

Supplemental Table 1: SVA elements on human chromosome 19

Full-length elements: 56

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee (primate panel)	poly A signal
H19_70	5652312 - 5654042	aaaaaat	+	19 (788 bp)	+	D	fixed		-	aataaa
H19_67	8710091 - 8712109	agaagg/tgcaga	+	24 (1033 bp)	+	D			+	aataaa
H19_66	9454300 - 9455983	aaaaagaaa	+	18 (722 bp)	+	C			-	ataaaa
H19_64	9650980 - 9652614	aaaaatagc	+	17 (675 bp)	+	D			+	ataaaa
H19_62	11247783 - 11249224	aaaaattaaaa	+	12 (460 bp)	+	B			+	aataaa
H19_58	12438919 - 12440536	agaatgctatcacg	+	16 (644 bp)	+	D			+	aataaa
H19_JR54	17889406 - 17891336	aaaatacaaaaatg	+ (22bp internal deletion)	24 (984 bp)	+	D			-	ataaaa
H19_53	17976816 - 17978472	tttttaa	+	18 (729 bp)	+	D			-	ataaaa
H19_52	19741249 - 19742692	aaaaaggtaaatta	+	11 (420 bp)	+	B			+	aataaa
H19_51	20035331 - 20037113	aaaaagaa	+	25 (986 bp)	+	D			+	ataaaa
H19_45	20147128 - 20148671	aattaatagggcttc	+	15 (586 bp)	+	D			+	aataaa
H19_55	20185643 - 20187137	aaatcccc	+	12 (488 bp)	+	D			-	aataaa
H19_44	20267564 - 20269090	aaaatttgcga	+	14 (543 bp)	+	D	het		-	aataca
H19_43	20926666 - 20928127	aatagtgtactt	+	12 (470 bp)	+	C			+	aataaa
H19_37	21186572 - 21188133	agtgatatta	+	15 (581 bp)	+	B			poly (G,O poly)	aataaa
H19_32	21233527 - 21236685	gttgatgca	+	50 (1965 bp)	+	A			+	aataaa
H19_31	21417606 - 21419069	gtagtaaaaa	+	14 (546 bp)	+	B	poly	Afr. 0.975 As. 0.8 Eur. 0.825 SA 1.0	poly*	aataaa
H19_155	21485206 - 21486762	aagaaattacc	+	16 (589 bp)	+	D			+	aataaa
H19_25	21588922 - 21590197	aaaaccagtcct	+	9 (337 bp)	+	B			-	aataaa
H19_24	21681936 - 21683377	gaaagtttaattgtt	+	11 (432 bp)	+	C	fixed		-	aataaa
H19_27	21694876 - 21697507	cttaaaaatt	+	35 (1458 bp)	+	E	poly	Afr. 0.425 As. 0.625 Eur. 0.6 SA 0.775	-	aataaa
H19_JR23	21749754 - 21751329	aaaacactgacca	+	14 (547 bp)	+	D			+	aataaa
H19_22	22080678 - 22083527	aaaacttagcaaaagtcacaatata	+	40 (1708 bp)	+	F	fixed		-	aataaa
H19_19	22295675 - 22297098	aaaaaattag	+	11 (418 bp)	+	B			+	aataaa
H19_18	22674437 - 22675996	aaaaagagcca	+	15 (587 bp)	+	C			+	aataaa
H19_12	22845509 - 22847127	aatgaacaaat	+	16 (649 bp)	+	D			-	aataaa
H19_10	23134863 - 23136508	ttatgca	+	18 (687 bp)	+	B			+	aataaa
H19_9	23264010 - 23265502	ataaaaaa	+	14 (540 bp)	+	B			poly	aataaa
H19_5	23393238 - 23394902	aaaaatgttatt	+	18 (673 bp)	+	C			poly	aataac
H19_3	23542495 - 23543624	aaaataattaaac	+	5 (201 bp)	+	A			+	aataaa
H19_2	23938009 - 23939783	ttactctt	+	22 (812 bp)	+	C			+	ataaaa
H19_112	39261213 - 39263397	agaagacaca	+	36 (1557 bp)	+	F	fixed		-	aataaa
H19_114	39314624 - 39316492	aagaatataattta	+	23 (943 bp)	+	D	fixed		-	aataaa
H19_115	40104161 - 40105819	ttatgtataa	+	17 (673 bp)	+	D	fixed		-	aattaa
H19_116	41251087 - 41252812	aaaaaatgcaatcctc	+	18 (728 bp)	+	C			+	aataaa
H19_117	41673188 - 41674918	aataataagtatatat	+	19 (764 bp)	+	D			-	aataaa
H19_106	43305749 - 43307533	aaaaaacaaaa	+	21 (830 bp)	+	D	fixed		-	aattaa
H19_140	43979629 - 43981321	aaaatacaaaaata	+	19 (781 bp)	+	D			-	aataaa
H19_138	44037403 - 44039234	aagaacactagcc	+	20 (807 bp)	+	D			-	aataaa
H19_134	45304930 - 45307615	aaaactattaattgg	+	37 (1576 bp)	+	F			-	aataaa
H19_133	45314938 - 45317166	aaaaaaaaaaca	+	30 (1277 bp)	+	F	poly	Afr. 0.825 As. 0.775 Eur. 0.95 SA 0.9	-	aataaa

