A single mutation in ACTR8 gene results in lineage-specific expression in primates

Se-Hee Choe
Korea Research Institute of Bioscience and Biotechnology

Sang-je Park
Korea Research Institute of Bioscience and Biotechnology

Hyeon-Mu Cho
Korea Research Institute of Bioscience and Biotechnology

Hye-Ri Park
Korea Research Institute of Bioscience and Biotechnology

Ja-Rang Lee
Korea Research Institute of Bioscience and Biotechnology

Young-Hyun Kim
Korea Research Institute of Bioscience and Biotechnology

Jae-Won Huh
Korea Research Institute of Bioscience and Biotechnology

Research article

Keywords: Alternative splicing, Alu, ACTR8, exonization, Lineage-specific

Abstract

Background: Alternative splicing (AS) generate various transcripts from a single gene and thus plays a significant role in transcriptomic diversity and proteomic complexity. Alu elements are primate-specific transposable elements (TEs) and can provide a donor or acceptor site for AS. In a study on TE-mediated AS, we recently identified a novel Alu Sz6-exonized ACTR8 transcript of a crab-eating monkey (Macaca fascicularis). In the present study, we sought to analyse the molecular mechanism of Alu Sz6-derived exonization in ACTR8 gene.

Results: We performed RT-PCR and genomic PCR using the tissue RNA and DNA samples from the crab-eating monkey and various other primates, including humans, to study Alu Sz6-derived exonization in ACTR8 gene and transcript variants expression. Alu Sz6 integration was estimated to have occurred before the divergence of simians and prosimians. The novel transcript was expressed only in Old world monkeys and apes, and humans. Lineage-specific expression of ACTR8 gene was due to a 'G' duplication in the Alu Sz6 sequences of Old world monkey and ape lineages. Six alternative transcripts (TV1-TV6) generated by various AS mechanisms were newly identified. Based on in-silico analysis, the alternative transcripts were transcribed into new isoforms with C-terminus deletion.

Conclusion: 'G' duplication together with TE exonization and AS via various mechanisms resulted in a different fate of ACTR8 gene expression during primate evolution.

Background

Alternative splicing (AS) is a molecular mechanism producing various transcripts and diverse proteins from a single gene and plays an important role in genomic and phenotypic complexity [1, 2]. In the human genome, more than 95% of pre-mRNAs undergo AS [3, 4]. AS mechanisms are classified into five types, including exon skipping, alternative 3' splice site (SS), alternative 5' SS, intron retention, and mutual exclusion [5, 6]. According to a recent studies, transposable elements (TEs) are closely related to the AS mechanisms [7–9]. A transcriptome study of the crab-eating monkey reported that 10% of AS events were associated with TEs [10].

TEs are mobile genetic elements that can change their position in the genome and thus can affect the sequence and structure of genes. Accordingly, they can modulate gene functions in a relatively short time and are considered as an evolutionary driving force [11]. They make up over 45% of the human genome and are categorized into two classes; Class Ⅰ TEs or retrotransposons, which require reverse transcription for their activation, and Class Ⅱ TEs of DNA transposons, which encode transposases responsible for their excision and insertion. Retrotransposons comprise approximately 42% of the human genome and include endogenous retroviruses (ERVs), long interspersed elements (LINEs), short interspersed elements (SINEs), as well as SINE-R, variable number of tandem repeats (VNTRs) and Alu elements (known as SVA elements for SINE/VNTR/Alu) [12]. Alu and SINE elements are the most successful of all TEs in the primate genome. Alu elements are slightly less than 300 base pairs (bp) in length, present at over than

1 million copies, and widely dispersed within introns and genes in the human genome [13]. Alu elements generally consist of distinct monomeric left and right arms, an A-rich linker, and a poly (A) tail. These features facilitate gene expression regulation by AS and alternative polyadenylation from Alu elements [14, 15]. Non-allelic Alu/Alu recombination can cause genomic instability and has contributed to primate genome divergence [16]. DNA rearrangements caused by Alu elements contribute to genetic disease [17, 18]. Thus, Alu elements are involved in transcription regulation and affect gene function through the generation of new isoforms.

In a study on TE-mediated AS, we recently identified a novel AluSz6-exonizaed ACTR8 transcript in a crab-eating monkey. Actr8 (Actin-Related Protein 8 Homolog, Arp8) is a member of the Actin superfamily and has 15–72% sequence identity with Actin, which is structurally and evolutionary similar [19]. As actin-related proteins (Arps) contain the ATP-binding pocket termed the actin fold, their functions are distinct from that of Actin [20]. Unlike Actin, Arps are predominantly located in the nucleus and have been associated with nucleosome remodeling, histone acetylation, histone variant exchange, transcription regulation, and DNA repair [21–23]. Actr8 is a key component of the INO80 complex, which has critical functions in DNA replication, repair, and recombination as well as in transcription and heterochromatin maintenance [24, 25]. Actr8 is involved in the ATPase activity of the INO80 complex and recruits the complex to the DNA damage sites, and mutation or deletion of Actr8 has effects similar to INO80 deletion [26]. Further, Actr8 is involved in transcriptional activation via regulating promoter architecture and contributes to the regulation of cellular function via cytoskeleton organization [27]. Despite the numerous functional studies on Actr8, information on the alternative transcripts and evolution of the ACTR8 gene is limited. Therefore, this study aimed to explain a series of mutational events in the ACTR8 gene during primate evolution.

Results

Structural analysis of the ACTR8 gene in various primates

We conducted a structural analysis of the *ACTR8* gene in nine primates, including humans, which revealed that *Alu*Sz6 is located in the 7th intron region in antisense orientation through the NCBI genome database (Fig. 1A). *ACTR8* gene is composed of 13 exons. The length of untranslated region (UTR) differs between species, but the open reading frame (ORF) region is highly conserved and encodes 624 amino acids. Remarkably, squirrel monkey *ACTR8* gene has 12 exons and encodes a short protein of 616 amino acids. Transcripts containing *Alu*Sz6 were not found in all primates. Next, we performed genomic PCR to determine the integration time of *Alu*Sz6 using the nine primates genomic DNA samples. Amplicons containing *Alu*Sz6 were detected in all the primates studied, including hominoids (human, chimpanzee, and gorilla), Old world monkeys (rhesus monkey, crab-eating monkey, and African green monkey), New world monkeys (marmoset and squirrel monkey), and prosimian (ring-tailed lemur) (Fig. 1B). These findings indicated that *Alu*Sz6 integrated into the primate genome before the divergence of simian and prosimian lineages.

Alternative transcripts containing the *Alu*-derived transcript of the *ACTR8* gene

To confirm the occurrence of *Alu*Sz6-derived exonization of *ACTR8* gene in the crab-eating monkey, reverse transcription (RT-)PCR was performed using two validation primer pairs (V1 and V2) (Fig. 2A and Additional file 1:Table. 1). The V1 primers were designed to identify transcript variants and the V2 primers were used to detect the transcripts containing the *Alu*-derived exon. In total, seven transcripts were identified in the crab-eating monkey; the V1 primers yielded five transcripts and the V2 primers two (Fig. 2B). Sequence analysis of the transcripts revealed that the variants originated from multiple AS events, including exon skipping, alternative 3' SS and 5' SS, intron retention, mutual exclusion, and *Alu*-exonization (Fig. 2C) [5, 6]. The TV1 transcript skips exon 8 and 9a through alternative 3' SS and is 19 bp longer than exon 9. The TV2 transcript has exon 7a and an *Alu*Sz6-derived exon, which were generated by mutual exclusion and *Alu*-exonization, respectively. TV3 and TV4 have the same *Alu*Sz6-derived exon, but carry exon 9 and exon 9a, respectively, through differential alternative 3' SSs. TV5 is generated by simultaneous *Alu*Sz6 exonization and intron retention. TV6 has longer *Alu*Sz6-derived exon due to a differential alternative 5' SS.

