Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes

Ryuichi Ono1,2,*, Masayuki Ishii2,*, Yoshitaka Fujihara3, Moe Kitazawa2, Takako Usami4, Tomoko Kaneko-Ishino5, Jun Kanno1, Masahito Ikawa3 & Fumitoshi Ishino2,6

The CRISPR/Cas system efficiently introduces double strand breaks (DSBs) at a genomic locus specified by a single guide RNA (sgRNA). The DSBs are subsequently repaired through non-homologous end joining (NHEJ) or homologous recombination (HR). Here, we demonstrate that DSBs introduced into mouse zygotes by the CRISPR/Cas system are repaired by the capture of DNA sequences deriving from retrotransposons, genomic DNA, mRNA and sgRNA. Among 93 mice analysed, 57 carried mutant alleles and 22 of them had long de novo insertion(s) at DSB-introduced sites; two were spliced mRNAs of Pcnt and Inadl without introns, indicating the involvement of reverse transcription (RT). Fifteen alleles included retrotransposons, mRNAs, and other sequences without evidence of RT. Two others were sgRNAs with one containing T7 promoter-derived sequence suggestive of a PCR product as its origin. In conclusion, RT-product-mediated DSB repair (RMDR) and non-RMDR repair were identified in the mouse zygote. We also confirmed that both RMDR and non-RMDR take place in CRISPR/Cas transfected NIH-3T3 cells. Finally, as two de novo MuERV-L insertions in C57BL/6 mice were shown to have characteristic features of RMDR in natural conditions, we hypothesize that RMDR contributes to the emergence of novel DNA sequences in the course of evolution.

Whole genome sequencing of a number of different mammalian species has established that approximately 50% of the mammalian genome is derived from transposable elements1–3. Retrotransposons, which mobilize via an RNA intermediate by a copy-and-paste mechanism, comprise the majority of mammalian transposable elements, whereas DNA transposons, which move via a cut-and-paste mechanism, comprise a relatively small fraction, having accumulated mutations that render them immobile4.
Retrotransposons can be subdivided into two classes, long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. LTR retrotransposons contain two LTRs, they encode the proteins Gag and Pol with activities similar to those of simple retroviruses such as protease, reverse transcriptase and integrase activities, but they lack an envelope (Env) gene. The LTRs are direct repeats that contain endonuclease and reverse transcriptase activities. The endonuclease and reverse transcriptase activities of non-LTR retrotransposons also have the ability to mobilize other non-autonomous short interspersed elements (SINEs) and certain classes of non-coding RNAs, which can result in the formation of processed pseudogenes.

Retrotransposons continue to sculpt mammalian genomes and behave as insertional mutagens, either by disrupting exons or by insertion into introns, leading to mis-splicing. Hence, retrotransposons occasionally have deleterious effects on host genes and thus organisms. However, a growing body of evidence suggests that retrotransposons and retrotransposon-derived genes have also acquired functions essential for host survival during mammalian evolution including placental formation, neurogenesis and gene regulation. Recent studies have shown that a large number of different retrotransposon families are highly transcribed in the mouse zygote, and in fact, they produce cDNAs by reverse transcription (RT).

Results

DSBs were introduced by CRISPR/Cas into mouse zygotes. Single guide RNAs (sgRNAs) were designed for each of the eight target genomic loci, and the DSB induction efficiency was validated with an EGxxFP system. We confirmed that all the sgRNAs were able to induce fluorescent cells at more than 30% efficiency. It was previously reported that more than 30% efficiency obtained with an EGxxFP system allows for the stable generation of mutant mice by injecting the animals into fertilized eggs. In this report, DSBs were introduced into 8 genomic loci of the mouse zygote, including Peg10-ORF1 and Peg10-ORF2 regions and the Cxx1a, Cxx1b, Rgag1, Rsp6a, Spaca5 and Ddx3y genes. We analysed the sequences of the DSB-induced sites to determine whether RMDR was at work. We also introduced DSBs into NIH-3T3 cells to assess the universality of the phenomenon, doing so with or without an RT-inhibitor to determine whether RMDR could be inhibited.

