



# LTRs of Endogenous Retroviruses as a Source of Tbx6 Binding Sites

Yukuto Yasuhiko, Yoko Hirabayashi and Ryuichi Ono \*

Division of Cellular and Molecular Toxicology, Biological Safety Research Centre, National Institute of Health Sciences, Tokyo, Japan

Retrotransposons are abundant in mammalian genomes and can modulate the gene expression of surrounding genes by disrupting endogenous binding sites for transcription factors (TFs) or providing novel TFs binding sites within retrotransposon sequences. Here, we show that a (C/T)CACACCT sequence motif in ORR1A, ORR1B, ORR1C, and ORR1D, Long Terminal Repeats (LTRs) of MaLR endogenous retrovirus (ERV), is the direct target of Tbx6, an evolutionary conserved family of T-box TFs. Moreover, by comparing gene expression between control mice (Tbx6 +/–) and Tbx6-deficient mice (Tbx6 –/–), we demonstrate that at least four genes, *Twist2*, *Pitx2*, *Oscp1*, and *Nfxl1*, are down-regulated with Tbx6 deficiency. These results suggest that ORR1A, ORR1B, ORR1C and ORR1D may contribute to the evolution of mammalian embryogenesis.

## OPEN ACCESS

### Edited by:

Tammy A. Morrish,  
Independent Investigator,  
United States

### Reviewed by:

Feng Zhang,  
Fudan University, China  
Dustin C. Hancks,  
University of Utah, United States  
Yong Zhang,  
Institute of Zoology (CAS), China

### \*Correspondence:

Ryuichi Ono  
onoryu@nihs.go.jp

### Specialty section:

This article was submitted to  
Cellular Biochemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 22 December 2016

Accepted: 23 May 2017

Published: 15 June 2017

### Citation:

Yasuhiko Y, Hirabayashi Y and Ono R  
(2017) LTRs of Endogenous  
Retroviruses as a Source of Tbx6  
Binding Sites. *Front. Chem.* 5:34.  
doi: 10.3389/fchem.2017.00034

**Keywords:** endogenous retroviruses, retrotransposon, transcription factors, evolution, TBX6

## INTRODUCTION

About half of the mammalian genome is occupied by DNA sequences derived from transposable elements (TEs) (Lander et al., 2001; Waterston et al., 2002; Lindblad-Toh et al., 2005; de Koning et al., 2011). Retrotransposons, which mobilize via an RNA intermediate by a copy-and-paste mechanism, comprise the majority of mammalian TEs, whereas DNA transposons, which move via a cut-and-paste mechanism, comprise a small fraction and have accumulated mutations that render them immobile (Deininger et al., 2003). Most TEs are nonfunctional and are regarded as genomic parasites or junk DNA; however, a growing body of evidence suggests that retrotransposons and retrotransposon-derived genes have acquired functions essential for host survival during mammalian evolution (Yoder et al., 1997; Levin and Moran, 2011; Hancks and Kazazian, 2012).

In some cases, open reading frames from TEs are domesticated as endogenous genes during mammalian evolution. For example, *Peg10* and *Rtl1*, derived from the gag and pol proteins of the Ty3/Gypsy type retrotransposon, which is similar to *Sushi-ichi*, are highly conserved in mammals and participate in placental formation (Ono et al., 2001, 2003, 2006; Sekita et al., 2008). Similar to the gag protein of *Sushi-ichi*, the other two of the eleven *Sushi-ichi* retrotransposon homolog (*Sirh*) family genes, *Sirh7/Ldoc1* and *Sirh11/Zcchc16*, encode ORF (Open-Reading frame); they are also involved in the determination of the timing of parturition and cognitive function in the brain, respectively (Ono et al., 2011; Naruse et al., 2014; Irie et al., 2015). *Syncytins/SYNCYTINs* (mouse/human) and *FEMATRIN* (cow), derived from the envelope of endogenous retrovirus (ERV), mediate cell-cell fusion to form the syncytiotrophoblast and induce fusion with bovine endometrial cells *in vitro* (Mi et al., 2000; Dupressoir et al., 2009, 2011; Nakaya et al., 2013). *Skin aspartic protease (SAPase)*, which has a retrovirus-like aspartic protease, plays important roles in the determination of the texture of the skin by modulating the degree of hydration by processing profilaggrin (Matsui et al., 2010).

Since the discovery of TEs, it has been posited that TEs may seed regulatory elements throughout genomes and drive phenotypic differences between species via changes in transcriptional output (McClintock, 1950; Britten and Davidson, 1969; Feschotte, 2008). It has become evident that many TEs, such as long terminal repeats (LTRs) of endogenous retroviruses (ERVs), contain TF binding sites and are associated with gene expression patterns. For example, MuERV-L LTRs function as alternative promoters for protein coding genes, including *Gata4* and *Tead4*, which are important for the specification of primitive endoderm and trophoblast, respectively, in two-cell embryos (Kigami et al., 2003; Evsikov et al., 2004; Macfarlan et al., 2012). It has also been reported that MuERV-L, exclusively expressed in two-cell embryos, is captured at double-strand break (DSB) sites introduced by the CRISPR/Cas9 system in mouse zygotes (Ono et al., 2015). Some of the intracisternal A-particle (IAP) retrotransposon insertions are known to induce *de novo* metastable epi-alleles, such as *agouti viable yellow* (*Avy*), *axin fused* (*AxinFu*) and *Cdk5rap* locus (Vasicek et al., 1997; Morgan et al., 1999; Druker et al., 2004). The stochastic nature of the establishment of the epigenetic state of the 5' LTR leads to variable expressivity of the adjacent genes. Both the sense and anti-sense LINE-1 (L1) promoter can drive L1 chimeric transcripts (Criscione et al., 2016). Moreover, AS071 and AS021, two AmnSINEs present in mammals as well as birds and reptiles, are enhancers of the genes *FGF8* (fibroblast growth factor 8), 178 kb from AS071, and *SATB2*, 392 kb from AS021 (Sasaki et al., 2008). Recently, it was reported that MER41, a primate-specific endogenized gammaretrovirus, is a source of interferon  $\gamma$  (IFNG)-inducible binding sites (Chuong et al., 2016).

In this study, we demonstrate a potential role for ORR1A (Origin-Region Repeat 1A), ORR1B, ORR1C, and ORR1D, LTRs of the MaLR (Mammalian-Apparent Long-Terminal Repeat Retrotransposon) endogenous retrovirus-like element, in controlling gene expression via *Tbx6* binding (Smit, 1993). Because *Tbx6* functions in the regulation of early embryogenesis, including anti-neural fate regulation in the presomitic mesoderm and later somite segmentation, ORR1A, ORR1B, ORR1C, and ORR1D may have played a role in the evolution of mammalian embryogenesis (Chapman and Papaioannou, 1998; Takemoto et al., 2011).

## RESULTS AND DISCUSSION

*Tbx6* belongs to an evolutionarily conserved family of T-box transcription factors (TFs), known to be involved in the neural-mesodermal fate determination of axial stem cells (Chapman and Papaioannou, 1998; Takemoto et al., 2011). Previously, we revealed that *Tbx6* directly activates the expression of *Mesp2*, a segmentation and polarization factor in somitogenesis, in a Notch signal-dependent manner (Yasuhiko et al., 2006). A ligand of Notch signal, Dll1, is also a direct target of *Tbx6*, implying that *Tbx6* participates in the regulation of the Notch signaling pathway (White and Chapman, 2005). The consensus core

sequence of *Tbx6* binding sites has been reported as CACACCT or AGGTGTBRNNNN (White and Chapman, 2005). In this study, we used (C/T)CACACCT as a consensus for both reports (White and Chapman, 2005; Yasuhiko et al., 2006).