H19_122	47805658 - 47807303	agaa	+	17 (698 bp)	+	D			-	aataaa
H19_121	47834773 - 47836580	aaaaatggcaaca	+	21 (851 bp)	+	C			+	aataaa
H19_104	49399228 - 49400897	aagaaatataaaaaata	+	15 (579 bp)	+	B			+	aataaa
H19_103	49684084 - 49685745	aaaaaaactgac	+	13 (526 bp)	+	D			-	aataaa
H19_152	52204180 - 52205744	aaaaatacaaaaattag	+	15 (584 bp)	+	D	fixed		-	aataaa
H19_100	55280737 - 55282430	aagaaaacaaacaacc	+	17 (667 bp)	+	D			-	ataaaa
H19_99	55335376 - 55337153		+	13 (523 bp)	+	D			-	aataaa
H19_98	55585642 - 55586789	aaaaaaaa	+	6 (222 bp)	+	B			+	aataaa
H19_93	5777819 - 57789436	aaaaatacaaaaattag	+	16 (631 bp)	+	D	poly	Afr. 0.5 As. 0.675 Eur. 0.55 SA 0.55	-	aataaa
H19_92	57871090 - 57872222	aacaat	+	7 (337 bp)	+	C-F (gene conversion)			+	aataaa
H19_91	57914802 - 57916942	aaaaaaag	+	36 (1005 bp)	+	B			poly (G poly)	aataaa
H19_88	58417296 - 58418739	aaaaaaaa	+	12 (464 bp)	+	D	fixed		-	aataaa
H19_87	58464444 - 58465845	gaaactagtc	+	10 (375 bp)	+	D	het		-	aataaa
H19_83	63349684 - 63351601	gagaggaagga	+	23 (935 bp)	+	D			-	ataaaa
H19_82	63607398 - 63609570	aaaatctgttctct	+	29 (1205 bp)	+	D			-	aataaa

5' truncated elements: 30

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee (primate panel)	poly A signal
H19_72	1449618 - 1451095	agacaag	+/Trunc.	18 (728 bp)	+	D			-	aataaa
H19_71	3949770 - 3950479	gaccagcctggcc	-	4 (161 bp)	+	B			+	aatgaa
H19_68	6464263 - 6464693	aatagctcggctatgtct	-	-	+/Trunc.	B			+	aataaa
H19_63	10449752 - 10449955	aagaaaatctgtgtctt	-	-	+/Trunc.	B			+	aataaa
H19_61	11274618 - 11275285	aaaaactctggccctca	-	3 (88 bp)	+	D			-	aataaa
H19_57	12593515 - 12594256	aaaacctccaacctc	-	3 (129 bp)	+	F	fixed		-	aataaa
H19_46	20123828 - 20124517	agaacgctgg	-	3 (140 bp)	+	C			-	aataaa
H19_42	21022800 - 21023541	aaqatatttatg	-	4 (152 bp)	+	C			+	aataaa
H19_36	21525276 - 21526800	aaaataagccagaaata	+/Trunc.	17 (682 bp)	+	D			+	aataaa
H19_21	22079324 - 22080694	aaaaatgcca	+/Trunc.	17 (838 bp)	+	C			+	aattta
H19_14	23108966 - 23109918	aaqctctcaggattgtcc	-	9 (389 bp)	+	E			-	aataaa
H19_7	23340207 - 23340951	aaagacttct	-	5 (204 bp)	+	D			+	aataaa
H19_6	23343981 - 23344689	aaaaaaaaaaaa	-	4 (163 bp)	+	C			+	aataaa
H19_15	41707252 - 41707479	aaaaactagcatg	-	-	+/Trunc.	ambiguous			-	aataaa
H19_137	43545953 - 43547234	aggaaataggctaacggg	+/Trunc.	14 (550bp)	+/del.	D			+	aataaa
H19_139	43989978 - 43990649	aaaataggcaggaaa	-	3 (112 bp)	+	A			+	aataaa
H19_135	44240973 - 44241322	taaaaatacaa	-	-	+/Trunc.	F	poly	Afr. 0.925 As. 1.0 Eur. 1.0 SA 1.0	-	aataaa
H19_131	45999384 - 45999655	aagaacctgtggc	-	-	+/Trunc.	ambiguous	fixed		-	aataaa
H19_127	46300548 - 46302048	agaaaaccccaacatctc	-	21 (911 bp)	+	F			-	aataaa
H19_125	46377452 - 46379613	aagataaaaaaaaaat	+/Trunc.	27 (1133 bp)	+	D			-	aataaa
H19_123	46963799 - 46964587	aaqaatgatctccca	-	5 (204 bp)	+	D			-	aataaa
H19_105	48744004 - 48744795	acaaacagtagcaggtt	-	5 (240 bp)	+	C or D			+	aaaaaa
H19_102	50116165 - 50117145	aaatgc/tgtgagccaca	-	11 (484 bp)	+	D			-	aataaa
H19_101	54052595 - 54053027	agaaactggcatgtct	-	-	+/Trunc.	C or D			+	aataaa
H19_97	55776892 - 55777580	ccaggc	-	5 (242 bp)	+	B/A			+	aataaa
H19_95	57147890 - 57148735	taaaatgcatccctt	-	8 (306 bp)	+	D	fixed		-	aataaa
H19_146	58782019 - 58782729	agaagccaagccagg	-	3 (152 bp)	+	D			-	aataaa
H19_86	59409950 - 59410648	aaaaacttatgattcc	-	4 (155 bp)	+	D			-	aataaa
H19_84	60111146 - 60111822	aaaaaaaaaaag	-	4 (144 bp)	+	B	fixed		-(G +)	aataaa
H19_85	60201757 - 60202363	aaqatgtaaaact	-	3 (161 bp)	+/Trunc.	ambiguous	fixed		-	aataaa

