	58	ENSG00000183349	ENSP00000329248						
100	100	0	151	ENSG00000183541	ENSP00000330792						
82.5	90	2	0	ENSG00000182243	ENSP00000331230						
99.8	100	0	248	ENSG00000182467	ENSP00000331250						

100	100	0	200	ENSG00000185852	ENSP00000331616						
100	100	0	185	ENSG00000185540	ENSP00000331893						
96.5	100	1	0	ENSG00000183730	ENSP00000332358						
99.5	100	0	84	ENSG00000182102	ENSP00000332890						
100	100	0	159	ENSG00000184994	ENSP00000333176						
IPR005162 (PF03732) Retrotransposon gag protein											
-	-	478	0	ENSG00000146743	ENSP00000275616	+			+		
21.9	33	136	0	ENSG00000171362	ENSP00000305039	+	~	+	+	~	+
-	-	9	0	ENSG00000181110	ENSP00000313658	+			+	*	*
79.8	2	10	0	ENSG00000183551	ENSP00000327693	+	+	+	+	+	+
IPR002514 (PF01527) Transposase IS3/IS911											
90.4	2	15	0	ENSG00000133980	ENSP00000256362	~			~		
IPR002156 (PF00075) RNase H											
94.1	100	0	0	ENSG00000129375	ENSP00000250275						
95.6	41	138	5	ENSG00000171865	ENSP00000313350	+	+	+	+	+	+
-	-	24	0	ENSG00000183021	ENSP00000331027		+	+			
IPR000477 (PF00078) RNA-directed DNA polymerase (Reverse transcriptase)											
99.8	100	0	634	ENSG00000132951	ENSP00000255303						
99.8	100	0	417	ENSG00000133610	ENSP00000255982						
99.7	100	0	604	ENSG00000134465	ENSP00000256882						
99.8	100	0	568	ENSG00000136265	ENSP00000258765						
99.9	100	0	314	ENSG00000138808	ENSP00000265501						
99.9	100	0	482	ENSG00000147595	ENSP00000276593		+				
99.9	100	0	515	ENSG00000149990	ENSP00000279479						
99.9	100	0	484	ENSG00000150298	ENSP00000279821						
99.8	100	0	344	ENSG00000162937	ENSP00000295039						
99.9	100	0	439	ENSG00000163142	ENSP00000295293						
99.9	100	0	262	ENSG00000164131	ENSP00000296539						
99.7	100	0	0	ENSG00000164862	ENSP00000297486						
99.9	100	0	428	ENSG00000165053	ENSP00000297727						
99.8	100	0	409	ENSG00000169095	ENSP00000306211						
99.8	100	0	368	ENSG00000169742	ENSP00000307486						
99.9	100	0	364	ENSG00000172747	ENSP00000308526						
99.9	100	0	179	ENSG00000174274	ENSP00000308846						
98.4	100	0	673	ENSG00000173902	ENSP00000309397						
-	-	22	0	ENSG00000164362	ENSP00000309572	+	~	~	+	~	~
99.8	100	0	303	ENSG00000174875	ENSP00000309869						
99.7	100	0	368	ENSG00000180990	ENSP00000313780						
99.7	100	0	418	ENSG00000178245	ENSP00000315143						
99.7	100	0	320	ENSG00000180452	ENSP00000316489						
92.8	98	1	0	ENSG00000176514	ENSP00000316555						
95.7	99	0	0	ENSG00000176004	ENSP00000316964						
99.8	100	0	437	ENSG00000176174	ENSP00000317865						
94.2	100	0	0	ENSG00000181329	ENSP00000318150		+				
90.8	100	0	0	ENSG00000176151	ENSP00000319824						
98.3	100	0	697	ENSG00000176397	ENSP00000320212						
99.3	95	1	53	ENSG00000179006	ENSP00000322241						
99.5	100	0	438	ENSG00000181549	ENSP00000324718						
93.3	100	1	1	ENSG00000178131	ENSP00000324719						
91.5	100	1	0	ENSG00000178861	ENSP00000325648						
99.9	100	0	505	ENSG00000179906	ENSP00000327325			+			
99.9	100	0	450	ENSG00000184652	ENSP00000327354						