At first, the *Tbx6* binding sequence motif, (C/T) CACACCT, was identified by whole genome *in silico* screening. Furthermore, we chose the *Tbx6* binding sequence, which has at least two more *Tbx6* binding sequences within the neighboring 100 bp upstream and/or downstream regions, because we previously demonstrated that higher enhancer activity of *Tbx6* was observed when there are more than three *Tbx6* binding sequences within a narrow region. As a result, 3500 potential *Tbx6* binding sites were identified, and a characteristic feature was revealed (**Figure 1A**; Supplementary Table 1).

Approximately 70.0% of potential *Tbx6* binding sites comprise repeat sequences (**Figure 1A**). Specifically, 85.7% of the potential *Tbx6*-binding-repeat sequences were within ORR1A, ORR1B, ORR1C, and ORR1D, LTRs of the MaLR that span 679 independent ORR1s-LTRs, while SINES and LINES represent only 2% of the *Tbx6*-binding repeat sequences (Bao et al., 2015; Supplementary Table 1).

There are 166,375 Repeatmasker annotated ORR1s, including partial sequences, in the mouse genome (MM10), and 20% of them have at least one *Tbx6* binding site (**Figure 1B**). In fact, the reference sequences of ORR1s-LTRs from Repbase, which are consensus sequences of ORR1s, have one or two *Tbx6* binding sequence motifs (**Figure 1C**). These data suggest that the tandem insertion of these LTRs or degenerated LTR sequences with more than three *Tbx6* binding sequence motifs might be good targets for *Tbx6* to bind *in vivo*. Furthermore, potential *Tbx6*-binding ORR1s have more than three *Tbx6* binding motifs within themselves or share the *Tbx6* binding motifs with neighboring sequences.

*Tbx6*-binding ORR1s more than 300 bp in length were selected, and the consensus sequences including three *Tbx6*-binding motifs and the absolute distance from each ORR1 to the nearest mouse reference gene were determined (**Figures 1D,E**; Supplementary Table 1). The strong interaction between *Tbx6* and the consensus sequence of *Tbx6*-binding ORR1s were confirmed by electrophoretic mobility shift assay (EMSA), while the interactions disappeared by introducing mutations into the *Tbx6*-binding motif one by one (**Figure 1F**). The finding that three *Tbx6* binding motifs rather than one or two *Tbx6* binding motifs have stronger binding affinity was comparable to our previous report (**Figure 1G**; Yasuhiko et al., 2008). As *Tbx6*-binding ORR1s were relatively enriched near gene transcription start sites (**Figure 1E**), *Tbx6* may contribute to regulating the expression level of nearby genes until reaching 60 kb-windows. Then, to explore the influence of ORR1A, ORR1B, ORR1C, and ORR1D on the regulation of gene expression by *Tbx6*, we compared the expression level of 9 genes that are randomly selected within 50 kb of potential *Tbx6* binding sites on ORR1A, ORR1B, ORR1C, and/or ORR1D in *Tbx6* (+/-) (control) and *Tbx6* (-/-) (*Tbx6* KO) embryos at 8.0 day post-coitus (dpc). Because *Tbx6* KO embryos have morphological abnormalities after 9 dpc, we used 8.0 dpc embryos in this study to exclude secondary effects from morphological abnormalities.

