

The genetic basis of tail-loss evolution in humans and apes

Bo Xia^{1,2*}, Weimin Zhang², Aleksandra Wudzinska², Emily Huang², Ran Brosh², Maayan Pour¹, Alexander Miller⁴, Jeremy S. Dasen⁴, Matthew T. Maurano², Sang Y. Kim⁵, Jef D. Boeke^{2,3,6*} and Itai Yanai^{1,3*}

¹ Institute for Computational Medicine, NYU Langone Health, New York, NY 10016, USA

² Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA

³ Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY 10016, USA

⁴ Department of Neuroscience and Physiology, NYU Langone Health, New York, NY 10016, USA

⁵ Department of Pathology, NYU Langone Health, New York, NY 10016, USA

⁶ Department of Biomedical Engineering, NYU Tandon School of Engineering, Brooklyn, NY, 11201, USA

* Correspondence: Bo.Xia@nyulangone.edu; Jef.Boeke@nyulangone.org; Itai.Yanai@nyulangone.org

The loss of the tail is one of the main anatomical evolutionary changes to have occurred along the lineage leading to humans and to the “anthropomorphous apes”^{1,2}. This morphological reprogramming in the ancestral hominoids has been long considered to have accommodated a characteristic style of locomotion and contributed to the evolution of bipedalism in humans^{3–5}. Yet, the precise genetic mechanism that facilitated tail-loss evolution in hominoids remains unknown. Primate genome sequencing projects have made possible the identification of causal links between genotypic and phenotypic changes^{6–8}, and enable the search for hominoid-specific genetic elements controlling tail development⁹. Here, we present evidence that tail-loss evolution was mediated by the insertion of an individual *Alu* element into the genome of the hominoid ancestor. We demonstrate that this *Alu* element – inserted into an intron of the *TBXT* gene (also called *T* or *Brachyury*^{10–12}) – pairs with a neighboring ancestral *Alu* element encoded in the reverse genomic orientation and leads to a hominoid-specific alternative splicing event. To study the effect of this splicing event, we generated a mouse model that mimics the expression of human *TBXT* products by expressing both full-length and exon-skipped isoforms of the mouse *TBXT* ortholog. We found that mice with this genotype exhibit the complete absence of a tail or a shortened tail, supporting the notion that the exon-skipped transcript is sufficient to induce a tail-loss phenotype, albeit with incomplete penetrance. We further noted that mice homozygous for the exon-skipped isoforms exhibited embryonic spinal cord malformations, resembling a neural tube defect condition, which affects ~1/1000 human neonates¹³. We propose that selection for the loss of the tail along the hominoid lineage was associated with an adaptive cost of potential neural tube defects and that this ancient evolutionary trade-off may thus continue to affect human health today.

The tail appendage varies widely in its morphology and function across vertebrate species^{4,5}.

For primates in particular, the tail is adapted to a range of environments, with implications for the animal's style of locomotion^{14,15}. The New World howler monkeys, for example, evolved a prehensile tail that helps the animal to grasp or hold objects while occupying arboreal habitats¹⁶. Hominoids – which include humans and the apes – however, are distinct among the primates in their loss of an external tail (**Fig. 1a**). The loss of the tail is inferred to have occurred ~25 million years ago when the hominoid lineage diverged from the ancient Old World monkeys (**Fig. 1a**), leaving only 3-4 caudal vertebrae to form the coccyx, or tailbone, in modern humans¹⁷. It has long been speculated that tail loss in hominoids has contributed to bipedal locomotion, whose evolutionary occurrence coincided with the loss of tail^{18–20}. Recent progress in developmental biology has led to the elucidation of the gene regulatory networks that underlie tail development^{9,21}. Specifically, the absence of the tail phenotype in the Mouse Genome Informatics has so far recorded 31 genes from the study of mutants and naturally occurring variants^{21,22} (**Supp. Table 1**). Expression of these genes is enriched in the development of the primitive streak and posterior body formation, including the core gene regulation network for inducing the mesoderm and definitive endoderm such as *Tbxt*, *Wnt3a*, and *Msgn1*. While these genes and their relationships have been studied, the exact genetic changes that drove the evolution of tail-loss in hominoids remain unknown, preventing an understanding of how tail loss affected other human evolutionary events, such as bipedalism.

Results

A hominoid-specific intronic *AluY* element in *TBXT*

We screened through the 31 human genes – and their primate orthologs – involved in tail development, with the goal of identifying a genetic variation associated with the loss of the tail in hominoids (**Supp. Table 1**). We first examined protein sequence conservation between the

hominoid genomes and its closest sister lineage, the Old World monkeys (Cercopithecidae).

However, we failed to detect candidate variants in hominoid coding sequences that might provide a genetic mechanism for tail-loss evolution (**Supp. File 1**). We next queried for hominoid-specific genomic rearrangements in the non-coding regions of genes related to tail development. Surprisingly, we found a hominoid-specific *Alu* element inserted in intron 6 of *TBXT* (**Fig. 1b and Supp. File 1**)^{10,11}. *TBXT* codes a highly-conserved transcription factor critical for mesoderm and definitive endoderm formation during embryonic development^{12,23–25}. Heterozygous mutations in the coding regions of the *TBXT* orthologs in tailed animals, such as mouse¹⁰, Manx cat²⁶, dog²⁷ and zebrafish²⁸, lead to the absence or reduced form of the tail, and homozygous mutants are typically not viable. Moreover, this particular *Alu* insertion is from the *AluY* subfamily, a relatively ‘young’ but not human-specific subfamily shared between the genomes of hominoids and Old World monkeys, the activity of which coincides with the evolutionary time when early hominoids lost their tails²⁹.

The *AluY* element in *TBXT* is not inserted in the vicinity of a splice site; rather, it is >500 bp from exon 6 of *TBXT*, the nearest coding exon. As such, it would not be expected to lead to an alternative splicing event, as found for other intronic *Alu* elements affecting splicing^{30–32}. However, we noted the presence of another *Alu* element (*AluSx1*) in the reverse orientation in intron 5 of *TBXT*, that is conserved in all simians. Together, the *AluY* and *AluSx1* elements form an inverted repeat pair (**Fig. 1b**). We thus posited that upon transcription, the simian-specific *AluSx1* element pairs with the hominoid-specific *AluY* element, forming a stem-loop structure in the *TBXT* pre-mRNA and trapping exon 6 in the loop (**Fig. 1c**). An inferred RNA secondary structure model supported an interaction between these two *Alu* elements³³ (**Fig. S1**). The secondary structure of the transcript may thus conjoin the splice donor and receptor of exons 5 and 7, respectively, and promote the skipping of exon 6, leading to a hominoid-specific and in-frame alternative splicing isoform, *TBXT-Δexon6* (**Fig. 1c**). Indeed, we validated the existence

of *TBXT*- Δ exon6 transcripts in human and its corresponding absence in mouse, which lacks the

Alu elements, using an embryonic stem cells (ESCs) *in vitro* differentiation system that induces

TBXT expression similar to that present in the primitive streak of the embryo (**Fig. S2**)^{34,35}.

Considering the high conservation of *TBXT* exon 6 and its potential transcriptional regulation function (**Fig. S3**), we thus hypothesized that in humans and apes, the *TBXT*- Δ exon6 isoform

protein disrupts tail elongation during embryonic development, leading to the reduction or loss of an external tail (**Fig. 1c**).

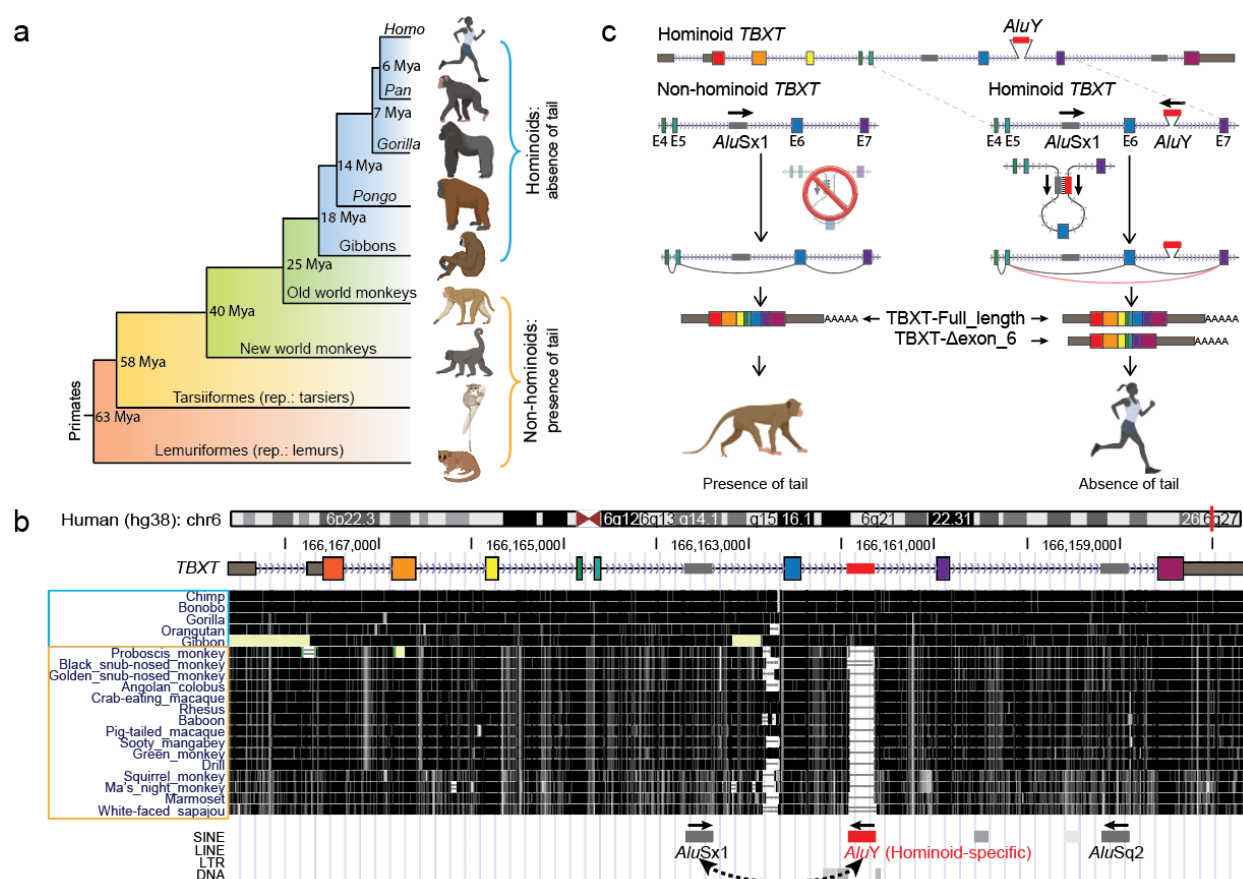


Fig. 1 | Evolution of tail loss in hominoids. a, Tail phenotypes across the primate phylogenetic tree. **b**,

UCSC Genome browser view of the conservation score through multi-species alignment at the *TBXT*

locus across primate genomes³⁶. The hominoid-specific *AluY* element is labelled in red. **c**, Schematic of

the hypothesized mechanism of tail-loss evolution in hominoids.