92.2	100	0	2	ENSG00000185724	ENSP00000327720						
52.1	41	1	0	ENSG00000185458	ENSP00000328349						
99.6	100	0	335	ENSG00000182848	ENSP00000328438						
99.7	100	0	392	ENSG00000182761	ENSP00000328912						
99.5	100	0	769	ENSG00000185103	ENSP00000329917						
93.8	3	0	0	ENSG00000185923	ENSP00000330169						
99.3	100	0	699	ENSG00000183006	ENSP00000330738						
99.6	100	0	273	ENSG00000184212	ENSP00000330790						
-	-	1	0	ENSG00000185469	ENSP00000331538				*	*	*
99.8	100	1	7	ENSG00000184853	ENSP00000331960						
94.4	100	0	11	ENSG00000184657	ENSP00000332029						
99.2	100	0	27	ENSG00000184379	ENSP00000332276	+	+	+			
90.6	100	0	0	ENSG00000185353	ENSP00000332609						
99.6	100	0	699	ENSG00000185146	ENSP00000332803						
99.7	100	0	338	ENSG00000182096	ENSP00000332935						
99.7	100	0	837	ENSG00000184455	ENSP00000333005						
79.9	46	6	0	ENSG00000184317	ENSP00000333216						
98.7	100	0	673	ENSG00000183188	ENSP00000333682						
96.5	95	0	110	ENSG00000184172	ENSP00000333700						
IPR002050 (PF00429) ENV polyprotein (coat polyprotein)											
94.2	100	100	4	ENSG00000135186	ENSP00000257638						
97.8	100	0	2	ENSG00000136549	ENSP00000259063	+	+	+			
94.2	100	0	1	ENSG00000163489	ENSP00000295716						
95.9	100	0	3	ENSG00000165377	ENSP00000298116						
94.3	100	0	2	ENSG00000166319	ENSP00000299405						
95.9	100	0	1	ENSG00000167229	ENSP00000300617						
93.7	100	0	0	ENSG00000178867	ENSP00000314789						
88.9	97	2	0	ENSG00000176524	ENSP00000316607						
94	100	1	0	ENSG00000176121	ENSP00000317677						
-	-	1	0	ENSG00000179135	ENSP00000320605			+			
96.7	98	3	3	ENSG00000180236	ENSP00000323110						
59.6	76	12	0	ENSG00000181590	ENSP00000325374						
95.7	100	0	3	ENSG00000179442	ENSP00000326747						
IPR001584 (PF00665) Integrase, catalytic domain											
94.1	-	92	0	ENSG00000092028	ENSP00000206466	+			+		
-	-	43	0	ENSG00000145723	ENSP00000274384	+	+	+	+	+	+
76.8	2	17	0	ENSG00000164531	ENSP00000297049				*	*	*
76.8	2	14	0	ENSG00000164531	ENSP00000297051				*	*	*
93.4	100	16	0	ENSG00000166331	ENSP00000299417						
93	99	1	0	ENSG00000174801	ENSP00000311069						
-	-	54	0	ENSG00000145723	ENSP00000313934	+	+	+	+	+	+
76.8	5	11	0	ENSG00000181497	ENSP00000314319				*	*	*
96	100	0	1	ENSG00000181725	ENSP00000314419	+					
-	-	24	0	ENSG00000183021	ENSP00000331027	+	+	+			
IPR008180 (PF00692) DeoxyUTP pyrophosphatase											
80	54	450	1	ENSG00000128951	ENSP00000249783	+	+	+	+	+	+
-	-	1	0	ENSG00000169128	ENSP00000303742						
93	99	1	0	ENSG00000174801	ENSP00000311069						
99.2	100	0	27	ENSG00000184379	ENSP00000332276	+	+	+			
35.1	83	0	0	ENSG00000183835	ENSP00000332786						
IPR003308 (PF02022) Integrase, N-terminal zinc-binding											
96	100	0	1	ENSG00000181725	ENSP00000314419	+					