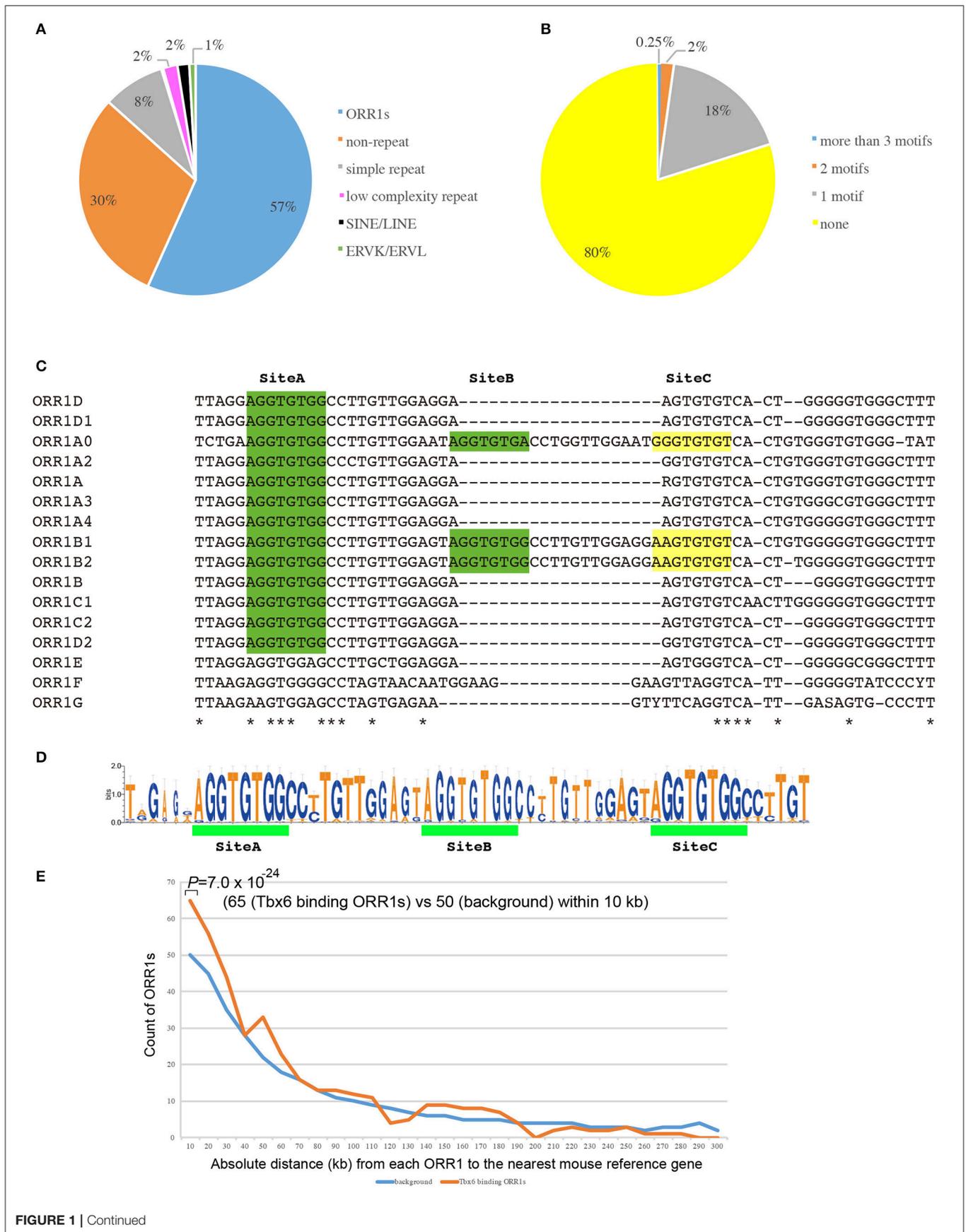
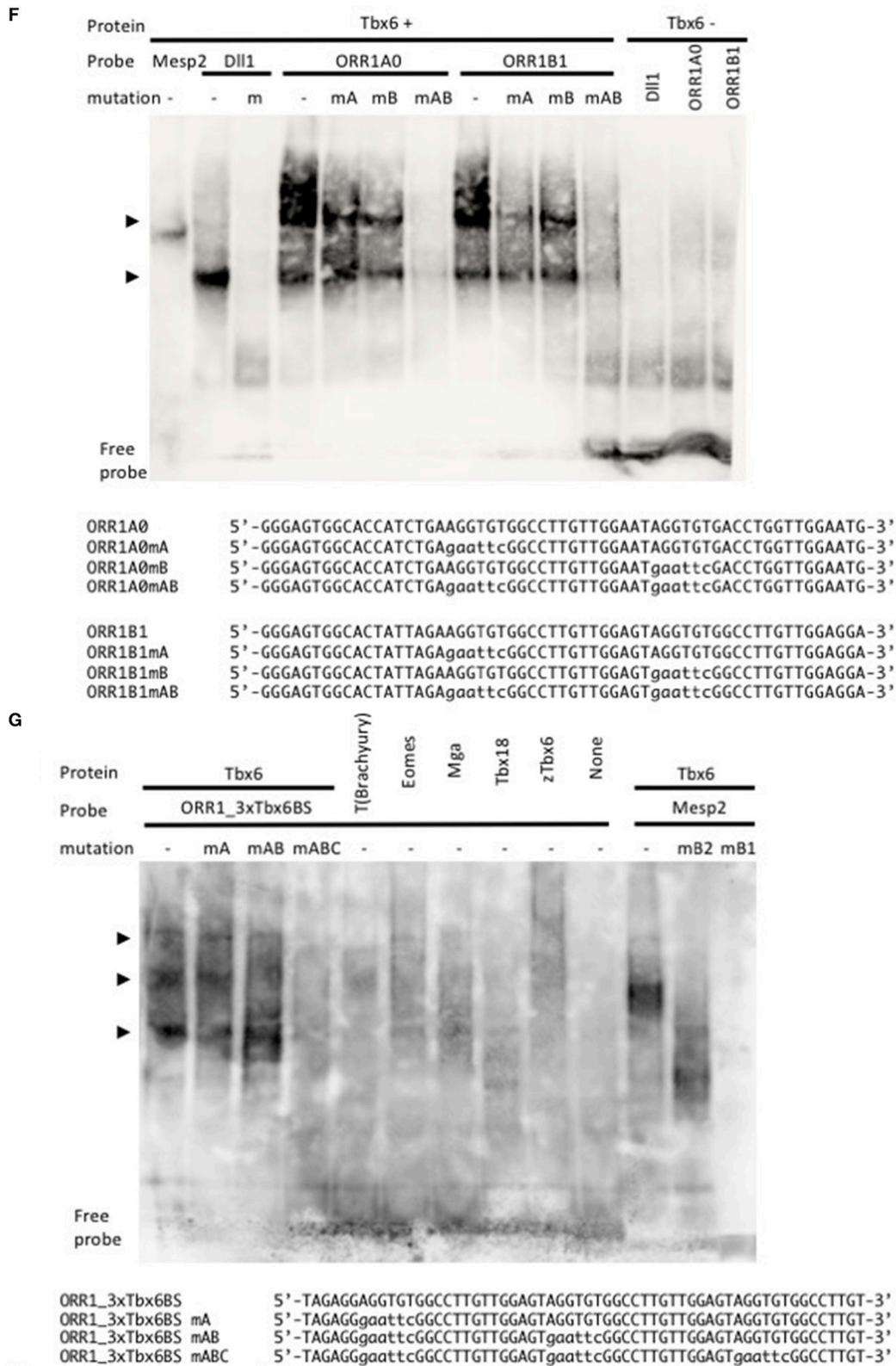


FIGURE 1 | Continued



**FIGURE 1 |** Tbx6 binding sequence motifs in LTRs of ORR1 ERV families. **(A)** Distribution of Tbx6 binding sequence motifs in the mouse genome (MM10). Of the 3,500 Tbx6 binding sequence motifs, 70% were repeat sequences, including TEs and simple repeats. ORR1, Long Terminal Repeats (LTRs) of MaLR endogenous *(Continued)*

**FIGURE 1 | Continued**

retrovirus (ERV), occupy 57% of the total Tbx6 binding sequence motifs. **(B)** Of all the ORR1 sequences in the mouse genome, 20% of ORR1s have at least one Tbx6 binding sequence. **(C)** DNA sequence comparison between ORR1A (rodentia ancestral shared), ORR1A0 (mus musculus), ORR1A2 (muridae), ORR1A3 (muridae), ORR1A4 (muridae), ORR1B (rodentia ancestral shared), ORR1B1 (mus musculus), ORR1B2 (mus musculus), ORR1C1 (rodentia ancestral shared), ORR1C2 (rodentia ancestral shared), ORR1D1 (rodentia ancestral shared), ORR1D2 (rodentia ancestral shared), ORR1E (rodentia ancestral shared), ORR1F (muridae) and ORR1G (muridae) LTRs. Identical sequences are indicated by asterisks. The Tbx6 binding sequence motif is indicated by green boxes. Yellow boxes are indicated as a corresponding region of the Tbx6 binding sequence motif "Site C" in **Figure 1D**. **(D)** Sequence logo of the the ORR1 LTRs that had more than three Tbx6 binding sequence motifs. Three tandem "AGGTGTGs," a Tbx6 binding sequence motif, are highly conserved between ORR1 LTRs, which have more than three Tbx6 binding sequence motifs. **(E)** Frequency histogram of the absolute distance from each ORR1 to the nearest mouse reference gene. The background expectation is derived from the genome-wide ORR1s distribution. Statistical significance of the observed enrichment within the first 10 kb of the nearest mouse reference gene was assessed by a binominal test. **(F)** Site A and site B sequences independently bind to Tbx6 in an electromobility shift assay (EMSA); however, the binding affinity is much higher with the presence of both sites A and B. Sequences of oligonucleotide probes were shown below the gel image. Mutated nucleotides were depicted in lower case. **(G)** Triple Tbx6 binding sequence motif shows the highest binding affinity to Tbx6, while other T-box TFs, including T (Brachyury), Eomes, Mga, Tbx18, and zebrafish Tbx6 (zTbx6), have no affinity. Arrowheads in **(F,G)**: Positions of the bands resulted from multiple Tbx6 binding to ORR1 sequences. Sequences of oligonucleotide probes were shown below the gel image. Mutated nucleotides were depicted in lower case.

As expected, four genes, *Twist2*, *Pitx2*, *Oscp1*, and *Nfxl1*, were down-regulated, although the expression of five other genes, *Enpep*, *Prdm2*, *Corin*, *Pdpr* and *Map4k4*, was not altered significantly (**Figure 2**). It has been reported that enhancer activity could be blocked by the epigenetic repressive marks of the neighboring regions, such as histone deacetylation and trimethylation of K9 and K27 on histone H3 (H3K27me3 and H3K9me3) or an insulator, a genetic boundary element blocking the interaction between enhancers and promoters (Roth et al., 2001; Schmidl et al., 2009; Greer and Shi, 2012; Downen et al., 2014). It might be possible that five genes whose expression levels were not altered by Tbx6 deficiency could be blocked by epigenetic modifications or unknown silencers.

