Supplemental Material

PCR primers for primer walking:

Cntrob_ERVwalk+641F	GCTAAAAGGTCTGCGGTCTG
Cntrob_walk2+1261F	ATCCCTCCTTAAAAGCCTTGAC
Cntrob_walk3+2059F	CTGATACAGCCATGCCACTC
Cntrob_walk4+2836F	AATCGCTCAAATGCTGCTTC
Cntrob_walk5+3641F	TGGGTCCCCTACAGAAAGGT
Cntrob_walk5-3648R	CTGAGGCAGTGACCTTTGGA
Cntrob_walk4-2936R	CGAGCCTTTTCACCAGTCAT
Cntrob_walk3-2154R	CAAATACAGGGCAGGGAAGA
Cntrob_walk2-1408R	GCACAAAGGAGGGAATCTTG
Cntrob_ERVwalk-734R	TTCAAAGAGGCCAGCCATAC

PCR primers to amplify the full-length RnERV-K8e sequences:

Chr 1	
RV-1-F	CCAGATGACATTTGGTGTTTTGTAG
RV-1-R	ACGCCAGAGGAAAACACTGTC
Chr 5	
RV-5-138F	GCAGGTTTCTCCAGGTGTAGG
RV-5-8502R	ATCTGACATGCTGCCTGGAT
Chr 6	
RV-6-431F	CCCCAAGGATAAGTTGCTGA
RV-6-9232R	GTGTTTTGTGGGGGTATTGG
Chr 8	
RV-8-544F	TCAATGTGGGCATTCTCATC
RV-8-9331R	GTGTTGCTCCGTACACCAGA
Chr 10	
Cntrob-11-139F	CTGGGAGCCACACTTAGGTC
Cntrob-c2-2150R	AACTCCTGTTGGTGCTGTCC
InversF1	GACGTGACACCCAGATCTAGACCA
LTR_3_F	CTGGGGCGGTACTATGCTAA
Chr 20	
F-Ch20 (IAPF)	GGTGATCATTGGGTGTCCACTGTAGGGAGCGG
R2-Ch20 (IAPR2)	GGAATTCTTAGTAGAACCTTAGCATCTGTCTGTG
Chr Un	
RV-Un-113F	TCAATCTGAACCCGAGTGAA
RV-Un-8751R	CTGGTATCGGGAAAGAAGCA

PCR primers used to determine polymorphic status of RnERV-K8e.

Chr 1 (expected fragment size: 460 bp)

RV-1-F	CCAGATGACATTTGGTGTTTTGTAG
retroPbsR	CGTACGGGCCACCACTTTGT

Chr 5 (expected fragment size: 461bp)

RV-5-138F	GCAGGTTTCTCCAGGTGTAGG
retroPbsR	CGTACGGGCCACCACTTTGT

Chr 6 (expected fragment size: 587 bp)

RV-6-431F	CCCCAAGGATAAGTTGCTGA
retroPbsR	CGTACGGGCCACCACTTTGT

Chr 7 (expected fragment size: 280 bp)

1st round	Chr 7-F1	CCATGGGATACCCATTTGTC
1st round	retroPbsR	CGTACGGGCCACCACTTTGT
2nd round	Ch7-F2	CCAGCAGATGGGGTAGATGT
2nd round	LTR-5-R	GCATGAACTCCCATTTAGCA

Chr 8 (expected fragment size: 474 bp)

RV-8-544F	TCAATGTGGGCATTCTCATC
retroPbsR	CGTACGGGCCACCACTTTGT

Chr 10 (expected fragment size: 226 bp)

1st round	Chr 10-F2	ATGCACATGTGGCTGTGTTT
1st round	retroPbsR	CGTACGGGCCACCACTTTGT
2nd round	Cntrob-11-139F	CTGGGAGCCACACTTAGGTC
2nd round	LTR-5-R	GCATGAACTCCCATTTAGCA

Chr 20 (expected fragment size: 366 bp)

1st round	IAP21R	AATGCCACCATATTCTGTACCC
1st round	InversF1	GACGTGACACCCAGATCTAGACCA
2nd round	IAPR2	GGAATTCTTAGTAGAACCTTAGCATCTGTCTGTG
2nd round	LTR-3-F	CTGGGGCGGTACTATGCTAA

