Title: Post-fertilization transcription initiation in an ancestral LTR retrotransposon drives lineage-specific genomic imprinting of *ZDBF2*

Running title: Genomic imprinting and LTR insertion

Hisato Kobayashi[1,*] Tatsushi Igaki[1], Soichiro Kumamoto[2,3], Keisuke Tanaka[4,5], Tomoya Takashima[1], Shunsuke Suzuki[6], Masaaki Hayashi[6], Marilyn B. Renfree[7], Manabu Kawahara[8], Shun Saito[8], Toshihiro Kobayashi[9,10,11], Hiroshi Nagashima[11], Hitomi Matsunari[11], Kazuaki Nakano[11], Ayuko Uchikura[11], Hiroshi Kiyonari[12], Mari Kaneko[12], Hiroo Imai[13], Kazuhiko Nakabayashi[14], Matthew C. Lorincz[15], Kazuki Kurimoto[1]

1 Department of Embryology, Nara Medical University, Kashihara, Nara 634-0813, Japan.

2 Department of Chemistry and Biochemistry, School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan

3 Division of Cancer and Senescence Biology, Cancer Research Institute, Institute for Frontier Science Initiative, Kanazawa University, Kanazawa, Ishikawa 920-1192, Japan

4 Department of Informatics, Tokyo University of Information Sciences, Wakaba, Chiba 265-8501, Japan

5 NODAI Genome Research Center, Tokyo University of Agriculture, Setagaya, Tokyo 156-8502, Japan

6 Department of Agriculture, Graduate School of Science and Technology, Shinshu University, Nagano 399-4598, Japan

7 School of BioSciences, University of Melbourne, Melbourne, Victoria 3010, Australia

8 Laboratory of Animal Genetics and Reproduction, Research Faculty of Agriculture,

Hokkaido University, Sapporo, Hokkaido 060-8589, Japan

9 Division of Mammalian Embryology, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, University of Tokyo, Minato, Tokyo 108-8639, Japan.

10 Center for Genetic Analysis of Behavior, National Institute for Physiological Sciences, Okazaki, Aichi 444-8787, Japan.

11 Meiji University International Institute for Bio-Resource Research, Kawasaki 214-8571, Japan

12 Laboratory for Animal Resources and Genetic Engineering, RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo 650-0047, Japan

13 Molecular Biology Section, Center for the Evolutionary Origins of Human Behavior, Kyoto University, Inuyama, Aichi 484-8506 Japan

14 Division of Developmental Genomics, Research Institute, National Center for Child Health and Development, Setagaya, Tokyo 157-8535, Japan

15 Life Sciences Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.

*Corresponding author: hiskobay@naramed-u.ac.jp (Hisato Kobayashi)

ABSTRACT

The imprinted *ZDBF2* gene is controlled by oocyte-derived DNA methylation, but its regulatory system is quite different from that of other canonically imprinted genes that are dependent on DNA methylation deposited in the gametes. At the *ZDBF2* locus, maternal DNA methylation in the imprinted differentially methylated region (DMR) does not persist after implantation. Instead, a transient transcript expressed in the early

embryo exclusively from the unmethylated paternal allele of the DMR, known as *GPR1-AS*, contributes to establishing secondary DMRs that maintain paternal expression of *ZDBF2* in the somatic lineage. While the imprinting of *ZDBF2* and its unique regulatory system are evident in humans and mice, whether this process is conserved in other mammals has not been addressed. Here, we show that the first exon of human *GPR1-AS* overlaps with that of a long terminal repeat (LTR) belonging to the MER21C subfamily of retrotransposons. Although this LTR family appears and is amplified in eutherians, the MER21C insertion into the *GPR1-AS* orthologous region occurred specifically in the common ancestor of Euarchontoglires, a clade that includes primates, rodents, and rabbits. Directional RNA sequencing of placental tissues from various mammalian species revealed *GPR1-AS* orthologs in rabbits and nonhuman primates, with their first exon embedded within the same ancestral LTR. In contrast, allele-specific expression profiling showed that cow and tammar wallaby, mammals outside the Euarchontoglires group, expressed both alleles in all tissues analyzed. Our previous studies showed that LTRs reactivated in oocytes drive lineage-specific imprinting during mammalian evolution. The data presented here suggest that LTR-derived sequence activation after fertilization can also contribute to the lineage-specific establishment of imprinted genes.

Keywords: genomic imprinting, LTR retrotransposon, mammalian evolution, placenta, RNA-seq

INTRODUCTION

Genomic imprinting is a well-established biological phenomenon that refers to the differential expression of imprinted genes depending on their parental origin. Epigenetic mechanisms, including DNA methylation, histone modification, and noncoding RNA molecules, regulate this process, establishing and maintaining parent-of-origin-specific gene expression patterns. Imprinted genes play vital roles in various biological processes such as fetal growth and development, placental function, metabolism, and behavior. These imprinted genes are often clustered within imprinted domains and regulated by epigenetic marks in specific imprinting control regions (ICRs). The inheritance of parent-of-origin-specific DNA methylation marks from the oocyte or sperm occurs in germline differentially methylated regions (germline DMRs), which are propagated in all somatic tissues of the next generation and function as ICRs (therefore, referred to as canonical imprinting). Imprinting patterns can vary significantly among species. Although some genes are imprinted in only one or a few species, others are conserved in many mammals. For example, nearly 200 imprinted genes have been identified in mice and humans. However, less than half of these genes overlapped in both species, suggesting the existence of many species- or lineage-specific imprinted genes (Tucci et al. 2019). The establishment of species-specific imprinting at specific loci can be driven by various evolutionary events, including mutations, differences in the function of DNA methyltransferases or the binding sites of proteins that control gene expression, the acquisition of new genes by retrotransposition, and the insertion of CpG-rich regions (CpG islands) that can serve as germline DMRs (Takahashi et al. 2019; Kobayashi 2021; Kaneko-Ishino and Ishino 2022). Our previous studies have also shown that oocyte-specific transcripts from upstream promoters, such as integrated LTR retrotransposons,

play a critical and fundamental role in the mechanism underlying the establishment of lineage-specific maternal germline DMRs (Brind'Amour et al. 2018; Bogutz et al. 2019). However, these genetic and genomic backgrounds alone do not explain the establishment of all mouse- or human-specific imprinted genes identified to date. Furthermore, in mammals other than humans and mice, the cataloging of imprinted genes has not been systematically addressed and species-specific imprinting events remain poorly understood.

Zinc finger DBF-type containing 2 (*ZDBF2*), a paternally expressed gene located on human chromosome 2q37.1, has been reported to regulate neonatal feeding behavior and growth in mice, although its molecular function remains unknown (Kobayashi et al. 2009; Glaser et al. 2022). There are at least three imprinted genes, *ZDBF2*, *GPR1* (alternatively named *CMKLR1*), and *ADAM23* around this locus (Hiura et al. 2010; Morcos et al. 2011). Recently, the *Zdbf2* gene was also shown to be paternally expressed in rats, a member of the family Muridae, similar to mice (Richard Albert et al. 2023)(**Supplemental Table S1**). Studies have shown that *ZDBF2* imprinting is regulated by two types of imprinted DMRs: a germline DMR located in the intron of the *GPR1* gene and a somatic (secondary) imprinted DMR upstream of *ZDBF2* (Kobayashi et al. 2012; Kobayashi et al. 2013; Duffie et al. 2014). The germline DMR is methylated on the maternal allele, leading to silencing of *GPR1* antisense RNA (*GPR1-AS*, alternatively named *CMKLR1-AS*) expression on that allele, whereas the paternal allele is hypomethylated, allowing for the expression of *GPR1-AS*. *GPR1-AS* is transcribed in the reverse direction of *GPR1* (in the same direction as *ZDBF2*) from the germline DMR. Although it is not known whether *GPR1-AS* encodes a protein, a fusion transcript of

Gpr1-as (*Platr12*) and *Zdbf2* (alternatively named *Zdbf2linc* or *Liz*; a long isoform of *Zdbf2*) has been identified in mice. This fusion transcript exhibited an imprinted expression pattern similar to that of human *GPR1-AS* (**Supplemental Table S1**). *Liz* directs DNA methylation at the somatic DMR, which competes with *ZDBF2* to repress the paternal allele (Greenberg et al. 2017). Germline DMRs that function as ICRs generally maintain uniparental methylation throughout development; however, germline DMRs at this locus are no longer maintained because the paternally derived allele is also methylated after implantation. The secondary DMR, rather than germline DMR, is thought to maintain the imprinted status of this locus in somatic cell lineages. The unique regulatory mechanism responsible for imprinting at the *ZDBF2* locus appears to be conserved between humans and mice; however, whether this conservation extends to other mammals remains unknown.

Here, we focus on *GPR1-AS*, a critical factor in the imprinting of *ZDBF2*, and identified orthologous transcripts of *GPR1-AS* in primates and rabbits using RNA-seq-based transcript analysis. The first exon of these orthologs overlapped with a common LTR retrotransposon, suggesting that LTR insertion during a local evolutionary event in their common ancestor led to the establishment of *ZDBF2* imprinting. In support of this hypothesis, *ZDBF2* expression is not imprinted in other mammals such as cattle and tammar wallabies, which lack this LTR. Our findings demonstrate that the insertion of an LTR-derived sequence that becomes active after fertilization triggered the evolution of lineage-specific imprinting of *ZDBF2* in a specific branch of the mammalian lineage.

RESULTS

Detection of *GPR1-AS* orthologs using RNA-seq data sets

GPR1-AS and *Liz* expression have been observed in placental tissues, blastocyst-to-gastrulation embryos, and/or ES cells in humans and mice (Kobayashi et al. 2009; Kobayashi et al. 2012; Kobayashi et al. 2013; Duffie et al. 2014). Public human tissue RNA-seq data sets show detectable *GPR1-AS* transcription in the placenta, albeit at a relatively low level (RPKM value of around 0.45–1.8), but significantly suppressed in other tissues (**Supplemental Fig. S1**) (Fagerberg et al. 2014; Duff et al. 2015). To identify *GPR1-AS* orthologs across different species, we performed transcript prediction analysis using short-read RNA-seq data from multiple mammalian placentas and extra-embryonic tissues.

First, we downloaded the placental RNA-seq data from 15 mammals: humans, bonobos, baboons, mice, golden hamsters, rabbits, pigs, cattle, sheep, horses, dogs, bats, elephants, armadillos, and opossums (Armstrong et al. 2017; Mika et al. 2022). These datasets were generated using conventional non-directional RNA-seq (also known as non-stranded RNA-seq), which does not retain information regarding the genomic element orientation of the sequenced transcripts. In addition, library preparation methods and the amount of data varied significantly among the datasets, ranging from millions to tens of millions (**Supplemental Table S2**). Although transcriptional analysis of these datasets identified transcripts such as *GPR1-AS* in human and baboon placentas, the structures of these predicted transcripts were so fragmented that their full-length forms could not be robustly determined (**Supplemental Fig. S1**). Furthermore, we did not observe such a transcript at the *Gpr1-Zdbf2* locus in the mouse placental data, where *Liz* is known to be expressed. This suggests that conventional RNA-seq datasets may

not be sufficient to accurately detect *GPR1-AS* in different species.

Next, we downloaded directional RNA-seq data (also known as strand-specific or stranded RNA-seq, which can identify the direction of transcription) from the human placenta, rhesus macaque trophoblast stem cells, and mouse embryonic placenta (Nesculea et al., 2014; Rosenkrantz et al. 2021). These RNA-seq datasets contain tens of millions of reads. In each of these deeply sequenced datasets, transcripts such as *GPR1-AS* (and *Liz* in mice) were detected, which were transcribed from the *GPR1* intron in the opposite direction to *GPR1* towards the *ZDBF2* gene (**Fig. 1**). Thus, *GPR1-AS* transcripts can be identified in placental tissues using deep directional RNA-seq data.