Our analysis revealed the rodent-specific ORR1 family of ERVs to be a source of Tbx6 binding sites. Furthermore, Tbx6-binding ORR1s are enriched near genes which might be associated with several biological process and molecular pathways (**Figure 3**). In the human genome, there are 2,927 potential TBX6 binding motifs; however, the majority of sites are not in LTRs but in simple repeat sequences or Alu (Supplementary Table 2). Although the source of Tbx6/TBX6 binding sequences is different between species, each mammalian species might shape their Tbx6/TBX6 binding sequence through mammalian evolution. Our analysis and other reports, including the primate-specific MER41 family as IFNG-inducible binding sites and AmnSINE1s as mammalian enhancers, raised the possibility that TE-derived regulatory elements influence lineage-specific mammalian evolution (Sasaki et al., 2008; Chuong et al., 2016).

## MATERIALS AND METHODS

### Bioinformatic Analyses

(C/T)CACACCT, Tbx6 binding sequence motifs, were identified in the mouse whole genome (MM10) and human whole genome (hg19) and filtered out when there were not two more Tbx6 binding sequences within the neighboring 100 bp upstream and/or downstream regions using gggenome (<https://gggenome.dbcls.jp>). All the TE sequences were downloaded from Repeatmasker truck (mouse:MM10/human:hg19) of the UCSC genome browser (<https://genome.ucsc.edu>). The

Intersect intervals program (<https://usegalaxy.org/>) was used to identify the TEs that have potential Tbx6 binding sequences using potential Tbx6 binding sites identified as a query against Repeatmasker annotated TEs. The ClosestBed program (<https://usegalaxy.org/>) was used to find the closest mouse reference genes (MM10) and to identify the absolute distance between the potential Tbx6 binding motif and its closest reference gene. These distances were grouped by 10 kb-bin sizes. The expected background was determined by randomly sampling an equal number of the remaining 78,042 annotated ORR1s that did not have more than three Tbx6 binding motifs. Sampling was repeated 100 times, and the mean number of elements was used as the expected value for comparison to the potential Tbx6 binding ORR1s. Statistical significance was determined for the first 10-kb bin by a binominal test as previously described (Chuong et al., 2016). Gene ontology of the closest reference genes within 50 kb-windows of potential Tbx6-binding motifs were determined by the GREAT program <http://bejerano.stanford.edu/great/public/html/index.php> (**Figure 3**). The consensus sequence of the potential Tbx6-binding motifs was identified by ClustalW program (for alignment: <http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja>) and Sequence Logo program (for generation of sequence logos: <http://weblogo.berkeley.edu>). A gene ontology/signal pathway panel of Tbx6-binding ORR1s.

### Electrophoretic Mobility Shift Assay (EMSA)

Full sequences of ORFs of mouse *Tbx6* (NM\_011538.2), *T* (*Brachyury*; NM\_009309.2), *Eomes* (NM\_010136.3), *Tbx18* (NM\_023814.4), and T-box-coding fragment of *Mga* (NM\_013720.2) were PCR amplified and cloned in pCS2+ (Rupp et al., 1994) vector. Expression vector pCS2-zTbx6 for zebrafish Tbx6 translation was a gift from Dr Hiroyuki Takeda (Terasaki et al., 2006). Transcription factors were *in vitro* transcribed and translated using TnT(R) Quick Coupled Transcription/Translation System (Promega) following the manufacturer's protocol. Sequences of DNA probes were as follows: Mutated nucleotides are designated in lower case. *Mesp2* and *Dll1* were positive controls for the assay and described in Yasuhiko et al. (2006) and