Generally, *Alu*-derived exonization transcripts exhibit tissue-specific expression patterns [28]. Therefore, we profiled *ACTR8* gene expression in various tissues of the crab-eating monkey, including cerebellum, cerebrum, heart, kidney, lung, pancreas, spleen, and testis. Specific RT-PCR primers for each transcript variant were designed based on the splice junction (Fig. 2C). RT-PCR analysis did not reveal tissue-specific *ACTR8* gene expression; the original transcript was ubiquitously expressed in all tissues evaluated, whereas the variants TV1-TV6 showed low or no expression (Fig. 2D). We further investigated *ACTR8* gene expression in the cerebellum of other primates (humans, rhesus monkey, African green monkey, marmoset, and squirrel monkey) using RT-PCR with transcript variant-specific primers (Fig. 2E). In humans and Old world monkeys, all transcript variants showed expression patterns similar to that in the crab-eating monkey. Remarkably, in New world monkeys, only the original transcript was expressed.

Thus, the original transcript was ubiquitously expressed in all species studied, whereas transcript variants showed lineage-specific expression.

Lineage-specific *ACTR8* gene transcript expression

Comparative sequence analysis of *Alu*Sz6-derived exon in nine primates was conducted by multiple sequence alignment (Fig. 3 and Additional file 1:Fig. S1). A novel 'G' duplication was found at the 5' SS of *Alu*Sz6-derived exon in Old world monkeys and apes, providing a new canonical 5' SS. In New world monkeys, this duplication was not present and hence, the canonical 5' SS was not created.

In the squirrel monkey, exon 2 and 3 were found to be longer than in the other primates (Fig. 1A). Therefore, we examined the splice site of each exon. Interestingly, squirrel monkey has specific sequence, 'TA' acceptor splice site, whereas Old world monkeys and apes have no 5' SS in this region (Additional file 1:Fig. S2). Marmoset and lemur have 'TA' acceptor splice site like the squirrel monkey, but we did not

experimentally confirm that whether marmoset and lemur have longer exon 2 and 3 (Additional file 1:Fig. S2).

The TV2 transcript carrying exon 7a showed lineage-specific expression. The splicing sites (donor and acceptor site) were well conserved in all primates evaluated (Additional file 1:Fig. S1), and the branch point was analysed using SVM-BP finder as it may be caused by differences in the surrounding sequences. Multiple candidate branch points, including "TTATAAGAT", were identified. This sequence was located 21 bp upstream of the 3' SS of exon 7a (Additional file 1:Fig. S1). Old world monkeys and apes, but not New world monkeys and prosimians, acquired this branch point. Probably, a lineage-specific mutual exclusion exon, exon 7a, may have been spliced due to the branch point difference (Additional file 1:Fig. S1).

Protein structure analysis of various *ACTR8* gene transcripts

To assess how the seven transcript variants identified in this study affect translation and protein function, *in silico* analysis was performed using an ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) and Pfam (https://pfam.xfam.org/). The original transcript encodes the full-length protein with 624 amino acids, including an ATP-binding site (amino acids (aa) 55-56, 288, and 290) and a nucleotide-binding site (aa 283-286) [29, 30]. The TV1 transcript encodes isoform 1 with 579 amino acids, TV2 encodes isoform 2 with 341 amino acids, and TV3-TV6 encode isoform 3 with 304 amino acids (Fig. 4). The truncated isoforms 2 and 3 lack the C-terminus due to exon 7a and a pre-termination codon created by *Alu*Sz6, respectively (Additional file 1:Fig. S3). Notably, the functional domains (ATP- and nucleotide-binding sites), which are the most crucial for *ACTR8* function, are well preserved in TV2-TV6. According to a previous study, the N-terminal region of *ACTR8* is critical for functional activity and N-terminal deletions have deleterious effects on the expressed protein, whereas deletions in the C-terminal region did not have such effects [29-31]. Based on our findings and previously reported experimental results [32, 33], we suggested that the *ACTR8* gene can produce a lineage-specific protein by *Alu*Sz6 integration and subsequent splicing events.

Discussion

Alu elements belong to the SINE family of TEs and are primate-specific repetitive elements. In humans, they make up 11% of the genome. Alu elements emerged 65 million years ago before the radiation of primates by a fusion of 5' and 3' ends of the 7 SL RNA gene [34]. Alu elements have been continuously inserted into host genomes, which has been contributed to an increase in transcriptome diversity and resulted in positive effects on primate evolution [35–38]. However, insertion of individual Alu elements can also negatively impact genome stability through the induction of gene mutation, potentially leading to genetic diseases and cancer [13, 36, 39, 40]. Indeed, when Alu elements integrated into the intron, it was known to have very little effect on the phenotype [41]. Alu-derived exons tend to be alternatively spliced at a higher frequency than the original exon [42, 43]. AluSz6 is classified as the second oldest Alu lineage, which is estimated to be ~ 30 million years old and contains 551,383 full-length copies in

humans [44]. Based on the reverse transcriptase-mediated integration of Alu elements, the integration time of locus-specific Alu elements can be estimated by comparative analysis of different primate lineages. Hence, Alu-derived exonized transcripts can be useful for estimating the transcript generation time and analysing the mechanism underlying the creation of new transcript variants.

Alu elements are generally inserted into different primate species in a lineage-specific manner [45–48]. The lineage-specific integration of Alu elements contribute to alternative splicing by providing splicing donor and acceptor sites [49, 50]. In our study, an integration event of AluSz6 into ACTR8 gene was identified in all primates evaluated. However, Alu-derived exonization occurred only in Old world monkeys and apes. Therefore, AluSz6 has existed in ACTR8 integrated into the ACTR8 gene has existed in an sleeping state as "Junk DNA" in the genome, and it has been expressed as awake "ACTR8 gene" in old world monkey and ape lineages due to the 'G' duplication, a tiny mutation (Fig. 3). This duplication was carefully re-confirmed by comparative analysis of the AluSz6 sequences of UCSC Genome Browser in different primate species (Additional file 1:Fig. S4). Duplication events can be induced by two molecular mechanisms: secondary unequal homologous recombination or DNA replication-derived errors [40, 51, 52]. We found that the single 'G' duplication in AluSz6 is a result of DNA replication and increased ACTR8 transcript diversity in Old world monkeys and apes, specifically through evolutionary selection. The original transcript (288 bp) was strongly expressed in comparison to the other transcripts (153 bp, 412 bp, 431 bp and 531 bp). The original transcript showed 13.1 fold intensity compared to the weakly expressed transcript, TV4 (Additional file 1:Fig. S5). These gaps in expression levels suggest that the original transcript is a constitutive transcript and the other transcript variants are the minor expressed forms. Moreover, in various tissues, other transcript variants except the original were either not or very weakly expressed (Fig. 2D, E).

