



Supplementary Figure 1. The frequency distribution of indel lengths. The x-axis represents the indel length and the frequency of each length (raw counts) is shown in y-axis, both axes are logarithmic.

Supplementary material

Table S1. Genes with 3n- indels that do not disrupt the reading frame in the chimpanzee transcript

Gene	Transcript	Annotated indels ^a	Verification in chimpanzee ^d	Macaque ^e
CU002_HUMAN	ENST00000339818	Ins H: aa184 (Ser) Del H: aa 325 (Leu)	Indel aa 325 confirmed	Both indels identical to C
CU045_HUMAN	ENST00000290130	Ins H/ Del C: aa 72 (Glu)	No sequence available	No available sequence
CU070_HUMAN ^b	ENST00000291634	Ins H/ Del C: aa 88-89 (Ser,Val)	Indel confirmed	No available sequence
CLIC6 ^b	ENST00000349499	Del C: aa 178-187 (Asp, Ser, Val, Glu, Ala, Glu, Gly, Arg, Val, Gly), aa 205-214 (Pro, Leu, Gly, Asp, Asn, Ile, Gln, Ala, Glu, Gly), aa 245-254 (Arg, Val, Gly, Asp, Ser, Val, Glu, Ala, Gly, Asp)	All three indels confirmed	All three indels identical to H
CLIC6 ^b	ENST00000338833	Del C: aa 178-187 (Asp, Ser, Val, Glu, Ala, Glu, Gly, Arg, Val, Gly), aa 205-214 (Pro, Leu, Gly, Asp, Asn, Ile, Gln, Ala, Glu, Gly), aa 245-254 (Arg, Val, Gly, Asp, Ser, Val, Glu, Ala, Gly, Asp)	All three indels confirmed	All three indels identical to H
IFNAR2	ENST00000342101	Ins H: aa 352 (Glu)	Indel confirmed	Indel identical to C
KRTAP10-4 ^b	ENST00000344145	Ins C: aa 168-172 (Pro, Val, Cys, Cys, Val), aa 196 (Gly), aa 203-205 (Leu, Ser, Cys), aa 250-254 (Ser, Ile, Cys, Cys, Lys) Del H: aa 273-277 (Ser, Ser, Cys,	Indels aa 168-172 and aa 196 confirmed.	Indel aa 273-277 identical to C, other indels identical to H

KRTAP13-4 ^b	ENST00000334068	Gln, Ser) Del H: aa 24-28 (Ser, Ser, Cys, Gly, Phe), aa 68- 77 (Ser, Tyr, Val, Glu, Ser, Ser, Pro, Cys, Gln, Tyr)	Both indels confirmed	Both indels identical to C
KRTAP15-6 ^b	ENST00000334067	Del C: aa 72 (Gln)	Indel confirmed	Indel identical to H
KRTAP19-7 ^b	ENST00000334849	Del C: aa 21 (Leu) Ins C: aa 32 (Gly), aa 38-39 (Cys, Gly)	Alignment unclear	All three indels identical to H
KRTAP6-3 ^{b,c}	ENST00000327378	Del H: aa 29-30 (Ser, Leu) Ins H/ Del C: aa 36 (Gly) Ins C: aa 47-48 (Tyr, Val)	Alignment unclear	Indel aa 29-30 identical to C, aa 47-48 identical to H
MCM3AP	ENST00000291688	Del C: aa 939 (Cys) Ins C: aa 1000 (Val)	Indel aa 1000 confirmed.	Both indels identical to H
Q9NSH7_HUMAN ^b	ENST00000332666	Del H: aa 112-113 (Arg, Pro)	No sequence available	Indel identical to C
Q9NSI5_HUMAN ^b	ENST00000328060	Del C: aa 92 (Gln) Ins H: aa 250 (Arg)	Both indels confirmed	Indel aa 92 identical to H, aa 250 identical to C
SLC19A1	ENST00000344609	Del C: aa 116 (Pro)	No sequence available	Indel identical to H
SLC19A1	ENST00000311124	Del C: aa 235 (Pro)	No sequence available	Indel identical to H
SON ^b	ENST00000321758	Del H: bp 7141	Indel confirmed	Indel identical to C
SYNJ1 ^b	ENST00000322229	Del C/ Ins H: aa 1296-1297 (Gln, Val) Ins H: aa 1357- 1358 (Asn, Val)	Indel aa 1357- 1358 confirmed.	Indel aa 1296- 1297 poor alignment, aa 1357-1358 identical C
TRPM2	ENST00000300481	Del C: aa15 (Glu)	Indel confirmed	Indel identical to H
TRPM2	ENST00000300482	Del C: aa15 (Glu)	Indel confirmed	Indel identical to H

^a Annotation of indels refer to the analysis of chimpanzee BAC clone sequences from RIKEN.

^b Indel not previously reported by Watanabe et al.

^c Not translated according to Ensembl.

^d Verified in the chimpanzee WGS assembly.

^e Sequence for macaque was obtained from the comparative track in the Ensembl browser (version 36).

Table S2. Primers used for re-sequencing of indels and substitutions causing premature stop codons.

Gene	Transcript(s)	Forward	Reverse
ABCG1	ENST00000340588	ATCCTGCACGTGGTCAGTC	CTCATCCTCACCTTCCTCCA
CU121_HUMAN	ENST00000332872	CGTTTCCTTAATGGCTGGAA	AGAGTCCAGGATGACGATGG
CU124_HUMAN	ENST00000327574	GTCACGTGGGCAGCTTCT	AGCTCACGGTGTGGAAAAAT
CU129_HUMAN	ENST00000331508	AAGTCCCATGTGCTGTCTC	CAGACTGTCCACTGCGAAGA
CU032_HUMAN	ENST00000327663	TGACTGGCATGAAGTATACAAACTATA	AACCTGGGCATACGGTCAT
CU090_HUMAN	ENST00000330490	TGCCAGCAATTGCATTTTAA	ATGCCATCCTCTCCAGCTAA
PIGP	ENST00000329667	CCTGGGGCTTTAGACAATC	GTCTGTGCGCCCCAGAACC
FAM3B	ENST00000337056	GAGCGTGCTCAAATCCAGTT	GCTTTCCAGGGCAAGTTGTA
Q6ZR03	ENST00000343732	CGGTTATCAACGTCCTCCTG	CTGGGAGGGAAGTCAGATCC
Q9NS10	ENST00000328881	CCATGACCAGGATTATGAACA	CGTTGTTGAGGTTGTCATCG