

Supplementary Table Ia: L1Xplorer: Human L1 Annotations, Classifications and References

Nr	Name	L1 Part	Description	References
1	arr260	ORF1 aa	ARR260 Checks the sequence for intactness of ARR260 locus in ORF1.	(1)
2	caiorf1	ORF1 nt	ORF1 CAI Calculates the Codon Adaptation Index of ORF1.	(2)
3	caiorf2	ORF2 nt	ORF2 CAI Calculates the Codon Adaptation Index of ORF2.	(2)
4	cpg	All nt	CPG Islands Annotates CPG islands found in the L1.	(3)
5	d145	ORF2 aa	D145 Checks the sequence for intactness of D145 in ORF2.	(4)
6	d205	ORF2 aa	D205 Checks the sequence for intactness of D205 in ORF2.	(4)
7	e43	ORF2 aa	E43 Checks the sequence for intactness of E43 locus in ORF2.	(4)
8	ensgenes	All nt	Genome Vicinity Determines neighbouring genes with intron structure from the correspondent Ensembl release (only for L1Base)	(5)
9	ensrepeats	-	Repeat Annotation Determines repeats within the L1 region from the correspondent Ensembl release (only for L1Base)	(6)
10	enssnp	All nt	Look for SNPs Adds SNPs from the correspondent Ensembl release to the annotation, and shows aminoacid exchanges (only for L1Base)	(5,7)
11	fadd700	ORF2 aa	FADD700 Checks the sequence for intactness of FADD700 locus in ORF2.	(4)

Nr	Name	L1 Part	Description	References
12	gccontent	All nt	G-C Content Calculates the %G-C of the L1 Element in a 50nt window.	(3)
13	hmkk1091	ORF2 aa	HMKK1091 Checks the sequence for intactness of HMKK1091 locus in ORF2.	(1)
14	i1220m	ORF2 aa	I1220 Checks the sequence for intactness of I1220 locus in ORF2. I to M mutation is known to limit retrotransposition rate.	(8)
15	intactness	-	Intactness Score Calculates an intactness measure for the L1 annotated. Each feature putatively contributing to activity awards for one point when conserved.	
16	l1mvspa	ORF2 nt	L1M/L1PA Discrimination Checks through a HMM a diagnostic part of ORF2, which discriminates L1M and L1PA families.	(6)
17	n14	ORF2 aa	N14 Checks the sequence for intactness of N14 locus in ORF2.	(4)
18	n147	ORF2 aa	N147 Checks the sequence for intactness of N147 in ORF2.	(4)
19	nkx25	5' UTR nt	TF nkx-2.5 Checks the sequences for intactness of the first putative nkx2-5 site.	Motif taken from (9), authors prediction
20	nkx25b	5' UTR nt	TF nkx-2.5B Checks the sequences for intactness of the second putative nkx2-5 site.	Motif taken from (9), authors prediction
21	orf1acont	ORF1 nt	ORF1 %A Calculates %A for ORF1	(10)

Nr	Name	L1 Part	Description	References
22	orf1tcont	ORF1 nt	ORF1 %T Calculates %T for ORF1	(10)
23	orf2acont	ORF2 nt	ORF2 %A Calculates %A for ORF2	(10)
24	orf2tcont	ORF2 nt	ORF2 %T Calculates %T for ORF2	(10)
25	orfclass	All nt	ORF StartStop Checks the ORFs of the L1 for presence of valid methionine start codons and stop codons. (M*,M*) indicates conserved start and stop codons in both ORFs.	(1)
26	orfsacont	ORF1&2 nt	ORF1&2 %A Calculates %A for ORF1&2	(10)
27	orfstcont	ORF1&2 nt	ORF1&2 %T Calculates %T for ORF1&2	(10)
28	polya	3' nt	Poly-A Identifies the position of the Poly-A signal and determines the length of the pure Poly-A tail, as well as the length of an estimated Poly-A Tail (containing mutations). Calculates as well a Kimura distance to a pure poly-A tail of the length of the estimated sequences. Distance is not available if mutation rate is too high.	(11)
29	polyasig	3' nt	Poly-A Signal Tests the sequence for an intact Poly-A signal. Poly-A signals that fulfill the consensus AATAAA or AATTAAA will be considered as valid signals.	(12)
30	r363g	ORF2 aa	R363 Checks the sequence for intactness of R363 locus in ORF2. R to G mutation is known to limit retrotransposition rate.	(8)
31	rekg235	ORF1 aa	REKG235 Checks the sequence for intactness of REKG235 locus in ORF1.	(1)
32	runx3_1	5' UTR nt	Runx3 Site Checks for the presence of an intact RUNX3 binding motif in the 5'UTR of the L1 Element.	(13)