Actr8 operates in linking to H2A via the C-terminal and the N-terminal binds to the DNA. Related studies have reported that there is no effect of C-terminal loss on functional activity [29, 33]. Interestingly, our newly identified transcript variants obtained the expressing ability and it was predicted to express a short protein with C-terminal deletion by the premature termination codon (PTC). To understand the lineage-specific evolution of Actr8 in primates, we identified and analysed Actr8 protein isoforms in various mammals (mouse, rat, dog, horse, cow, and pig) and we identified Actr8 protein isoforms (Fig. 5). As expected, Old world monkeys and apes expressed various isoforms, whereas other species did not have the diverse isoform or the C-terminal deletion isoform. This result explain the appearance of the novel ACTR8 protein in primates.

Taken together, our findings indicate that the AluSz6 integration event occurred prior to primate radiation. The birth of tiny 'G' duplication and creation of the Alu-derived exon provided lineage-specific splicing site in the Old world monkeys and apes during the primate evolution of ACTR8 gene (Fig. 6). At the same time, Old world monkeys and apes acquired a branch-point of "TATATAAGAT" sequence, and these species finally gained lineage-specific transcripts. In addition, New world monkeys have lineage-specific transcripts because of specific sequences that function as splice donor sites.

Conclusion

The current study revealed step-by-step evolutionary events in ACTR8 gene that contributed to transcriptome diversity and the generation of novel isoforms of this gene in primates.

Methods

Total RNA and genomic DNA extraction

The total RNA samples isolated from different human tissues including the cerebellum, cerebrum, heart, kidney, lung, liver, spleen, and testis were purchased from Clontech Laboratories, Inc. The total RNA of non-human primates was extracted from the specified tissues: the crab-eating monkey (cerebellum, cerebrum, heart, kidney, lung, pancreas, spleen, and testis), the African green monkey (cerebrum), and the common marmoset (cerebrum). The RNA samples were isolated using the RNeasy® Plus Mini kit (Qiagen), according to the manufacturer's instructions. The DNA contamination was eliminated using the genomic DNA eliminator column and RNase-free DNase (Qiagen). The RNA concentration and purity (A260/A280) were determined with the ND-1000 spectrophotometer (NanoDrop Technologies) and the integrity of RNA was further confirmed by running the samples on an agarose gel. About 500 ng of the total RNA was used for complementary DNA synthesis using the GoScript Reverse Transcription System (Promega) according to the manufacturer's instructions. The primate genomic DNA samples (HU: *Homo sapiens*, CH: *Pan Troglodytes*, GO: *Gorilla gorilla*, RH: *Macaca mulatta*, CR: *Macaca fascicularis*, AGM: *Cercopithecus aethiops*, MAR: *Callithrix jacchus*, SQ: *Saimiri sciureus*, RTL: *Lemur catta*) and all the non-human primate samples were provided by the National Primate Research Center (NPRC) of Korea.

RT-PCR and genomic PCR

ACTR8 transcripts were analysed by RT-PCR amplification using the following primer pairs: V1 (S: 5'-AGG AGT CTG TGT GTG CCA-3', AS: 5'-GAA GCC CAG AGA TGT CCT GA-3'); V2 (S: 5'-TTT AAG GTG TTG CAG AAG AAG AC-3', AS: 5'-TTT AAG GTG TTG CAG AAG AAG AC-3'). RT-PCR was conducted using the ExPrime Taq premix (Genetbio) under the following conditions: pre-denaturation for 5 min, 30~35 cycles at 94°C for 30 sec, 57−61°C for 30 sec, 72°C for 30~40 sec, and a final elongation at 72°C for 5 min. To analyse the integration time of the AluSz6 element during the primate evolution, genomic DNAs from different primate lineages were amplified using specific primer sets designed from well conserved regions (S: 5'-TAT GCT GTG TGG AGG ATG GG-3', AS: 5'-CTG GCA TTC TCT GTA AGG GAA C-3'). The genomic PCR was carried out under the following conditions: 94°C for 5 min, 30 cycles at 94°C for 30 sec, at 56°C for 30 sec, at 72°C for 90 sec, and a final elongation step at 72°C for 5 min. The GAPDH gene was used as a standard control and analysed using specific primer pairs (S: 5'- GAA ATC CCA TCA CCA TCT TCC AGG-3', AS: 5'- GAG CCC CAG CCT TCT CCA TG- 3') designed based on the human *GAPDH* sequence.

Molecular cloning and sequencing

For the validation and sequencing of the PCR products, all PCR products were separated on a 1.5% agarose gel, purified with the Gel SV Extraction kit (GeneAll), and cloned into the RBC T&A Cloning Vector. The cloned DNA was isolated using the Plasmid DNA Mini-prep kit (GeneAll) and sequenced by Macrogen. The validation of primate DNA samples and alternative transcripts was performed as mentioned above.

Acquisition and quantitative analysis of PCR-gel images

The separated PCR products on the ethidium bromide-stained gels were scanned using the gel image software (Vision-capt, Vilber). Band intensity was calculated by the volume and relative intensity compared to the volume %.

Branch-point analysis

Splicing is catalysed by the spliceosome, a large RNA-protein complex, and this complex binds to the branch point site, which are short sequences at the upstream of the acceptor splice sites (3'SS). As the branch point sequences are associated to the cryptic and canonical 3'SSs, branch point prediction is important in pre-mRNA splicing. The SVM-BPfinder program (http://regulatorygenomics.upf.edu/Software/SVM_BP/) [53] was used to analyse the branch point site in the 7th intron sequence of the *ACTR8* gene.

Comparative in-silico analysis of Actr8 protein in different species

Actr8 protein sequences were collected from the NCBI data base in FASTA format for the following different species: human, chimpanzee, gibbon, crab-eating monkey, rhesus monkey, African green monkey, marmoset, rat, dog, horse, cow, and pig. The BioEdit program was used for multiple sequence comparison. Pairwise alignments were used to determine the identity and similarity between each species.

Abbreviations

AS: Alternative splicing; Arps: Actin-related proteins; bp: base pairs; ERVs: Endogenous retroviruses; LINEs: Long interspersed elements; ORF: Open reading frame; PTC: Premature termination codon; SINEs: Short interspersed regions; 3' SS: 3' splice site; TE: Transposable elements; UTR: Untranslated region; VNTRs: Variable number of tandem repeats

Declarations

Ethics approval and consent to participate

All the animal procedures were conducted based on the guidelines of the Institutional Animal Care and Use Committee (KRIBB-AEC-18144) of Korea Research Institute of Bioscience and Biotechnology (KRIBB).

Consent for publication

Not applicable.

Availability of data and materials

All figures and tables generated in this study are available in this article and its additional files.

Competing interests

Authors declare that they have no competing interests.

Funding

This research was supported by a grant from the KRIBB Research Initiative program (KGM4622013, KGM5282022).

Authors' contributions

SHC, SJP, HMC, HRP, JRL conducted all research procedures and analyses in the study, and SHC wrote the manuscript. SHC, SJP and HMC contributed equally to this work. SHC and JWH designed the project. SHC, SJP and JRL performed the experiments. SHC, HMC, HRP and JRL analyzed the data. YHK and JWH provided supervision to the research. SHC drafting the whole manuscript. SHC, SJP, HMC, YHK, and JWH contributed the interpretation of the data and the revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Affiliations

[1]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju 28116, Korea

[2]Department of Functional Genomics, KRIBB School of Bioscience, Korea University of Science & Technology (UST), Daejeon 34113, Korea

[3]Primate Resource Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Jeongeup 56216, Korea

Corresponding author

* Corresponding author: Dr. Young-Hyun Kim

Fax. +82 43 240 6309

Tel . +82 43 240 6332

E-mail: kyh@kribb.re.kr

* Corresponding author: Dr. Jae-Won Huh

Fax. +82 43 240 6309

Tel . +82 43 240 6327

E-mail: huhjw@kribb.re.kr

References

1. Keren H, Lev-Maor G, Ast G: **Alternative splicing and evolution: diversification, exon definition and function**. *Nat Rev Genet* 2010, **11**(5):345-355.

2. Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO: **Alternative splicing and the evolution of phenotypic novelty**. *Philos Trans R Soc Lond B Biol Sci* 2017, **372**(1713).

3. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing**. *Nat Genet* 2008, **40**(12):1413-1415.

4. Lee C, Wang Q: **Bioinformatics analysis of alternative splicing**. *Brief Bioinform* 2005, **6**(1):23-33.

5. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing**. *Nat Rev Genet* 2002, **3**(4):285-298.

6. E Z, Wang L, Zhou J: **Splicing and alternative splicing in rice and humans**. *BMB Rep* 2013, **46**(9):439-447.

7. Kim E, Goren A, Ast G: **Alternative splicing: current perspectives**. *Bioessays* 2008, **30**(1):38-47.

8. Kelley DR, Hendrickson DG, Tenen D, Rinn JL: **Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions**. *Genome Biol* 2014, **15**(12):537.

9. Abascal F, Tress ML, Valencia A: **Alternative splicing and co-option of transposable elements: the case of TMPO/LAP2alpha and ZNF451 in mammals**. *Bioinformatics* 2015, **31**(14):2257-2261.

10. Huh JW, Kim YH, Park SJ, Kim DS, Lee SR, Kim KM, Jeong KJ, Kim JS, Song BS, Sim BW *et al*: **Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (Macaca fascicularis) for biomedical research**. *BMC Genomics* 2012, **13**:163.

11. Sultana T, Zamborlini A, Cristofari G, Lesage P: **Integration site selection by retroviruses and transposable elements in eukaryotes**. *Nat Rev Genet* 2017, **18**(5):292-308.

12. Hedges DJ, Batzer MA: **From the margins of the genome: mobile elements shape primate evolution**. *Bioessays* 2005, **27**(8):785-794.

13. Deininger P: **Alu elements: know the SINEs**. *Genome Biol* 2011, **12**(12):236.

14. Ule J: **Alu elements: at the crossroads between disease and evolution**. *Biochem Soc Trans* 2013, **41**(6):1532-1535.

15. Chen LL, Yang L: **ALUternative Regulation for Gene Expression**. *Trends Cell Biol* 2017, **27**(7):480-490.

16. Cordaux R, Batzer MA: **The impact of retrotransposons on human genome evolution**. *Nat Rev Genet* 2009, **10**(10):691-703.

17. Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D: **Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation**. *Proc Natl Acad Sci U S A* 1991, **88**(3):815-819.

18. Kuehnen P, Krude H: **Alu elements and human common diseases like obesity**. *Mob Genet Elements* 2012, **2**(4):197-201.

19. Muller J, Oma Y, Vallar L, Friederich E, Poch O, Winsor B: **Sequence and comparative genomic analysis of actin-related proteins**. *Mol Biol Cell* 2005, **16**(12):5736-5748.

20. Vieu E, Hernandez N: **Actin's latest act: polymerizing to facilitate transcription?** *Nat Cell Biol* 2006, **8**(7):650-651.

21. Kandasamy MK, McKinney EC, Deal RB, Meagher RB: **Arabidopsis ARP7 is an essential actin-related protein required for normal embryogenesis, plant architecture, and floral organ abscission**. *Plant Physiol* 2005, **138**(4):2019-2032.

22. Dion V, Shimada K, Gasser SM: **Actin-related proteins in the nucleus: life beyond chromatin remodelers**. *Curr Opin Cell Biol* 2010, **22**(3):383-391.

23. Kristo I, Bajusz I, Bajusz C, Borkuti P, Vilmos P: **Actin, actin-binding proteins, and actin-related proteins in the nucleus**. *Histochem Cell Biol* 2016, **145**(4):373-388.

24. Oma Y, Harata M: **Actin-related proteins localized in the nucleus: from discovery to novel roles in nuclear organization**. *Nucleus* 2011, **2**(1):38-46.

25. Tosi A, Haas C, Herzog F, Gilmozzi A, Berninghausen O, Ungewickell C, Gerhold CB, Lakomek K, Aebersold R, Beckmann R *et al*: **Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex**. *Cell* 2013, **154**(6):1207-1219.

26. Chen M, Shen X: **Nuclear actin and actin-related proteins in chromatin dynamics**. *Curr Opin Cell Biol* 2007, **19**(3):326-330.

27. Ford J, Odeyale O, Shen CH: **Activator-dependent recruitment of SWI/SNF and INO80 during INO1 activation**. *Biochem Biophys Res Commun* 2008, **373**(4):602-606.

28. Mersch B, Sela N, Ast G, Suhai S, Hotz-Wagenblatt A: **SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements**. *BMC Genet* 2007, **8**:78.

29. Saravanan M, Wuerges J, Bose D, McCormack EA, Cook NJ, Zhang X, Wigley DB: **Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex**. *Proc Natl Acad Sci U S A* 2012, **109**(51):20883-20888.

30. Gerhold CB, Winkler DD, Lakomek K, Seifert FU, Fenn S, Kessler B, Witte G, Luger K, Hopfner KP: **Structure of Actin-related protein 8 and its contribution to nucleosome binding**. *Nucleic Acids Res* 2012, **40**(21):11036-11046.

31. Brahma S, Ngubo M, Paul S, Udugama M, Bartholomew B: **The Arp8 and Arp4 module acts as a DNA sensor controlling INO80 chromatin remodeling**. *Nat Commun* 2018, **9**(1):3309.

32. Aoyama N, Oka A, Kitayama K, Kurumizaka H, Harata M: **The actin-related protein hArp8 accumulates on the mitotic chromosomes and functions in chromosome alignment**. *Exp Cell Res* 2008, **314**(4):859-868.

33. Osakabe A, Takahashi Y, Murakami H, Otawa K, Tachiwana H, Oma Y, Nishijima H, Shibahara KI, Kurumizaka H, Harata M: **DNA binding properties of the actin-related protein Arp8 and its role in DNA repair**. *PLoS One* 2014, **9**(10):e108354.

34. Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW: **Base sequence studies of 300 nucleotide renatured repeated human DNA clones**. *J Mol Biol* 1981, **151**(1):17-33.

35. Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL: **Alu insertion polymorphisms for the study of human genomic diversity**. *Genetics* 2001, **159**(1):279-290.

36. Batzer MA, Deininger PL: **Alu repeats and human genomic diversity**. *Nat Rev Genet* 2002, **3**(5):370-379.

37. Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA: **Recently integrated Alu elements and human genomic diversity**. *Mol Biol Evol* 2003, **20**(8):1349-1361.

38. Lee JR, Kim YH, Park SJ, Choe SH, Cho HM, Lee SR, Kim SU, Kim JS, Sim BW, Song BS *et al*: **Identification of Alternative Variants and Insertion of the Novel Polymorphic AluYl17 in TSEN54 Gene during Primate Evolution**. *Int J Genomics* 2016, **2016**:1679574.