DSBs were repaired by the capture of retrotransposon sequences. Detailed characterization of the de novo insertions at the target DSB sites in the Peg10-ORF2 coding region revealed that two of the animals (Peg10-ORF2-#8 and Peg10-ORF2-#18) had 327-bp and 357-bp insertions of the murine endogenous retrovirus cDNA at the site of DSBs. MuERV-L, also known as the MERVL or Env4 Pol protein coding region (Fig. 2a,c). MuERV-L is an endogenous retrovirus that is one of the most abundant transcripts in the 2-cell stage embryo. In each case, there were small overlapping nucleotides called “microhomologies” between the inserted retrotransposon and the DSB-induced target site, and truncations of both the
5′ and 3′ regions of the inserted retrotransposon including the LTRs were present, suggest that these LTR retrotransposons had not been integrated by typical replicative retrotransposition. Furthermore, the Peg10-ORF2-#17 animal was found to have a 950-bp insertion of a partial internal region and a truncated LTR of the retrovirus-like element MaLR, which is the second most abundant retrotransposon transcript in the 2-cell stage mouse embryo. To the best of our knowledge, this is the first direct evidence of the introduction of MuERV-L and MaLR retrotransposons at a specifically desired genomic locus in mouse zygotes. Because MaLR does not encode any known protein and its means of propagation in the genome is unknown, it has been suggested that the RT activity of MuERV-L might be the means of its propagation. Insertions of partial retrotransposon sequences with microhomologies were also observed at all of the target DSB sites introduced (Fig. 2d–f, Supplementary Fig. 2a–c,e–g–j). Although the DSB-induced loci were different, the same types of endogenous retroviruses were inserted at each locus, indicating that the capture of retrotransposon sequences may occur at any DSB site in the mouse zygote. Furthermore, there was a case in which an allele had multiple retrotransposon insertions at the same DSB site (Fig. 2e). Each of the junction sequences between the retrotransposons has 1–5 bp of complete overlapping microhomology. DSBs were repaired by the capture of RT-mediated cDNA. In addition to retrotransposon sequences there were also mRNA sequence insertions at the DSB-induced loci. De novo insertions at Peg10 ORF2, Peg10 ORF1 and Spac5 included mRNA sequence-derived partial sequences of the Pcnt (Pericentrin) gene (Fig. 3a), Inadl (Inad-like) gene (Fig. 3b), Cpd (carboxypeptidase D) gene (Supplementary Fig. 2b), Tpm3 (tropomyosin 3, gamma) gene (Supplementary Fig. 2c), Zfp609 (zinc finger protein 609) gene (Supplementary Fig. 2d), Actr2 (ARP2 actin-related protein 2) gene (Supplementary Fig. 2j) and Peg10 gene (Fig. 2a, Supplementary Fig. 2a). Expression of the Pcnt, Inadl, Cpd, Tpm3, Zfp609, Actr2 and Peg10 genes was confirmed at the 2- to 16-cell embryonic stages, and the inserted partial Pcnt and Inadl sequences correspond to the 6803–7599 bp and 1527–1836 bp regions of the full length Pcnt and Inadl mRNAs, respectively, skipping Pcnt intron 30–32 and Inadl intron 11–14, demonstrating that
these partial *Pcnt* and *Inadl* insertions are mediated by RMDR. The sequences flanking the insertions have no polyA tails for any of the captured genes, but short microhomology is present for *Pcnt*, *Cpd*, *Tpm3*, *Zfp609*, *Actr2* and *Peg10*, supporting the notion that the cDNA gene formation is not mediated by conventional TPRT pathways but rather by RMDR (Fig. 4c,d).

The mouse with the *Inadl* insertion was produced in the process of obtaining knock-in (KI) mice with a point mutation in the CCHC zinc finger domain of Peg10 ORF1 by CRISPR/Cas co-injection with DNA oligos. The DNA oligos have 53 bp and 80 bp homologous regions and a T to G point mutation near the DSB site. As a result, 5 pups (#53, #54, #55, #61 and #62 in Table S1) were born with the KI
allele (mosaicism) and 5 pups (#52, #55, #61, #62 and #63 in Table S1) had captured DNA sequences (the sequences of #62 and #63 were not determined). In mouse #55, the KI allele and the captured allele were independent (mosaicism); however, in mouse #61, the 5′ end of the DSB-induced site was repaired...
Figure 4. Possible mechanisms of RMDR, and RMDR under natural conditions. DSBs (the orange triangles) induced by CRISPR/CAS (blue sphere) are shown undergoing repair by NHEJ (a), non-RT-product-mediated DSB repair (non-RMDR) (b), RMDR with pre-existing cDNA (c) and RMDR with direct RT (d). Most of the DSBs induced by CRISPR/Cas are repaired by NHEJ (a), while certain DSBs are repaired by the capture of other sequences (b–d). (c) A pre-existing cDNA (red bar) generated by RT anneals with both DNA ends of a DSB site, which is repaired (RMDR with double microhomologies). (d) mRNA anneals with one DSB end with microhomology, and cDNA is synthesized by RT machinery. Two murine-specific truncated MuERV-L insertions were identified by comparing rodent genomes. Pre-integration sequences (indicated in blue text) were deduced from other available rodent genomes. (e) Murine-specific truncated MuERV-L (52824976–52825534 bp; chromosome 9) was integrated with two 2-bp microhomologies at both the 5' and 3' ends. (f) C57BL/6 strain-specific truncated MuERV-L (191922408–191922914 bp; chromosome 1) was integrated with a 27-bp microhomology (6-bp mismatches and a 5-bp insertion).
by RMDR (Inadl mRNA) and the 3′ end of the DSB-induced site was repaired by HR at the same allele, indicating that RMDR and HR might alternatively repair DSBs (Fig. 3b). We also tried to produce KI mice in the Peg10 ORF2 DSG protease domain, however no KI mice were obtained in this experiment.