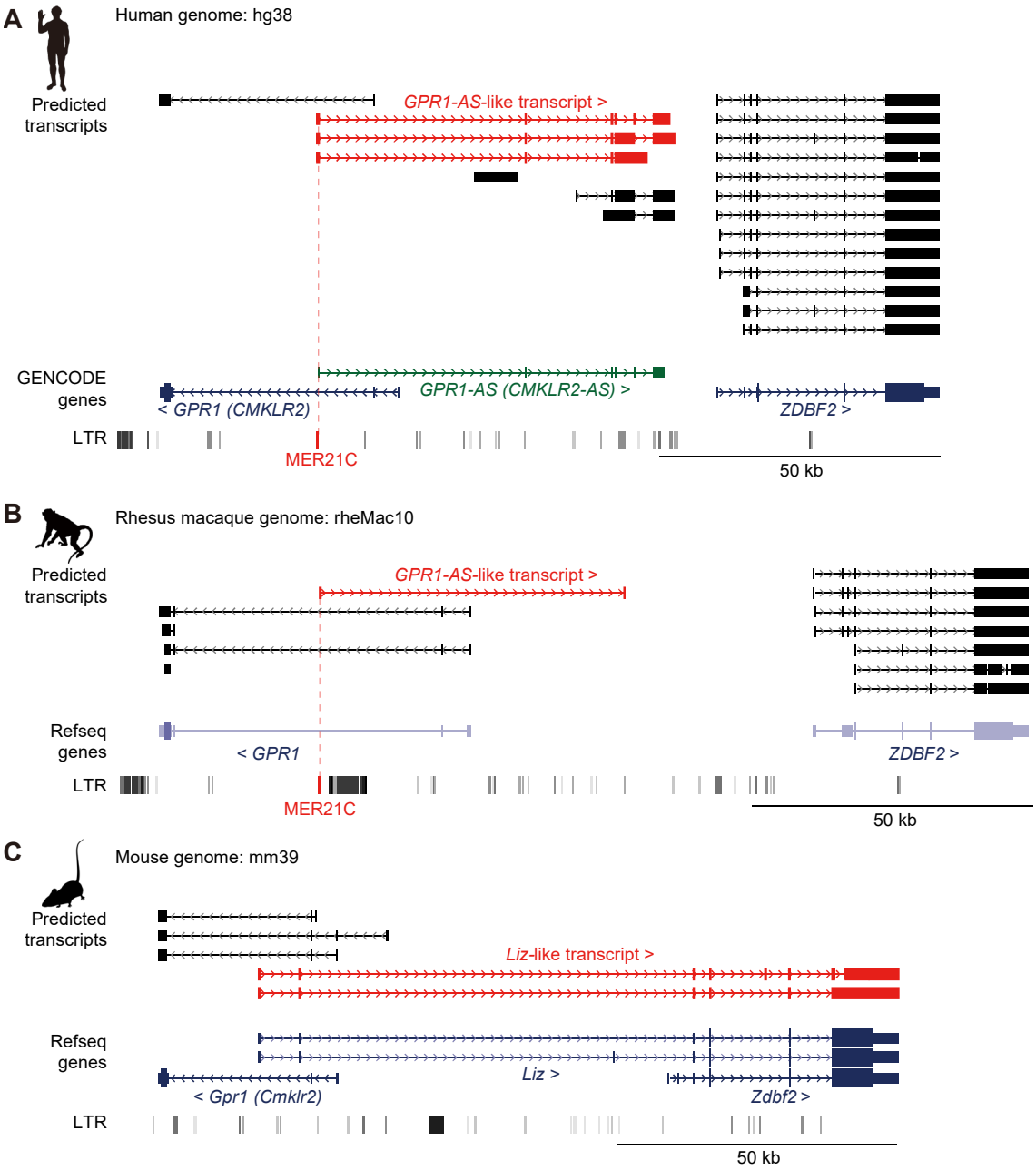


Fig. 1 Identification of *GPR1-AS* orthologs from public placental transcriptomes

UCSC Genome browser screenshots of the *GPR1-ZDBF2* locus in humans (**A**), rhesus macaques (**B**) and mice (**C**). Predicted transcripts were generated using public directional placental RNA-seq datasets (accession numbers: SRR12363247 for humans, SRR1236168 for rhesus macaques, and SRR943345 for mice) using the Hisat2-StringTie2 programs. Genes annotated from GENCODE or RefSeq databases and LTR

retrotransposon positions are also displayed. Among the gene lists, only the human reference genome includes an annotation for *GPR1-AS* (highlighted in green). *GPR1-AS*-like transcripts and MER21C retrotransposons are highlighted in red.

To identify *GPR1-AS*-like transcripts in other mammalian species, we generated deep directional RNA-seq datasets from the whole placenta of chimpanzees and the trophectoderm (TE) of rabbit, bovine, pig, and opossum post-gastrulation embryos, which yielded a total of 50–100 million reads (**Supplemental Table S2**). Among these datasets, *GPR1-AS* orthologs were identified in chimpanzees and rabbits but were not observed in cattle, pigs, or opossums (**Fig. 2**). Additionally, we performed directional RNA-seq of embryonic proper tissues (embryonic disc: ED) from individual animal embryos but did not identify any *GPR1-AS*-like transcripts in any of these samples (**Supplemental Fig. S2**). Thus, rabbit *GPR1-AS* is expressed only in placental lineages. Identification of *GPR1-AS* in rabbits and several primate species indicates that *GPR1-AS* transcripts appear in the common ancestor of these mammalian lineages, which belong to a group called Euarchontoglires.

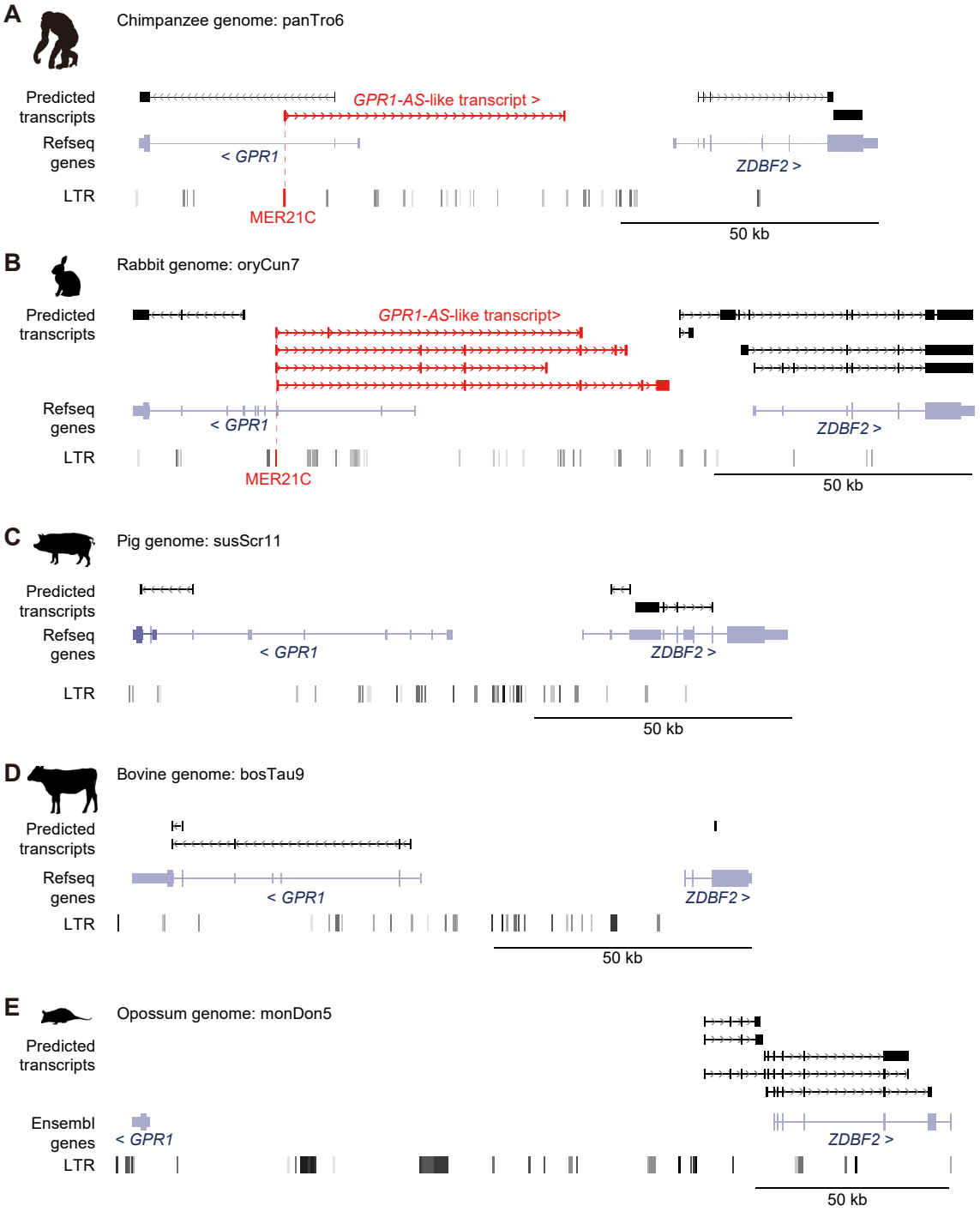


Fig. 2 Identification of *GPR1-AS* orthologs from original placental and extra-embryonic transcriptomes

Predicted transcripts were generated from placental and extra-embryonic directional RNA-seq datasets of chimpanzee (**A**), rabbit (**B**), pig (**C**), cow (**D**), and opossum (**E**) with

the Hisat2-StringTie2 programs. Genes annotated from RefSeq or Ensembl databases and their LTR positions are also shown. *GPR1-AS*-like transcripts and MER21C retrotransposons are highlighted in red.

Origin of *ZDBF2* imprinting and *GPR1-AS* transcript

Since *GPR1-AS*/*Liz* is essential for *ZDBF2* imprinting, it is assumed that *ZDBF2* imprinting system via *GPR1-AS* transcription is not established in other eutherians and marsupials in which *GPR1-AS* is absent (Greenberg et al. 2017). To test this hypothesis, we performed allele-specific expression analysis of the *ZDBF2* gene in tammar wallabies and cattle, which are mammals outside the Euarchontoglires group. We identified available SNPs at the 3'-UTR of *ZDBF2* in each animal, and Sanger-sequencing-based polymorphism analysis showed the expression of both alleles of *ZDBF2* in all analyzed bovine and wallaby tissues (**Fig. 3**)(**Supplemental Table S1**). These data support the notion that the imprinting of *ZDBF2* is established only among Euarchontoglires, similar to the appearance of *GPR1-AS*.