AluY* insertion in *TBXT* induces alternative splicing, and requires interaction with *AluSx1

To test whether both *AluY* and *AluSx1* are required to induce the hominoid-specific alternatively spliced isoform of *TBXT*, we used CRISPR/Cas9 in human ESCs to individually delete the hominoid-specific *AluY* element and – in a separate line – its potentially interacting counterpart *AluSx1* (**Fig. 2a, S4a**). Again, we adapted the hESC *in vitro* differentiation system to mimic the *TBXT* expression in the embryo (**Fig. S2**)³⁴. We found that deleting the *AluY* almost completely eliminated the generation of the *TBXT-Δexon6* isoform transcript (**Fig. 2b**, middle). Similarly, deleting the interacting partner, *AluSx1*, sufficed to repress this alternatively spliced isoform (**Fig. 2b**, right). These results support the notion that the hominoid-specific *AluY* insertion induces a novel *TBXT-Δexon6* AS isoform, through an interaction with the neighboring *AluSx1* element (**Fig. 2c**, top).

Interestingly, we found that wild-type differentiated hESCs also express a minor, previously unannotated transcript that excludes both exon 6 and exon 7, leading to a frameshift and early truncation at the protein level (**Fig. 2b**, left, and **S4b**). Whereas deleting *AluY* slightly enhanced the abundance of this *TBXT-Δexon6&7* transcript, deleting *AluSx1* in intron 5 completely eliminated this transcript (**Fig. 2b**). This may be best explained by a secondary interaction of the *AluSx1* element with a distal *AluSq2* element in intron 7. In this scenario, the secondary interaction would occur at a lower probability than the *AluY-AluSx1* interaction pair (**Fig. 2c**, bottom). These results further support an interaction among intronic transposable elements affecting splicing of the conserved *TBXT* regulator (**Fig. 2c, S4b**).

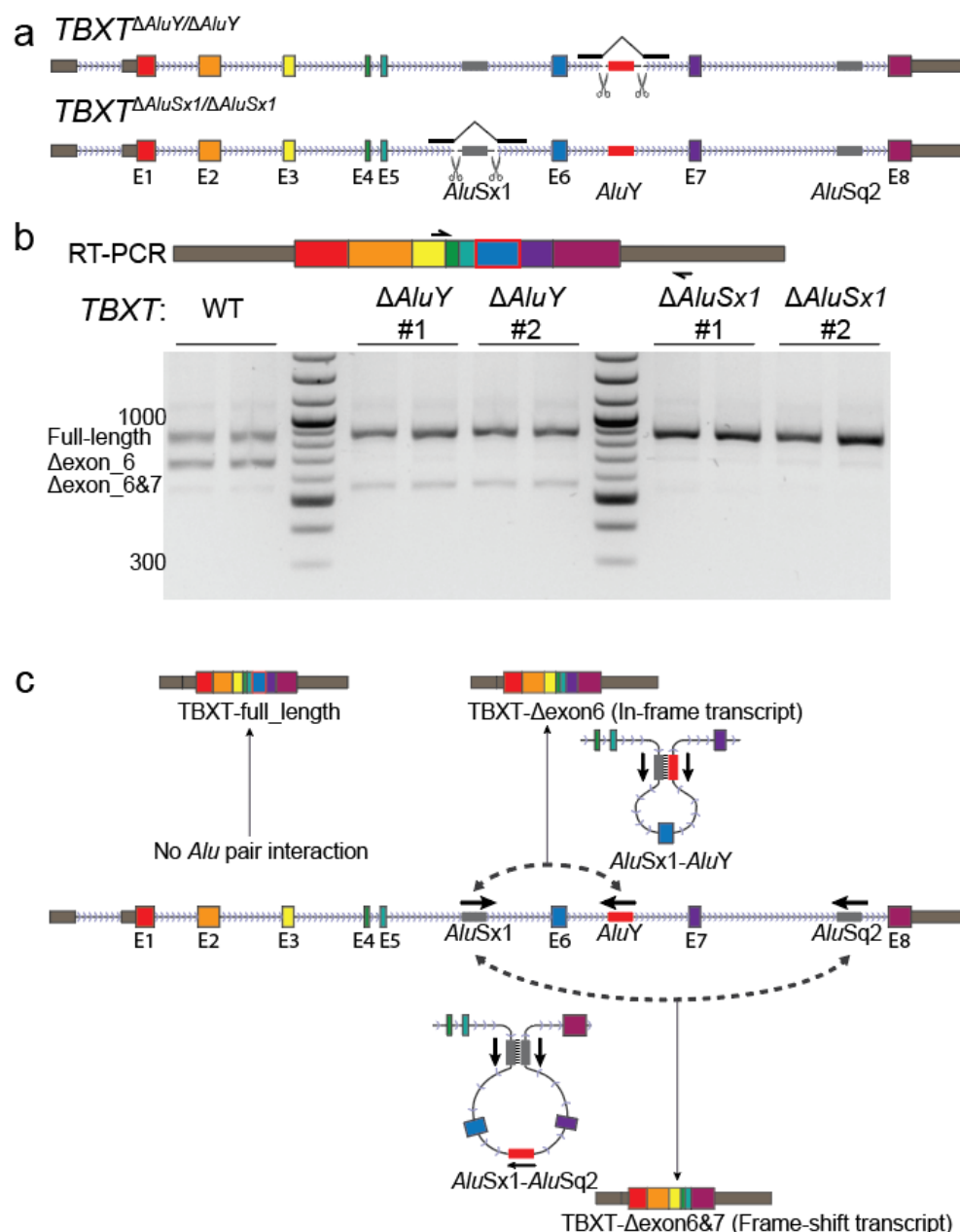


Fig. 2 | Both *AluY* and *AluSx1* are required for *TBXT* alternative splicing. **a**, CRISPR-generated homozygous knock-outs of the *AluY* element in intron 6 and (in a separate line) *AluSx1* element in intron 5 of *TBXT*. **b**, RT-PCR results of *TBXT* transcripts isolated from differentiated hESCs of wild-type, Δ *AluY*, and Δ *AluSx1* genotypes. Each genotype (Δ *AluY* and Δ *AluSx1*) was analyzed by two independent replicate clones. **c**, A schematic of *Alu* interactions and the corresponding *TBXT* transcripts in human, indicating that an *AluY*-*AluSx1* interaction leads to the *TBXT*- Δ exon6 transcript. The *TBXT*- Δ exon6&7 transcript may stem from an *AluSx1*-*AluSq2* interaction.

***Tbxt-Δexon6* is sufficient to induce tail loss in mice**

To test whether the *TBXT-Δexon6* isoform is sufficient to induce tail loss, we generated a heterozygous mouse *Tbxt*^{Δexon6/+} model (**Fig. 3a**). *TBXT* is highly conserved in vertebrates and human and mouse protein sequences share 91% identity with a similar exon/intron architecture¹¹. We could thus simulate a *TBXT-Δexon6* isoform by deleting exon 6 in mice, forcing splicing of exons 5 with exon 7. The *Tbxt*^{Δexon6/+} heterozygous mouse thus mimics the *TBXT* gene in human, which expresses both full-length and Δexon6 splice isoforms (**Fig. 2b and 3b-c**).

Studying the phenotypes of the *Tbxt*^{Δexon6/+} mice, we found that simultaneous expression of both isoforms led to strong but heterogeneous tail morphologies, including no-tail and short-tail phenotypes (**Fig. 3d-e, S5**). Specifically, 21 of the 63 heterozygous mice showed tail phenotypes, while none of their 35 wild-type littermates showed phenotypes (**Table 1**). The incomplete penetration of phenotypes among the heterozygotes was stable across generations and founder lines: no-/short-tailed (*Tbxt*^{Δexon6/+}) parent can give birth to long-tailed *Tbxt*^{Δexon6/+} mice, whereas long-tailed (*Tbxt*^{Δexon6/+}) parents can give birth to pups with varied tail phenotypes (**Table 1, Fig. S5**), providing further evidence that the presence of *TBXT-Δexon6* suffices to induce tail loss.

To control for the possibility that zygotic CRISPR targeting induced off-targeting DNA changes at the *Tbxt* locus, we performed Capture-seq covering the *Tbxt* locus and ~200kb of both upstream and downstream flanking regions³⁷ (**Fig. S6**). Capture-seq did not detect any off-targeting at the *Tbxt* locus across three independent founder mice, supporting our conclusion that the observed tail phenotype from the *Tbxt*^{Δexon6/+} mice derived from the *Tbxt-Δexon6* isoform.

Table 1. Genotype and phenotype analyses of the *Tbxt* ^{Δ exon6/+} intercrossed F2 pups.