3' truncated elements: 8

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee (primate panel)	poly A signal	length of TD	source locus of TD
H19_59	12333471 - 12334718	aaaaattgga	+	15 (582 bp)	+/Trunc.	B or C			+	aagaaa		
H19_20	22677184 - 22678531	aaaatgtaaatg	+	18 (710 bp)	+/Trunc.	B			+	tataac		
H19_8	23568763 - 23570131	aaaaaagaaccagga	+	17 (641 bp)	+/Trunc.	D	fixed		-	aatgga		
H19_1	23911344 - 23912514	gttttc	+/Trunc.	19 (884bp)	+/Trunc.	D			-	aagaaa		
H19_111	35131293 - 35132708	aagtaac	+	17 (664 bp)	+/Trunc.	C			+	aacaga		
H19_129	46162570 - 46165303	aaaaaattataaaggg	+	48 (2004 bp)	+/Trunc.	E			-	aagaaa		
H19_147	58114024 - 58115614	aaaaaaaaaaaa	+	16 (624 bp)	+/Trunc.	E	poly	Afr. 0.275 As. 0.55 Eur. 0.125 SA 0.25	-	aacaga		
H19_142	62685545 - 62687083	aaaaattc	+	18 (739 bp)	+/Trunc.	D	poly	Afr. 0.8 As. 0.65 Eur. 0.75 SA 0.75	-	aaagaa	24	Chr.9:33761217 - 33761239

SVAs with 3' transduced sequences: 9

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee (primate panel)	poly A signal	length of TD	source locus of TD
H19_4	6305771 - 6307759	aaaaacttaaaaaa	+	18 (740 bp)	+	D			-	aataaa	220	Chr.2:85659919 - 85660140
H19_13	11647999 - 11649763	aaaaaactgaagctg	+	18 (711 bp)	+	B			+	cataaa	45	Chr.15:70755113 - 70755163
H19_17	15334153 - 15335952	agaacacagatttcag	+	22 (900 bp)	+	D			-	aataaa	11	
H19_23	20109058 - 20110682	aagaaaaagctcag	+	11 (422 bp)	+	D			-	aataaa	209	Chr.19:12181213 - 12181427
H19_26	21580829 - 21582129	ccttcctctgg	-	11 (435 bp)	+	D			-	aataata	297	Chr.6:14214579 - 14214872
H19_54	21838914 - 21841564	aaaaacta	+	29 (1215 bp)	+	F			-	aataaa	446	Chr.11:105262106 - 105262551
H19_73	23507304 - 23509050	aagagtactcaggtg	+	15 (589 bp)	+	D			+	aataaa	115	Chr.1:39567915 - 39568024
H19_120	47870808 - 47872687	agaaatggattggtta	-	9 (366 bp)	+	D			-	aataaa	944	Chr.2:85627851 - 85628799
H19_130	58005895 - 58008338	aaaaaaaga	+	24 (965 bp)	+	E	fixed		-	attaaa	462	Chr.14:65277341 - 65277807

SVAs with 5' transduced sequences: 6 (7 elements if 3'truncated H19_142 is included)