Chr Un (expected fragment size: 331 bp)

1st round	Chr Un-F1	ACAATGGCCACGGAGATTAG
1st round	retroPbsR	CGTACGGGCCACCACTTTGT
2nd round	RV-Un-113F	TCAATCTGAACCCGAGTGAA
2nd round	LTR-5-R	GCATGAACTCCCATTTAGCA

The following primers were to analyze the aberrant splice products:

Cntrob_c3_2053F	ex10	GCGGAGGTTCAAGAGAACAG
Cntrob_c2_2150R	ex11	AACTCCTGTTGGTGCTGTCC

The following primers were used for inverse PCR:

1st round LTR5R	GCATGAACTCCCATTTAGCA
1st round inverseF5	CTCGCCAATCTCTGGAGGACAA
2nd round retroLtrR	ACGTCAGGTGCAAAACTCCAC
2nd round retroLTRF	GGCGCGGTAACACTTAAAGCTGTAA

The following primers were used to generate probes to the Northern blot:

env1-F	CTCTACCTCCTGTTTGCGTGATGTT
env6-R	ACAAAGGAGGGAATCTTGACCAT
gag7-F	CACCTCCGTACAACCCTCAACATAAAA
gag6-R	TGTATGGTGTTATGGCATTCCTACA

Primers for Taqman PCR:

ERVK8e_gag-ANYF	CGGCGTAAGCTCAAAGTTGAAA
ERVK8e_gag-ANYR	CAACGATCGCACTCATCAAGAAAA
probe (FAM-MGB)	
ERVK8e_gag-ANYM2	CTTGACAGGATTTTCC
ERVK8e_env-ANYF	ACAGAATAGCTCTACCCGTCGTA
ERVK8e_env-ANYR	TTGTTCTATGGAATGGACTCTGACATG
probe (FAM-MGB)	
ERVK8e_env-ANYM1	CTTACCCGCTTGCCCC

Accession numbers of retroviral sequences used in phylogenetic studies as follows: MLV, AF033811; WDSV, NC001867; HIV-1, M15654; HERV-K10, M14123; *HERV-K9I HML3, HERV-K HML5;* MMTV(β2), M15122; MusD(β7), BK001485; SRV-1(β6), M11841; RnSRV-1(β6), M11841; RnERV-β6, NW_047390; RnERV-β5, AC106444; RnERV-β4, NW_047390; JSRV(β3), M80216; RnERV- β1, BX883042; *RnIAP* ; MmIAP, M17551*; MmIAPLTR1a_I*; Mm*IAPLTR3_I;* MmIAPEYI, X87638; MaIAP, M10134, Eker rat (Tsc2), U23776.1; Mm_NT_039515 from 2357163-2363952; MmERV-K10c, NT_039617 from 3650419-4489977 bp; RSV, AF033808; HTLV-I, *GgERV-K10;* NC001436. Repbase IDs are given in italic and could be looked up at www.girinst.org.

Genomic location of the full-length copies of RnERV-K8e:

The displayed regions based on the rat genome assembly RGSC v3.1 for Chr 1 Ch01 ref|NW_047563.2 (16485803-16477478 bp); for Chr 5 Ch05 ref|NW_047711.2 (31373351-31365021 bp); for Chr 7, Ch07 ref|NW_001084856.1 (4445102- 4453430); for Chr 8 Ch08 ref|NW_047797.1 (2950056-2958376 bp), for Chr 6 Ch06 ref|NW_047762.2 (30981005-30972694 bp). The element on Chr 20 was placed to, Ch20 emb|BX883050.1 (220893-229224 bp) strain Brown Norway (BN/ssNHsd), (GenBank BX883050). Remaining elements were placed to contig NW_047907.1 (1641297-1640869 bp) to unknown chromosome (Chr Un).

The following primers were used in the Transposon Display:

BfaI linker(+)	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
BfaI linker(-)	p-TAGTCCCTTAAGCGGAG-Amino
1st round linker	GTAATACGACTCACTATAGGGC
1st round retroPbsR	CTCGCCAATCTCTGGAGGACAA
2nd round Nested	AGGGCTCCGCTTAAGGGAC
2nd round retroLtrR	TTGGAGCTGTAACACTTGGAGCTGT

Detailed Protocol of the Transposon Display

1. Restriction digest of genomic DNA

Restriction mix (volumes to be multiplied by the # of samples)

10x NEB4 buffer 5 µl

BfaI (10 U/ml) 2 µl

dH_2O 43 µl

DNA 500 ng

Total 50 µl

Incubate for 3 hours at 37°C.