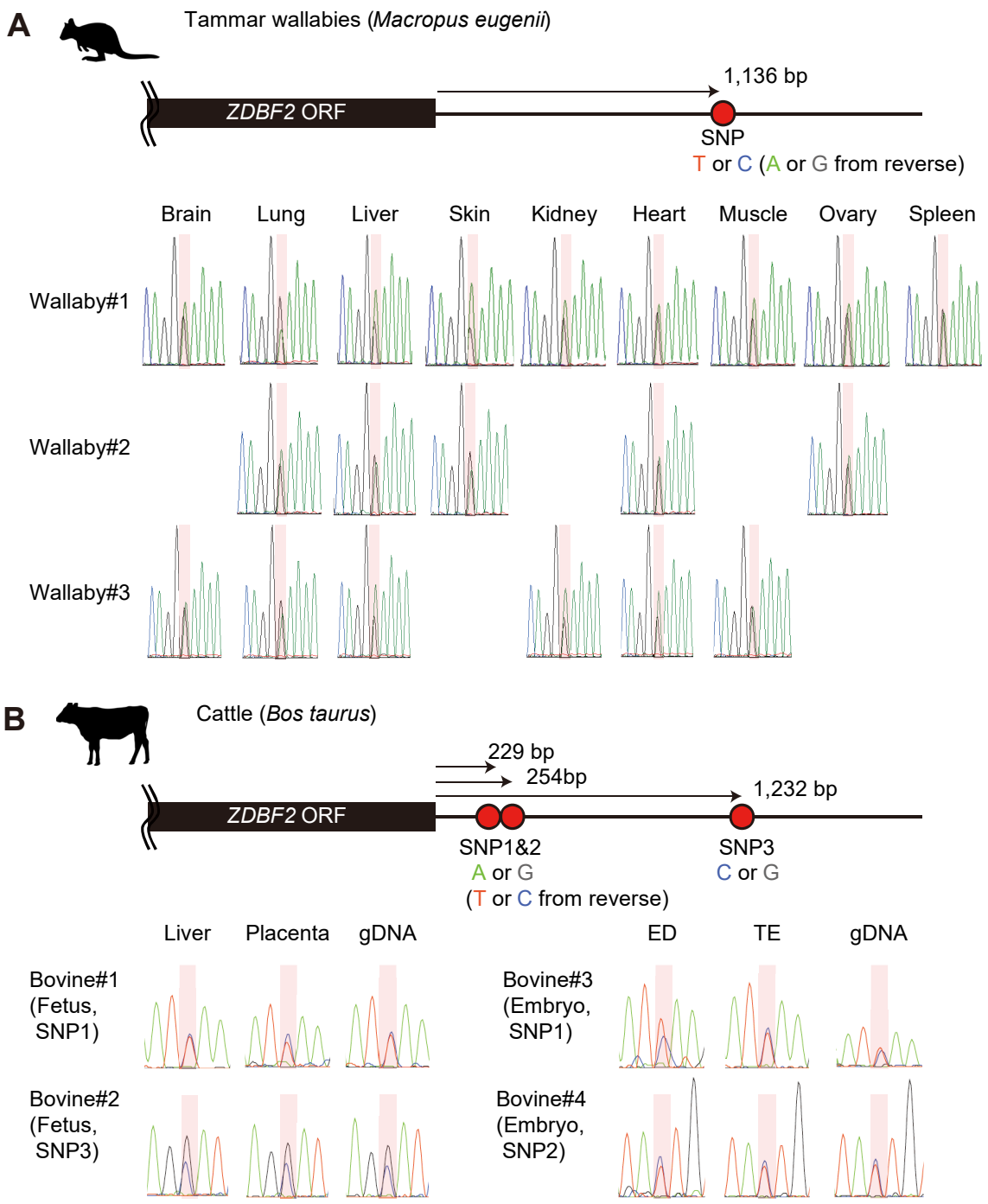


Fig. 3 Allele-specific RT-PCR sequencing of *ZDBF2* in tammar wallabies and cattle

One and three heterozygous genotypes were used to distinguish between parental alleles in adult tissues form tammar wallabies (**A**) and fetal/embryonic tissues from cattle (**B**), respectively. Primers were designed to amplify the 3'-UTR regions of *ZDBF2* orthologs and detect SNPs. Each SNP position are highlighted in red. Reverse primers

were also used for Sanger sequencing.

Notably, the first exon of all *GPR1-AS* orthologs, except for mouse *Liz,* overlapped with MER21C, a member of the LTR retrotransposon subfamily (**Figs. 1,2**)(**Supplemental Fig. S1**). The MER21C retrotransposon has been present in eutherian genomes throughout evolution; however, among the mammals tested, only primate and rabbit genomes have MER21C in the *GPR1* intron, where the first exon of *GPR1-AS* is located. Therefore, we compared the LTR sequence location information of the synteny region of the *GPR1* and *GPR1-AS* loci to examine traces of MER21C insertion in several mammalian genomes, excluding marsupials. Among eutherians, we found the insertion of MER21C (or MER21B retrotransposons that showed high similarity to MER21C) in the introns of *GPR1* in five groups of Euarchontoglires: primates, colugos, treeshrews, rabbits, and rodents (excluding mice, rats, and hamsters among rodents) (**Fig. 4**).

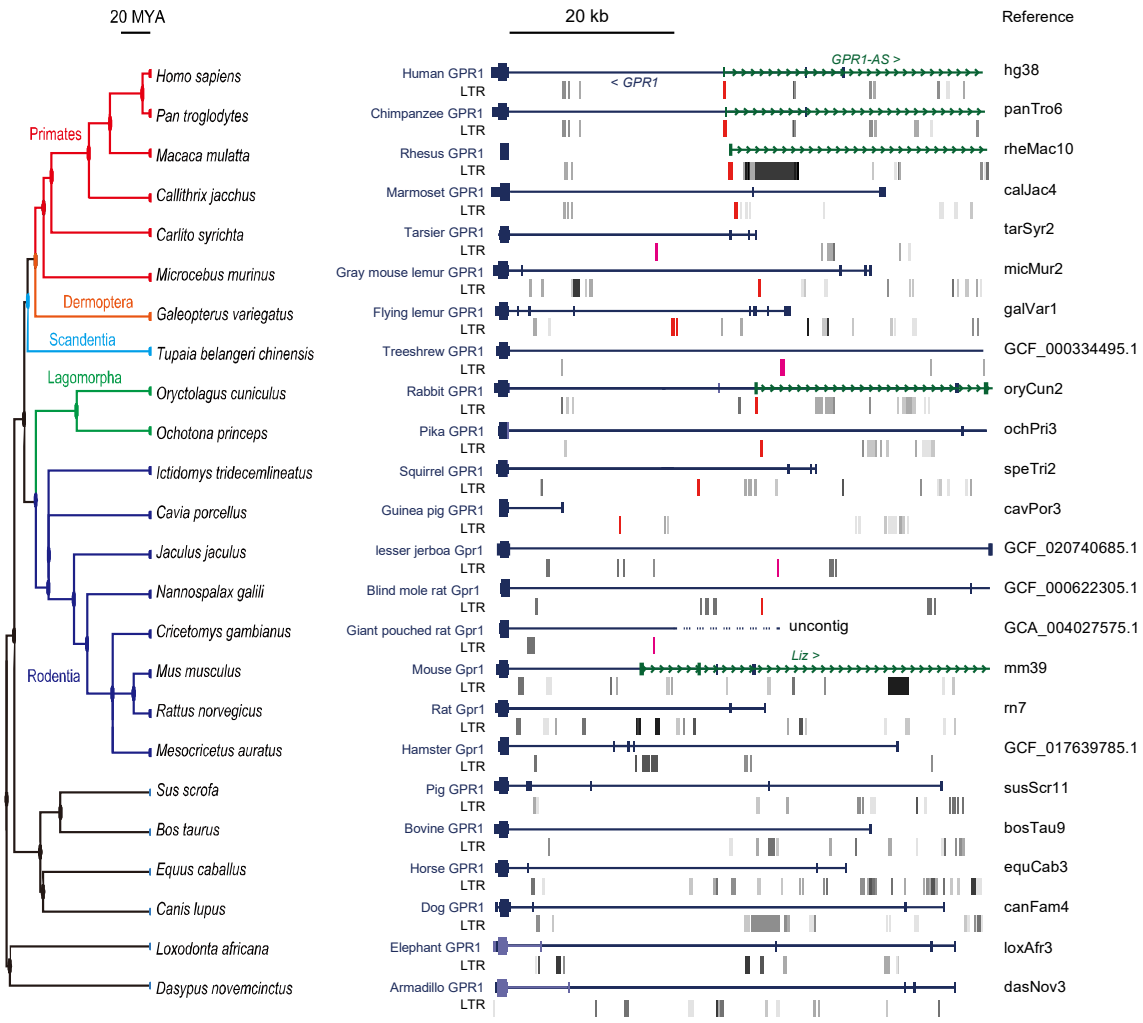


Fig. 4 Multi-species comparison of LTR retrotransposon locations at *GPR1* locus

A total of 24 mammalian genomes were compared, including six primates (human, chimpanzee, rhesus macaque, marmoset, tarsier, and gray mouse lemur), one colugo (flying lemur), one treeshrew (Chinese treeshrew), two rabbits (rabbit and pika), eight rodents (squirrel, guinea pig, lesser jerboa, blind mole rat, giant pouched rat, mouse, rat, and golden hamster), and six eutherians (pig, cow, horse, dog, elephant, and armadillo). The locations of MER21C and MER21B retrotransposons, which exhibited 88% similarity between their consensus sequences through pairwise alignment, are highlighted in red and purple, respectively.

We also compared MER21C sequences that overlapped with the first exon of *GPR1-AS* in each primate (345 bp, 339 bp, and 506 bp from humans, chimpanzees, and rhesus macaques, respectively) and rabbits (218 bp), the first exon of *Liz* in mice (317 bp), and the common sequence of MER21C (938 bp). Pairwise alignment revealed high similarities between primate sequences (71–72%) and the consensus MER21C sequence compared to rabbit (60%) and mouse sequences (46%)(**Supplemental Fig. S3**). Multiple sequence alignments and phylogenetic tree reconstruction reveal that the rabbit and mouse sequences deviate to a greater extent from the consensus MER21C sequence than the cluster formed by the primate sequences, indicating that a greater number of mutations have accumulated in the non-primate lineage of Euarchontoglires since their divergence from Primates (**Fig. 5A**). This observation aligns with the shorter generation time of lagomorphs and rodents compared to primates (Wu and Li 1985; Huttley et al. 2007).

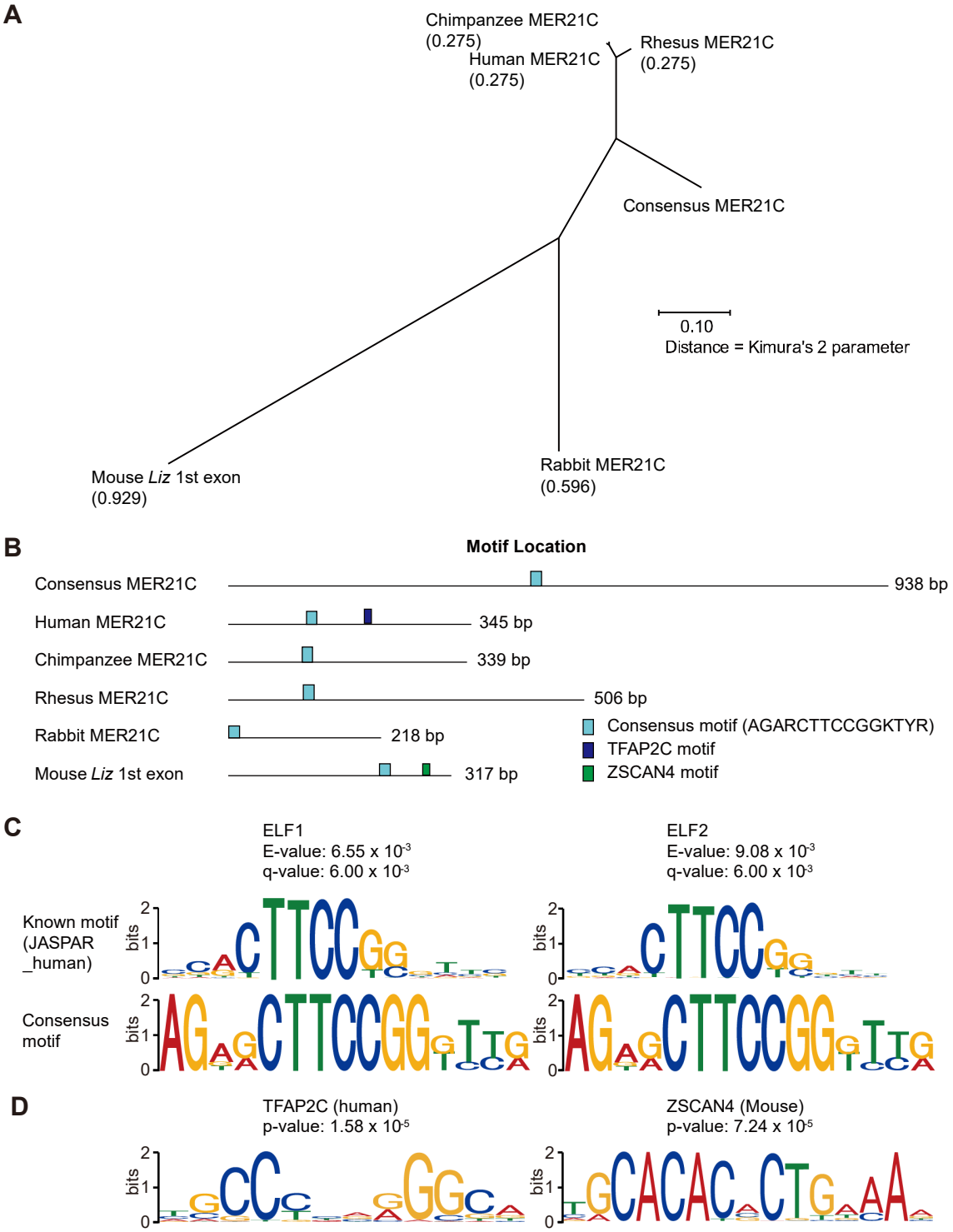


Fig. 5 Comparison of MER21C-derived sequences overlapping the first exon of *GPR1-AS* orthologs

(**A**) Phylogenetic tree of MER21C-derived sequences estimated by multiple sequence

alignment (MSA) using multiple sequence comparison by log-expectation (MUSCLE) program. (**B**) Positions of common and unique cis-acting elements at each sequence. (**C**) Motif structures of the common region that contains E74-like factor 1 and 2 (ELF1 and ELF2) binding motifs. (**D**) Motif structures of transcription factor AP-2 gamma (TFAP2C) and Zinc finger and SCAN domain containing 4 (ZSCAN4).