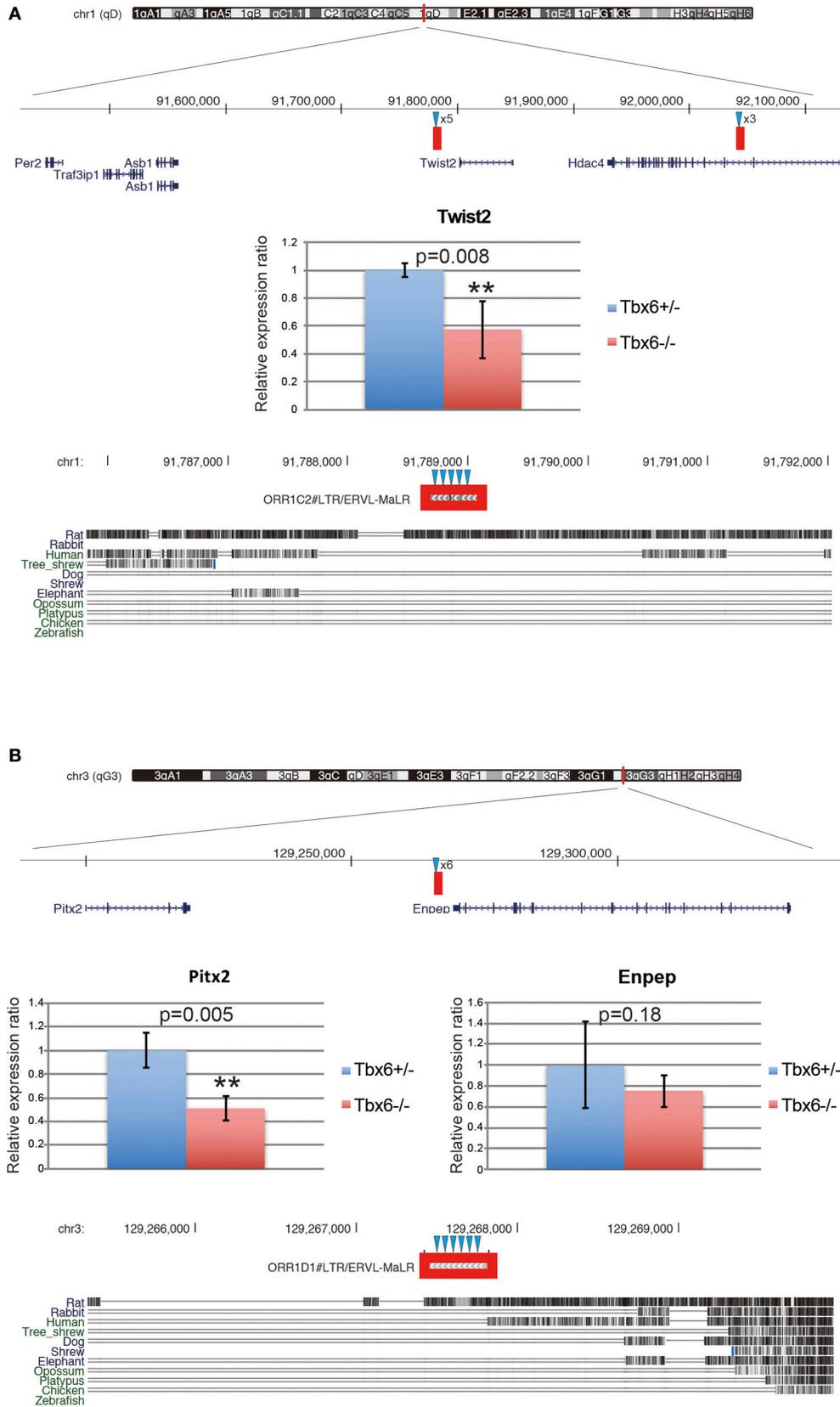


FIGURE 2 | Continued

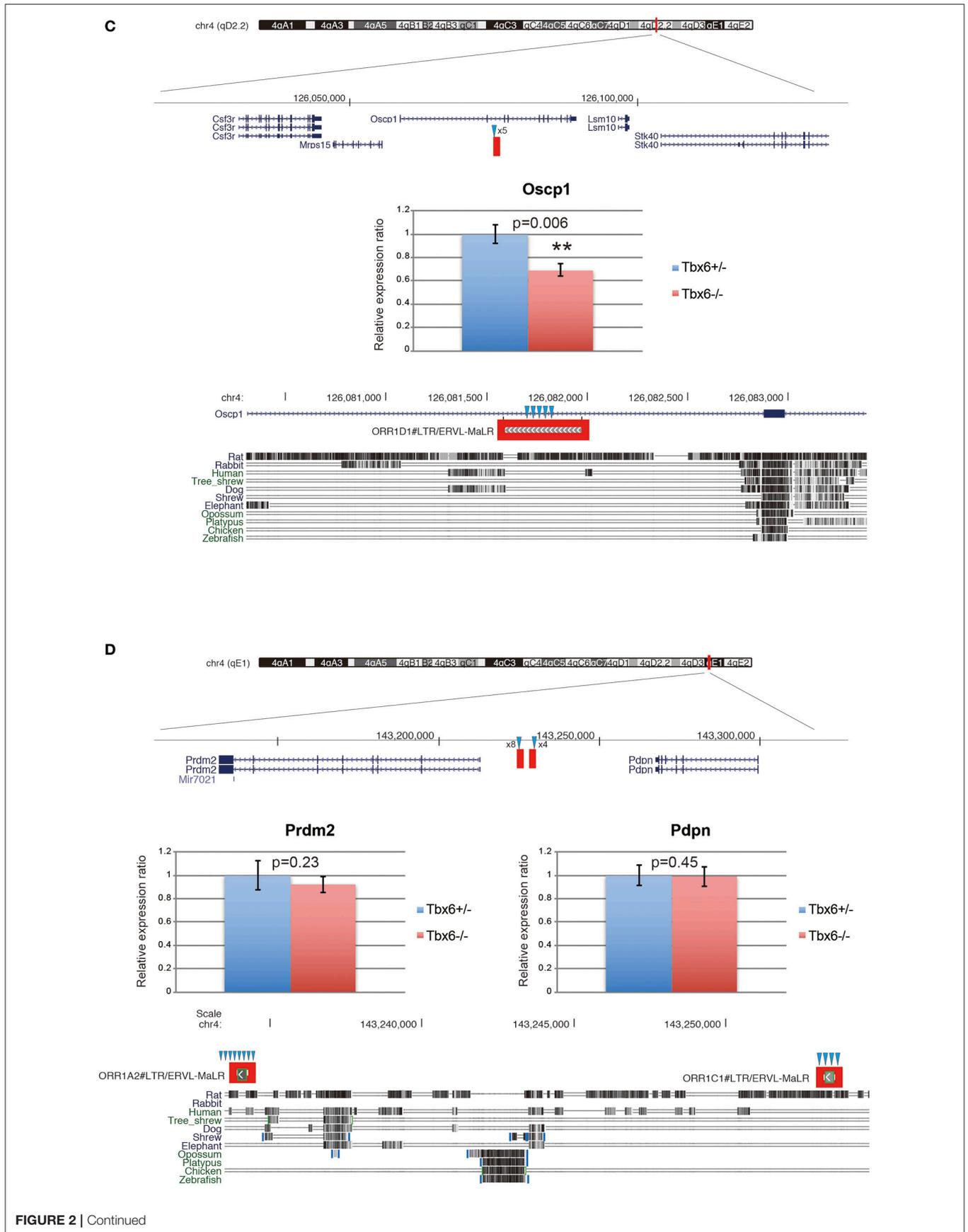


FIGURE 2 | Continued

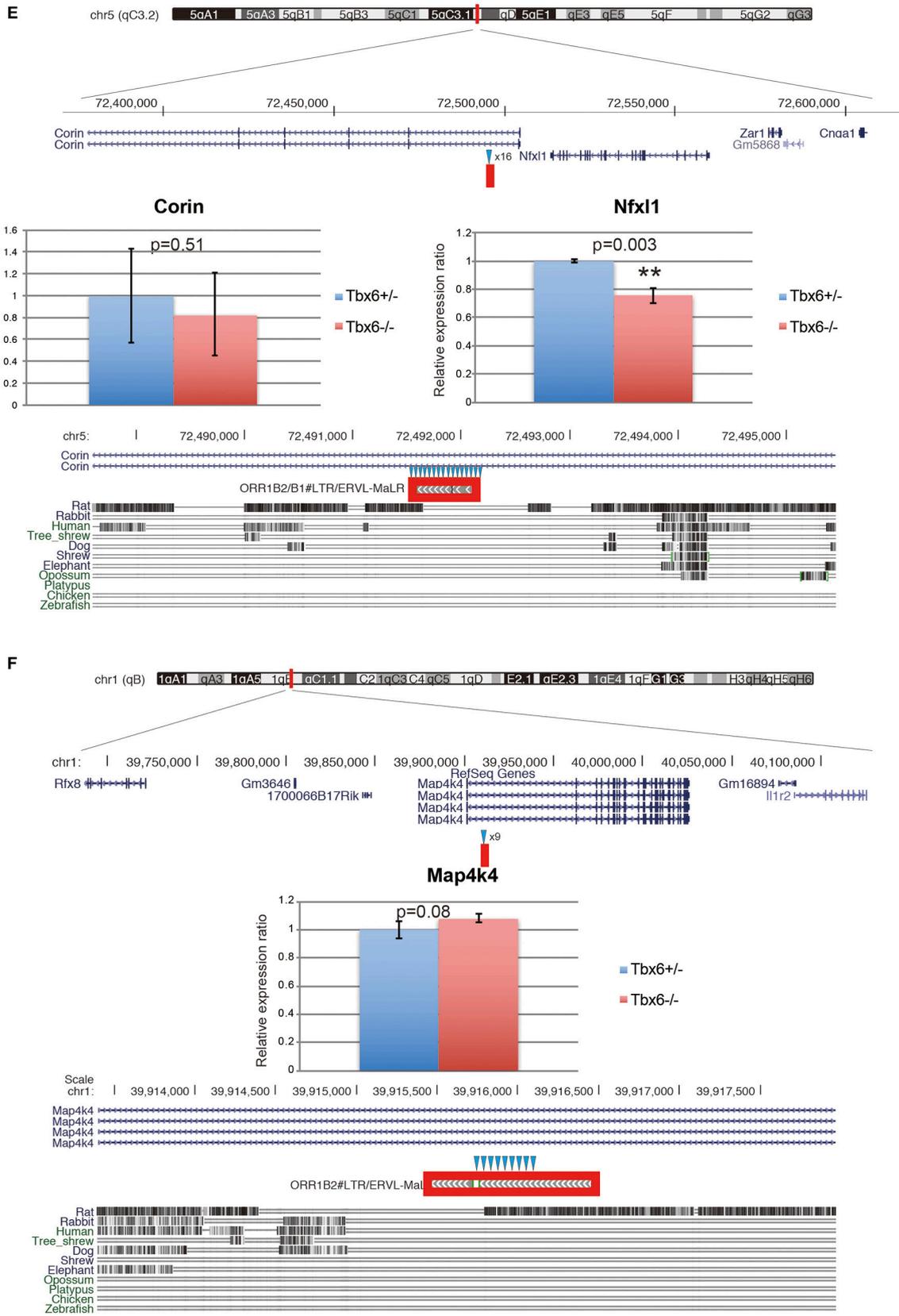