**DSBs repaired by the capture of injected DNA templates.** Two Peg10-ORF1-sgRNA sequences were captured at DSB sites (Fig. 3c,d) by injection of *in vitro* transcribed sgRNA into mouse zygotes. One with T7 promoter sequence at its 5′ end was derived from the PCR product for *in vitro* transcription, which was not eliminated through the purification of *in vitro* transcribed sgRNA, demonstrating that non-RT-product-mediated DSB repair (non-RMDR) is also at work (Fig. 3c). Another captured sgRNA sequence was derived from PCR product (non-RMDR) or RT-mediated cDNA (RMDR) (Fig. 3d).

The global transcription level is very low at the one–cell stage until the two–cell stage [33,34]. Therefore, the transcription from CRISPR-Cas DNA plasmid may be very low compared with injected *in vitro* transcribed RNA, making RNA injection into the cytoplasm more efficient than plasmid injection [35]. High DSB activity might be the reason why the injected DNA templates were captured only in mice with DSBs induced by CRISPR/Cas RNA injection.

**Possible mechanism of RMDR.** The data show that RMDR is at work in mouse zygotes, at least in the case of two spliced mRNAs (Fig. 3a,b). At the same time, non-RMDR is also at work in mouse zygotes, at least in one of the CRISPR/Cas RNA injected mice (Fig. 3c). These data suggest that DNA fragments in the nucleus, whether generated by RT (RMDR) or not (non-RMDR), are captured at DSB sites. In all other cases, we cannot determine whether the captured sequences were derived from RT. However, RMDR is more likely than non-RMDR because 20% of the captured sequences are derived from exons compared to only 1% of exon sequences in the whole genome. The enrichment of exons here favours the idea that they are of cDNA origin. Furthermore, *in silico* analysis shows that MuERV-L and MaLR sequences occupy only 0.04% and 0.01% of the mouse genome, respectively; however, 20% and 33.3% of captured sequences in DSB-induced mice are MuERV-L and MaLR, respectively, suggesting that the capture of MuERV-L, MaLR and mRNA sequences is most likely mediated by RMDR. Therefore, it is highly plausible that DSBs are repaired not only by classical NHEJ and HR but also by RMDR, in mouse zygotes.