To identify common cis-motifs, we compared these sequences with known transcription factor-binding motifs using the TOMTOM program in the MEME suite. Through this analysis, we found a common region that contained an ETS family transcription factor (ELF1 and ELF2) binding site, which had significant matches with E-values < 0.01 and q-values < 0.01 (**Fig. 5B,5C**). Additionally, we identified the TFAP2 and ZSCAN4 binding motifs, defined as p-values = 1.58×10^{-5} and 7.24×10^{-7}, from the JASPAR database at the first exon of human *GPR1-AS* and mouse *Liz*, respectively (**Fig. 5B,5D**). Since TFAP2C and ZSCAN4C are activated in the placental lineage and preimplantation embryos, respectively (**Supplemental Fig. S4**), these transcription factors may play a role in the transcriptional activation of *GPR1-AS* or *Liz* during embryogenesis by promoting LTR-derived promoter activity (Zhang et al. 2019; Papuchova and Latos 2022).

***GPR1-AS* transcription and LTR reactivation in human**

GPR1-AS is reported to be expressed only briefly during embryogenesis, both before and after implantation, with the exception of the placental lineage. For instance, human *GPR1-AS* has been identified in ES cells, and mouse *Liz* expression has been observed in blastocyst-to-gastrulation embryos but not in gametes (Kobayashi et al. 2012; Kobayashi et al. 2013; Duffie et al. 2014). Therefore, it is likely that the expression of

GPR1-AS/*Liz* is activated during preimplantation. To investigate the timing of *GPR1-AS* activation during embryonic development, we obtained public human RNA-seq data from the oocyte to blastocyst stages and performed expression analysis and transcript prediction (Kai et al. 2022; Zou et al. 2022b). Among the datasets from oocytes; zygotes; 2-, 4-, and 8-cell embryos; inner cell mass (ICM); and TE from blastocysts, *GPR1-AS* was only detected at the 8-cell stage, ICM, and TE (**Fig. 6**). Hence, the expression of *GPR1-AS* initiates at the 8-cell stage in humans. However, at the *GPR1-ZDBF2* locus, both *GPR1* and *ZDBF2* were expressed in oocytes, but only *GPR1* was downregulated after fertilization and could not be detected after the 8-cell stage (**Fig. 6**)(**Supplemental Fig. S4**). We also examined the LTR expression at each stage and found that MER21C expression increased from the from 4-cell to 8-cell stage and peaked at the TE, which coincided with the appearance of *GPR1-AS* (**Supplemental Fig. S4, S5**). Thus, *GPR1-AS* was shown to begin expression from the preimplantation embryo stage, concurrently with the reactivation of the MER21C retrotransposon.

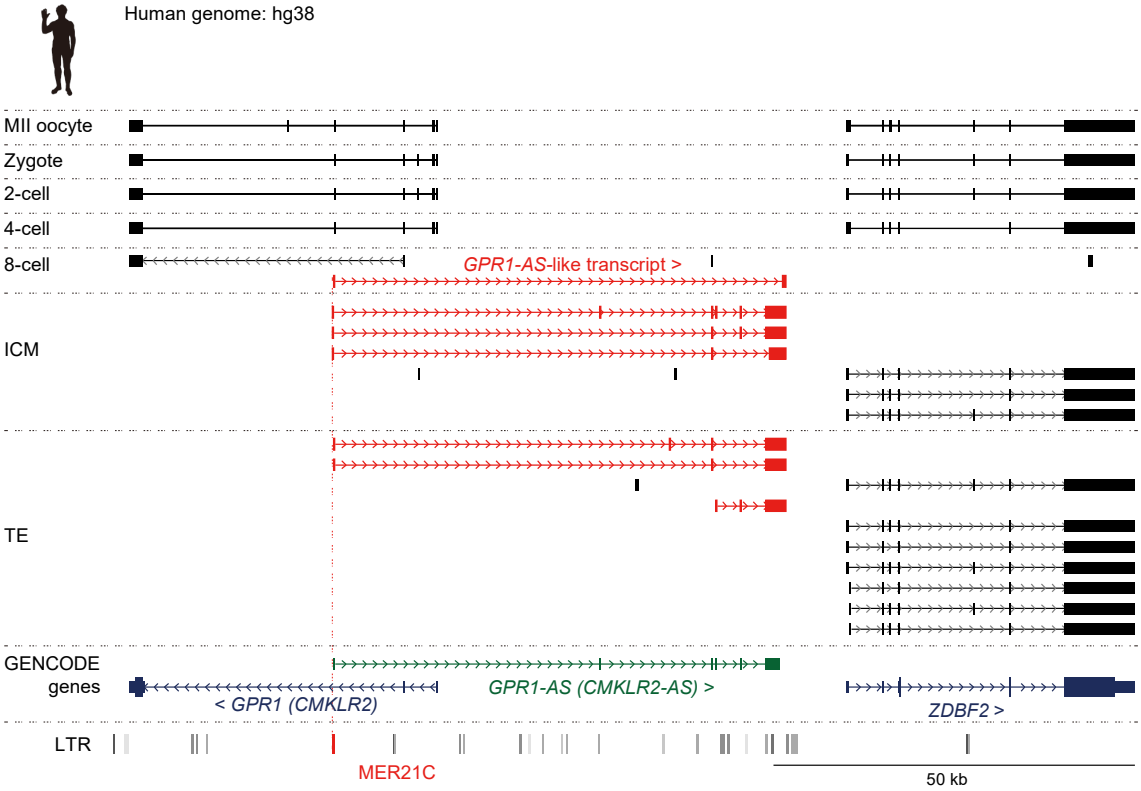


Fig. 6 Initiation of *GPR1-AS* transcription before implantation

Genome browser screenshots of the *GPR1-ZDBF2* locus in humans during preimplantation stages: MII oocyte, zygote, 2-, 4-, and 8-cell, inner cell mass (ICM) and trophectoderm (TE) from blastocyst. Predicted transcripts were generated using publicly available placental full-length RNA-seq datasets.

Next, we examined the epigenomic modifications of LTR-derived sequences located in the first exon of human *GPR1-AS* and mouse *Liz* using the ChIP-Atlas database (Zou et al. 2022a) (**Supplemental Fig. S5**). Both sequences that constitute the oocyte-derived germline DMR exhibit an enrichment of H3K4me3 in sperm or male germ cells, which is consistent with DNA hypomethylation in sperm. These sequences also show an enrichment of H3K27ac in pluripotent stem cells (such as ES and iPS cells). Although the first exon of mouse *Liz* has not been computationally determined as an LTR

(according to the RepBase database), the mouse sequence is also enriched for H3K9me3, which regulates retrotransposons in multiple cell types and tissues. Additionally, human and mouse germline DMRs coincided with the TFAP2C peak (trophoblast stem cells) and ZSCAN4C peak (ES cells), respectively, and aligned with the binding motif predictions found in the JASPAR database. These data indicate that the first exons of human *GPR1-AS* and mouse *Liz* exhibit similar epigenomic modification variations from gamete to embryonic development, and their promoter activities are triggered after fertilization, although the underlying mechanisms may differ (**Supplemental Fig. S7**).

DISCUSSION

The *ZDBF2* gene, the last canonical imprinting gene identified in mice and humans, was recently shown to regulate neonatal feeding behavior (Glaser et al. 2022). This discovery has garnered attention owing to its implications for behavioral phenotypes, aligning with the conflicting hypothesis of genomic imprinting mechanisms. The conflict hypothesis proposes that the evolutionary significance of genomic imprinting mechanisms lies in the functional conflicts between paternally and maternally expressed genes, resulting from the conflicting genomic survival strategies of the father, child, and mother (Moore and Haig 1991). *ZDBF2* does not directly influence cell differentiation or ontogeny, such as embryogenesis, but rather affects individual growth through an indirect pathway involving feeding behavior. In humans, “imprinting diseases”, such as Prader-Willi syndrome and Angelman syndrome, cause psychiatric disorders and growth retardation (Nicholls 2000). Considering that many imprinted genes show specialized expression patterns in the brain, *ZDBF2* is particularly interesting as a gene that can induce behavioral effects.

However, the evolutionary origin of *ZDBF2* imprinting had remained unclear. In this study, we report that *ZDBF2* does not exhibit imprinted expression in cattle and tammar wallabies, mammalian species that do not belong to the Euarchontoglires. In contrast, we identified *GPR1-AS* orthologs in all the Euarchontoglires species examined, including humans, chimpanzees, baboons, rhesus macaques, and rabbits. We propose that the first exon of *GPR1-AS* is derived from the MER21C retrotransposon and that the insertion of this LTR at the *GPR1* intragenic region in the common ancestor of Euarchontoglires was crucial for establishing the imprinting status of the *ZDBF2* gene (**Supplemental Fig. S7**).

Two paternally expressed imprinted genes, *PEG10*/*SIRH1* and *PEG11*/*RTL1*/*SIRH2* have been identified in mammals. They encode GAG-POL proteins of *sushi-ichi* LTR retrotransposons and are essential for mammalian placenta formation and maintenance (Kaneko-Ishino and Ishino 2012). The presence of these genes provides evidence that LTR insertions have played a significant role in genomic imprinting and placentation throughout evolution. In contrast, while the *ZDBF2* gene is present in all vertebrate genomes, it is only imprinted in mammals belonging to the Euarchontoglires. This suggests that a nonimprinted gene acquired a new imprint through LTR insertion. Indeed, we previously demonstrated that transcriptional activity of specific LTRs in oocytes results in species-specific imprinting of numerous genes in rodents and primates (Bogutz et al. 2019). These LTRs act as alternative promoters and produce unique transcripts in the oocyte, which we referred to as LITs (LTR-initiated transcripts). Such transcription leads to a high level of methylation in the transcribed region, similar to gene body methylation, and is responsible for extensive interspecies differences in DNA methylation

(Brind'Amour et al. 2018). The major difference between these LTRs and the LTR retrotransposon focused on in this study is whether (re)activation occurs before or after fertilization. Some LITs become active in oocytes and establish oocyte-derived gDMRs of species-specific imprinted genes, such as *Impact* and *Slc38a4* genes in mice. Deletion of these LTRs causes loss of imprinting in mouse (rodent)-specific imprinted genes (Bogutz et al. 2019). *GPR1-AS* is a (likely noncoding) RNA whose transcription is initiated from an LTR-derived sequence, similar to the LITs found in oocytes; however, transcription starts after fertilization. In mice, *Liz*, a long isoform of *Zdbf2* and putative *Gpr1-as* ortholog, is essential for establishing a somatic secondary DMR upstream of *Zdbf2* (Greenberg et al. 2017). Although the first exon of *Liz* does not overlap with an annotated LTR, MER21C (or MER21B, a closely related LTR family) insertions are found in the homologous region in most rodents, with the exception of mice, rats, and hamsters. Our analyses suggest that the MER21C in these rodents has accumulated a number of mutations and in turn is no longer recognized as a MER21C element using optimal pairwise alignment algorithms such as RepeatMasker. Based on these observations, we propose that *Liz* and *GPR1-AS* are bona fide LITs with a MER21C-derived sequence serving as an alternative promoter.