Genotypes	Total F2 pups	Pups with tail phenotypes	Tail phenotypes			Intercross (type 1)*	Intercross (type 2)
			No-tail	Short-tail	Kinked-tail		
<i>Tbxt</i> ^{Δexon6/Δexon6}	0	0	0	0	0	0	0
<i>Tbxt</i> ^{Δexon6/+}	63	21	4	9	8	17(7)**	46(14)
<i>Tbxt</i> ^{+/+}	35	0	0	0	0	7(0)	28(0)

Note: *: Type 1 intercrossing: at least one of the parent mice is no-/short-tailed. Type 2 intercrossing: both parent mice are long-tailed.

** : Numbers in parentheses indicate the number of pups with tail phenotypes.

Homozygous removal of *Tbxt*- Δ exon6 is lethal

The human *TBXT* gene expresses a mixture of *TBXT*- Δ exon6 and *TBXT*-full length transcripts – induced as we inferred by the *AluY* insertion and interaction with *AluSx1* – while mouse *Tbxt* only expresses the full length *Tbxt*. Thus, we next inquired into the mode by which homozygous *TBXT*- Δ exon6 mutation (*Tbxt* ^{Δ exon6/ Δ exon6}) affects development. Intercrossing the *Tbxt* ^{Δ exon6/+} mice across multiple litters and replicated in different founders, we failed to produce viable homozygotes (**Table 1**). Dissecting intercrossed embryos at E11.5 showed that homozygotes either arrested development at ~E9 or developed with spinal cord malformations that consequently led to death at birth (**Fig. S7**). We noted that the *Tbxt* ^{Δ exon6/ Δ exon6} embryos showed malformations of the spinal cord similar to spina bifida in humans. Together, while the *Tbxt* ^{Δ exon6/+} mice present incomplete penetrance of the tail phenotypes requires further investigation, these results indicate that the *TBXT*- Δ exon6 isoform, which in human is induced by the intronic *AluY*-*AluSx1* interaction, may indeed be the key driver of tail-loss evolution in hominoids.

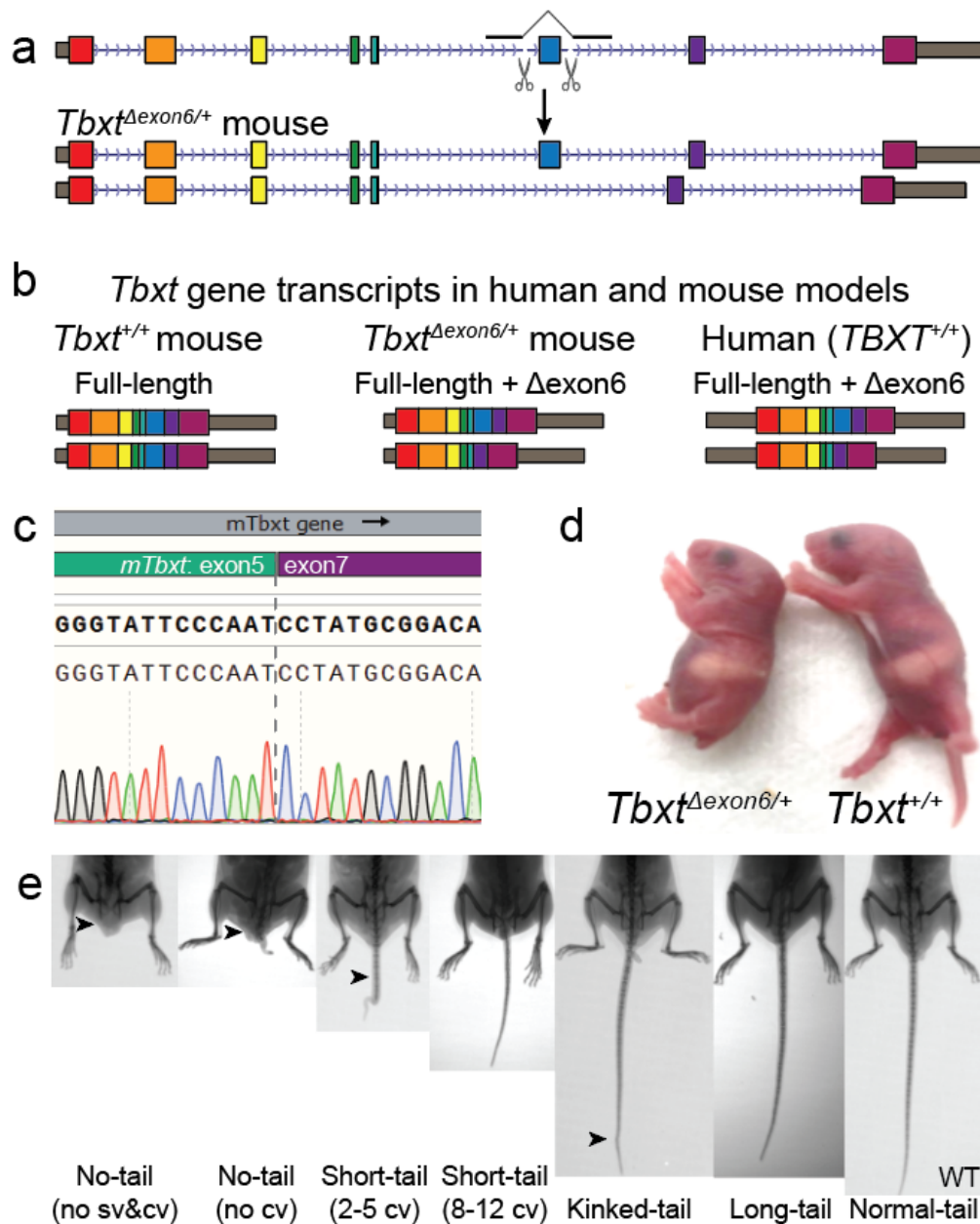


Fig. 3 | *TBXT*- Δ exon6 isoform is sufficient to induce tail-loss phenotype. **a**, CRISPR design for generating a *Tbxt*^{Δexon6/+} heterozygous mouse model. **b**, *Tbxt*^{Δexon6/+} mouse mimics *TBXT* gene expression products in human. **c**, Sanger-sequencing of the *Tbxt* RT-PCR product shows that deleting exon 6 in *Tbxt* leads to correct splicing by fusing exon 5 and 7. **d**, A representative *Tbxt*^{Δexon6/+} founder mouse (day 1) showing an absence of the tail. Two additional founder mice are shown in Figure S5. **e**, *Tbxt*^{Δexon6/+} heterozygous mice display heterogeneous tail phenotypes varied from absolute no-tail to long-tails. sv, sacral vertebrae; cv, caudal vertebrae.

Discussion

We have presented evidence that tail-loss evolution in hominoids was driven by the intronic insertion of an *AluY* element. As opposed to disrupting a splice site, we inferred that this element interacts with a neighboring (simian-shared) *AluSx1* element in the neighboring intron, leading to an alternatively spliced isoform which skips an intervening exon (**Fig. 1c**). Experimental deletion of *AluY* or its interacting counterpart eliminates such *TBXT* alternative splicing in differentiated hESCs in the primitive streak state (**Fig. 2**).

Alternative splicing mediated by *Alu*-pairing in the *TBXT* gene demonstrates how an interaction between intronic transposable elements can dramatically modulate gene function to affect a complex trait. The human genome contains ~1.8 million copies of SINE transposons – including ~1 million *Alus* – of which more than 60% are intronic³⁸. Systematically searching for such interactions may lead to the identification of additional functional roles by which these elements impact human development and disease. Interestingly, inverted *Alu* pairs have been found to contribute to the biogenesis of exonic circular RNAs (circRNA)³⁹ through ‘backsplicing’. Thus, it is an interesting possibility that the interactions between paired transposable elements may create functional splice variants and circRNA isoforms from the same genetic locus.

We found that expressing the *Tbxt-Δexon6* transcript – along with the full-length transcript – in mice was sufficient to induce no-tail phenotypes, though with incomplete penetrance (**Fig. 3** and **Table 1**). It is possible that a heterogeneity of tail phenotypes also existed in the ancestral hominoids upon the initial *AluY* insertion. Thus, while tail-loss evolution in hominoids may have been initiated by the *AluY* insertion, additional genetic changes may have then acted to stabilize the no-tail phenotype in early hominoids (**Fig. S8**). Such a set of genetic events would explain how a change to the *AluY* in modern hominoids would not result in the re-appearance of the tail.

The specific evolutionary advantage for the loss of the tail is not clear, though it likely involved enhanced locomotion in a non-arboreal lifestyle. We can assume however that the selective advantage must have been very strong since the loss of the tail may have included an evolutionary trade-off of neural tube defects, as demonstrated by the presence of spinal cord malformations in the *Tbxt* ^{$\Delta exon6/\Delta exon6$} mutant at E11.5 (**Fig. S7**). Interestingly, mutations leading to neural tube defect and/or sacral agenesis have been detected in the coding and noncoding regions of the *TBXT* gene^{40–43}. We thus speculate that the evolutionary tradeoff involving the loss of the tail – made ~25 million years ago – continues to influence health today. This evolutionary insight into a complex human disease may in the future lead to the design of therapeutic strategies.

Acknowledgments

We thank Naoya Yamaguchi, Eric Wang, John Shin, Susan Liao, Huiyuan Zhang, and the members of the Yanai and Boeke labs for constructive comments and suggestions. We thank Megan Hogan and Raven Luther for sequencing assistance, and Michael Ceriello and Ahmad Naimi for assistance with the mice work. This work was supported in part by the NHGRI RM1 HG009491 to J.B., and by the NYU Grossman School of Medicine with funding to I.Y. MTM is partially funded by NIH grant R35GM119703. B.X. was partially supported by the NYSTEM pre-doctoral fellowship (C322560GG).