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee (primate panel)	poly A signal	length of TD	source locus of TD
H19_56	13673729 - 13676405	aaaagcctcattca	+/Trunc.	28 (1178 bp)	+	F			-	aataaa	385 (MER115/AluJo) 108 (AluSc) 364 (MAST-2)	Chr.9:130695439 - 130695822 Chr.9:33445301 - 33445422 Chr.1:45981401 - 45981764
H19_38	20503705 - 20506000	aaaatcaacaaca	+	30 (1239 bp)	+	D			-	aataaa	56	Chr.3:129867196 - 129867251
H19_11	22899898 - 22903267	gaaattactgtc	+/Trunc.	57 (2812 bp)	+	F			-	aataaa	283(MAST-2)	Chr.1:45981481 - 45981764
H19_132	45845099 - 45847858	taaaaatacaaaaa	+	22 (902 bp)	+	D			-	attaaa	881(L1-AluSp-AluSx-L1)	Chr.15:38760679 - 38761563
H19_153	50842212 - 50844334	aaaatatattg	+	18 (758 bp)	+	D			-	attaaa	390(AluSx-AluSg-FAM)	Chr.10:73300276 - 73301023
H19_76	58380982 - 58383779	aaaaaaaaacaaaacc	+	40 (1624 bp)	+	E			-	aataaa	32 18	Chr.6:111582462 - 111582493 Chr.10:43238004 - 43238021

SVAs with 5' and 3' transduced sequences: 3

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Chimpanzee (primate panel)	poly A signal	length of 5' TD length of 3' TD	source locus of 5' TD source locus of 3' TD
H19_47	20120146 - 20122599	aaattatgggagtc	+/Trunc.	32 (1368 bp)	+	F		-	aataaa	76 (MAST-2)	Chr.1:45981689 - 45981764
										428 (47-299AluSp-82)	Chr.10:101586803 - 101587235
H19_39	20376997 - 20379861	gaaaactgaatc	+/Trunc.	35 (1483 bp)	+	F		-	aataaa	362 (MAST-2)	Chr.1:45981403 - 45981764
										431 (47-302AluSp-82)	Chr.10:101586803 - 101587235
H19_108	35080630 - 35084954	aaaaataaaaatta	+/Trunc.	64 (2835 bp)	+	F		-	aataaa	148 (134 AluSc-14)	Chr.9:33445289 - 33445422
										364 (MAST-2)	Chr.1:45981401 - 45981764
										432 (52-300 AluSp-80)	Chr.10:101586803 - 101587235

SVA 2 elements: 4

Locus name	Position on chr.19	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Allele freq.	Chimpanzee	poly A signal
H19_65	9460757 - 9461255	aaqt	-	9 (368 bp)	-	-			+ (Chr.19)	attaaa
H19_35	21238980-21240142	tgga	-	25 (1068 bp)	-	-			+ (Chr.19)	attaaa
H19_124	46713172-46713880	attatc(c/t)tctcccc	-	14 (575bp)	-	-			? (Chr.7)	attaaa
H19_94	57217330-57217690	aagactgcctcctt	-	5 (185 bp)	-	-			+ (Chr.19)	attaaa

Source elements for 5' transduced SVAs on chromosome 19

Locus name	Position	TSD	Alu-like	VNTR - number (length)	SINE-R	Subfamily	Poly-morphic state	Chimpanzee	poly A signal	length of 5'TD	source locus of 5'TD
H1_3	1:26605984 - 26608163	agttaatatgtg	+	27 (1122bp)	+	D		-	aataaa	24	Chr.9:33761217 - 33761239
H3_2	3:15178506 - 15180806	caaatggcctttg	+	29 (1212bp)	+	D		-	aataaa	30	Chr.9:33761211 - 33761239
H10_10	10:43238007 - 43238853	aaaattaatctg	+	3 (112 bp)	+/only 3'part	E		-	aataaa		
H10_3	10:73298405 - 73300291	aacatgattagctc	+	21 (889 bp)	+	D		-	aataaa		
H15_4	15:38761549 - 38763760	aaaatactctgatgc	+	30 (1255 bp)	+	D		-	attaaa		

SVAs elements tested in primate and human population panels are highlighted in yellow. Elements tested in primate panel only are highlighted in light blue. Abbreviations: trunc. - truncated; het - element found almost exclusively in heterozygous state; poly - polymorphic; G - Gorilla; O - Orangutan; poly* - PCR results for element H19_31 indicate absence in chimpanzees, BLAST search using the preintegration site against the wgs database indicates presence/absence polymorphism in chimpanzees. +* - PCR results indicate presence of element H19_20 in chimpanzees. However, this result could not be confirmed using the UCSC genome browser or BLAST against the wgs database. Allele frequencies are given for the SVA-containing allele in case of presence/absence polymorphisms. Ethnic groups are as follows: Afr. - African; Eur. - European; As. - Asian; SA - South American