FIGURE 2 | Continued

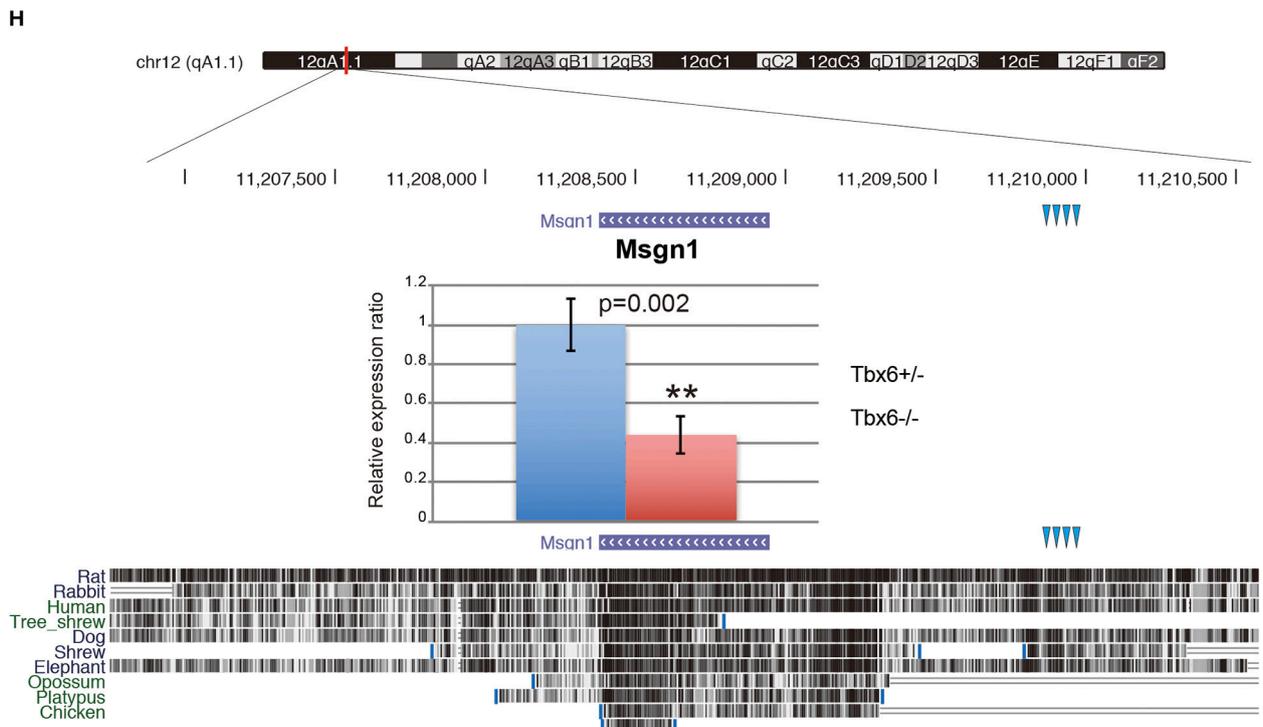
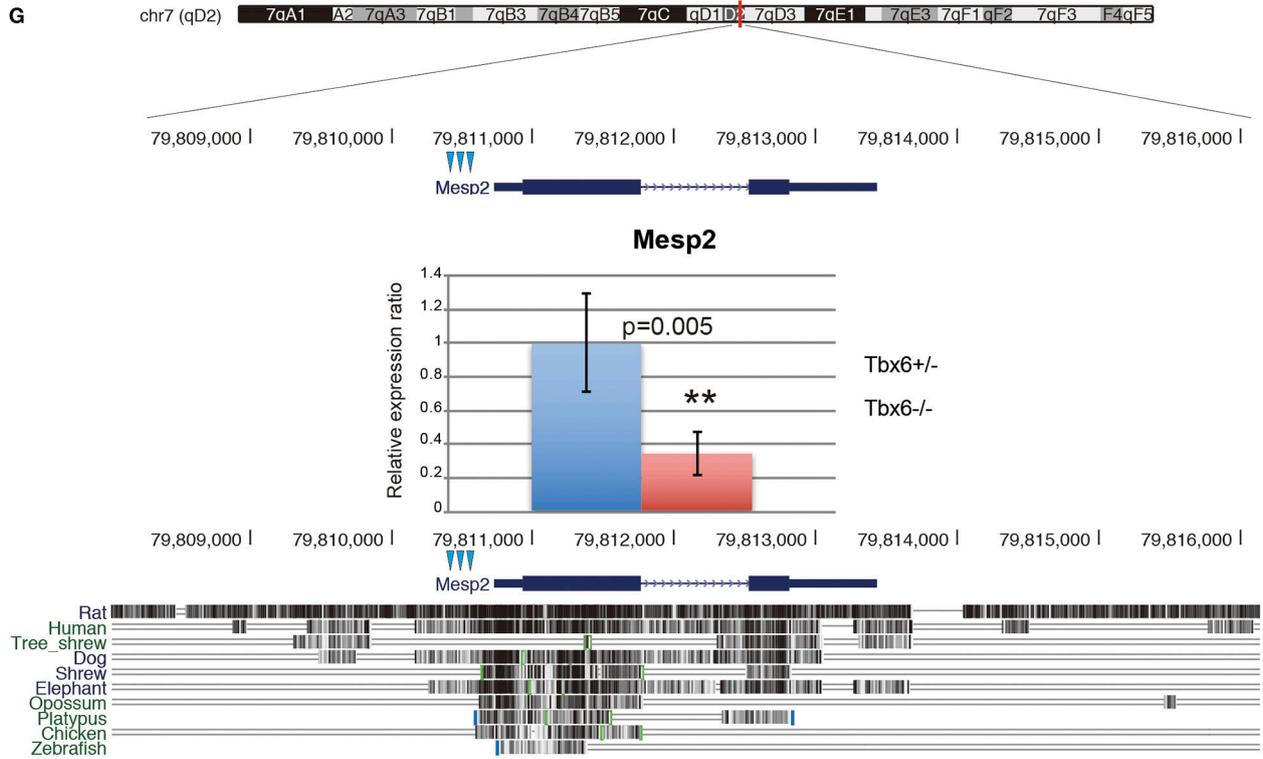
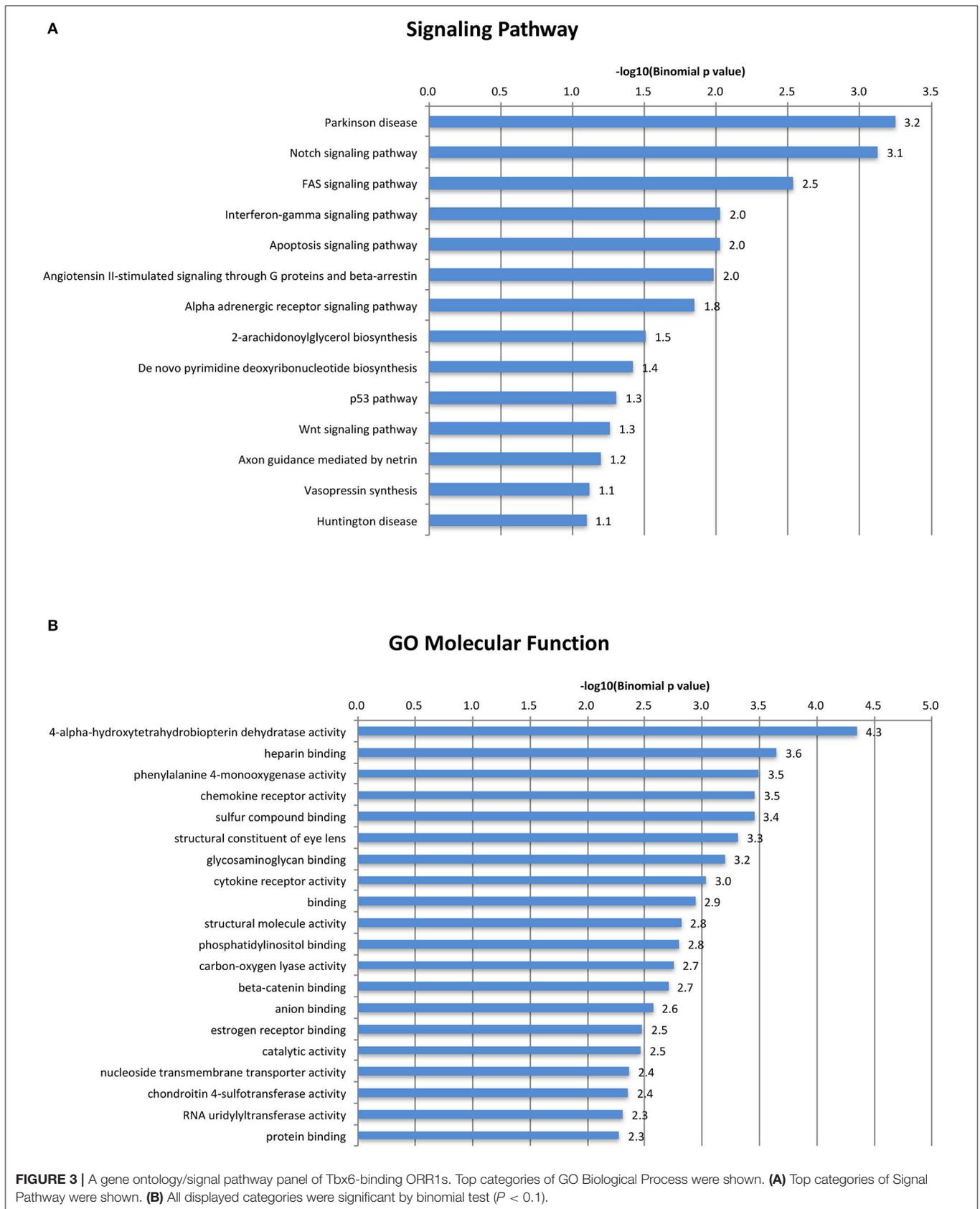