IPR003036 (PF02093) Core shell protein Gag P30											
93.6	100	0	0	ENSG00000172672	ENSP00000311588	+	+	+			
-	-	8	0	ENSG00000177999	ENSP00000320203						
IPR003656 (PF02892) BED finger											
-	-	105	0	ENSG00000100426	ENSP00000216268	+	~	~		~	
-	-	35	0	ENSG00000132846	ENSP00000255198	~				~	
90.7	6	77	0	ENSG00000169084	ENSP00000305414						
-	-	8	0	ENSG00000177494	ENSP00000321370					*	*
IPR004295 (PF03056) Env gp36 protein, HERV/MMTV type											
97.3	100	0	16	ENSG00000162741	ENSP00000294791						
95.8	100	23	0	ENSG00000172398	ENSP00000304323	+					
99	100	0	37	ENSG00000174924	ENSP00000312633						
93.7	100	1	0	ENSG00000184934	ENSP00000333690						
IPR004875 (PF03184) CENP-B protein											
76.7	7	154	0	ENSG00000125817	ENSP00000246014	+	+	+	+	+	+
-	-	34	0	ENSG00000140993	ENSP00000268674					*	*
72.8	3	150	0	ENSG00000143157	ENSP00000271372	+	+	+	+	+	+
89.7	3	37	0	ENSG00000164296	ENSP00000296736					*	*
92.4	7	90	0	ENSG00000171567	ENSP00000302140						
-	-	12	0	ENSG00000169989	ENSP00000305677	+	+	+	+	+	+
-	-	23	0	ENSG00000173825	ENSP00000308354	+				+	
94.5	-	301	1	ENSG00000143442	ENSP00000313204	+	~	~	+	~	~
-	-	40	0	ENSG00000179886	ENSP00000315906	+	+	+	+	+	+
-	-	27	0	ENSG00000180346	ENSP00000317170	+		+	+		
-	-	3	0	ENSG00000180850	ENSP00000318421						
-	-	18	0	ENSG00000180873	ENSP00000326586	+	+	+	+	+	+
-	-	19	0	ENSG00000183340	ENSP00000333350		+	+		+	
IPR006695 (PF04218) CENP-B, N-terminal DNA-binding											
76.7	7	154	0	ENSG00000125817	ENSP00000246014	+	+	+	+	+	+
-	-	34	0	ENSG00000140993	ENSP00000268674					*	*
44.4	9	142	0	ENSG00000141378	ENSP00000269135	~	~	~	~	~	~
89.7	3	37	0	ENSG00000164296	ENSP00000296736					*	*
92.4	7	90	0	ENSG00000171567	ENSP00000302140						
-	-	12	0	ENSG00000169989	ENSP00000305677	+	+	+	+	+	+
-	-	23	0	ENSG00000173825	ENSP00000308354	+				+	
-	-	40	0	ENSG00000179886	ENSP00000315906	+	+	+	+	+	+
-	-	27	0	ENSG00000180346	ENSP00000317170	+		+	+		
-	-	18	0	ENSG00000180873	ENSP00000326586	+	+	+	+	+	+
-	-	19	0	ENSG00000183340	ENSP00000333350		+	+		+	
IPR008906 (PF05699) HAT dimerisation											
-	-	105	0	ENSG00000100426	ENSP00000216268	+	+	+		+	
-	-	18	0	ENSG00000106479	ENSP00000223210					*	*
98.6	100	142	57	ENSG00000137492	ENSP00000260045	~	~	~	~	~	~
76.8	2	17	0	ENSG00000164531	ENSP00000297049					*	*
69.6	4	37	0	ENSG00000157657	ENSP00000303982	~	+	~	~	+	~
90.7	6	77	0	ENSG00000169084	ENSP00000305414						
IPR004191 (PF02920) Tn916 integrase, N-terminal DNA binding											
95.3	30	174	0	ENSG00000110841	ENSP00000228425	~	~	~	~	~	~
83.9	1	190	0	ENSG00000166387	ENSP00000299492	+	~	~	+	~	~

**Supp Table 4.** ‘Rescue list’ of vertebrate RETRA genes. A) The list of 35 human Ensembl genes with RETRA domains having mouse or rat orthologs conserved in synteny (sorted by InterPro accession number). B) The list of the orthologous genes in studied vertebrates, including syncytin1 and syncytin2 present only in hominoids. C) FASTA formatted cDNA sequences of the true human RETRA genes.

A)

<b>InterPro (Pfam)</b>	<b>Description</b>	<b>Human genes</b>
IPR000477 (PF00078)	RNA-directed DNA polymerase (Reverse transcriptase)	ENSG00000164362 ENSG00000185469
IPR001584 (PF00665)	Integrase, catalytic domain	ENSG00000092028 ENSG00000145723 ENSG00000164531 ENSG00000181497
IPR001888 (PF01359)	Transposase, type 1	ENSG00000170364
IPR002156 (PF00075)	RNase H	ENSG00000171865
IPR002514 (PF01527)	Transposase IS3/IS911	ENSG00000133980
IPR003656 (PF02892)	BED finger	ENSG00000100426 ENSG00000132846 ENSG00000177494
IPR004191 (PF02920)	Tn916 integrase, N-terminal DNA binding	ENSG00000110841 ENSG00000166387
IPR004244 (PF02994)	L1 transposable element	ENSG00000162606
IPR004875 (PF03184)	CENP-B protein	ENSG00000125817 ENSG00000140993 ENSG00000143157 ENSG00000164296 ENSG00000169989 ENSG00000173825 ENSG00000143442 ENSG00000179886 ENSG00000180346 ENSG00000183340 ENSG00000180873
IPR005162 (PF03732)	Retrotransposon gag protein	ENSG00000146743 ENSG00000171362 ENSG00000181110 ENSG00000183551
IPR008180 (PF00692)	DeoxyUTP pyrophosphatase	ENSG00000128951
IPR008906 (PF05699)	HAT dimerisation	ENSG00000100426 ENSG00000106479 ENSG00000137492 ENSG00000164531 ENSG00000157657
IPR006695 (PF04218)	CENP-B, N-terminal DNA-binding	ENSG00000141378