Author contributions

B.X. conceived the project. B.X., J.D.B. and I.Y. designed the experiments with contribution from W.Z. and S.Y.K. B.X. led and conducted most of the experimental and analysis components, with contribution from W.Z., A.W., R.B. and M.P. E.H., R.B. and M.T.M. contributed to the

capture-seq validation. A.M. and J.S.D. helped with embryo analysis work. J.D.B. and I.Y.

supervised the study. B.X. drafted the manuscript. B.X., J.D.B. and I.Y. edited the manuscript with contribution from all authors.

Competing interests

J.D.B is a Founder and Director of CDI Labs, Inc., a Founder of and consultant to Neochromosome, Inc, a Founder, SAB member of and consultant to ReOpen Diagnostics, LLC and serves or served on the Scientific Advisory Board of the following: Sangamo Inc., Modern Meadow Inc., Sample6 Inc., Tessera Therapeutics Inc. and the Wyss Institute. The other authors declare no competing interests.

Methods

2

Gene/protein sequence analysis

4 Protein sequences were downloaded from NCBI, and analyzed by the MUSCLE
algorithm using MEGA X software with default settings⁴⁴. Multiple species gene sequence
6 alignment and analysis were done through the Ensembl Comparative Genomics (release 104)
module⁴⁵, using available species of hominoids (human, chimpanzee, bonobo, gorilla,
8 orangutan and gibbon) and Old World monkeys (macaque, crab-eating macaque, pig-tailed
macaque, olive baboon, drill, black snub-nosed monkey, golden snub-nosed monkey, Ma's night
10 monkey). The identified candidate gene (*TBXT*) was then visualized through the UCSC genome
browser³⁶ (Figure 1), highlighting the multiple sequence alignment mapping to the human
12 genome (hg38) sequences.

RNA secondary structure prediction

14 RNA secondary structure prediction of the *TBXT* intron5-exon6-intron6 sequence was
performed using the ViennaRNA RNAfold Web Services (<http://rna.tbi.univie.ac.at/>). The
16 algorithm calculates the folding probability using minimum free energy (MFE) matrix with default
algorithm calculates the folding probability using minimum free energy (MFE) matrix with default
18 parameters. In addition, the calculation included the partition function and base pairing
probability matrix.

20

Human ESCs culture and differentiation

22 Human ESCs (WA01, also called H1, from WiCell Research Institute) were cultured with
StemFlex Medium (Gibco, Cat. No. A3349401) in a feeder cell-free condition. Cells were grown
24 on tissue culture-grade plates coated with hESC-qualified Geltrex (Gibco, Cat. No: A1413302).
Geltrex was 1:100 diluted in DMEM/F-12 (Gibco, Cat. No. 11320033) supplemented with 1X

GlutaMax (100X, Gibco, Cat. No. 35050061) and 1% Penicillin-Streptomycin (Gibco, Cat. No. 15070063). Before seeding hESCs, the plate was treated with Geltrex working solution in a tissue culture incubator (37°C and 5% CO₂) for at least 1h.

Human ESC maintenance and culturing condition were performed according to the manufacturer's protocol of StemFlex Medium. Briefly, StemFlex complete medium was made by combining StemFlex basal medium (450mL) with 50mL of StemFlex supplement (10X), plus 1% Penicillin-Streptomycin. Each Geltrex-coated well on a 6-well plate was seeded with 200K cells for ~80% confluence in 3-4 days. Human ESCs were cryopreserved in PSC Cryomedium (Gibco, Cat. No. A2644601). The culturing medium was supplemented with 1X RevitaCell (100X, Gibco, Cat. No. A2644501, which is also included in the PSC Cryomedium kit) when cells had gone through stressed condition, such as freezing-and-thawing or nucleofection. RevitaCell supplemented medium was replaced with regular StemFlex complete medium on the second day. hESCs grown under RevitaCell condition might become stretched, but would recover after replacing to the normal StemFlex complete medium.

The human ESC differentiation assay to induce a gene expression pattern of primitive streak was adapted from Xi *et al*³⁴. On day -1, freshly cultured hESC colonies were dissociated into clumps (2-5 cells) with Versene buffer (with EDTA, Gibco, Cat. No. 15040066). The dissociated cells were seeded on Geltrex-coated 6-well tissue culture plates at 25,000 cells/cm² (0.25M per well in the 6-well plates) in StemFlex complete medium. Differentiation to the primitive streak state was initiated on the next day (day 0) by switching StemFlex complete medium to basal differentiation medium. Basal differentiation medium (50mL) was made with 48.5mL DMEM/F-12, 1% GlutaMax (500uL), 1% ITS-G (500uL, Gibco Cat. No. 41400045), and 1% penicillin-streptomycin (500 µL), and supplemented with 3µM GSK3 inhibitor CHIR99021 (10µL of 15mM stock solution in DMSO. Tocris, Cat. No. 4423). The cells were collected at differentiation day 1 to 3 for downstream experiments, which confirmed the expression fluctuations of mesoderm genes (*TBXT* and *MIXL1*) in a 3-day differentiation period (Fig. S3)³⁴.

2 **Mouse ESC culture and differentiation**

Mouse ESCs derived from C57BL/6J strain background were cultured in a feeder cell-free condition. Cells were grown on tissue culture-grade plates coated with mESC-qualified gelatin. Before plating cells, the plastic tissue culture-treated plates were coated with 0.1% gelatin (EMD Millipore ES-006-B) at room temperature for at least 30min, followed by switching to mESC medium and warming up the medium at 37°C and 5% CO₂ incubator for at least 30min.

The feeder cell-free mESC culturing medium, also called '80/20' medium, comprises 80% 2i medium and 20% mESC medium by volume. 2i medium was made from a 1:1 mix of Advanced DMEM/F-12 (Gibco, Cat. No. 12634010) and Neurobasal-A (Gibco, Cat. No. 10888022), 1X N-2 supplement (Gibco, Cat. No. 17502048), 1X B-27 supplement (Gibco, Cat. No. 17504044), 1X Glutamax (Gibco, Cat. No. 35050061), 0.1 mM Beta-Mercaptoethanol (Gibco, Cat. No. 31350010), 1000 units/mL LIF (Millipore, Cat. No. ESG1107), 1 µM MEK1/2 inhibitor (Stemgent, Cat. No. PD0325901), and 3 µM GSK3 inhibitor CHIR99021 (Tocris, Cat. No. 4423). mESC medium was made from Knockout DMEM (Gibco, Cat. No. 10829018), containing 15% Fetal Bovine Serum (GeminiBio, Cat. No. 100-106), 0.1 mM Beta-Mercaptoethanol, 1X MEM Non Essential Amino Acids (Gibco, Cat. No. 11140050), 1X Glutamax, 1X Nucleosides (Millipore, Cat. No. ES-008-D) and 1000 units/mL LIF.

mESC differentiation for inducing *Tbxt* gene expression was adapted from Pour *et al* in a feeder cell-free condition⁴⁶. Cells were first plated in 80/20 medium for 24 hours on a gelatin-coated 6-well plate, followed by switching to N2/B27 medium without LIF or 2i for another 2-day culturing. The N2/B27 medium (50mL) is made with 18mL Advanced DMEM/F-12, 18mL Neurobasal-A, 9mL Knockout DMEM, 2.5mL Knockout Serum Replacement (Gibco, Cat. No. 10828028), 0.5mL N-2 supplement, 1mL B-27 supplement, 0.5mL Glutamax (100X), 0.5mL Nucleosides (100X), and 0.1 mM Beta-Mercaptoethanol. Then the N2/B27 medium was

supplemented with 3 μ M GSK3 inhibitor CHIR99021 for induced differentiation (day 0). The cells were collected at differentiation day 1 to 3 for downstream experiments, which showed consistent results of *Tbxt* gene expression fluctuations in a 3-day differentiation period.

CRISPR targeting

All guide RNAs of the CRISPR experiments were designed using CRISPOR algorithm through its predicted target sites integrated in the UCSC genome browser⁴⁷. Guide RNAs were cloned into the pX459V2.0-HypaCas9 plasmid (AddGene plasmid #62988), or its custom derivative by replacing the puromycin resistance gene to blasticidin resistance gene. Guide RNAs in this study were designed in pairs to delete the intervening sequences. The sequence and targeting sites of the guide RNAs were listed below:

	PAM	Guide RNA sequence	Locus (hg38/mm10)
Delete AluY	TGG	AGACTGTGCCCCACTCTCGGG	chr6:166161988-166162010
	TGG	GATAGACCATAAAGATCCCC	chr6:166161557-166161579
Delete AluSx1	GGG	CACAGTAGTTGTCCCGCTAG	chr6:166163730-166163752
	AGG	GAATGGGGGGGAGCTTAAACC	chr6:166163290-166163312
Delete Tbxt-exon6	TGG	ATTTTCGGTTCTGCAGACCGG	chr17:8438362-8438384
	GGG	CAAGATGCTGGTTGAACCAG	chr17:8438920-8438942

All oligos (plus Goldern-Gate assembly overhangs) were synthesized from Integrated DNA Technologies (IDT) and ligated into empty pX459V2.0 vector following standard Golden Gate Assembly protocol using BbsI restriction enzyme (NEB, Cat. No. R3539). The constructed plasmids were purified from 3mL *E. coli* cultures using ZR Plasmid MiniPrep Purification Kit (Zymo Research, Cat. No. D4015) for sequence verification. Plasmids delivered into ESCs were purified from 250mL *E. coli* cultures using PureLink HiPure Plasmid Midiprep Kit (Invitrogen, Cat. No. K210005). To facilitate DNA delivery to ESCs through nucleofection, the purified plasmids were resolved in Tris-EDTA buffer (pH 7.5) for a concentration of at least 1 μ g/ μ L in a sterile hood.