Transposable elements are recognized by DNA-binding factors and their co-repressors and are suppressed through epigenetic and post-transcriptional regulation in both germline and somatic tissues to protect host genome stability. For example, LTR-retrotransposons are bound by KRAB-ZFPs, which recruit the KAP1/SETDB1 co-repressor complex, promoting deposition of the repressive histone modification H3K9me3 (Matsui et al. 2010; Karimi et al. 2011). However, numerous retrotransposons are activated and robustly expressed during maternal to zygotic transition, where

dynamic epigenetic reprogramming occurs. Indeed, a subset of transposable elements have been co-opted by the host during this window of development, including in support of embryonic development (Sakashita et al. 2023). In humans, retrotransposons are activated at the 8-cell stage and gradually downregulated in later developmental stages, coinciding with embryonic genome activation. In addition to the full-length endogenous retrovirus (ERV) sequences with partial coding potential, approximately 85% of human ERVs (HERVs) exist as solitary LTRs, known as “solo LTRs,” which provide a rich source of cis-regulatory elements for gene expression during human embryonic development. Retrotransposons located near host genes can act as alternative promoters or regulatory modules, such as enhancers, to activate embryonic genes (Grow et al. 2015; Hashimoto et al. 2021). The discovery of the *GPR1-AS* as a transcript that initiates in an ancestral LTR reveals a new aspect of the functional role of these ectopic regulatory elements, with activation in this case occurring after fertilization.

In conclusion, this study reveals that the origin of imprinting of *ZDBF2* can be traced back to the emergence of a common ancestor of Euarchontoglires, with insertion of an LTR promoter for *GPR1-AS* playing a critical role in establishing imprinting of this locus, including in mice and humans. Although the house mouse is used widely in the laboratory, and the *Mus musculus* genome has been extensively characterized, the retroviral origin of this promoter would not have been identified if this analysis had been performed solely on mice, as the annotated exon 1 of *Liz* (alternatively named *Gpr1-as, Platr12*, or *Zdbf2linc*) is highly degenerate and in turn not recognized as an LTR in this species. Thus, comparative analysis of multi-omics data, including genomes, transcriptomes, and epigenomes, across species was critical for the identification of the central role of a

MER21C LTR in imprinting of *ZDBF2*. This study adds to a growing list of LTR elements that have been domesticated in mammals, with their transcriptional activity serving as a mechanism for the genesis of imprinting, including during gametogenesis and in this case following fertilization.

METHODS

Ethical approval for animal work

Animal experiments (chimpanzee, rabbit, pig, cow, and opossum) were conducted in accordance with the guidelines of the Science Council of Japan and approved by the Institutional Animal Care and Use Committee of the University of Tokyo; the National Institute for Physiological Sciences, RIKEN Kobe Branch, Meiji University; Shinshu University; Hokkaido University and the Primate Research Institute (PRI) of Kyoto University. Experimental procedures of tammar wallaby conformed to Australian National Health and Medical Research Council (2013) guidelines and were approved by the Animal Experimentation Ethics Committees of the University of Melbourne.

Animals for transcriptome analysis

Placental samples from chimpanzees (*Pan troglodytes verus*) were provided by Kumamoto Zoo and Kumamoto Sanctuary via the Great Ape Information Network. The samples were stored at −80 °C before use. Total RNA was isolated from the placenta using an AllPrep DNA/RNA Mini Kit (Qiagen, Netherland). Bovine (*Bos taurus*, Holstein cow) embryos were prepared by in vitro oocyte maturation, fertilization (IVF), and subsequent in vitro embryo culture (Akizawa et al. 2018). Briefly, presumptive IVF zygotes were denuded by pipetting after 12 h of incubation and cultured up to the

blastocyst stage in mSOFai medium at 38.5 °C in a humidified atmosphere of 5% CO_2 and 5% O_2 in air for 8 days (D8). D8 blastocysts were further cultured on agarose gel for 4 dayss as described previously (Akizawa et al. 2018; Saito et al. 2022). The embryo proper (embryonic disc: ED) and TE portions of IVF D12 embryos were mechanically divided using a microfeather blade (Feather Safety Razor, Japan) under a stereomicroscope. Total RNA from ED and TE samples was isolated using the ReliaPrepTM RNA Cell Miniprep System (Promega, MA, USA) according to the manufacturer's instructions. Rabbit embryos were prepared on embryonic day 6.75 (E6.75) by mating female Dutch and male JW rabbits (Kitayama Labes, Japan). Pig embryos (*Sus scrofa domesticus*: Landrace pig) at E15 and opossum embryos at E11 were prepared (Kiyonari et al. 2021). Individual embryo proper and TE were mechanically divided using a 27-30G needle and tweezer under a stereomicroscope with 10% FCS/M199 -medium in rabbits, HBSS medium in pigs, and 3% FCS/PBS in opossums. Total RNA was isolated from individual tissues using the RNeasy Mini and Micro Kit (Qiagen). The quality of the total RNA samples was assessed using an Agilent 2100 Bioanalyzer system (Thermo Fisher Scientific, MA, USA) and high-quality RNA samples (RIN≧7) were selected for RNA-seq library construction.

Strand-specific RNA library preparation and sequencing

Embryonic total RNA (10 ng) from opossums and pigs, 5 ng of placental total RNA from chimpanzees, and 1 ng of embryonic total RNA from rabbits and cattle were reverse transcribed using the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio, Japan) according to the manufacturer's protocols, where Read 2 corresponds to the sense strand due to template-switching reactions. RNA-seq libraries

were quantified by qPCR using the KAPA Library Quantification Kit (Nippon Genetics, Japan). All libraries were pooled and subjected to paired-end 75 bp sequencing (paired-end 76 nt reads with the first 1 nt of Read 1 and the last 1 nt of Read 2 trimmed) using the NextSeq500 system (Illumina, CA, USA). For each library, the reads were trimmed using Trimmomatic to remove two nucleotides from the 5' end of the Read 2.

RNA-seq data set download

Strand-specific RNA-seq datasets from the human bulk placenta, rhesus macaque trophoblast stem cells, and mouse bulk placenta at E16.5, were downloaded from the accession numbers SRR12363247 and SRR12363248 for humans, SRR1236168 and SRR1236169 for rhesus macaques, and SRR943345 for mice (Necsulea et al. 2014; Rosenkrantz et al. 2021). Non-directional RNA-seq data from placentas or extra-embryonic tissues of 15 mammalian species, including humans, bonobos, baboons, mice, golden hamsters, rabbits, pigs, cattle, sheep, horses, dogs, bast, elephants, armadillos, and opossums, were downloaded (Armstrong et al. 2017; Mika et al. 2022). Full-length RNA-seq (based on the smart-seq method) data from human oocytes, zygotes, 2-, 4-, and 8-cell, ICM, and TE were downloaded from previous studies (Kai et al. 2022; Zou et al. 2022b). The corresponding accession numbers are shown in **Supplemental Table S2**.

RNA-seq data processing

All FASTQ files were quality-filtered using Trimmomatic and mapped to individual genome references (**Supplemental Table S2**) using Hisat2 with default parameters and Stringtie2 with default parameters and “-g 500” option. Transcripts per kilobase million

(TPM) was calculated using StringTie2 with “-e” option and the GENCODE GTF file, selecting lines containing the "Ensembl_canonical" tag. Read counts of human transposable element were computed at the subfamily level using the TEcount. The read counts were normalized by the total number of mapped reads to retrotransposons in each sample as Reads per kilobase million (RPKM).

Data visualization

All GTF files (StringTie2 assembled transcripts) were uploaded to the UCSC genome browser, and the predicted transcripts, annotated genes (GENCODE, Ensembl, or RefSeq), and LTR locations were visualized. The expression profiles of human *GPR1*, *GPR1-AS*, and *ZDBF2* in the somatic tissues (Fagerberg et al. 2014; Duff et al. 2015), and these imprinted genes, the associated transcription factors (*TFAP2C*, *ZSCAN4*, *ELF1*, and *ELF2*), and LTR subfamilies in early embryo (Kai et al. 2022; Zou et al. 2022b) were downloaded or calculated, and visualized as heatmaps using the Morpheus software. The estimated evolutionary distance between the selected mammals was visualized as a physiological tree with a branch length of millions of years (MYA) using Timetree. The pairwise alignment, multiple sequence alignment, and physiological tree of MER21C sequences overlapping *GPR1-AS* were conducted with Genetyx software (Genetyx, Japan) using the MSA and MUSCLE programs. The common cis-motif was identified and compared with known transcription factor binding motifs from JASPAR CORE database using the XSTREAM and TOMTOM programs in the MEME suite. JASPAR motifs, logos, and *p*-values were obtained from the JASPAR hub tracks of the UCSC genome browser and JASPAR database. LTR positions of human (hg19) and mouse (mm10) were downloaded from UCSC genome browser and re-uploaded to

Integrative Genomics Viewer with epigenetic patterns from ChIP-Atlas (Zou et al. 2022a) and oocyte/sperm DNA methylomes (Brind'Amour et al. 2018).

Allele-specific expression analysis

Tammar wallabies (*Macropus eugenii*) of Kangaroo Island origin were maintained in our breeding colony in grassy, outdoor enclosures. Lucerne cubes, grass and water were provided ad libitum and supplemented with fresh vegetables. Pouch young were dissected to obtain a range of tissues. DNA/RNA was extracted from multiple tissues of the Tammar wallaby using TRIzol Reagent (Thermo Fisher Scientific), and cDNA synthesis was performed using Transcriptor First Strand cDNA Synthesis Kit (Roche, Switzerland). PCR was performed using TaKaRa Ex Taq Hot Start Version (TaKaRa) with primers designed at 3'UTR of wallaby *ZDBF2* according to the manufacturer's instructions. Fetal tissues of pregnant healthy Holstein cattle were obtained from a local slaughterhouse. DNA/RNA was extracted from bovine fetus (liver), placentas, and *in vitro* cultured embryos using ISOGEN (Wako, Japan), and cDNA synthesis and PCR were performed using SuperScript IV One-Step RT-PCR System (Thermo Fisher Scientific) with primers designed at 3'UTR of bovine *ZDBF2* according to the manufacturer's instructions. Target regions were also amplified using genomic DNA and Takara EX Taq Hot Start Version (Takara Bio). Sanger sequencing was performed using a conventional Sanger sequencing service (Fasmac, Japan) and ABI 3130xl Genetic Analyzer (Thermo Fisher Scientific). Sanger sequencing results were visualized using 4Peaks. Primers used in this analysis are listed in **Supplemental Table S3**.

Softwares used

Trimmomatic（http://www.usadellab.org/cms/?page=trimmomatic)

Hisat2（http://daehwankimlab.github.io/hisat2/)

StringTie2（https://ccb.jhu.edu/software/stringtie/)

TEcount (https://github.com/bodegalab/tecount/)

Integrative Genomics Viewer (IGV: https://software.broadinstitute.org/software/igv/)

4Peaks (https://nucleobytes.com/4peaks/)

Web-tools used

UCSC Genome Browser（https://genome.ucsc.edu/)

Morpheus（https://software.broadinstitute.org/morpheus/)

Timetree（http://timetree.org/)

MEME Suite (https://meme-suite.org/meme/)

JASPAR (https://jaspar.genereg.net/)

ChIP-Atlas（https://chip-atlas.org/)

DATA ACCESS

All raw sequencing data generated in this study were submitted as FASTQ files to the NCBI SRA（https://www.ncbi.nlm.nih.gov/sra). The accession numbers are shown in **Supplemental Table S2**.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

H. K. and K. K. were supported by MEXT KAKENHI, Grant Numbers JP21H02382 and JP20H00471, respectively. H.K. and H.I. were supported by The Cooperative Research Programs of the NODAI Genome Research Center at Tokyo University of Agriculture, and the Primate Research Institute and Wildlife Research Center at Kyoto University. H.K. received additional support from the Mitsubishi Foundation and the Takeda Science Foundation. Additionally, we acknowledge Mr. Yasuyuki Osada (Kitayama Laboratories, Japan) for his technical assistance in handling the rabbit embryos.

Author Contribution

Designed the study and wrote the manuscript: H. Kobayashi

Data analysis: H. Kobayashi., T.I

NGS data processing: S.K., K.T., T.T

Wallaby sample preparation and data analysis: S. Suzuki, M.H., M.B.R.

Bovine sample preparation: S. Saito, M. K.

Rabbit sample preparation: T.K.

Pig sample preparation: H.N., H.M., K. Nakano, A.U.