Qiagen purification, dilution with 30 µl buffer.

2. Ligation of the linker primer to the fragmented genomic DNA
Ligation mix (volumes to be multiplied by the # of samples)
10x ligation buffer 10 µl
dH_2O 48.3 µl
T4 DNA Ligase (40 U/µl) 1.7 µl
DNA 30 µl
Adapter (5 µM): 10 µl
Total 100 µl
Incubate for 10 hours at 16°C.
Purification by QIAquick PCR purification kit (QIAgen).

3. Pre-amplification
PCR mix (volumes to be multiplied by the # of samples)
10x PCR buffer 2.5 µl
$MgCl_2$ 2 µl
dNTPs (5 mM) 0.65 µl
Linker primer (5 µM) 2 µl
retroPbR primer (5 µM) 2 µl
Taq polymerase 0.25 µl
dH_2O 12.6 µl
DNA (15-fold dilution) 3 µl
Total 25 µl

PCR profile for the pre-amplification:
2' @ 72 °C
3' @ 94 °C
24 cycles of:
30" @ 94 °C
30" @ 58 °C
5' @ 72 °C

4. Primer labeling for the hot PCR
retroLtrR primer (50 pmol/µl) 6 µl
P^{32} (3000 Ci/mmol) 12 µl
10x PNK buffer 5 µl
T4 PNK (diluted) 4 U (=0.5 µl)
dH_2O 26.5 µl
Total 50 µl

5. Hot PCR
PCR mix (volumes to be multiplied by the # of samples)
10x PCR buffer 2.5 µl
$MgCl_2$ 2 µl
dNTPs (10 mM) 0.65 µl
retroLtrR* primer (0.5 pmol/µl) 2 µl

nested primer (5 µM) 2 µl
Taq polymerase 0.25 µl
dH_2O 10.6 µl
Total 20 µl

Add the 20 µl of PCR mix to 5 µl of a 15x diluted pre-amplification sample for a PCR volume of 25 µl.
PCR profile for the hot PCR:
3' @ 94 °C
13 cycles of:
30" @ 94 °C
30" @ 65 °C (with ΔT –0.7 °C per cycle)
60" @ 72 °C
followed by 18 cycles of:
30" @ 94 °C
30" @ 58 °C
60" @ 72 °C
5' @ 72 °C

6. Gel analysis
After PCR amplification, 25 µl of formamide loading buffer is added and samples are analyzed on 3.5% polyacrylamide gels using the Sequi-Gen system (Bio-Rad).

Phylogenetic characterization of the RnERV-K8e family
DotPlots in Fig. 2B were created using BLAST2 algorithm, with default parameters (window size = 11, threshold = 50, (Mismatch scores 1, -2), BLAST searches (BLASTN 2.2.19+) using the LTR sequences of the different subfamilies were carried out on the rat or mouse genome database (gpipe/10116/all_contig), gpipe/10090/all_contig, respectively at http://www.ncbi.nlm.nih.gov).

Characterization of RnERV-K8e-specific transcripts
Statistical significance was assessed using ANOVA on ΔCt values, post-hoc comparison by Tukey honest significant difference test for unequal N (see p values in the graph Fig. 4A).

The AFLP study was following the protocol published in (Bonne et al. 2003)
AFLP Adapters and Primers Used in This Study