2 DNA delivery

DNA delivery into human/mouse ESCs for CRISPR/Cas9 targeting were performed using a Nucleofector 2b Device (Lonza, Cat. No. BioAAB-1001). Human Stem Cell Nucleofector Kit 1 (Cat. No. VPH-5012) and mESC Nucleofector Kit (Lonza, Cat. No. VVPH-1001) were used for delivering DNA into human and mouse ESCs, respectively. ESCs were double-fed the day before the nucleofection experiment to maintain a superior condition.

Before performing nucleofection on human ESCs, 6cm tissue culture plates were treated with 0.5µg/cm² rLaminin-521 in a 37°C and 5% CO₂ incubator for at least 2h. rLaminin-521-treated plates give better viability when seeding hESCs as single cells. Cultured human ESCs were then washed with PBS, and dissociated into single cells using TrypLE Select Enzyme (no phenol red. Gibco, Cat. No. 12563011). One million hES single cells were nucleofected using program A-023 according to the manufacturer's instruction of the Nucleofector 2b device. Transfected cells were transferred on the rLaminin-521-treated 6cm plates with pre-warmed StemFlex complete medium supplementing with 1X RevitaCell but not Penicillin-Streptomycin. Antibiotic selection was performed 24h after nucleofection with puromycin (Gibco, Cat. No. A1113802).

Mouse ESCs were dissociated into single cells using StemPro Accutase (Gibco, Cat. No. A1110501) and five million cells were transfected using program A-023 according to the manufacturer's instruction. Cells were plated onto gelatin-treated 10cm plates, followed by antibiotic selection 24h after nucleofection with blasticidin (Gibco, Cat. No. A1113903).

Together with the pX459V2.0-HypaCas9-gRNA plasmids for nucleofection, a single-strand DNA oligo were co-delivered for micro homology-induced deletion of the targeted sites⁴⁸. These ssDNA sequences were synthesized from IDT through its Ultramer DNA Oligo service, including phosphorothioate bond modification on the three bases of each end. Detailed sequence information was listed below ("|" indicates a junction site):

Delete <i>TBXT-AluY</i>	A*C*T*ACACCAGTGATTCTCCAAATACGGTGCCCAGGCCAGAGCCTCAGC ACCACCC CCCTGGCCTAATTCTCTCTATAACACTGTATATGTCTAGACTTA ATTTTGTC*T*G*G
Delete <i>TBXT-AluSx1</i>	C*A*G*T*GCTGCTGGAGAATTGTTAGTAGTTTGGAAATTGAAGCCACAGTAGTTGT CCCGC ACCAGGAGAGTGAGCAGTAAAAGGGTCTACCCCAGCTAGGAAGGCACCT CCCGTC*T*C*T
Delete <i>Tbxt-exon6</i>	T*T*T*ATTCTAGAGCCCATTAAACATATCACTCCTGCTCACTTGGTAGAAAGCCACCG CAGGGGTCCCCAAGGAGGCTTTCAATTCATATCCATGTGCCTCAGAACATG*C*C* C

2 Genotyping of CRISPR/Cas9-targeted sites were performed through PCR following standard protocol. The genotyping primers were listed below:

Target site	Orientation	Sequence
Delete <i>TBXT-AluY</i> : read through	F	CAGCCAGGCTCAAGAATTCC
	R	GACTTCCTAACCCAATAAGGTCC
Delete <i>TBXT-AluY</i> : left junction	F	CAGCCAGGCTCAAGAATTCC
	R	GTGTTCTAATATTGGAGCATGC
Delete <i>TBXT-AluSx1</i> : read through	F	TCCTAGGCTGATTGAACAACCAG
	R	CAAGGCAGGTGAGCTTTCC
Delete <i>TBXT-AluSx1</i> : left junction	F	TCCTAGGCTGATTGAACAACCAG
	R	TTAAGCTCCCCCATTC
Delete <i>Tbxt-exon6</i> : read through	F	GCAGTCTGAGTCCTACCTGTG
	R	TGTCAGTCTGGTTCTACACCTGGAGAGTCTTTGATC
Delete <i>Tbxt-exon6</i> : left junction	F	GCAGTCTGAGTCCTACCTGTG
	R	GTACAGGACCTACTTGGAGAGC
Delete <i>Tbxt-exon6</i> : right junction	F	GACAGGACTGAGTCTCAAGC
	R	TGTCAGTCTGGTTCTACACCTGGAGAGTCTTTGATC

4
6 **Splicing isoforms detection**
Total RNAa were collected from the undifferentiated or differentiated cells of both human
8 and mouse ESCs, using standard column-based purification kit (QIAGEN RNeasy Kit, Cat. No.
74004). DNase treatment was applied during the purification to remove any potential DNA
10 contamination. Following extraction, RNA quality was checked through electrophoresis based
on the ribosomal RNA integrity. Reverse transcription was performed with 1µg of high-quality
12 total RNA for each sample, using High-Capacity RNA-to-cDNA™ Kit (Applied Biosystems, Cat.
No. 4387406). DNA oligos used for PCR/RT-PCR/RT-qPCR were listed below:

Target site	Orientation	Sequence
Human <i>TBXT</i> : exon3-8 (Fig. 2B)	F	GGTGACTGCTTATCAGAACGAGGAG
	R	TACTGAGGCTGCATTTCTTCTTAACC
Human <i>TBXT</i> (Fig. S2)	F	CCTCATAGCCTCATGGACACCTG
	R	TCTTAACCTGAGACTGCCACTGG
Human <i>MIXL1</i> (Fig. S2)	F	GGCGTCAGAGTGGGAAATCC
	R	GGCAGGCAGTTCACATCTACC
Human <i>ACTB1</i> (Fig. S2)	F	CACCATTGGCAATGAGCGGTTC
	R	AGGTCTTTGCGGATGTCCACGT
Human <i>TBXT</i> : exon4-7 (Fig. S2)	F	CAGAACGAGGAGATCACAGCTC
	R	GGTACTGACTGGAGCTGGTAGG
Mouse <i>Tbxt</i> : exon4-7 (Fig. S2)	F	CCAGAATGAGGAGATTACAGCCCT
	R	GGATACTGGCTAGAGCCAGTAGG
Mouse <i>Actb1</i> (Fig. S2)	F	CATTGCTGACAGGATGCAGAAGG
	R	TGCTGGAAGGTGGACAGTGAGG

Mouse work

All mouse work was done following NYULH's animal protocol guidelines. The *Tbxt* ^{Δ exon6/+} heterozygous mouse model was generated through zygotic microinjection, using an experimental protocol adapted from Yang *et al*⁴⁹. Briefly, Cas9 mRNA (MilliporeSigma, Cat. No. CAS9MRNA), synthetic guide RNAs, and single-stranded DNA oligo were co-injected into the 1-cell stage zygotes following the described procedures⁴⁹. Synthetic guide RNAs were ordered from Synthego as their custom CRISPR revolution sgRNA EZ Kit, with the same targeting sites as used in the CRISPR deletion experiment of mouse ESCs (AUUUCGGUUCUGCAGACCGG and CAAGAUGCUGGUUGAACCAG). The co-injected single-stranded DNA oligo is the same as above mentioned as well. Processed embryos were then *in vitro* cultured to the blastomeric stage, followed by embryo transferring to the pseudopregnant foster mothers. Following zygotic microinjection and transferring, founder pups were screened based on their abnormal tail phenotypes. DNA samples were collected through ear punches at day ~21 for genotyping.

Upon confirming the heterozygous genotype (*Tbxt* ^{Δ exon6/+}), founder mice were backcrossed with wild-type C57B/6J mice for generating heterozygous F1 pups. Due to the

varied tail phenotypes, intercrossing between F1 heterozygotes were performed in two categories: type1 intercrossing includes at least one parent being no-/short-tailed, whereas type 2 intercrossing were mated between two long-tailed F1 heterozygotes. Both types of intercrossing produced heterogeneous tail phenotypes in F2 *Tbxt* ^{Δ exon6/+} pups, confirming the incomplete penetrance of tail phenotypes, and the absence of homozygotes (*Tbxt* ^{Δ exon6/ Δ exon6}), as summarized in Table 1. To confirm the embryonic phenotypes in homozygotes, embryos were dissected at E11.5 gestation stage from the timed pregnant mice through the standard protocol. Adult mice (>12 weeks) were anesthetized for X-ray imaging of vertebra using a Bruker In-Vivo Xtreme IVIS imaging system.

Capture-seq genotyping

Capture-seq, or targeted sequencing of the loci of interest, was performed as previously described³⁷. Conceptually, capture-seq uses custom biotinylated probes to pull down the genomic loci of interest from the standard whole-genome sequencing libraries, thus enabling sequencing of the specific genomic loci in a much higher depth while reducing the cost.

Genomic DNA were purified from mESCs or ear punches of founder mice using Zymo Quick-DNA Miniprep Plus Kit (Cat. No. D4068) according to manufacturer's instruction. DNA sequencing libraries compatible for Illumina sequencers were prepared following standard protocol. Briefly, 1µg of gDNA was sheared to 500-900 base pairs in a 96-well microplate using the Covaris LE220 (450 W, 10% Duty Factor, 200 cycles per burst, and 90-s treatment time), followed by purification with a DNA Clean and Concentrate-5 Kit (Zymo Research, Cat. No. D4013). Sheared and purified DNA were then treated with end repair enzyme mix (T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase, NEB, Cat. No. M0203, M0210 and M0201, respectively), and A-tailed using Klenow 3'-5'exo- enzyme (NEB, Cat. No.

M0212). Illumina sequencing library adapters were subsequently ligated to DNA ends, followed

by PCR amplification with KAPA 2X Hi-Fi Hotstart Readymix (Roche, Cat. No. KR0370).

Custom biotinylated probes were prepared as bait through nick translation, using BAC DNA and/or plasmids as the template. The probes were prepared to comprehensively cover the whole locus. We used BAC lines RP24-88H3 and RP23-159G7, purchased from BACPAC Genomics, to generate bait probes covering mouse *Tbxt* locus and ~200kb flanking sequences in both upstream and downstream regions. The pooled whole-genome sequencing libraries were hybridized with the biotinylated baits in solution, and purified through streptavidin-coated magnetic beads. Following pull-down, DNA sequencing libraries were quantified with Qubit 3.0 Fluorometer (Invitrogen, Cat. No. Q33216) using a dsDNA HS Assay Kit (Invitrogen, Cat. No. Q32851). The sequencing libraries were subsequently sequenced on an Illumina NextSeq 500 sequencer in paired-end mode.