39. Chenais B: **Transposable elements in cancer and other human diseases**. *Curr Cancer Drug Targets* 2015, **15**(3):227-242.

40. Kim S, Cho CS, Han K, Lee J: **Structural Variation of Alu Element and Human Disease**. *Genomics Inform* 2016, **14**(3):70-77.

41. Deininger PL, Batzer MA: **Alu repeats and human disease**. *Mol Genet Metab* 1999, **67**(3):183-193.

42. Sorek R, Ast G, Graur D: **Alu-containing exons are alternatively spliced**. *Genome Res* 2002, **12**(7):1060-1067.

43. Sorek R: **The birth of new exons: mechanisms and evolutionary consequences**. *RNA* 2007, **13**(10):1603-1608.

44. Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE: **Active Alu retrotransposons in the human genome**. *Genome Res* 2008, **18**(12):1875-1883.

45. Huh JW, Kim YH, Kim DS, Park SJ, Lee SR, Kim SH, Kim E, Kim SU, Kim MS, Kim HS *et al*: **Alu-derived old world monkeys exonization event and experimental validation of the LEPR gene**. *Mol Cells* 2010, **30**(3):201-207.

46. Huh JW, Kim YH, Lee SR, Kim DS, Park SJ, Kim H, Kim JS, Song BS, Kim HS, Chang KT: **Four different ways of alternative transcripts generation mechanism in ADRA1A gene**. *Genes Genet Syst* 2010, **85**(1):65-73.

47. Kim YH, Choe SH, Song BS, Park SJ, Kim MJ, Park YH, Yoon SB, Lee Y, Jin YB, Sim BW *et al*: **Macaca specific exon creation event generates a novel ZKSCAN5 transcript**. *Gene* 2016, **577**(2):236-243.

48. Cho HM, Park SJ, Choe SH, Lee JR, Kim SU, Jin YB, Kim JS, Lee SR, Kim YH, Huh JW: **Cooperative evolution of two different TEs results in lineage-specific novel transcripts in the BLOC1S2 gene**. *BMC Evol Biol* 2019, **19**(1):196.

49. Park SJ, Kim YH, Lee SR, Choe SH, Kim MJ, Kim SU, Kim JS, Sim BW, Song BS, Jeong KJ *et al*: **Gain of a New Exon by a Lineage-Specific Alu Element-Integration Event in the BCS1L Gene during Primate Evolution**. *Mol Cells* 2015, **38**(11):950-958.

50. Lee JR, Park SJ, Kim YH, Choe SH, Cho HM, Lee SR, Kim SU, Kim JS, Sim BW, Song BS *et al*: **Alu-Derived Alternative Splicing Events Specific to Macaca Lineages in CTSF Gene**. *Mol Cells* 2017, **40**(2):100-108.

51. Khitrinskaia I, Stepanov VA, Puzyrev VP: **[Alu repeats in the human genome]**. *Mol Biol (Mosk)* 2003, **37**(3):382-391.

52. Chen L, Zhou W, Zhang L, Zhang F: **Genome architecture and its roles in human copy number variation**. *Genomics Inform* 2014, **12**(4):136-144.

53. Corvelo A, Hallegger M, Smith CW, Eyras E: **Genome-wide association between branch point properties and alternative splicing**. *PLoS Comput Biol* 2010, **6**(11):e1001016.

Supplemental Information Note

Additional files legends

Fig. S1 Multiple sequence alignment of the insertion *Alu*Sz6 region in various primates in the *ACTR8* gene. Sequences obtained from the genomic PCR product (Fig. 1B). Black line boxes indicate exons and sequences between boxes are intron regions. Two forward slashes indicate the region of omitted intron sequence for good looking data. The sky blue region represents the full-length of *Alu*Sz6. 3' splice sites and 5' splice sites in the *Alu*SZ6 region are not painted in sky blue. The black-shade region indicates the branch point site and new world monkey and prosimians have deletion in this branch point region.

Fig. S2 Multiple sequence alignment of the region from exon 2 to exon 4 in various primates in the *ACTR8* gene. Genomic sequence from the NCBI database were aligned in various primates. Donor splice site (5' splice site) and acceptor splice site (3' splice site) are indicated as blue and red dotted line boxes, respectively.

Fig. S3 Multiple sequence alignment of the products of RT-PCR using the validation primer. The RT-PCR products were sequenced and these sequences were analyzed by multiple nucleotide alignment. The sky

blue region indicates the *Alu*Sz6-derived exon region. The red line box represents the premature termination codon (PTC).

Fig. S4 Multiple sequence alignment of *Alu*Sz6 of *ACTR8* gene. *Alu*Sz6 sequences of various primates were searched using online resources (UCSC Genome browser). In this figure, the *Alu*Sz6 sequences were shown in the forward direction (5' to 3'), whereas the main figure is shown in the reverse direction as inserted into genome. The blue line box indicates the point mentioned as 'G' duplication. *Alu*Sz6 of RTL could not be analyzed for unenrolled reason, but we confirmed *Alu*SZ6 sequence of RTL through the experiment.

Fig. S5 Analysis of the *ACTR8* gene PCR band intensity. (A) Green lanes (1-5) were PCR positive bands by V1 primer from Fig. 2B. Vision-capt software selected 5 bands and generated peaks of the corresponding PCR band. The PCR band intensity was generated from these peaks. (B) Vision-capt software measured the volume based on height and area of the positive bands, respectively.