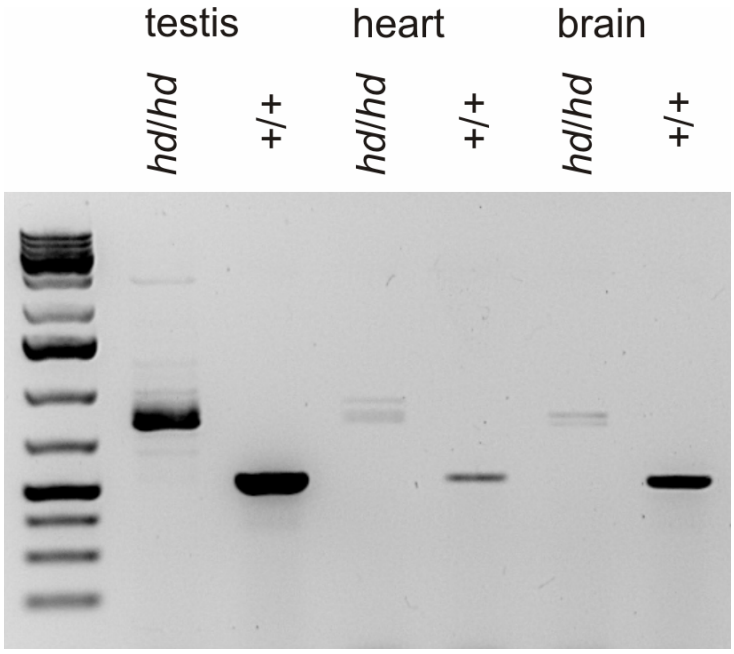
Adapters and primers used	Adapter and primer sequences
EcoRI-adapter1	5'-CTCGTAGACTGCGTACC-3'
EcoRI-adapter2	5'-AATTGGTACGCAGTCTAC-3'
EcoRI-A	5'-GACTGCGTACCAATTCA-3'
EcoRI-AAG	5'-GACTGCGTACCAATTCAAG-3'
EcoRI-ACA	5'-GACTGCGTACCAATTCACA-3'
EcoRI-ATA	5'-GACTGCGTACCAATTCATA-3'
TaqI-adapter1	5'-CGGTCAGGACTCAT-3'
TaqI-adapter2	5'-GACGATGAGTCCTGAC-3'
TaqI-A	5'-GATGAGTCCTGACCGAA-3'
TaqI-AAC	5'-GATGAGTCCTGACCGAAAC-3'
TaqI-AAG	5'-GATGAGTCCTGACCGAAAG-3'
TaqI-AGA	5'-GATGAGTCCTGACCGAAGA-3'
TaqI-AGT	5'-GATGAGTCCTGACCGAAGT-3'
TaqI-ATC	5'-GATGAGTCCTGACCGAATC-3'
TaqI-ATG	5'-GATGAGTCCTGACCGAATG-3'

Primer combinations: EcoRI-AAG/TaqI-AAG, EcoRI-AAG/TaqI-ATC, EcoRI-AAG/TaqI-AAC, EcoRI-AAG/TaqI-AGT, EcoRI-ACA/TaqI-AAC, EcoRI-ACA/TaqI-AGA, EcoRI-ACA/TaqI-AGT, EcoRI-ACA/TaqI-ATC, EcoRI-ATA/TaqI-AAC, EcoRI-ATA/TaqI-AAG, EcoRI-ATA/TaqI-ATG, EcoRI-ATA/TaqI-AGT.

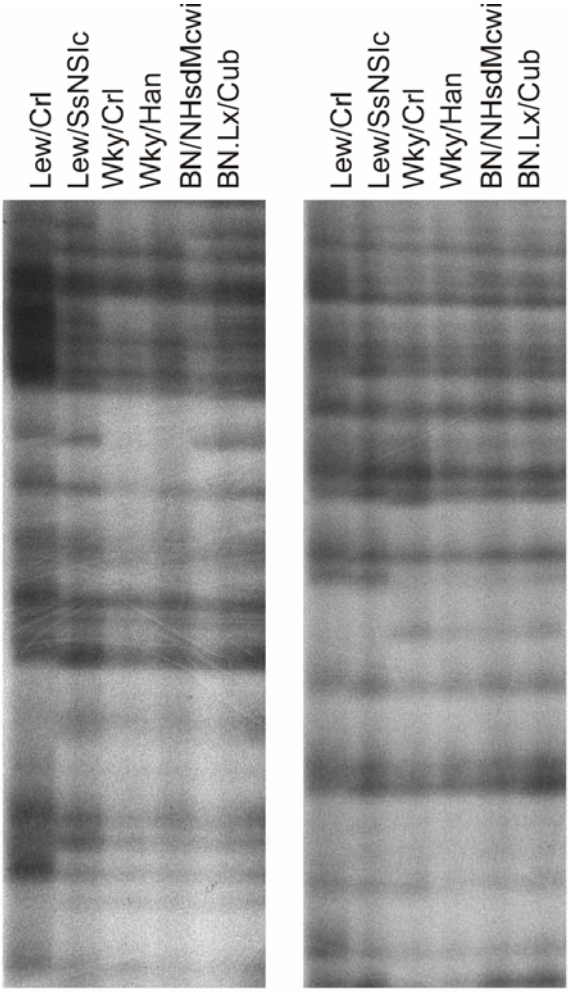
Supplemental Figure S1. Expression of both wild-type and alternatively spliced mutant centrobin is highest in the testis. The alternatively spliced centrobin transcripts of similar size could be detected from heart and brain. Possibly, due to the lower transcription levels in spleen, kidney and liver, no convincing signals were detectable.