Sequencing results were demultiplexed with Illumina bcl2fastq v2.20 requiring a perfect match to indexing BC sequences. Low quality reads/bases and Illumina adapters were trimmed with Trimmomatic v0.39. Reads were then mapped to mouse genome (mm10) using bwa 0.7.17. The coverage and mutations in and around *Tbxt* locus were checked through visualization in a mirror version of UCSC genome browser.

Supplementary Figures:

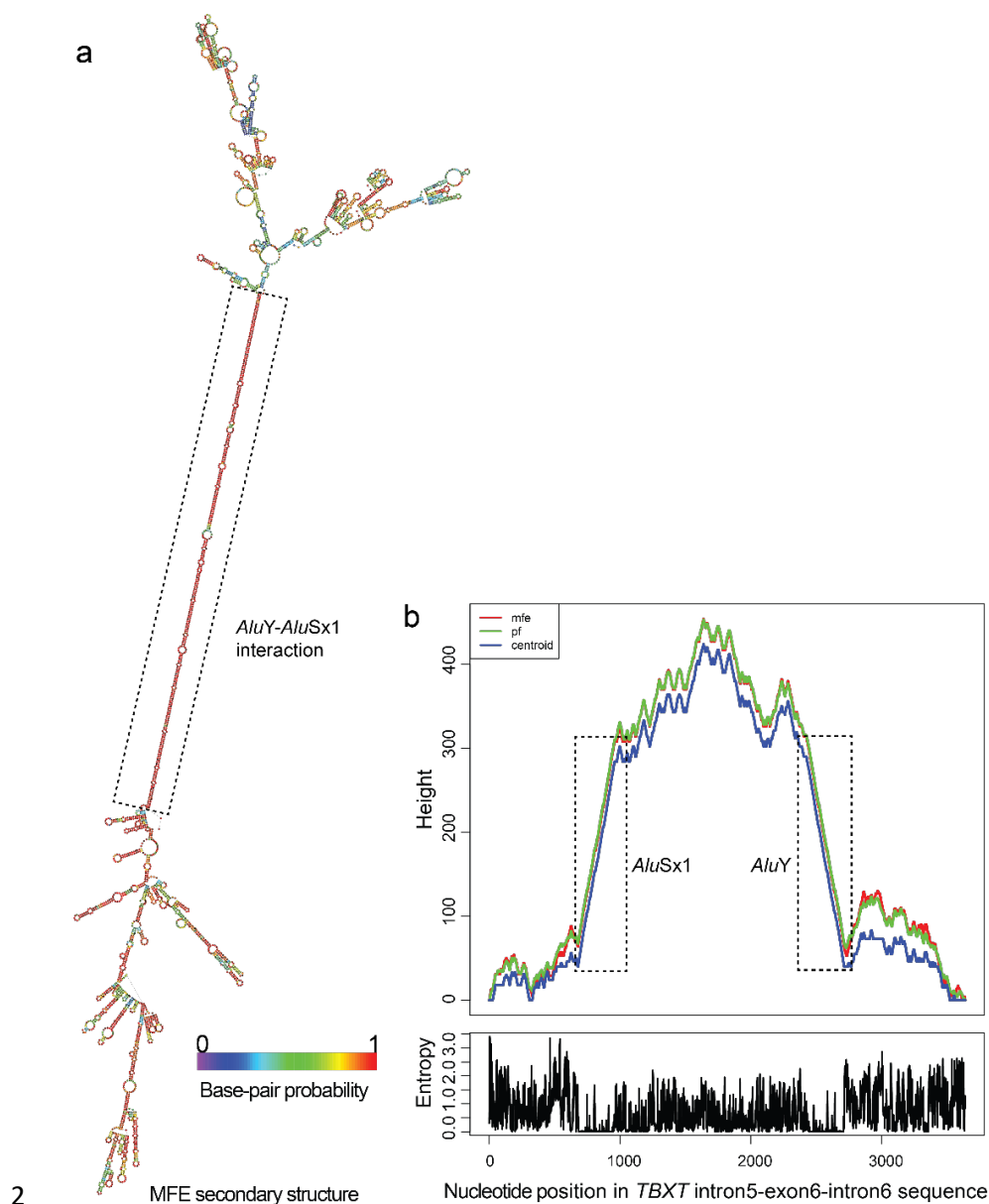


Fig. S1 | RNA structure prediction using RNAfold algorithm of the ViennaRNA package³³.

a, Predicted RNA secondary structure of the *TBXT* intron5-exon6-intron6 sequence. The paired *AluY-AluSx1* region is highlighted. **b**, Mountain plot of the RNA secondary structure prediction, showing the 'height' in predicted secondary structure across the nucleotide positions. Height is computed as the number of base pairs enclosing the base at a given position. Overall, the *AluSx1* and *AluY* regions are predicted to form helices with high probability (low entropy).

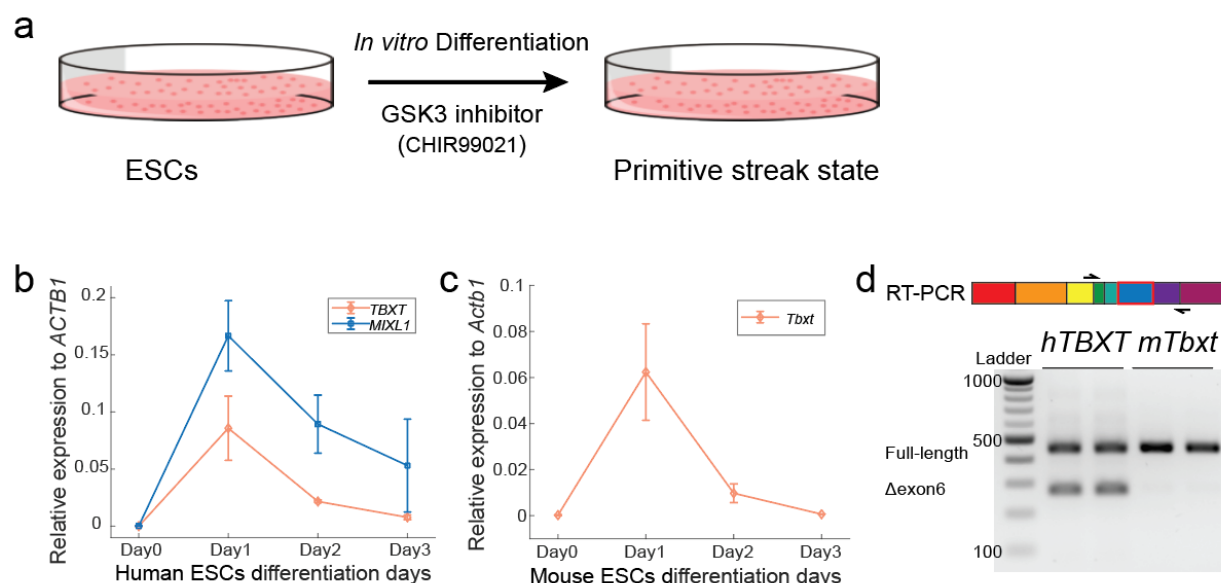
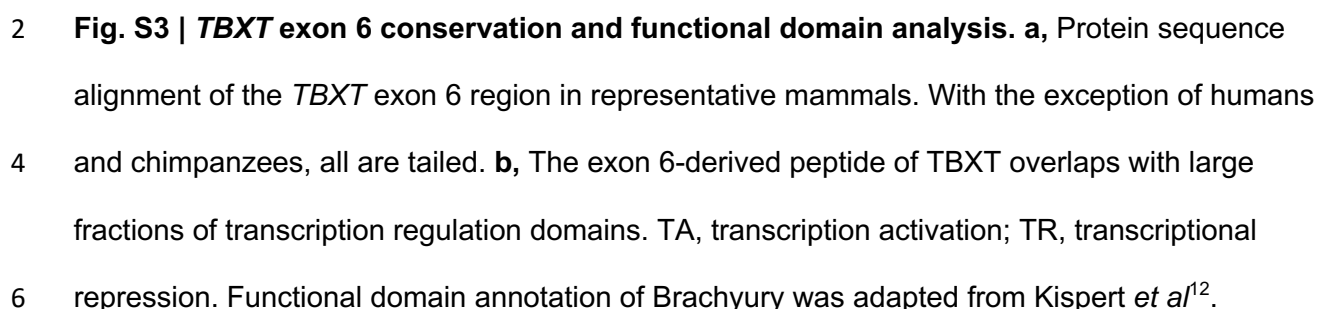


Fig. S2 | Studying *TBXT* expression isoforms using primitive streak *in vitro* differentiation. **a**, Human and mouse ESCs *in vitro* differentiation for inducing *TBXT* expression. Human and mouse ESCs differentiation assay was adapted from Xi *et al*³⁴ and Pour *et al*³⁵, respectively. **b**, Quantitative RT-PCR of *TBXT* and *MIXL1* expression during hESC differentiation, indicating correct induction of mesodermal gene expression program³⁴. **c**, Quantitative RT-PCR of *Tbx1* expression during mESC differentiation. **d**, RT-PCR of *TBXT/Tbx1* transcripts in human and mouse, highlighting a unique Δ exon6 splicing isoform in human.