Figures
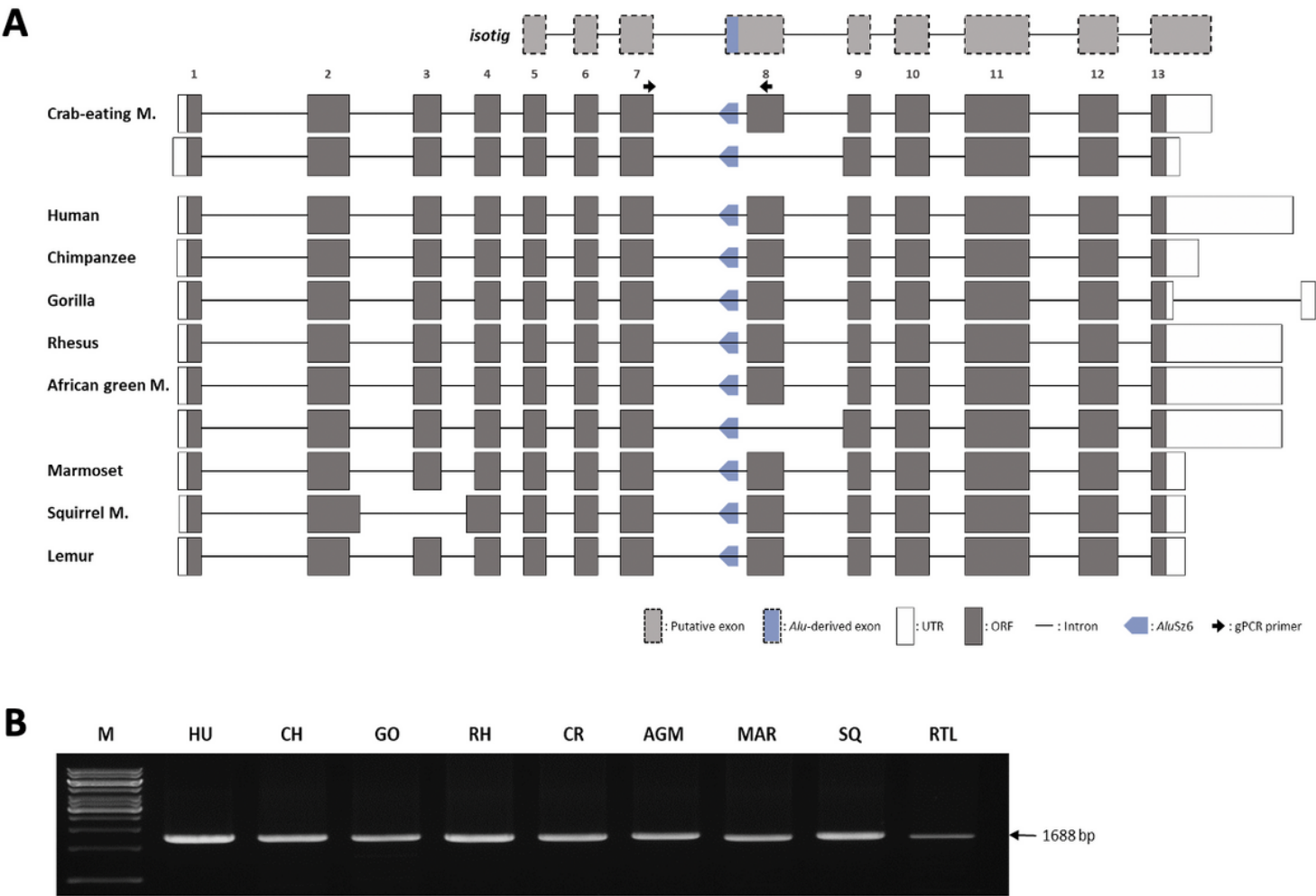
Figure 1

Structural analysis of ACTR8 gene and AluSz6 integration analysis in various primates. (A) Isotig is a putative transcript from the transcriptome study of the crab-eating monkey. Dotted boxes represent the putative exon and the Alu-derived exon. Open and closed boxes represent the UTR region and ORF region, respectively. In all primates, the antisense-oriented AluSz6 element is located between exon 7 and exon 8 of the ACTR8 gene. Thick arrows indicate the location of the genomic PCR primer. This figure is schematic and not to scale. (B) Genomic PCR amplification for AluSz6 integration analysis in the ACTR8 gene using various primate samples. M indicates the size marker. Primate DNA samples are abbreviated as follows: HU: human (Homo sapiens); CH: chimpanzee (Pan troglodytes); GO: gorilla (Gorilla gorilla); RH: rhesus monkey (Macaca mulatta); CR: crab-eating monkey (Macaca fascicularis); AGM: African green monkey (Chlorocebus aethiops); MAR: Marmoset (Callithrix jacchus); SQ: squirrel monkey (Saimiri sciureus); RTL: ring-tailed lemur (Lemur catta).
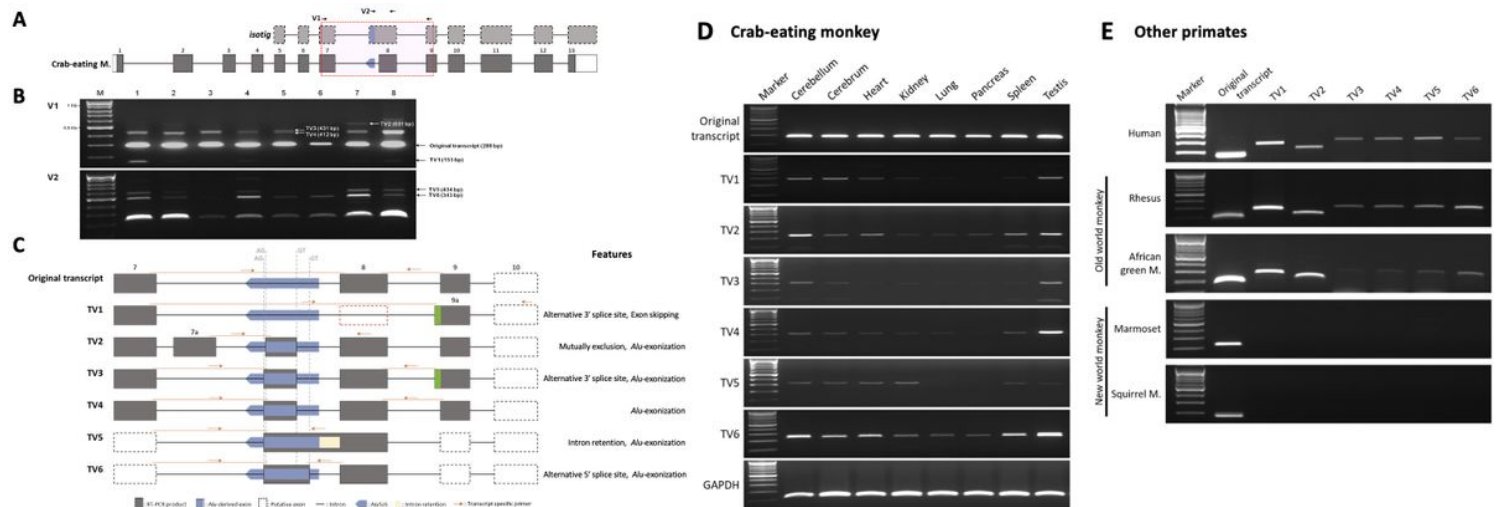
Figure 2