Previous studies demonstrated that MuERV-L exhibits two-cell specific expression, suggesting that DSB repair by MuERV-L insertion tends to occur at the 2-cell embryonic stage or later if the capture of MuERV-L at DSB sites is in fact mediated by the RT activity of MuERV-L [26–29]. In this study, the amount of the inserted MuERV-L PCR band was mostly less than 30% of the total PCR products in each mouse, suggesting that RMDR occurred in a single allele of an individual cell in 2-cell or later stage embryos, whereas the MuERV-L insertion Cxx1b→#24 apparently occurred at the one cell embryo stage because there is no allele other than the MuERV-L insertion allele (Figs 1c,2e). The coincidence between these two events, MuERV-L and MaLR being two of the most abundantly expressed transcripts at the 2-cell stage, while MuERV-L and MaLR are two of the most frequent insertions at DSB-induced loci in this study, also indicates that the capture of MuERV-L and MaLR is mediated by transcriptional level-dependent RMDR rather than non-RMDR (Fig. 3e). There are previous reports similarly related to RMDR. They involve a DSB repair mechanism affected by the endonuclease-independent (ENi) retrotransposition of an artificial human L1 reporter [52,56,57]. These ENi retrotranspositional features include a lack of target site duplications (TSDs) and frequent truncations at both the 5′ and 3′ ends of the artificial L1 reporter in NHEJ-deficient CHO cells, but the ENi retrotranspositions have no microhomology [52,56]. Because RMDR is often associated with microhomology and may be mediated by an LTR-type retrotransposon, i.e., MuERV-L, the mechanism of RMDR might be different from that of ENi retrotransposition mediated by L1, a non-LTR retrotransposon. It was previously reported that mRNA in 2-cell stage embryo undergoes RT in mice [27,58]. Therefore, we propose RMDR with pre-existing cDNA (Fig. 4c) and RMDR with direct RT (Fig. 4d), although the detailed mechanisms of how RNA is reverse transcribed are unknown and the possibility of non-RMDR cannot be ruled out (Fig. 4b). If a cDNA becomes annealed with both of the DSB DNA ends via microhomologies, the DSB is repaired by the filling in of the missing base pairs (RMDR with a double microhomology) (Fig. 4c, Supplementary Fig. 2g). If a cDNA is annealed with only one of two DSB DNA ends, the cDNA and the other DSB end are repaired by NHEJ (RMDR with a single microhomology) (Supplementary Fig. 3a, Figs 2b–d,f,3a,d, Supplementary Fig. 2d,e,h,i). Most of the junction sequences in the multiple retrotransposons and mRNA insertions have one to five microhomologous nucleotide sequences (Fig. 2e, Supplementary Fig. 2a,j), indicating that these multiple insertions were mediated by sequential-RMDR (s-RMDR) (Supplementary Fig. 3b,c). Although we were unable to firmly establish the mechanism without performing further experiments, there does exist a capture process of retrotransposons and/or mRNA sequences at DSB sites in mouse zygote RMDR.

**RMDR could be inhibited by an RT inhibitor in cultured cells.** Because two cDNAs with skipped introns were inserted into DSB-induced sites, it is clear that at least 2 of the 30 captured sequences were mediated by RMDR. To assess the possibility that the other insertions were also from cDNA by RT, we introduced DSBs into an NIH-3T3 cell line by transfecting a CRISPR/Cas plasmid (pX330–Peg10–ORF1, including both sgRNA and hCas9 genes) and a pTracer-CMV/Bsd plasmid (including the Blasticidin S resistance gene), and performing Blasticidin S drug selection, with or without the RT inhibitor.
azidothymidine (AZT), which is known to inhibit human and mouse L1 retrotransposition in HeLa cells.59

Extra unidentified PCR products larger than the expected length were observed in CRISPR-Cas transfected cells regardless of the presence of the RT-inhibitor. The ratio of these extra products was reduced by the addition of AZT to the culture medium (Supplementary Fig. 4a–c). Sequencing analysis of the extra PCR bands revealed that insertions of mRNA, retrotransposons, and transfected plasmids (both pX330-Peg10-ORF1 and pTracer-CMV/Bsd) were observed in the absence of AZT (Supplementary Fig. 5), whereas insertions of only plasmid DNA and genomic DNA sequences were observed with AZT (Supplementary Fig. 6). The capture of plasmids (both pX330-Peg10-ORF1 and pTracer-CMV/Bsd) was observed in 35.4% (without AZT) and 75% (with AZT), and the inserted regions include the plasmid vector backbone (not the gene body), suggesting that it is mediated by non-RMDR (Supplementary Fig. 7). One of the cDNA insertion sequences without AZT was the Tubulin folding cofactor B (Tubc) gene, skipping introns 1–3, demonstrating that RMDR occurs not only in mouse zygotes but also in cultured cells (Supplementary Fig. 5f). Approximately 53% of the captured sequences were occupied with MuERV-L and MaLR retrotransposons in mouse zygotes, while the retrotransposons in the NIH-3T3 cell line included L1 (12.9%) and ERV1/2 (12.9%) (Fig. 3e, Supplementary Fig. 7a). This difference in the species of incorporated retrotransposons might reflect their cell type-specific expression levels. The capture of plasmid DNA was not influenced by AZT, suggesting a non-RMDR mechanism. In any case, DNA fragments in the nuclei, whether generated by RT (RMDR) or not (non-RMDR), are captured at DSB sites. The reason why plasmid DNA is not captured in mouse zygotes may be that the zygote has sufficiently high RT activity to produce an excessive amount of cDNA compared to the exogenous plasmid DNA.