Opossum sample preparation: H. Kiyonari, M. K.

Chimpanzee sample preparation: H.I, K. Nakabayashi.

Study conception: H. Kobayashi, M.C.L.

Funding: H. Kobayashi, K.K.

REFERENCES

Akizawa H, Kobayashi K, Bai H, Takahashi M, Kagawa S, Nagatomo H, Kawahara M. 2018. Reciprocal regulation of TEAD4 and CCN2 for the trophectoderm development of the bovine blastocyst. *Reproduction* **155**: 563-571.

Armstrong DL, McGowen MR, Weckle A, Pantham P, Caravas J, Agnew D, Benirschke K, Savage-Rumbaugh S, Nevo E, Kim CJ et al. 2017. The core transcriptome of mammalian placentas and the divergence of expression with placental shape. *Placenta* **57**: 71-78.

Bogutz AB, Brind'Amour J, Kobayashi H, Jensen KN, Nakabayashi K, Imai H, Lorincz MC, Lefebvre L. 2019. Evolution of imprinting via lineage-specific insertion of retroviral promoters. *Nat Commun* **10**: 5674.

Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, Kamio A, Bogutz A, Koike T, Karimi MM, Lefebvre L et al. 2018. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat Commun* **9**: 3331.

Duff MO, Olson S, Wei X, Garrett SC, Osman A, Bolisetty M, Plocik A, Celniker SE, Graveley BR. 2015. Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature* **521**: 376-379.

Duffie R, Ajjan S, Greenberg MV, Zamudio N, Escamilla del Arenal M, Iranzo J, Okamoto I, Barbaux S, Fauque P, Bourc'his D. 2014. The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes Dev* **28**: 463-478.

Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K et al. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**: 397-406.

Glaser J, Iranzo J, Borensztein M, Marinucci M, Gualtieri A, Jouhanneau C, Teissandier A, Gaston-Massuet C, Bourc'his D. 2022. The imprinted Zdbf2 gene finely tunes control of feeding and growth in neonates. *Elife* **11**.

Greenberg MV, Glaser J, Borsos M, Marjou FE, Walter M, Teissandier A, Bourc'his D. 2017. Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat Genet* **49**: 110-118.

Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY et al. 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**: 221-225.

Hashimoto K, Jouhilahti EM, Tohonen V, Carninci P, Kere J, Katayama S. 2021.

Embryonic LTR retrotransposons supply promoter modules to somatic tissues. *Genome Res* **31**: 1983-1993.

Hiura H, Sugawara A, Ogawa H, John RM, Miyauchi N, Miyanari Y, Horiike T, Li Y, Yaegashi N, Sasaki H et al. 2010. A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic Acids Res* **38**: 4929-4945.

Huttley GA, Wakefield MJ, Easteal S. 2007. Rates of genome evolution and branching order from whole genome analysis. *Mol Biol Evol* **24**: 1722-1730.

Kai Y, Mei H, Kawano H, Nakajima N, Takai A, Kumon M, Inoue A, Yamashita N. 2022. Transcriptomic signatures in trophectoderm and inner cell mass of human blastocysts classified according to developmental potential, maternal age and morphology. *PLoS One* **17**: e0278663.

Kaneko-Ishino T, Ishino F. 2012. The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front Microbiol* **3**: 262.

Kaneko-Ishino T, Ishino F. 2022. The Evolutionary Advantage in Mammals of the Complementary Monoallelic Expression Mechanism of Genomic Imprinting and Its Emergence From a Defense Against the Insertion Into the Host Genome. *Front Genet* **13**: 832983.

Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**: 676-687.

Kiyonari H, Kaneko M, Abe T, Shiraishi A, Yoshimi R, Inoue KI, Furuta Y. 2021. Targeted gene disruption in a marsupial, Monodelphis domestica, by CRISPR/Cas9 genome editing. *Curr Biol* **31**: 3956-3963 e3954.

Kobayashi H. 2021. Canonical and Non-canonical Genomic Imprinting in Rodents. *Front Cell Dev Biol* **9**: 713878.

Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T. 2012. Imprinted DNA methylation reprogramming during early mouse embryogenesis at the Gpr1-Zdbf2 locus is linked to long cis-intergenic transcription. *FEBS Lett* **586**: 827-833.

Kobayashi H, Yamada K, Morita S, Hiura H, Fukuda A, Kagami M, Ogata T, Hata K, Sotomaru Y, Kono T. 2009. Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. *Genomics* **93**: 461-472.

Kobayashi H, Yanagisawa E, Sakashita A, Sugawara N, Kumakura S, Ogawa H, Akutsu H, Hata K, Nakabayashi K, Kono T. 2013. Epigenetic and transcriptional features

of the novel human imprinted lncRNA GPR1AS suggest it is a functional ortholog to mouse Zdbf2linc. *Epigenetics* **8**: 635-645.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927-931.

Mika K, Whittington CM, McAllan BM, Lynch VJ. 2022. Gene expression phylogenies and ancestral transcriptome reconstruction resolves major transitions in the origins of pregnancy. *Elife* **11**.

Moore T, Haig D. 1991. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* **7**: 45-49.

Morcos L, Ge B, Koka V, Lam KC, Pokholok DK, Gunderson KL, Montpetit A, Verlaan DJ, Pastinen T. 2011. Genome-wide assessment of imprinted expression in human cells. *Genome Biol* **12**: R25.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635-640.

Nicholls RD. 2000. The impact of genomic imprinting for neurobehavioral and developmental disorders. *J Clin Invest* **105**: 413-418.

Papuchova H, Latos PA. 2022. Transcription factor networks in trophoblast development. *Cell Mol Life Sci* **79**: 337.

Richard Albert J, Kobayashi T, Inoue A, Monteagudo-Sanchez A, Kumamoto S, Takashima T, Miura A, Oikawa M, Miura F, Takada S et al. 2023. Conservation and divergence of canonical and non-canonical imprinting in murids. *Genome Biol* **24**: 48.

Rosenkrantz JL, Gaffney JE, Roberts VHJ, Carbone L, Chavez SL. 2021. Transcriptomic analysis of primate placentas and novel rhesus trophoblast cell lines informs investigations of human placentation. *BMC Biol* **19**: 127.

Saito S, Akizawa H, Furukawa E, Yanagawa Y, Bai H, Takahashi M, Kawahara M. 2022. Generation of viable calves derived from developmentally mature blastocysts produced by on-gel culture. *J Reprod Dev* **68**: 330-334.

Sakashita A, Kitano T, Ishizu H, Guo Y, Masuda H, Ariura M, Murano K, Siomi H. 2023. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet* **55**: 484-495.

Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, Turelli P, Imbeault M, Ferguson-Smith AC, Trono D. 2019. ZNF445 is a primary regulator of genomic imprinting. *Genes Dev* **33**: 49-54.

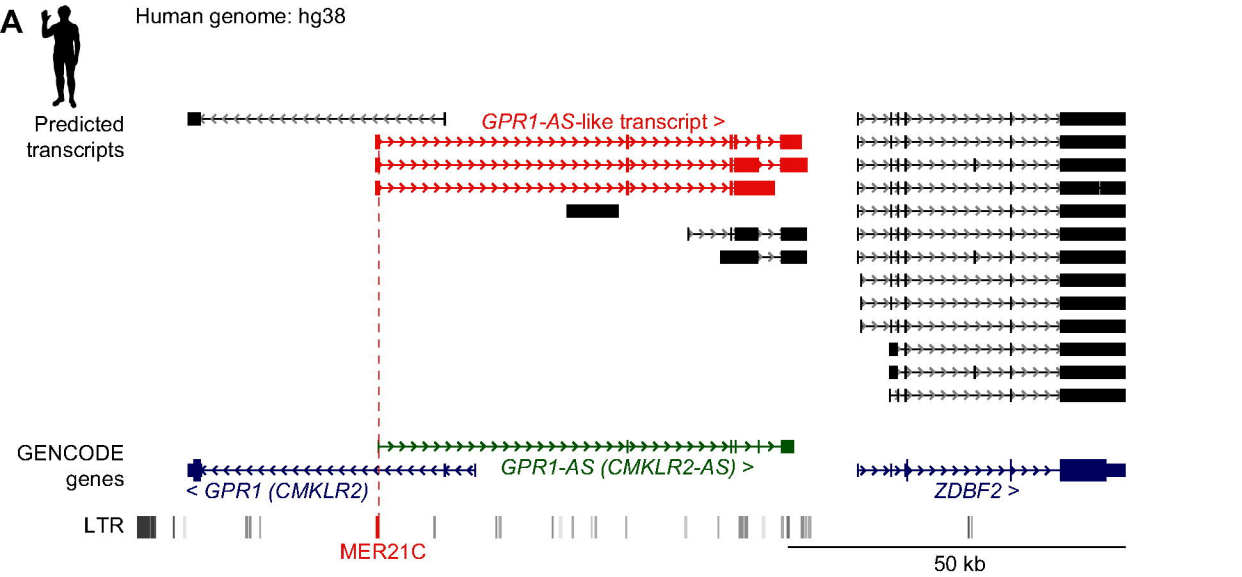
Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Erice Imprinting G. 2019. Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**: 952-965.

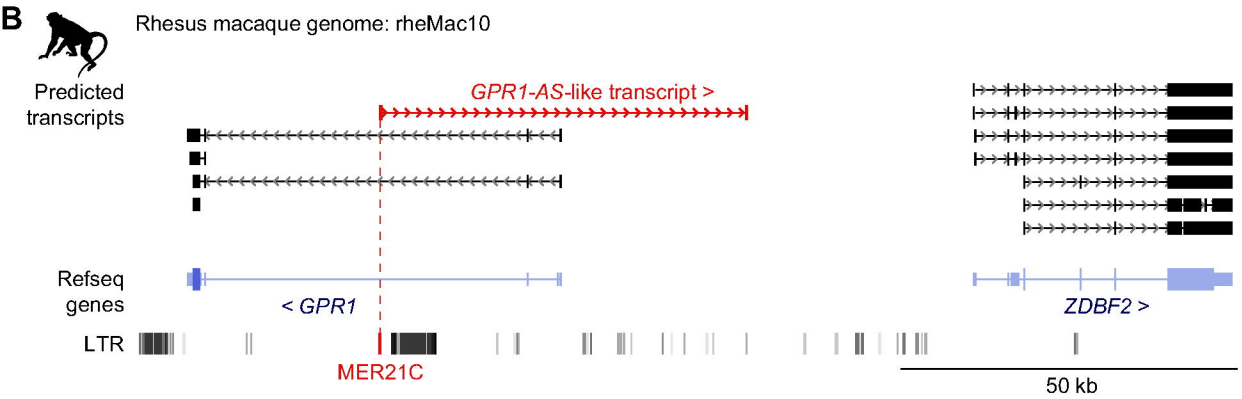
Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A* **82**: 1741-1745.

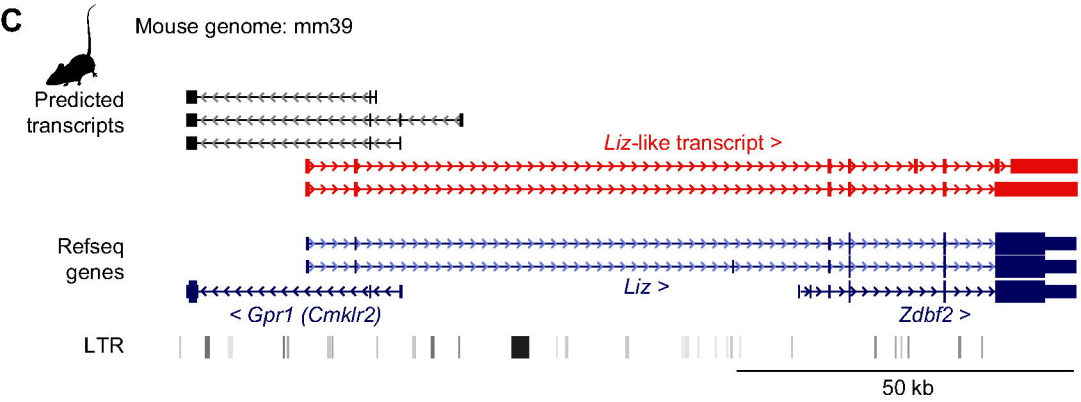
Zhang W, Chen F, Chen R, Xie D, Yang J, Zhao X, Guo R, Zhang Y, Shen Y, Goke J et al. 2019. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res* **47**: 8485-8501.