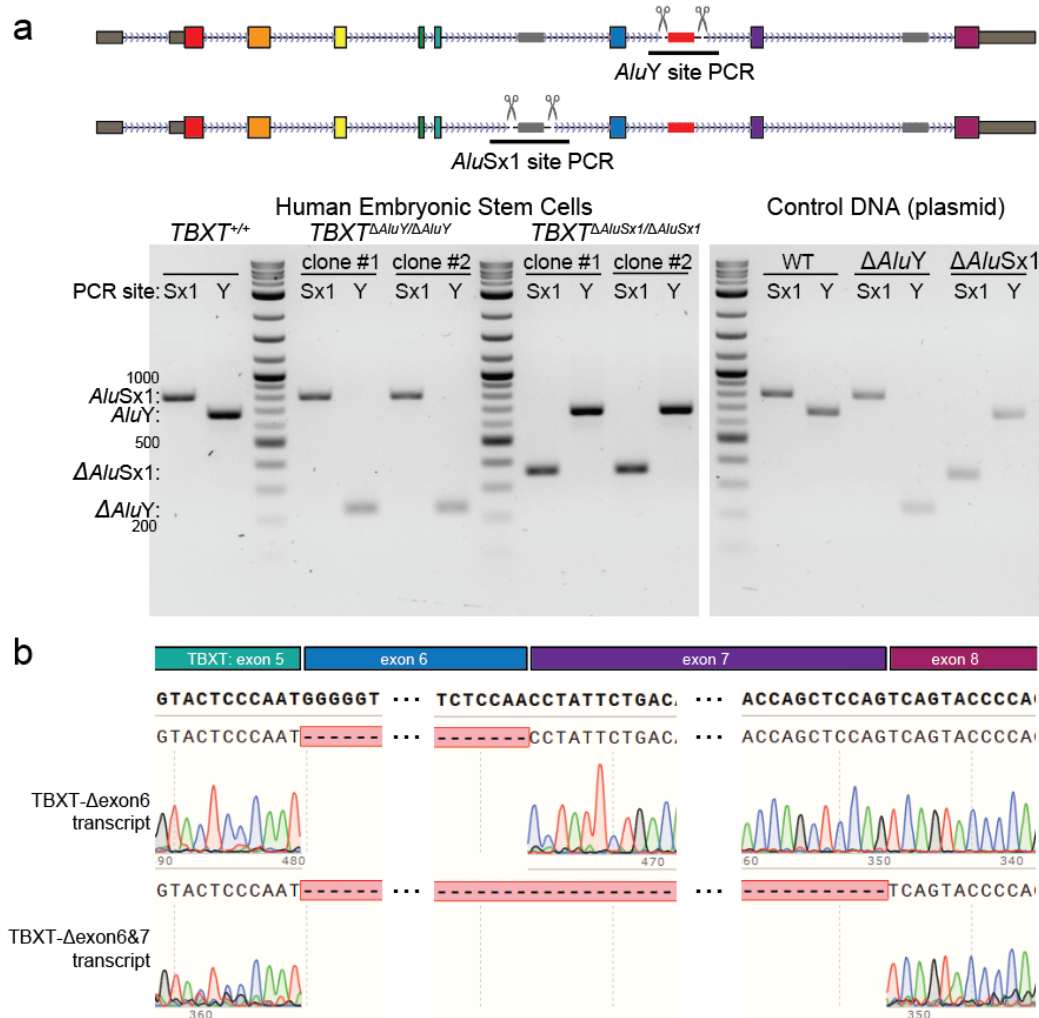


Fig. S4 | Validation of hESC CRISPR-deletion clones and the *TBXT* expression isoforms.

a, PCR validation of the hESC clones with deletions of *AluY* or *AluSx1* in *TBXT*. PCR validation for each clone or control samples were performed in pairs, each amplifying the *AluSx1* locus (Sx1) or the *AluY* locus (Y), respectively, with primers that bind the two flanking sequences of the deleted region. Each genotype included two independent clones of *AluY* deletion or *AluSx1* deletion, corresponding to the two replicates in Figure 2B. **b**, Sanger sequencing of the *TBXT*- Δ exon6 and *TBXT*- Δ exon6&7 transcripts detected in Figure 2B. The sequencing results were aligned to the full length *TBXT* mRNA sequence.

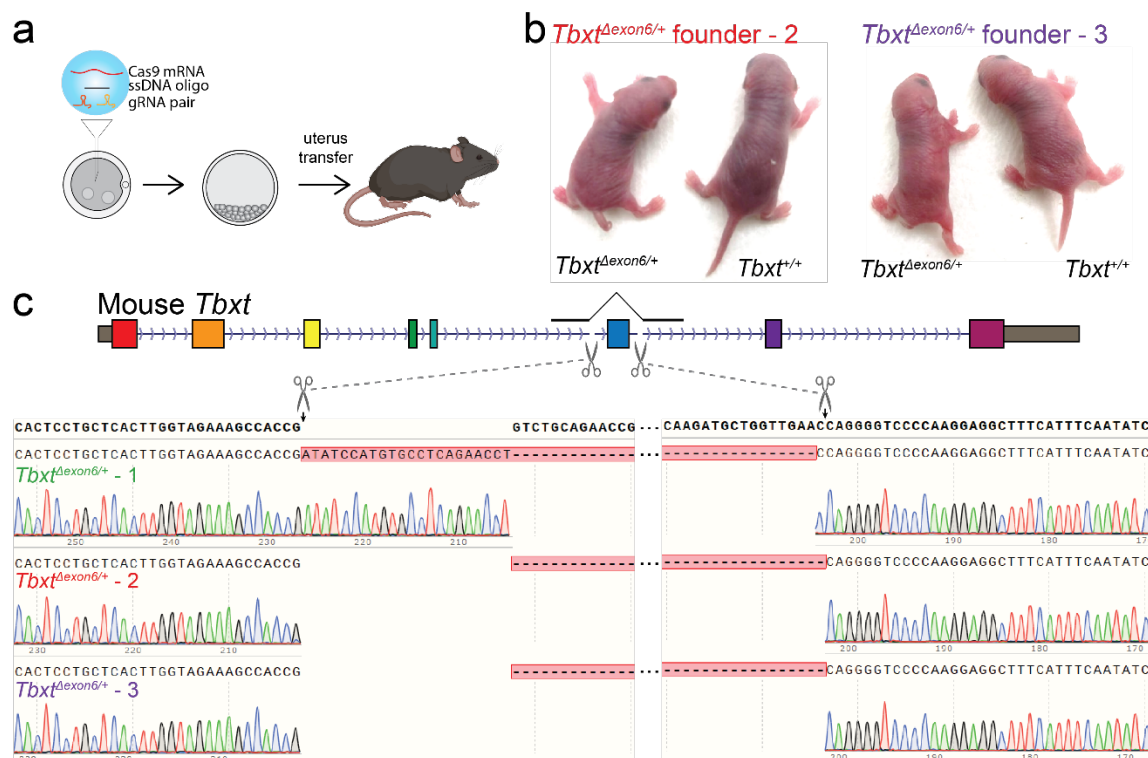


Fig. S5 | *Tbxt*^{Δexon6/+} founder mice generated through CRISPR/Cas9 targeting in the zygotes. **a**, Schematic of zygotic injection of CRISPR/Cas9 reactions. **b**, Two *Tbxt*^{Δexon6/+} founder mice (in addition to the one shown in Fig. 3) indicating an absence or reduced form of the tail (Founders 2 & 3). **c**, Sanger sequencing of the exon 6-deleted allele isolated from the genomic DNA of *Tbxt*^{Δexon6/+} founder mice. Founder 1 had an unexpected insertion of 23 base pairs at the CRISPR cutting site in the original intron 5 of *Tbxt*. Both founder 2 and 3 had the exact fusion between the two CRISPR cutting sites in introns 5 and 6.

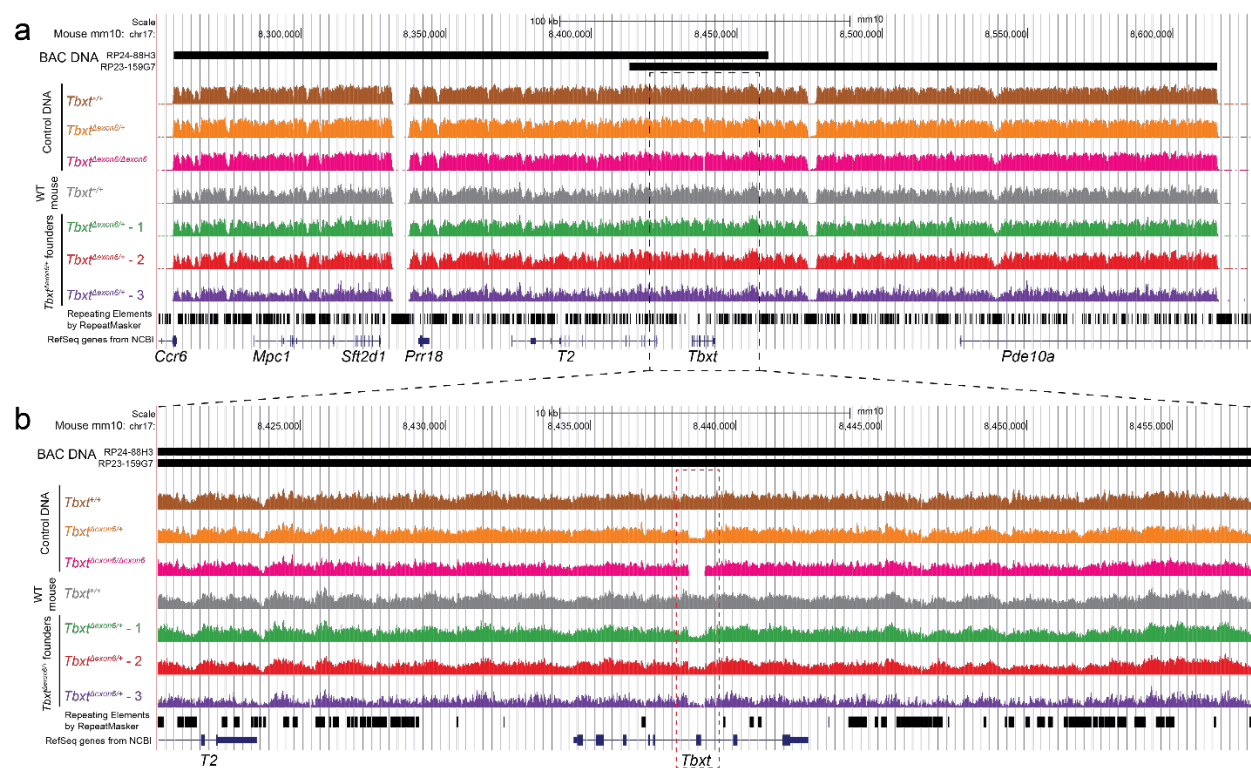


Fig. S6 | Capture-seq at the *Tbx1* locus of founder mice did not detect off target