RMDR could be functional under natural conditions. Finally, to determine whether RMDR occurs under natural conditions, we screened potential MuERV-L insertions in the mouse genome. As it is necessary to predict the pre-integration DNA sequence to identify microhomologies, two murine-specific MuERV-L insertions with both 5′ and 3′ truncations were identified by comparative analysis of rodent genomes. One insertion was a murine-specific truncated Gag-Pol region of MuERV-L with two 2-bp overlapping microhomologies at both DSB ends (Fig. 4e). The other insertion was a C57BL/6 mice strain-specific truncated Gag region of MuERV-L with a 27-bp microhomology (6-bp mismatches and 5-bp insertion) (Fig. 4f). Although this insertion has 10-bp TSDs, these TSDs were not generated by endogenous MuERV-L integrase activity, but by other DSB events. This is because MuERV-L retrotranspositions cause random 5bp (rarely 6bp) TSDs when they retrotranspose (Supplementary Table 3). As these insertional features are identical with those of RMDR, we hypothesize that RMDR contributes to the emergence of novel DNA sequences in the course of evolution.

Discussion
It is perhaps one of the greatest mysteries of biological evolution how retrotransposons, endogenous retroviruses (ERVs) and their remnant DNA sequences have come to occupy one half of the mammalian genome. Recently, these sequences have drawn attention as one of the ostensible driving forces of genomic evolution. In this report, we demonstrated that DSBs introduced into mouse zygotes by the CRISPR/Cas system were repaired by the capture of retrotransposons and other genomic DNA, with evidence in some cases of reverse-transcribed mRNA sequences and even exogenous single guide RNA (sgRNA) sequences at DSB sites. RMDR in the mouse zygote was confirmed in at least 2% of the CRISPR-Cas injected mice in this study. Moreover, three alleles were shown to generate novel long-range fusion proteins between Peg10-ORF2 and truncated MuERV-L (Peg10 ORF2-#18) and between Peg10-ORF2 and truncated Pcnt (Peg10 ORF2-#2) in the DSB-introduced mice (Supplementary Fig. 8). Therefore, DSB repair by CRISPR-Cas injection into mouse zygotes has the potential to generate novel gene sequences. In nature, DSBs result from both exogenous insults (e.g., reactive oxygen species, irradiation, chemical agents, ultraviolet light) and endogenous cellular events (e.g., transposition, meiotic double strand break formation). Apart from its frequency, extrapolation of our findings here on the consequences of DSBs leads us to conclude that RMDR may contribute to the generation of novel gene sequences under certain natural conditions. In fact, by comparing rodent genomes, we found that two de novo MuERV-L insertions in wild-type C57BL/6 mice show features characteristic of RMDR. Although we could not exclude the possibility of DNA recombination, we consider that this finding is compatible with DSB repair by the capture of retrotransposon sequences occurring in natural conditions (Fig. 4e,f). Thus, we propose the hypothesis that RMDR has contributed to the evolution of the mammalian genome.

Methods
Animals. All animal studies were conducted in accordance with the guidelines approved by the animal care committee of Tokyo Medical Dental University, Osaka University and National Institute of Health Sciences (NIHS). Animals were allowed access to a standard chow diet and water ad libitum and were housed in a pathogen-free barrier facility with a 12L:12D cycle.

Plasmid preparation. To construct the pCAG-EGxxFP validation plasmid, the N-terminal and C-terminal EGFP coding regions were PCR-amplified and placed under a ubiquitous CAG promoter.
with the multicloning sites (MCS) BamHI, NheI, PstI, SalI, EcoRI, and EcoRV. The ~500 bp genomic fragments containing the sgRNA target sequence were PCR-amplified and placed in the MCS of the pCAG-EGxxFP validation plasmid. The plasmids expressing hCas9 and sgRNA were prepared by ligating oligos into the BbsI site of pX330 (http://www.addgene.org/42230/). The 20bp sgRNA recognition sequences are shown below.

- **Peg10-ORF1-sgRNA3** (5’- TGCTCTCTCTGTGGCCATAGG-3’)
- **Peg10-ORF2-sgRNA4** (5’- GGTGAACCTGGTTCATCCATG-3’)
- **Cxx1a-b-sgRNA** (5’- ACAGGATTTTGGTTCCCGCCG-3’)
- **Rgag1-sgRNA** (5’- TGGTGTGGATGTCACTCCAAGAGG-3’)
- **Ddx3y-sgRNA** (5’- GGCTGGACCTCTGTGGCCAG-3’)
- **RsgRNA** (5’- GCATGAAGGTCTGCAGCATTG-3’)
- **Spaca5-sgRNA** (5’- GCATGAAGGTCTGCAGCATTG-3’)
- **Cxx1b/F1-ORF1-sgRNA4** (5’- GTCCGAGCTATGATTGATTC-3’)

The oligo DNAs for co-injection with the plasmid into mouse zygotes are shown below.