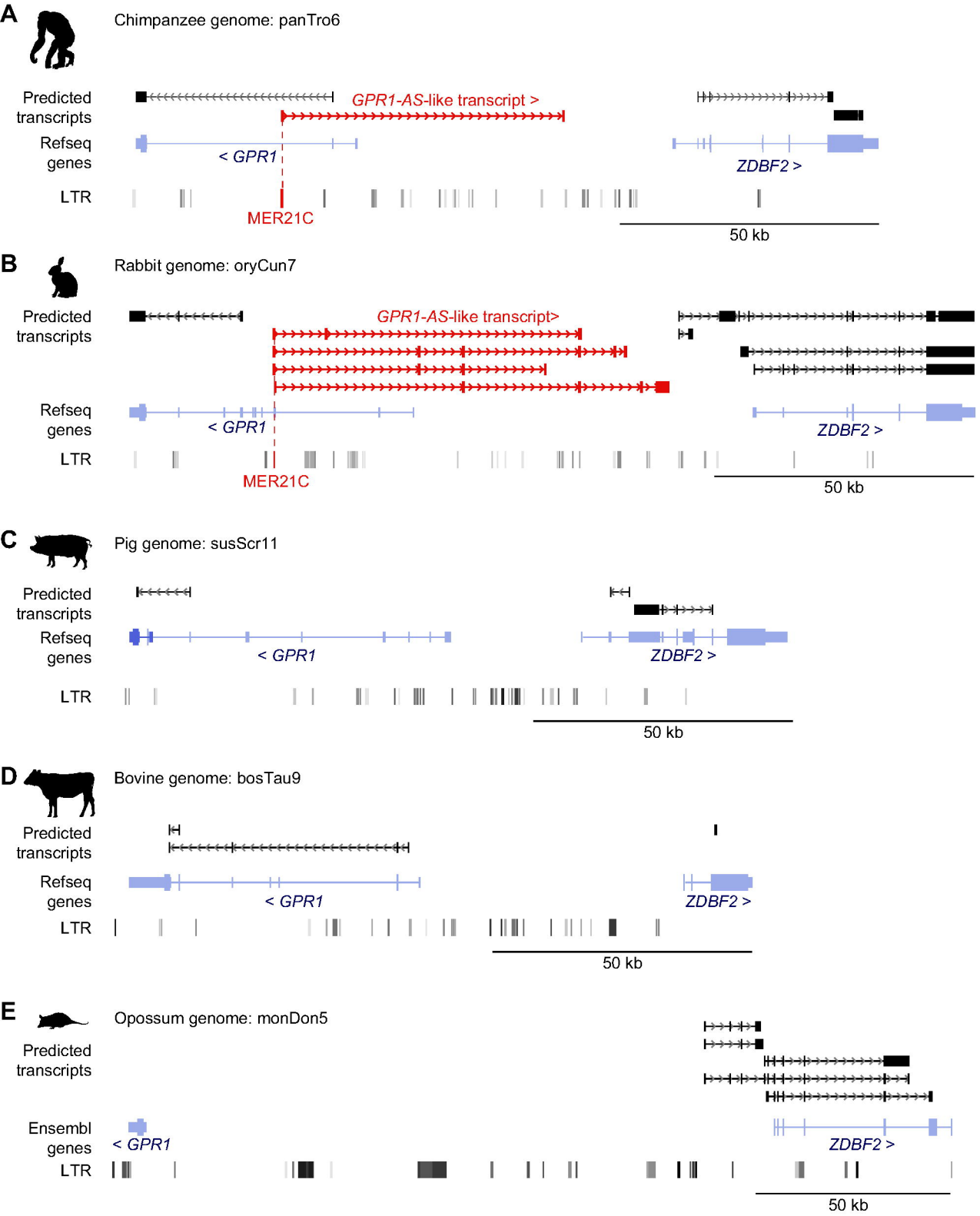
Zou Z, Ohta T, Miura F, Oki S. 2022a. ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res* **50**: W175-W182.

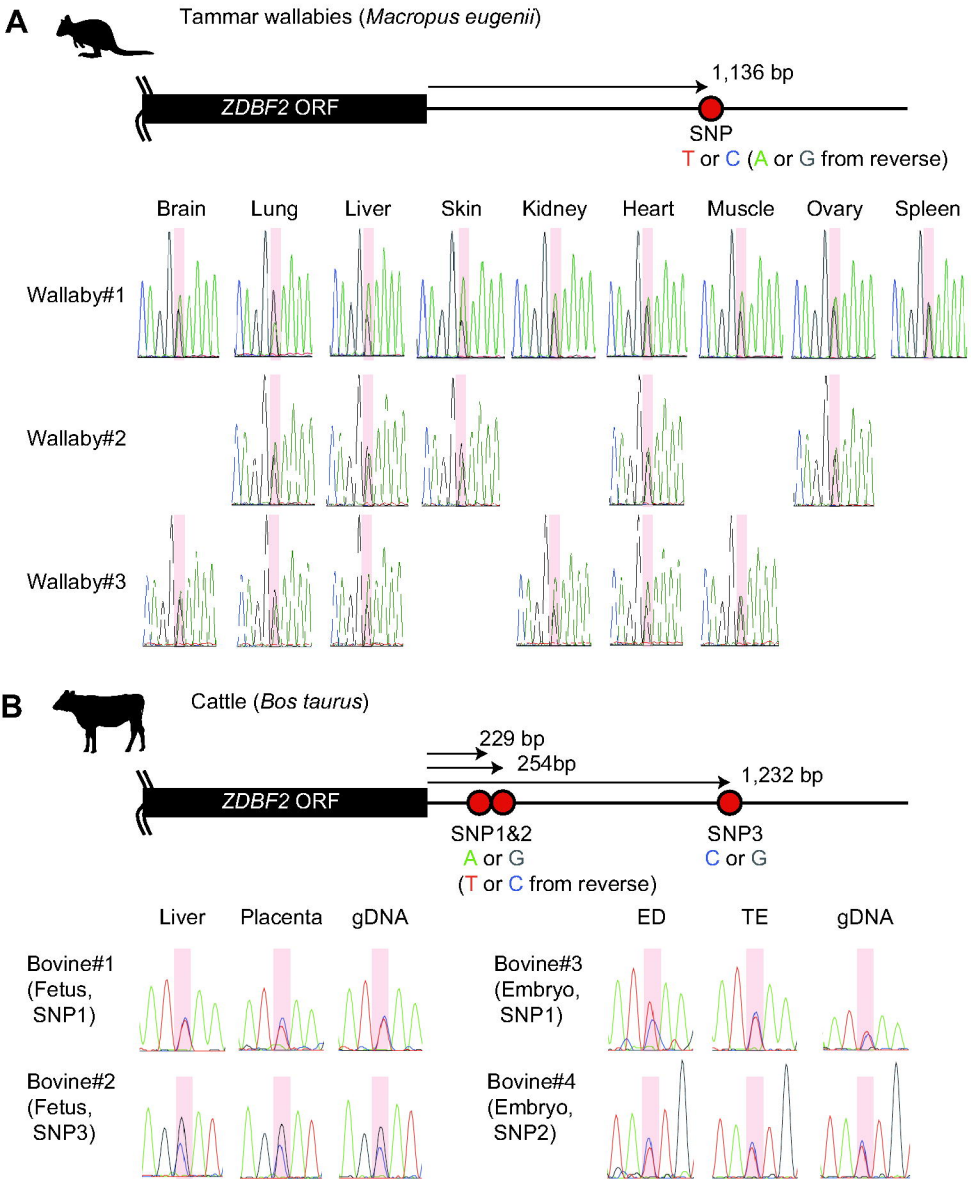
Zou Z, Zhang C, Wang Q, Hou Z, Xiong Z, Kong F, Wang Q, Song J, Liu B, Liu B et al. 2022b. Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science* **378**: abo7923.

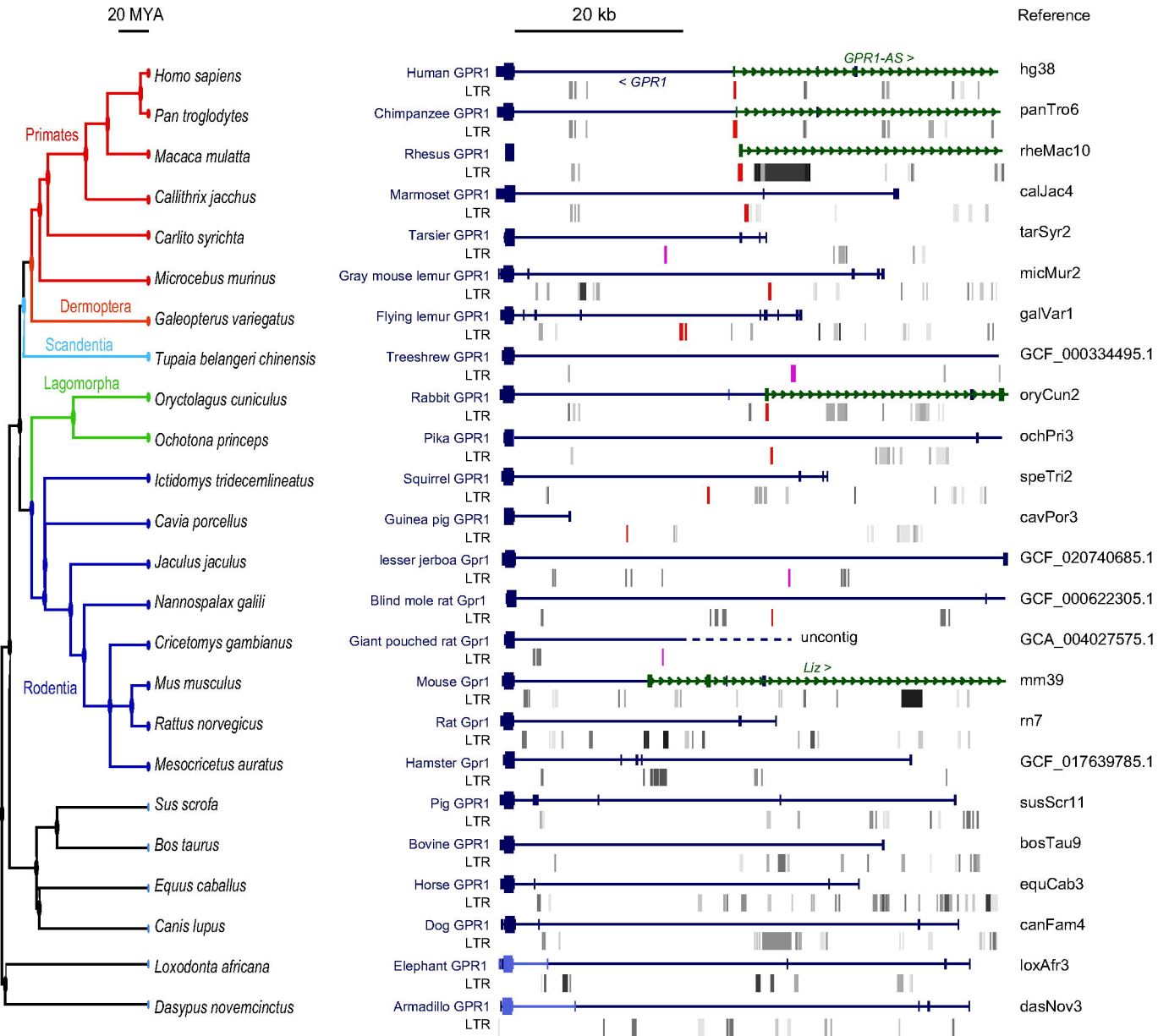

A
Human genome: hg38
Predicted transcripts
GPR1-AS-like transcript >
GENCODE genes
< GPR1 (CMKLR2)
GPR1-AS (CMKLR2-AS) >
ZDBF2 >
LTR
MER21C
50 kb