ACTR8 gene transcript variants in primates using RT-PCR and sequence analysis. (A) The AluSz6 element integrated transcript was validated using the V1 and V2 primer pairs. Gray and dotted boxes represent the coding exon and putative exon, respectively. Horizontal arrows indicate the V1 and V2 primer location. (B) RT-PCR was performed using the tissue cDNA of the crab-eating monkey and the PCR product was electrophoresed by agarose gel electrophoresis and stained with ethidium bromide. A total of 7 positive bands were observed. 1: Cerebellum; 2: Cerebrum; 3:Heart; 4:Kidney; 5:Lung; 6: Pancreas; 7: Spleen; 8: Testis (C) Seven transcript variants were identified. Gray boxes represent the coding exon and PCR product. Vertical dashed lines represent the 3'(AG) and 5'(GT) splicing site. The green box indicates the exon 9a that is longer than exon 9. The orange-colored arrows represent validation of RT-PCR primers of each transcript variant (D) In the crab-eating monkey, RT-PCR was performed using various tissues and specific primer pairs. GAPDH (120 bp) indicates the positive control. (E) Human, old world monkey (rhesus, African green monkey), and new world monkey (marmoset, squirrel monkey) cerebellar cDNA were used for validation of the transcript variant expression.
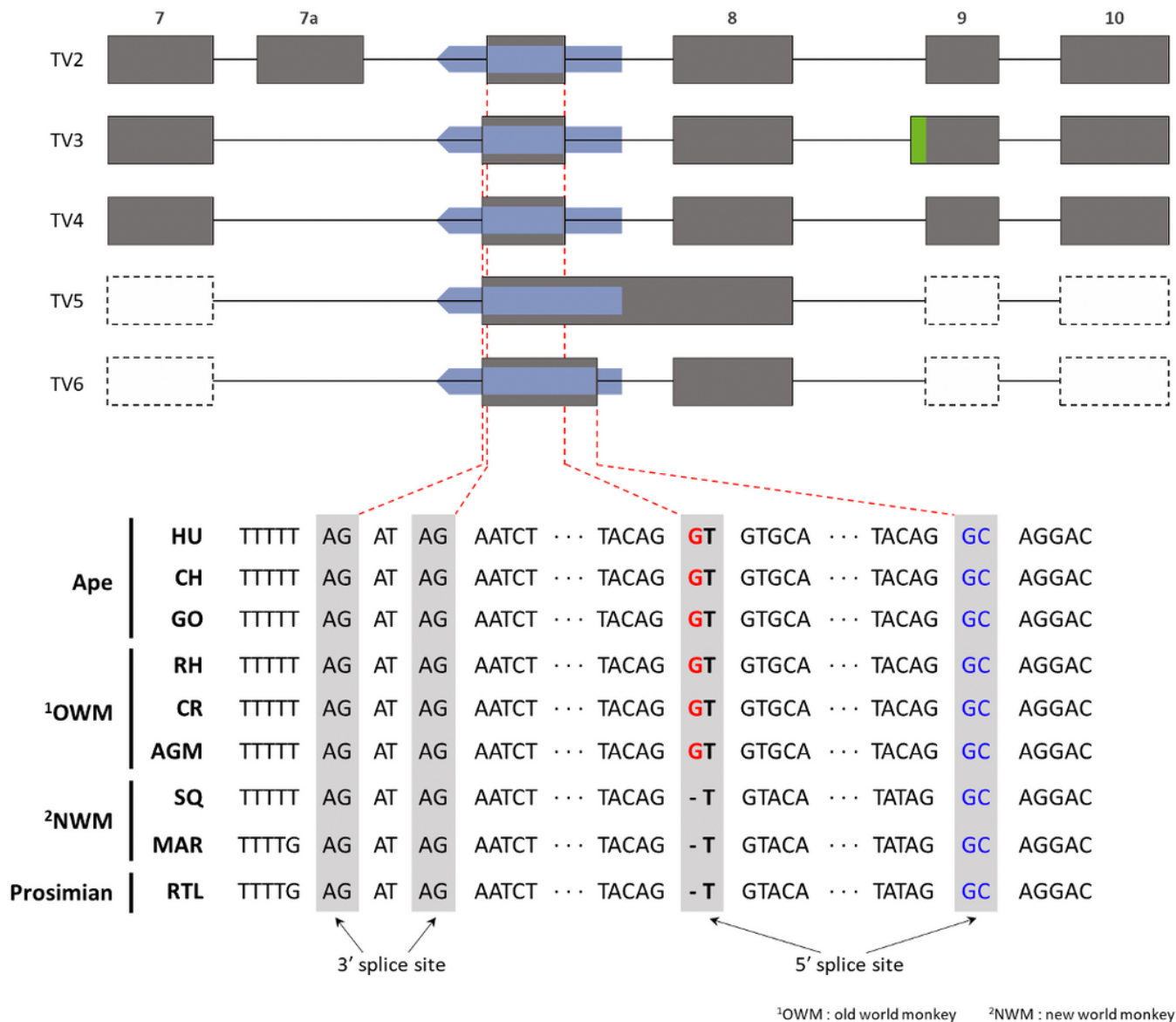
Figure 3

AluSz6 integration analysis in the ACTR8 gene in various primates. Boundary sequence of potential 3' and 5' splicing sites in various primates. Gray boxes indicate the splicing site by AluSz6 element. Primates are abbreviated as follows: HU: human (Homo sapiens); CH: chimpanzee (Pan troglodytes); GO: gorilla (Gorilla gorilla); RH: rhesus monkey (Macaca mulatta); CR: crab-eating monkey (Macaca fascicularis); AGM: African green monkey (Chlorocebus aethiops); MAR: Marmoset (Callithrix jacchus); SQ: squirrel monkey (Saimiri sciureus); RTL: ring-tailed lemur (Lemur catta)
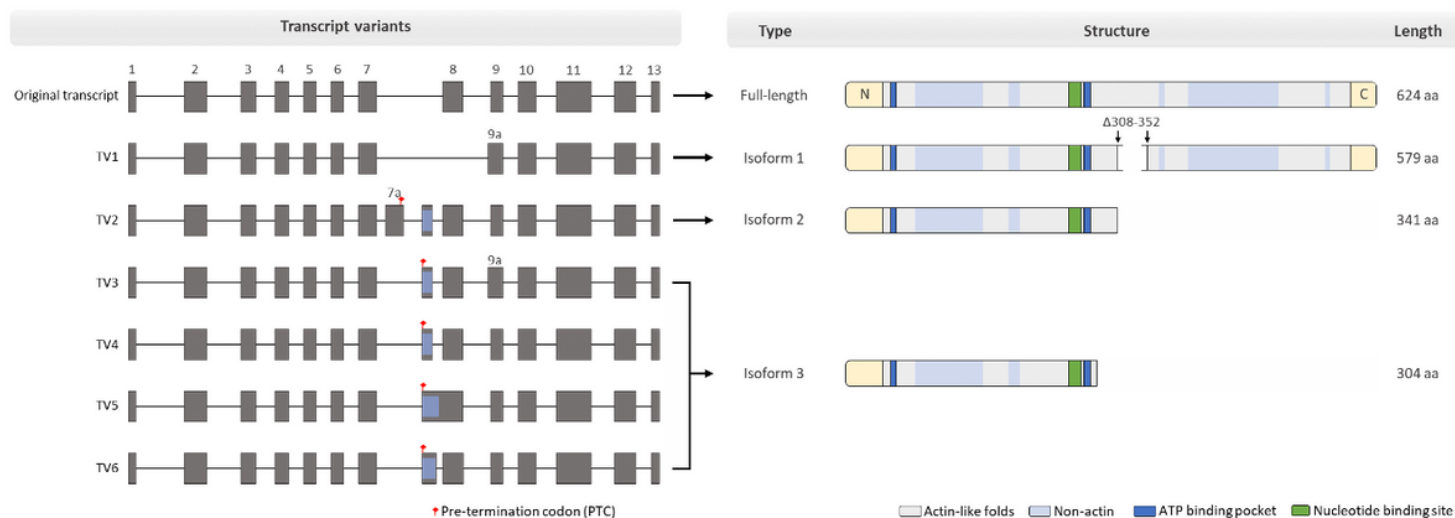
Figure 4