Nr	Name	L1 Part	Description	References
33	runx3_3	5' UTR nt	Runx3 ASP Checks for the presence of an intact RUNX3 Anti-Sense-Promoter binding motif in the 5'UTR of the L1 Element.	(13)
34	s12591	ORF2 aa	S1259 Checks the sequence for intactness of S1259 locus in ORF2. S to L mutation is known to limit retrotransposition rate.	(8)
35	sdh228	ORF2 aa	SDH228 Checks the sequence for intactness of SDH228 locus in ORF2.	(4)
36	sry1	5' UTR nt	SRY Site 1 Checks for the presence of an intact first SRY1 binding motif in the 5'UTR of the L1 Element.	(14)
37	sry2	5' UTR nt	SRY Site 2 Checks for the presence of an second first SRY1 binding motif in the 5'UTR of the L1 Element.	(14)
38	sss1096	ORF2 aa	SSS1096 Checks the sequence for intactness of SSS1096 locus in ORF2.	(1)
39	t192	ORF2 aa	T192 Checks the sequence for intactness of T192 in ORF2.	(4)
40	ta	3' UTR nt	Ta SSVs Inspects the family determining Ta locus of the L1 element identified (positions 5930-2, 6015, relative to L1.2, M80343).	(15-18)
41	ta01	ORF2 nt	Ta0/Ta1 SSVs Inspects the family determining Ta0-1 locus in ORF2 of the L1 element identified (positions 5536,5539 relative to L1.2, M80343)	(15-17)
42	tand	5' UTR nt	Ta1-nd/d Checks for the family determining Tad/nd deletion in the 5'UTR of the L1 element identified (positions 74-5, relative to L1.2, M80343)	(16)
43	tsdsearch	All nt	find TSDs Searches for target-site-duplications flanking the L1 Element. Only TSDs of a length greater than 10 will be considered as valid TSDs.	(19,20)

Nr	Name	L1 Part	Description	References
44	y115	ORF2 aa	Y115 Checks the sequence for intactness of Y115 locus in ORF2.	(4)
45	ypakls282	ORF1 aa	YPAKLS282 Checks the sequence for intactness of YPAKLS282 locus in ORF1.	(1)
46	yy1	5' UTR nt	YY1 BoxA+BoxA Checks for the presence of an intact YY1 binding motif at the very beginning of the L1 Element.	(21-23)

Supplementary Table Ib: L1Xplorer: Mouse L1 Annotations, Classifications and References

Nr	Name	L1 Part	Description	References
1	ensrepeats	-	Repeat Annotation	(6)
2	intactness	-	Intactness Score Calculates an intactness measure for the L1 annotated. Each feature putatively contributing to activity awards for one point when conserved.	
3	polya	3' nt	Poly-A Identifies the position of the Poly-A signal and determines the length of the pure Poly-A tail, as well as the length of an estimated Poly-A Tail (containing mutations). Calculates as well a Kimura distance to a pure Poly-A tail of the length of the estimated sequences. Distance is not available when mutation rate is too high.	(11)
4	polyasig	3' nt	PolyA Signal Test the sequence for an intact Poly-A signal. Poly-A signals that fulfill the consensus AATAAA or AATTAAA will be considered as valid signals.	(12)
5	sim6642	ORF1 nt	66/42 Monomers Searches the L1 for Mouse L1 ORF1 monomers (LPR, aka 66-42-42 monomers).	(24)
6	simaf	5' UTR nt	Search Monomers Searches the L1 for Mouse L1 promoter monomers. Detected Monomers are A-Monomer I, A-Monomer II, A-Monomer III, A-Monomer IV, A-Monomer V, A-Monomer VI, T.F-Monomer, F-Monomer, G.F-Monomer.	(24,25)
7	orfclass	All nt	ORF StartStop Checks the ORFs of the L1 for presence of valid methionine start codons and stop codons. (M*,M*) indicates conserved start and stop codons in both ORFs.	
8	gccontent	All nt	G-C Content Calculates the %G-C of the L1 Element in a 50nt window.	(3)
9	enssnp	All nt	Look for SNPs Adds SNPs from the correspondent Ensembl release to the annotation, and shows aminoacid exchanges (only for L1Base)	(5,7)