FIGURE 2 | Continued





protocol was approved by the animal welfare committee of National Institute of Health Sciences (NIHS; No.41). Animals had access to a standard chow diet and water *ad libitum* and were housed in a pathogen-free barrier facility with a 12L:12D cycle.

A Tbx6 conditional knockout mouse (Tbx6<sup>fl</sup>) was generated using the ES cell line TT2 and maintained in an ICR background (Yagi et al., 1993). Briefly, exon 3-5, encoding the T-box DNA binding domain of Tbx6, was flanked by a pair of loxP sites and knocked into the *Tbx6* locus by homologous recombination. The PGK-neo selection marker was removed by the FLP-FRT system to obtain Tbx6<sup>fl</sup> mice. For cDNA preparation, embryos (8.0 days post-coitus) were obtained by crossing female CAG-Cre/Tbx6<sup>fl</sup>/+ hybrid heterozygotes onto male Tbx6<sup>fl</sup>/fl homozygotes. Embryos were genotyped by PCR using allantois genomic DNA, and total RNA were prepared using an RNeasy mini kit (QIAGEN). Total RNA was pooled from 5 (Tbx6<sup>+/+</sup>) and 4 (Tbx6<sup>-/-</sup>) 8.0dpc sibling embryos in the same litter. The sequences of primers for real-time RT-PCR were as follows:

*Twist2*\_forward, 5'-TGTCGCTCCCACTAGC-3';  
*Twist2*\_reverse, 5'-TGTCAGGTGCCGAAAGTC-3';  
*Pitx2*\_forward, 5'-GGCAGTCACCCTGGGAAG-3';  
*Pitx2*\_reverse, 5'-GCCGACACTAGTTTGCACACA-3';  
*Enpep*\_forward, 5'-CCTGCTTTACGACCCCCTAC-3';  
*Enpep*\_reverse, 5'-TTAGCCACAAGTCGTCCCAC-3';  
*Oscp1*\_forward, 5'-GACTCTGCCGCTGCTCT-3';  
*Oscp1*\_reverse, 5'-TCGTCCATGAACTTCCTGTTGA-3';  
*Prdm2*\_forward, 5'-GCTTCGAGGACTTCCAGAGG-3';  
*Prdm2*\_reverse, 5'-TGGTTTAGTGCCCGACAC-3';  
*Pdprn*\_forward, 5'-AGGTGCTACTGGAGGGCTTA-3';  
*Pdprn*\_reverse, 5'-GCTGAGGTGGACAGTTCCTC-3';  
*Nfxl1*\_forward, 5'-AGAACCTCCTCAGTTGCTGC-3';  
*Nfxl1*\_reverse, 5'-AAGGGGCATTCACCAGGATG-3';  
*Corin*\_forward, 5'-GATATGTTTACGAAACGGCCC-3';  
*Corin*\_reverse, 5'-CGCTCCTGTCTGCTCTCAAG-3';  
*Map4k4*\_forward, 5'-TTCCGGCCTCTCAAGCCT-3';  
*Map4k4*\_reverse, 5'-TCCCAGACTCCTCACTGGAG-3';  
*Mesp2*\_forward, 5'-ACCCTACACCAGTCCCTGAGAAA-3';  
*Mesp2*\_reverse, 5'-GGTCTGGAGACACAGAAAGACT-3';  
*Msgn1*\_forward, 5'-GCCAGAAAGGCAGCAAAGTC-3';  
*Msgn1*\_reverse, 5'-AGACAGGCGGCAGTAATTC-3';  
*β-actin*\_forward, 5'-CTGTGAGTCGCGTCCA-3';  
*β-actin*\_reverse, 5'-ACGATGGAGGGGAATACAGC-3';

## REFERENCES

- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11. doi: 10.1186/s13100-015-0041-9
- Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* 165, 349–357. doi: 10.1126/science.165.3891.349
- Chapman, D. L., and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene *Tbx6*. *Nature* 391, 695–697.
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351, 1083–1087. doi: 10.1126/science.aad5497
- Criscione, S. W., Theodosakis, N., Micevic, G., Cornish, T. C., Burns, K. H., Neretti, N., et al. (2016). Genome-wide characterization of

Primers were designed using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) tool to amplify 70–150 base pair (bp) fragment separated by at least one intron (>500 bp), except *Msgn1* (single exon gene). PCR reaction was performed using SYBR(R) Premix Ex Taq(TM) II (Takara RR820S) following the manufacturer's protocol, with PCR cycle as follows: 1 cycle of 95°C 30 s, 40 cycles of 95°C 5 s and 60°C 30 s.