Schematic of the genomic and protein structures of the ACTR8 gene. Full-length ACTR8 of human and crab-eating monkey encodes 624 amino acids containing the actin, ATP-binding, and nucleotide binding sites. The identified transcript variants encoded isoform 1 and isoform 2-3 lacking the C-terminal.
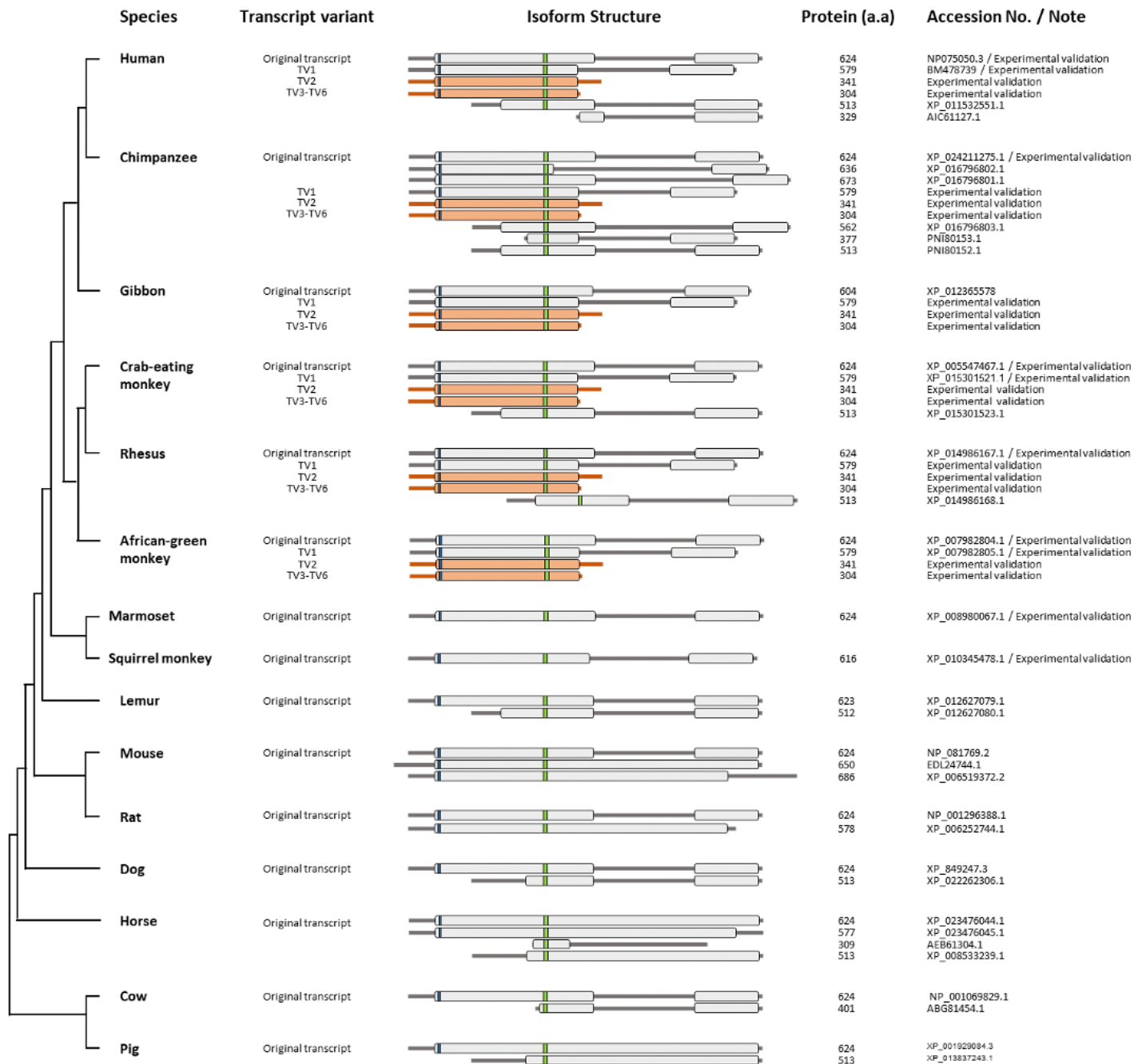
Figure 5

Schematic representation of the experimental methods and in silico analysis of the ACTR8 isoform in primates, mouse, rat, dog, horse, cow, and pig. Phylogenetic analysis of the ACTR8 isoform in mammals. Orange represents the truncated isoform with C-terminal deletion.
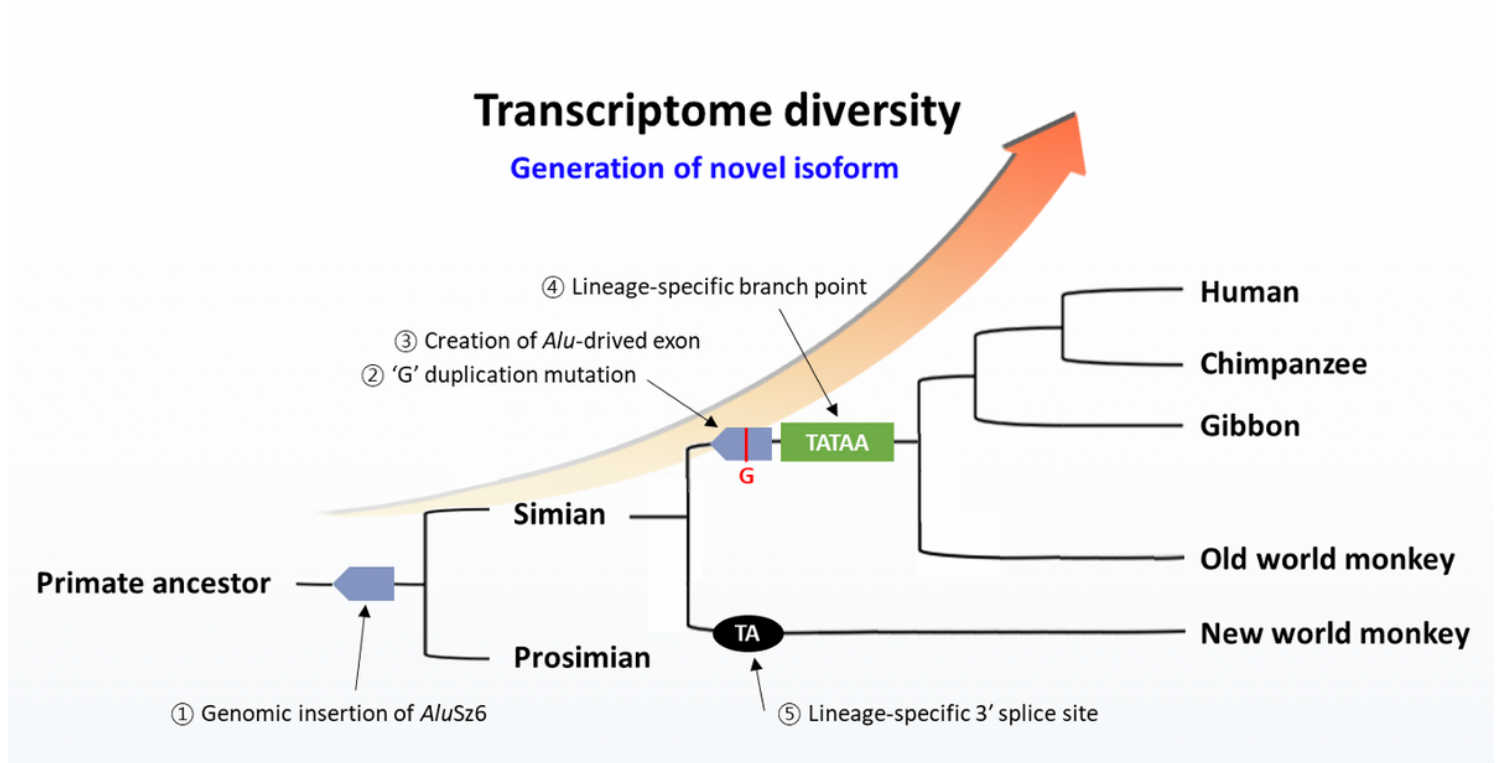
Figure 6

Schematic representation of the evolutionary event in primate ACTR8 genes. Events during the primate evolution of the ACTR8 gene are summarized.