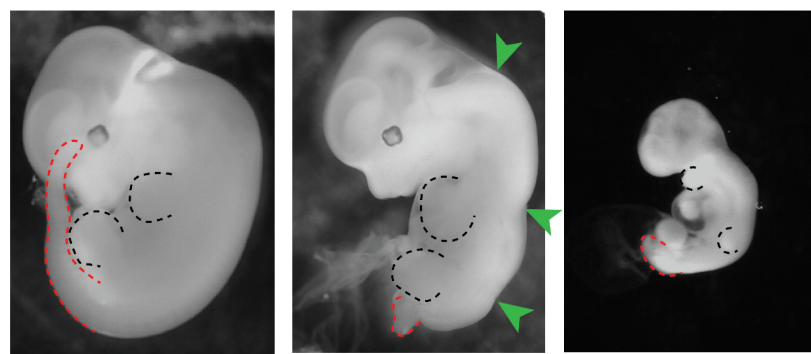
mutations. a, Capture-seq of the founder mice using baits generated from bacterial artificial

chromosomes (RP24-88H3 and RP23-159G7). The shallow-covered regions are typically repeat

sequences in the mouse genome and are consistent across samples. Control DNAs were

obtained from wild-type or exon6-deleted mESCs through CRISPR targeting. **b**, A zoom-in view

of the Capture-seq results at the *Tbx1* locus, highlighting the CRISPR-deleted exon 6 region.

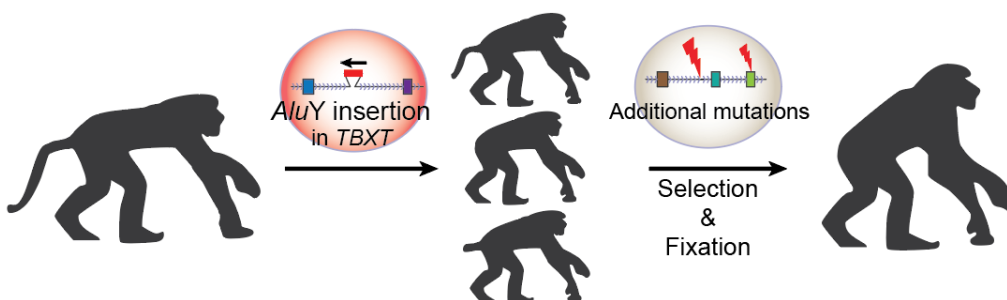


2 *Tbxt*^{+/+} *Tbxt*^{Δexon6/Δexon6}(#1) *Tbxt*^{Δexon6/Δexon6}(#2)

Fig. S7 | Analysis of *Tbxt*^{Δexon6/Δexon6} embryos at the E11.5 stage. *Tbxt*^{Δexon6/Δexon6} embryos

4 either develop spinal cord defects (middle) that die at birth or arrest at approximately stage E9
of development (right). Red and black dashed lines mark the embryonic tail regions and limb
6 buds, respectively. Green arrowheads in the middle panel indicate malformed spinal cord
regions.

8



2 **Fig. S8 | A model for tail-loss evolution in the early hominoids.** The *AluY* insertion in *TBXT*
 marked an early genetic event that initiated tail-loss evolution in the hominoid common
 4 ancestor. Additional genetic changes may have then acted to stabilize the no-tail phenotype in
 the ancient hominoids.

6

References

1. Darwin, C. *The descent of man, and selection in relation to sex*. (degruyter.com, 2008).
2. *Human evolution: an introduction to man's adaptations*. (Routledge, 2017).
doi:10.4324/9780203789537
3. Hunt, K. D. The evolution of human bipedality: ecology and functional morphology. *J. Hum. Evol.* **26**, 183–202 (1994).
4. Williams, S. A. & Russo, G. A. Evolution of the hominoid vertebral column: The long and the short of it. *Evol Anthropol* **24**, 15–32 (2015).
5. Hickman, G. C. The mammalian tail: a review of functions. *Mamm. Rev.* **9**, 143–157 (1979).
6. Rogers, J. & Gibbs, R. A. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* **15**, 347–359 (2014).
7. Rhesus Macaque Genome Sequencing and Analysis Consortium *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
8. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
9. Kimelman, D. Tales of tails (and trunks): forming the posterior body in vertebrate embryos. *Curr Top Dev Biol* **116**, 517–536 (2016).
10. Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617–622 (1990).
11. Edwards, Y. H. *et al.* The human homolog T of the mouse T(Brachyury) gene; gene structure, cDNA sequence, and assignment to chromosome 6q27. *Genome Res.* **6**, 226–233 (1996).
12. Kispert, A., Koschorz, B. & Herrmann, B. G. The T protein encoded by Brachyury is a tissue-specific transcription factor. *EMBO J.* **14**, 4763–4772 (1995).
13. Wilde, J. J., Petersen, J. R. & Niswander, L. Genetic, epigenetic, and environmental contributions to neural tube closure. *Annu. Rev. Genet.* **48**, 583–611 (2014).
14. Sehner, S., Fichtel, C. & Kappeler, P. M. Primate tails: Ancestral state reconstruction and determinants of interspecific variation in primate tail length. *Am. J. Phys. Anthropol.* **167**, 750–759 (2018).
15. Russo, G. A. Postsacral vertebral morphology in relation to tail length among primates and other mammals. *Anat Rec (Hoboken)* **298**, 354–375 (2015).
16. Lemelin, P. Comparative and functional myology of the prehensile tail in New World monkeys. *J Morphol* **224**, 351–368 (1995).
17. Narita, Y. & Kuratani, S. Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. *J Exp Zool B Mol Dev Evol* **304**, 91–106 (2005).

18. Young, N. M., Wagner, G. P. & Hallgrímsson, B. Development and the evolvability of human limbs. *Proc. Natl. Acad. Sci. USA* **107**, 3400–3405 (2010).
19. Pontzer, H., Raichlen, D. A. & Rodman, P. S. Bipedal and quadrupedal locomotion in chimpanzees. *J. Hum. Evol.* **66**, 64–82 (2014).
20. Bauer, H. R. Chimpanzee bipedal locomotion in the Gombe National Park, East Africa. *Primates* (1977).
21. Mallo, M. The vertebrate tail: a gene playground for evolution. *Cell Mol. Life Sci.* **77**, 1021–1030 (2020).
22. Smith, C. L. & Eppig, J. T. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* **1**, 390–399 (2009).
23. Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343**, 657–659 (1990).
24. Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. & McMahon, A. P. T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev.* **13**, 3185–3190 (1999).
25. Tasic, J. *et al.* Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat. Cell Biol.* **21**, 1518–1531 (2019).
26. Buckingham, K. J. *et al.* Multiple mutant T alleles cause haploinsufficiency of Brachyury and short tails in Manx cats. *Mamm. Genome* **24**, 400–408 (2013).
27. Haworth, K. *et al.* Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm. Genome* **12**, 212–218 (2001).
28. Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B. & Nüsslein-Volhard, C. no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009–1015 (1994).
29. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379 (2002).
30. Wallace, M. R. *et al.* A de novo Alu insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866 (1991).
31. Lev-Maor, G. *et al.* Intronic Alus influence alternative splicing. *PLoS Genet.* **4**, e1000204 (2008).
32. Payer, L. M. *et al.* Alu insertion variants alter mRNA splicing. *Nucleic Acids Res.* **47**, 421–431 (2019).
33. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
34. Xi, H. *et al.* In Vivo Human Somatogenesis Guides Somite Development from hPSCs. *Cell Rep.* **18**, 1573–1585 (2017).

35. Pour, M. *et al.* Emergence and patterning dynamics of mouse definitive endoderm. *SSRN Journal* (2021). doi:10.2139/ssrn.3848112
36. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
37. Brosh, R. *et al.* A versatile platform for locus-scale genome rewriting and verification. *Proc. Natl. Acad. Sci. USA* **118**, (2021).
38. International Human Genome Sequencing Consortium *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
39. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
40. Shaheen, R. *et al.* T (brachyury) is linked to a Mendelian form of neural tube defects in humans. *Hum. Genet.* **134**, 1139–1141 (2015).
41. Shields, D. C. *et al.* Association between historically high frequencies of neural tube defects and the human T homologue of mouse T (Brachyury). *Am. J. Med. Genet.* **92**, 206–211 (2000).
42. Morrison, K. *et al.* Genetic mapping of the human homologue (T) of mouse T(Brachyury) and a search for allele association between human T and spina bifida. *Hum. Mol. Genet.* **5**, 669–674 (1996).
43. Postma, A. V. *et al.* Mutations in the T (brachyury) gene cause a novel syndrome consisting of sacral agenesis, abnormal ossification of the vertebral bodies and a persistent notochordal canal. *J. Med. Genet.* **51**, 90–97 (2014).
44. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
45. Herrero, J. *et al.* Ensembl comparative genomics resources. *Database (Oxford)* **2016**, (2016).
46. Pour, M. *et al.* Emergence and patterning dynamics of mouse definitive endoderm. *SSRN Journal* (2021). doi:10.2139/ssrn.3848112
47. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
48. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8**, 753–755 (2011).
49. Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Protoc.* **9**, 1956–1968 (2014).