Supplemental Figure S2. AFLP analyses. Examples of autoradiographs of AFLP fingerprint analysis of the Lew/Crl, Lew/SsNSIc, WKY/Crl, WKY/Han, BN/NHsdMcwi, and BN.Lx/Cub strains using primer combinations EcoR-ATA/Taq-AGT (left)and EcoR-ATA/Taq-AAG (right). The table below shows the estimated divergence using 12 different primer combinations (see Supplemental Material and Methods). Only clear bands were scored.

Supplemental Figure S1



Supplemental Figure S2



#	Strain	# of total bands	# of div. bands	% divergence
6	BN.Lx/Cub	446	1	0.2
7	BN/NHsdMcwi	446		
14	Wky/Han	447	5	1.1
17	Wky/Crl	447		
4	Lew/SsNSlc	440	5	1.1
5	Lew/Crl	440		

Supplemental Table S1. Sequences of identified cryptic splice signals provided by the ERV-K8e insertion into intron 10 of the *Cntrob* gene.

ERV position*	site type	junction sequence**
332	acceptor	aacacttAGggctgg
366	acceptor	tcaacaaAGaaggtc
380	acceptor	ctcctgcAGtaaagg
531	donor	gttcccgGTaaggga
557	donor	tctggagGTgagcag
3751	acceptor	ctccgcaAGatagcg
3755	acceptor	gcaagatAGcgtgcg
3835	donor	cctcaagGTatggca
5190	acceptor	cttcaacAGaaattt
5871	acceptor	tttccgcAGaatcaa
5983	donor	aatacagGTtaattc
6413	donor	gggcaagGTaactat

* splicing occurs at 3´ end of the nucleotide in position listed (for donor sites it is the last base of exon, for acceptor sites the last base of intron).

** junction consensus sequences GT and AG are capitalized

Supplemental Table S2. Sequence variants in the ERV-K8e full-length copies (ERV and flanking LTR`s).

Chr 1	75, 94, 162, 176(-1), 227(-1), 246(-2), 276(+1), 301(-1), 329(-1), 340(-1), 346, 424, 1209(+1), 8089, 8245
Chr 5	2377(stop), 3002(-1)
Chr 6	15, 46, 187, 208(-3),218(-43), 424, 879(stop), 1280(-1), 3095(stop), 3134(stop), 5717(-1),6732(+1), 8035, 8210, 8313, 8323
Chr 7	68(+12), 92, 196(-1), 240, 425, 1215(+2), 2044(-35), 2496(stop), 3095(stop), 3134 (stop), 5567(stop), 5934(stop), 7998(-1), 8090, 8143(-13), 8189, 8204(-16), 8302, 8329
Chr 8	46, 398, 1215(+1), 2629(-1), 3060(stop), 3538(stop), 4143(stop), 4268(stop), 5001(+1), 5897(stop), 7258(-5), 8238
Chr Un	2676(+2), 2749(-1), 3679(stop)
Chr 10	1919(stop)

The number of nucleotides inserted (+) or deleted (-) are indicated in the brackets, basepairs without further explanation indicates point mutations.

Supplemental References

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, and Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110:** 462-467.

Bonne AC, den Bieman MG, Gillissen GF, Kren V, Krenova D, Bila V, Zidek V, Kostka V, Musilova A, Pravenec M, et al. 2003. Genetic map of AFLP markers in the rat (Rattus norvegicus) derived from the H x B/Ipcv and B x H/Cub sets of recombinant inbred strains. *Biochem Genet.* **41:** 77-89.