Nr	Name	L1 Part	Description	References
10	ensgenes	All nt	Genome Vicinity Determines neighbouring genes with intron structure from the correspondent Ensembl release (only for L1Base)	(5)
11	cpg	All nt	CPG Islands Annotates CPG islands found in the L1.	(3)
12	tsdsearch	All nt	find TSDs Searches for target-site-duplications flanking the L1 Element. Only TSDs of a length greater than 10 will be considered as valid TSDs.	(19,20)
13	sim6642	ORF1 nt	66/42 Monomers Searches the L1 for Mouse L1 ORF1 monomers (LPR, aka 66-42-42 monomers).	(24)
14	orf1acont	ORF1 nt	ORF1 %A Calculates %A for ORF1	(10)
15	orf1tcont	ORF1 nt	ORF1 %T Calculates %T for ORF1	(10)
16	caiorf1	ORF1 nt	ORF1 CAI Calculates the Codon Adaptation Index of ORF1.	(2)
17	orfsacont	ORF1&2 nt	ORF1&2 %A Calculates %A for ORF1&2	(10)
18	orfstcont	ORF1&2 nt	ORF1&2 %T Calculates %T for ORF1&2	(10)
19	orf2acont	ORF2 nt	ORF2 %A Calculates %A for ORF2	(10)
20	orf2tcont	ORF2 nt	ORF2 %T Calculates %T for ORF2	(10)
21	caiorf2	ORF2 nt	ORF2 CAI Calculates the Codon Adaptation Index of ORF2.	(2)
22	l1mvspa	ORF2 nt	L1M/L1PA Discrimination Checks through a HMM a diagnostic part of ORF2, which discriminates L1M and L1PA families.	(6)

Supplementary Table Ic: L1Xplorer: Rat L1 Annotations, Classifications and References

Nr	Name	L1 Part	Description	References
1	ensrepeats	-	Repeat Annotation	(6)
2	intactness	-	Intactness Score Calculates an intactness measure for the L1 annotated. Each feature putatively contributing to activity awards for one point when conserved.	
3	polya	3' nt	Poly-A Identifies the position of the Poly-A signal and determines the length of the pure Poly-A tail, as well as the length of an estimated Poly-A Tail (containing mutations). Calculates as well a Kimura distance to a pure Poly-A tail of the length of the estimated sequences. Distance is not available when mutation rate is too high.	(11)
4	polyasig	3' nt	PolyA Signal Test the sequence for an intact Poly-A signal. Poly-A signals that fulfill the consensus AATAAA or AATTAAA will be considered as valid signals.	(12)
5	mlvi2	3' UTR nt	L1.3/L1.4/L1mlvi2 Discriminates via SSVs between L1.3, L1.4 and L1mlvi2.	(26)
6	orfclass	All nt	ORF StartStop Checks the ORFs of the L1 for presence of valid methionine start codons and stop codons. (M*,M*) indicates conserved start and stop codons in both ORFs.	
7	gccontent	All nt	G-C Content Calculates the %G-C of the L1 Element in a 50nt window.	(3)
8	enssnp	All nt	Look for SNPs Adds SNPs from the correspondent Ensembl release to the annotation, and shows aminoacid exchanges (only for L1Base)	(5,7)
9	ensgenes	All nt	Genome Vicinity Determines neighbouring genes with intron structure from the correspondent Ensembl release (only for L1Base)	(5)

Nr	Name	L1 Part	Description	References
10	cpg	All nt	CPG Islands Annotates CPG islands found in the L1.	(3)
11	tsdsearch	All nt	find TSDs Searches for target-site-duplications flanking the L1 Element. Only TSDs of a length greater than 10 will be considered as valid TSDs.	(19,20)
12	orf1acont	ORF1 nt	ORF1 %A Calculates %A for ORF1	(10)
13	orf1tcont	ORF1 nt	ORF1 %T Calculates %T for ORF1	(10)
14	caiorf1	ORF1 nt	ORF1 CAI Calculates the Codon Adaptation Index of ORF1.	(2)
15	orfsacont	ORF1&2 nt	ORF1&2 %A Calculates %A for ORF1&2	(10)
16	orfstcont	ORF1&2 nt	ORF1&2 %T Calculates %T for ORF1&2	(10)
17	orf2acont	ORF2 nt	ORF2 %A Calculates %A for ORF2	(10)
18	orf2tcont	ORF2 nt	ORF2 %T Calculates %T for ORF2	(10)
19	caiorf2	ORF2 nt	ORF2 CAI Calculates the Codon Adaptation Index of ORF2.	(2)
20	l1mvspa	ORF2 nt	L1M/L1PA Discrimination Checks through a HMM a diagnostic part of ORF2, which discriminates L1M and L1PA families.	(6)