## Statistical Analyses

Statistical significance for qPCR was assessed using a two-tailed unpaired Student's *t*-test with a threshold of  $p < 0.1$ .

## ETHICS STATEMENT

The animal facility of the National Institute of Health Sciences was approved by the Japan Health Sciences Foundation since 2008. All animal studies were conducted in accordance with the guidelines approved by the animal welfare committee of the National Institute of Health Sciences (NIHS; No. 41).

## AUTHOR CONTRIBUTIONS

RO conceived of the project. YY, YH, and RO participated in the experimental design. RO performed most analyses. YY produced Tbx6 KO mice and performed EMSA and RT-PCR. RO wrote the manuscript. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by grants from Grant-in-Aid for Scientific Research (C) (26430183) and the Research on Regulatory Science of Pharmaceuticals and Medical Devices from Japan Agency for Medical Research and Development, AMED to RO. We are grateful to Hiroyuki Takeda (University of Tokyo) for providing *zTbx6* cDNA clones, and to Satoshi Kitajima and Eriko Ikeno for technical assistance in generating Tbx6 KO mouse.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fchem.2017.00034/full#supplementary-material>

human L1 antisense promoter-driven transcripts. *BMC Genomics* 17:463. doi: 10.1186/s12864-016-2800-5

- Deininger, P. L., Moran, J. V., Batzer, M. A., and Kazazian, H. H. Jr. (2003). Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* 13, 651–658. doi: 10.1016/j.gde.2003.10.013
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- Downen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi: 10.1016/j.cell.2014.09.030
- Druker, R., Bruxner, T. J., Lehrbach, N. J., and Whitelaw, E. (2004). Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res.* 32, 5800–5808. doi: 10.1093/nar/gkh914

- Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., et al. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12127–12132. doi: 10.1073/pnas.0902925106
- Dupressoir, A., Vernochet, C., Harper, F., Guegan, J., Dessen, P., Pierron, G., et al. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1164–E1173. doi: 10.1073/pnas.1112304108
- Evsikov, A. V., de Vries, W. N., Peaston, A. E., Radford, E. E., Fancher, K. S., Chen, F. H., et al. (2004). Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* 105, 240–250. doi: 10.1159/000078195
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/nrg2337
- Greer, E. L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–357. doi: 10.1038/nrg3173
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Irie, M., Yoshikawa, M., Ono, R., Iwafune, H., Furuse, T., Yamada, I., et al. (2015). Cognitive Function Related to the Sirh11/Zcchc16 Gene Acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* 11:e1005521. doi: 10.1371/journal.pgen.1005521
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* 68, 651–654. doi: 10.1095/biolreprod.102.007906
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. doi: 10.1038/nrg3030
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819. doi: 10.1038/nature04338
- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., et al. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927–931. doi: 10.1038/nature08858
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* 36, 344–355. doi: 10.1073/pnas.36.6.344
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi: 10.1038/35001608
- Morgan, H. D., Sutherland, H. G., Martin, D. I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318. doi: 10.1038/15490
- Nakaya, Y., Koshi, K., Nakagawa, S., Hashizume, K., and Miyazawa, T. (2013). Fematin-1 is involved in fetomaternal cell-to-cell fusion in Bovinae placenta and has contributed to diversity of ruminant placentation. *J. Virol.* 87, 10563–10572. doi: 10.1128/JVI.01398-13
- Naruse, M., Ono, R., Irie, M., Nakamura, K., Furuse, T., Hino, T., et al. (2014). Sirh7/Ldoc1 knockout mice exhibit placental P4 overproduction and delayed parturition. *Development* 141, 4763–4771. doi: 10.1242/dev.114520
- Ono, R., Ishii, M., Fujihara, Y., Kitazawa, M., Usami, T., Kaneko-Ishino, T., et al. (2015). Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes. *Sci. Rep.* 5:12281. doi: 10.1038/srep12281
- Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., et al. (2001). A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. *Genomics* 73, 232–237. doi: 10.1006/geno.2001.6494
- Ono, R., Kuroki, Y., Naruse, M., Ishii, M., Iwasaki, S., Toyoda, A., et al. (2011). Identification of tammar wallaby SIRH12, derived from a marsupial-specific retrotransposition event. *DNA Res.* 18, 211–219. doi: 10.1093/dnares/dsr012
- Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., et al. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38, 101–106. doi: 10.1038/ng1699
- Ono, R., Shiura, H., Aburatani, H., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2003). Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. *Genome Res.* 13, 1696–1705. doi: 10.1101/gr.906803
- Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* 70, 81–120. doi: 10.1146/annurev.biochem.70.1.81
- Rupp, R. A., Snider, L., Weintraub, H. (1994). Xenopus embryos regulate the nuclear localization of XMyoD. *Genes Dev* 8, 1311–1323. doi: 10.1101/gad.8.11.1311
- Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., et al. (2008). Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4220–4225. doi: 10.1073/pnas.0709398105
- Schmidl, C., Klug, M., Boeld, T. J., Andreesen, R., Hoffmann, P., Edinger, M., et al. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* 19, 1165–1174. doi: 10.1101/gr.091470.109
- Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., et al. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat. Genet.* 40, 243–248. doi: 10.1038/ng.2007.51
- Smit, A. F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21, 1863–1872. doi: 10.1093/nar/21.8.1863
- Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E., et al. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* 470, 394–398. doi: 10.1038/nature09729
- Terasaki, H., Murakami, R., Yasuhiko, Y., Shin, I. T., Kohara, Y., Saga, Y., et al. (2006). Transgenic analysis of the medaka mesp-b enhancer in somitogenesis. *Dev. Growth Differ.* 48, 153–168. doi: 10.1111/j.1440-169X.2006.00853.x
- Vasicek, T. J., Zeng, L., Guan, X. J., Zhang, T., Costantini, F., and Tilghman, S. M. (1997). Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics* 147, 777–786.
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi: 10.1038/nature01262
- White, P. H., and Chapman, D. L. (2005). Dll1 is a downstream target of Tbx6 in the paraxial mesoderm. *Genesis* 42, 193–202. doi: 10.1002/gene.20140
- Yagi, T., Tokunaga, T., Furuta, Y., Nada, S., Yoshida, M., Tsukada, T., et al. (1993). A novel ES cell line, TT2, with high germline-differentiating potency. *Anal. Biochem.* 214, 70–76. doi: 10.1006/abio.1993.1458
- Yasuhiko, Y., Haraguchi, S., Kitajima, S., Takahashi, Y., Kanno, J., and Saga, Y. (2006). Tbx6-mediated Notch signaling controls somite-specific Mesp2 expression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3651–3656. doi: 10.1073/pnas.0508238103
- Yasuhiko, Y., Kitajima, S., Takahashi, Y., Oginuma, M., Kagiwada, H., Kanno, J., et al. (2008). Functional importance of evolutionally conserved Tbx6 binding sites in the presomitic mesoderm-specific enhancer of Mesp2. *Development* 135, 3511–3519. doi: 10.1242/dev.027144
- Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340. doi: 10.1016/S0168-9525(97)01181-5

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Yasuhiko, Hirabayashi and Ono. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.