B)

HUMAN	MOUSE	RAT	FUGU
ENSG0000092028	ENSMUSG0000022222	ENSRNOG00000024898	not predicted
ENSG00000100426	ENSMUSG00000044530	ENSRNOG00000004588	FRUP00000150793
ENSG00000106479	not predicted	ENSRNOG00000024795	FRUP00000147504
ENSG00000110841	ENSMUSG00000016487	ENSRNOG00000001832	FRUP00000143052
ENSG00000125817	ENSMUSG00000045541	ENSRNOG00000021246	not predicted
ENSG00000128951	ENSMUSG00000027203	ENSRNOG00000007221	FRUP00000133083
ENSG00000132846	ENSMUSG00000041995	not predicted	not predicted
ENSG00000133980	ENSMUSG00000051302	not predicted	FRUP00000152306
ENSG00000137492	ENSMUSG00000030753	ENSRNOG00000015808	FRUP00000132351
ENSG00000140993	not predicted	not predicted	not predicted
ENSG00000141378	ENSMUSG00000048773	ENSRNOG00000004288	FRUP00000157833
ENSG00000143157	ENSMUSG00000040596	ENSRNOG00000003859	FRUP00000153965
ENSG00000143442	ENSMUSG00000038902	ENSRNOG00000020962	FRUP00000139000
ENSG00000145723	ENSMUSG00000026333	ENSRNOG00000011962	FRUP00000164780
ENSG00000146743	ENSMUSG00000048670	not predicted	FRUP00000160001
ENSG00000157657	ENSMUSG00000028358	ENSRNOG00000006386	FRUP00000127092
ENSG00000157657	ENSMUSG00000028358	ENSRNOG00000006719	FRUP00000127092
ENSG00000157657	ENSMUSG00000052775	ENSRNOG00000006386	FRUP00000127092
ENSG00000157657	ENSMUSG00000052775	ENSRNOG00000006719	FRUP00000127092
ENSG00000162606	not predicted	not predicted	not predicted
ENSG00000164296	not predicted	not predicted	not predicted
ENSG00000164362	ENSMUSG00000021611	ENSRNOG00000016699	FRUP00000149754
ENSG00000164531	not predicted	not predicted	not predicted
ENSG00000181497	same as ENSG00000164531, merged in v34		
ENSG00000166387	ENSMUSG00000036528	ENSRNOG00000019686	not predicted
ENSG00000169989	ENSMUSG00000047819	ENSRNOG00000010741	not predicted
ENSG00000170364	ENSMUSG00000034639	ENSRNOG00000006806	FRUP00000151488
ENSG00000171362	ENSMUSG00000046204	ENSRNOG00000009815	FRUP00000134837
ENSG00000171865	ENSMUSG00000020630	ENSRNOG00000008584	FRUP00000131271
ENSG00000173825	ENSMUSG00000044390	ENSRNOG00000023318	not predicted
ENSG00000177494	not predicted	not predicted	not predicted
ENSG00000179886	ENSMUSG00000047166	ENSRNOG00000008686	not predicted
ENSG00000180346	ENSMUSG00000049232	not predicted	not predicted
ENSG00000180873	ENSMUSG00000046380	ENSRNOG00000005873	not predicted
ENSG00000181110	ENSMUSG00000042217	ENSRNOG00000022175	not predicted
ENSG00000183340	not predicted	ENSRNOG00000005704	not predicted
ENSG00000183551	ENSMUSG00000049191	ENSRNOG00000003255	not predicted
ENSG00000185469	not predicted	not predicted	FRUP00000153975
ENSG00000169084	not predicted	not predicted	FRUP00000129385
ENSG00000135186	not predicted	not predicted	not predicted
ENSG00000181590	not predicted	not predicted	not predicted

C) The human cDNA sequences are available for download from:

[http://www.bork.embl.de/Docu/RETRA/Rescue\\_Human\\_RETRA\\_cDNA\\_FASTA.txt](http://www.bork.embl.de/Docu/RETRA/Rescue_Human_RETRA_cDNA_FASTA.txt)