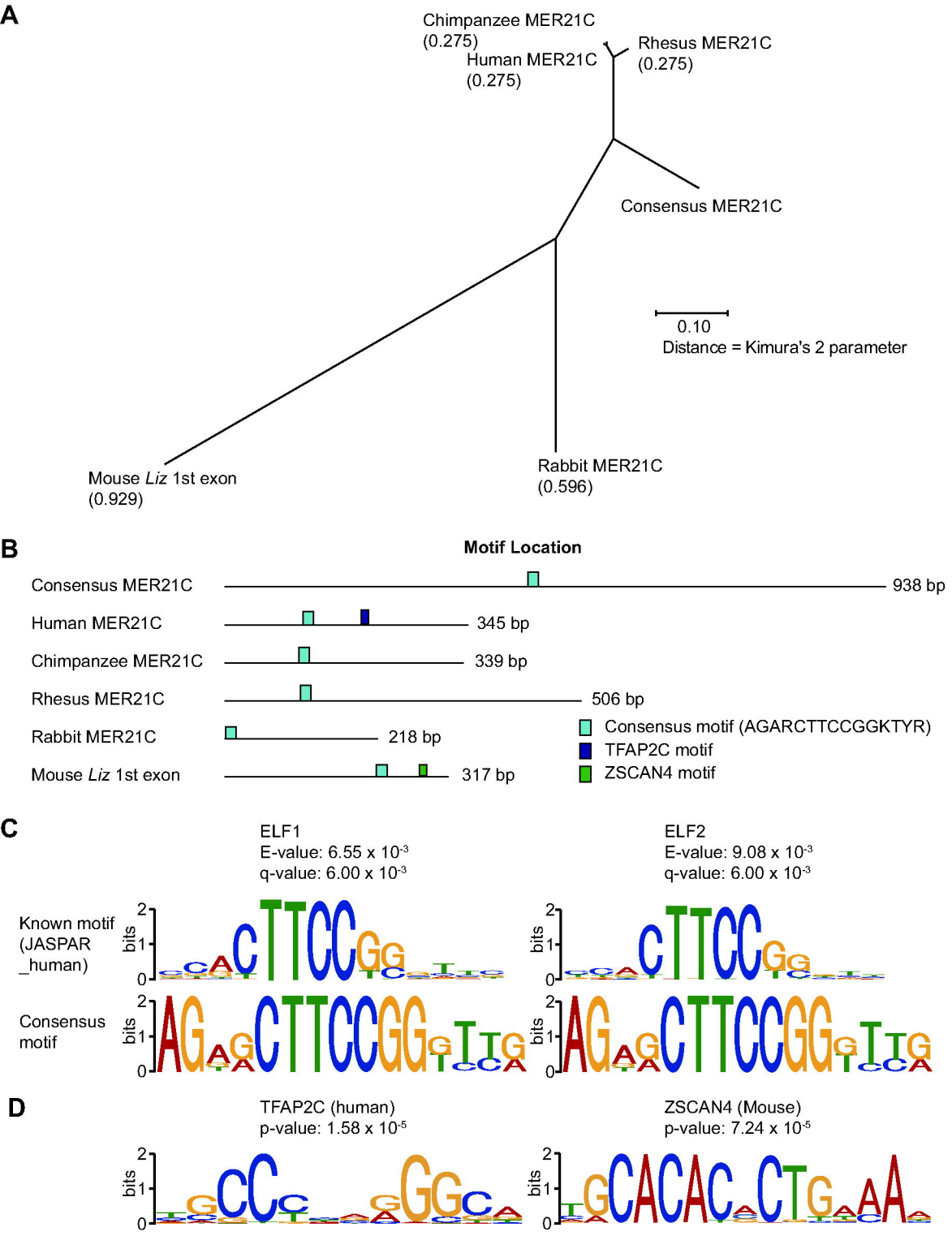

B
Rhesus macaque genome: rheMac10
Predicted transcripts
GPR1-AS-like transcript >
Refseq genes
< GPR1
ZDBF2 >
LTR
MER21C
50 kb

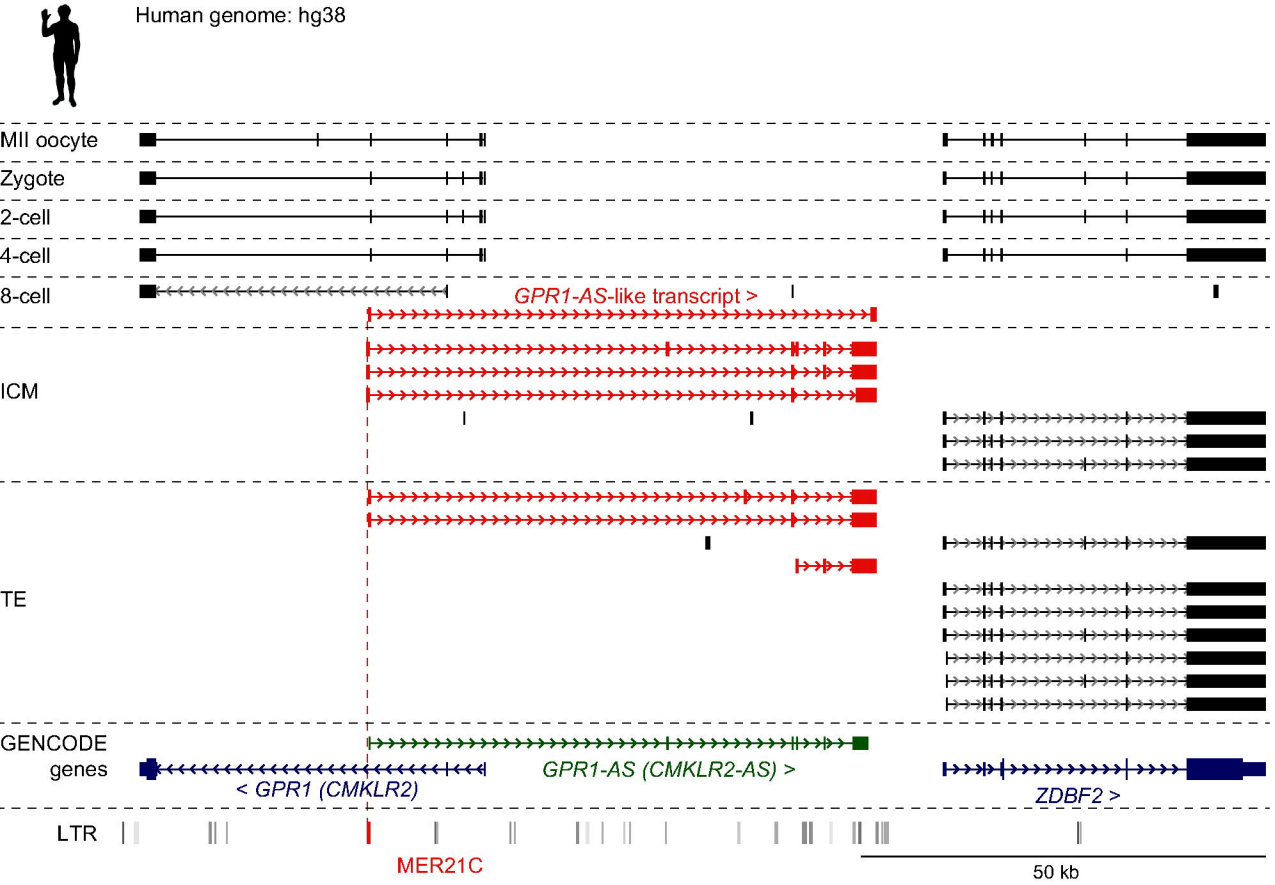

C
Mouse genome: mm39
Predicted transcripts
Liz-like transcript >
Refseq genes
Liz >
< Gpr1 (Cmklr2)
Zdbf2 >
LTR
50 kb


A
Chimpanzee genome: panTro6
Predicted transcripts
GPR1-AS-like transcript >
Refseq genes
< GPR1
ZDBF2 >
LTR
MER21C
50 kb
B
Rabbit genome: oryCun7
Predicted transcripts
GPR1-AS-like transcript>
Refseq genes
< GPR1
ZDBF2 >
LTR
MER21C
50 kb
C
Pig genome: susScr11
Predicted transcripts
Refseq genes
< GPR1
ZDBF2 >
LTR
50 kb
D
Bovine genome: bosTau9
Predicted transcripts
Refseq genes
< GPR1
ZDBF2 >
LTR
50 kb
E
Opossum genome: monDon5
Predicted transcripts
Ensembl genes
< GPR1
ZDBF2 >
LTR
50 kb


A
Tammar wallabies (Macropus eugenii)
1,136 bp
ZDBF2 ORF
SNP
T or C (A or G from reverse)
Brain
Lung
Liver
Skin
Kidney
Heart
Muscle
Ovary
Spleen
Wallaby#1
Wallaby#2
Wallaby#3
B
Cattle (Bos taurus)
229 bp
254bp
1,232 bp
ZDBF2 ORF
SNP1&2
A or G
(T or C from reverse)
SNP3
C or G
Liver
Placenta
gDNA
ED
TE
gDNA
Bovine#1 (Fetus, SNP1)
Bovine#2 (Fetus, SNP3)
Bovine#3 (Embryo, SNP1)
Bovine#4 (Embryo, SNP2)


20 MYA
20 kb
Reference
Primates
Dermoptera
Scandentia
Lagomorpha
Rodentia
Homo sapiens
Pan troglodytes
Macaca mulatta
Callithrix jacchus
Carlito syrichta
Microcebus murinus
Galeopterus variegatus
Tupaia belangeri chinensis
Oryctolagus cuniculus
Ochotona princeps
Ictidomys tridecemlineatus
Cavia porcellus
Jaculus jaculus
Nannospalax galili
Cricetomys gambianus
Mus musculus
Rattus norvegicus
Mesocricetus auratus
Sus scrofa
Bos taurus
Equus caballus
Canis lupus
Loxodonta africana
Dasypus novemcinctus
GPR1-AS >
< GPR1
Human GPR1
LTR
hg38
Chimpanzee GPR1
panTro6
Rhesus GPR1
rheMac10
Marmoset GPR1
calJac4
Tarsier GPR1
tarSyr2
Gray mouse lemur GPR1
micMur2
Flying lemur GPR1
galVar1
Treeshrew GPR1
GCF_000334495.1
Rabbit GPR1
oryCun2
Pika GPR1
ochPri3
Squirrel GPR1
speTri2
Guinea pig GPR1
cavPor3
lesser jerboa Gpr1
GCF_020740685.1
Blind mole rat Gpr1
GCF_000622305.1
Giant pouched rat Gpr1
uncontig
GCA_004027575.1
Liz >
Mouse Gpr1
mm39
Rat Gpr1
rn7
Hamster Gpr1
GCF_017639785.1
Pig GPR1
susScr11
Bovine GPR1
bosTau9
Horse GPR1
equCab3
Dog GPR1
canFam4
Elephant GPR1
loxAfr3
Armadillo GPR1
dasNov3


A
Chimpanzee MER21C
(0.275)
Human MER21C
(0.275)
Rhesus MER21C
(0.275)
Consensus MER21C
0.10
Distance = Kimura's 2 parameter
Mouse Liz 1st exon
(0.929)
Rabbit MER21C
(0.596)
B
Motif Location
Consensus MER21C
938 bp
Human MER21C
345 bp
Chimpanzee MER21C
339 bp
Rhesus MER21C
506 bp
Rabbit MER21C
218 bp
Mouse Liz 1st exon
317 bp
Consensus motif (AGARCTTCCGGKTYR)
TFAP2C motif
ZSCAN4 motif
C
ELF1
E-value: 6.55 x 10-3
q-value: 6.00 x 10-3
ELF2
E-value: 9.08 x 10-3
q-value: 6.00 x 10-3
Known motif
(JASPAR
_human)
Consensus
motif
bits
D
TFAP2C (human)
p-value: 1.58 x 10-5
ZSCAN4 (Mouse)
p-value: 7.24 x 10-5


Human genome: hg38
MII oocyte
Zygote
2-cell
4-cell
8-cell
GPR1-AS-like transcript >
ICM
TE
GENCODE
genes
< GPR1 (CMKLR2)
GPR1-AS (CMKLR2-AS) >
ZDBF2 >
LTR
MER21C
50 kb