References

1. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H., Jr. (1996) High frequency retrotransposition in cultured mammalian cells *Cell*, **87**, 917-927.
2. Sharp, P.M. and Li, W.H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications *Nucleic Acids Res*, **15**, 1281-1295.
3. Furano, A.V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons *Prog Nucleic Acid Res Mol Biol*, **64**, 255-294.
4. Feng, Q., Moran, J.V., Kazazian, H.H., Jr. and Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition *Cell*, **87**, 905-916.
5. Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J. *et al.* (2004) Ensembl 2004 *Nucleic Acids Res*, **32 Database issue**, D468-470.
6. Smit, A.F., Toth, G., Riggs, A.D. and Jurka, J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences *J Mol Biol*, **246**, 401-417.
7. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation *Nucleic Acids Res*, **29**, 308-311.
8. Lutz, S.M., Vincent, B.J., Kazazian, H.H., Jr., Batzer, M.A. and Moran, J.V. (2003) Allelic heterogeneity in LINE-1 retrotransposition activity *Am J Hum Genet*, **73**, 1431-1437.
9. Durocher, D., Chen, C.Y., Ardati, A., Schwartz, R.J. and Nemer, M. (1996) The atrial natriuretic factor promoter is a downstream target for Nkx-2.5 in the myocardium *Mol Cell Biol*, **16**, 4648-4655.
10. Han, J.S. and Boeke, J.D. (2004) A highly active synthetic mammalian retrotransposon *Nature*, **429**, 314-318.
11. Roy-Engel, A.M., Salem, A.H., Oyeniran, O.O., Deininger, L., Hedges, D.J., Kilroy, G.E., Batzer, M.A. and Deininger, P.L. (2002) Active Alu element "A-tails": size does matter *Genome Res*, **12**, 1333-1344.
12. Boeke, J.D. (1997) LINEs and Alus--the polyA connection *Nat Genet*, **16**, 6-7.
13. Yang, N., Zhang, L., Zhang, Y. and Kazazian, H.H., Jr. (2003) An important role for RUNX3 in human L1 transcription and retrotransposition *Nucleic Acids Res*, **31**, 4929-4940.
14. Tchenio, T., Casella, J.F. and Heidmann, T. (2000) Members of the SRY family regulate the human LINE retrotransposons *Nucleic Acids Res*, **28**, 411-415.
15. Salem, A.H., Myers, J.S., Otieno, A.C., Watkins, W.S., Jorde, L.B. and Batzer, M.A. (2003) LINE-1 preTa elements in the human genome *J Mol Biol*, **326**, 1127-1146.

16. Boissinot, S., Chevret, P. and Furano, A.V. (2000) L1 (LINE-1) retrotransposon evolution and amplification in recent human history *Mol Biol Evol*, **17**, 915-928.
17. Boissinot, S., Entezam, A., Young, L., Munson, P.J. and Furano, A.V. (2004) The insertional history of an active family of L1 retrotransposons in humans *Genome Res*, **14**, 1221-1231.
18. Ovchinnikov, I., Rubin, A. and Swergold, G.D. (2002) Tracing the LINEs of human evolution *Proc Natl Acad Sci U S A*, **99**, 10522-10527.
19. Szak, S.T., Pickeral, O.K., Landsman, D. and Boeke, J.D. (2003) Identifying related L1 retrotransposons by analyzing 3' transduced sequences *Genome Biol*, **4**, R30.
20. Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D. and Boeke, J.D. (2002) Molecular archeology of L1 insertions in the human genome *Genome Biol*, **3**, research0052.
21. Kurose, K., Hata, K., Hattori, M. and Sakaki, Y. (1995) RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system *Nucleic Acids Res*, **23**, 3704-3709.
22. Becker, K.G., Swergold, G.D., Ozato, K. and Thayer, R.E. (1993) Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element *Hum Mol Genet*, **2**, 1697-1702.
23. Hyde-DeRuyscher, R.P., Jennings, E. and Shenk, T. (1995) DNA binding sites for the transcriptional activator/repressor YY1 *Nucleic Acids Res*, **23**, 4457-4465.
24. Goodier, J.L., Ostertag, E.M., Du, K. and Kazazian, H.H., Jr. (2001) A novel active L1 retrotransposon subfamily in the mouse *Genome Res*, **11**, 1677-1685.
25. Schichman, S.A., Adey, N.B., Edgell, M.H. and Hutchison, C.A., 3rd. (1993) L1 A-monomer tandem arrays have expanded during the course of mouse L1 evolution *Mol Biol Evol*, **10**, 552-570.
26. Usdin, K., Chevret, P., Catzeflis, F.M., Verona, R. and Furano, A.V. (1995) L1 (LINE-1) retrotransposable elements provide a "fossil" record of the phylogenetic history of murid rodents *Mol Biol Evol*, **12**, 73-82.