#1: AACCTACAGTTCTGTCCCGAAACATTCTCTCACCCACAAGATTTGGAAACATATAAACGCGTAAACTCTCGTATGACAGATTGACGTGACGAAGT

#2: CTGCCAAGAGAAAGAGAGAGAGCCGAAAATGAAATGCTGTCTCTACTCGGAGCATGGAGGCGACATTCGGCCAGCGGAAGAC

HEK293T transfection and EGxxFP system. Five hundred ng of the pCAG-EGxxFP-target were mixed with 500 ng of pX330 with/without the sgRNA sequences and then introduced into 4 × 10⁶ HEK293T cells/well in a six well plate using Lipofectamine LTX (Life Technologies). The ratio of EGFP fluorescence positive cells/all cells (Hoechst 33342 positive nucleus) was monitored using a fluorescence microscopy EVOS cell counting system (Life Technologies) 48 hrs after transfection.

One-cell Embryo Injection. B6D2F1 and C57BL/6J female mice were superovulated and IVF was carried out using B6D2F1 and C57BL/6J male mice, respectively. pX330 plasmids with or without oligo DNA to generate knock in mice or mutant mice were injected into the pronucleus of fertilized eggs at the indicated concentrations. hCas9 mRNA and Peg10-ORF1-sgRNA were injected into the cytoplasm of fertilized eggs at the indicated concentration. The eggs were cultivated in KSOM overnight, then transferred into the oviducts of pseudopregnant ICR females.

Retrotransposition analysis. Identification of RMDR alleles was performed by the BLASTN program from the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/) and CENSOR program from the
Introduction of DSBs into NIH-3T3 cells with or without the RT-inhibitor. Two μg of the pX330 with/without Peg10-ORF1 sgRNA sequences were mixed with 500 ng of pTracer-CMV/Bsd and then introduced into 2 × 10⁵ NIH-3T3 cells/well in a six well plate using Lipofectamine LTX (Life Technologies). 24 hour after transfection, cells separated and cultured under two conditions for 2 days, one containing 10μg/mL Blasticidin S (Life Technologies) and the other 10μg/mL Blasticidin S and 50μg/mL Azithromycin (AZT) (Sigma). Five days after transfection, cells were collected and genomic DNA was extracted. Subsequent PCR products with 1500 and 15 bp internal markers were resolved and quantified by using the Agilent DNA 1000 kit (Agilent Technologies).

Identification of natural RMDR alleles. Identification of truncated MuERV-L sequences was performed using the BLASTN program (http://www.ncbi.nlm.nih.gov) against the mouse genome using full length MuERV-L (GenBank ID: Y12713) as a query. Among the truncated MuERV-L sequences, two MuERV-L sequences were identified as a murine specific insertion by comparing the sequences with other rodent genomes.

References
1. Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921, doi: 10.1038/35057062 (2001).
2. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562, doi: 10.1038/ nature01262 (2002).
3. Lindblad-Toh, K. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819, doi: 10.1038/nature04338 (2005).
4. Deininger, P. L., Moran, J. V., Batzer, M. A. & Kazazian, H. H., Jr. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13, 651–658 (2003).
5. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12, 615–627, doi: 10.1038/nrg3031 (2010).
6. Kajikawa, M. & Okada, N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111, 433–444 (2002).
7. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35, 41–48, doi: 10.1038/ng1223 (2003).
8. Hancks, D. C., Goodier, J. L., Mandal, P. K., Cheung, L. E. & Kazazian, H. H., Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. Human molecular genetics 20, 3386–3400, doi: 10.1093/hmg/dda245 (2011).
9. Garcia-Perez, J. L., Doucet, A. J., Bucheton, A., Moran, J. V. & Gilbert, N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. Genome Res 17, 602–611, doi: 10.1101/gr.5870107 (2007).
10. Gilbert, N., Lutz, S., Morrish, T. A. & Moran, J. V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Molecular and cellular biology 25, 7780–7795, doi: 10.1128/MCB.25.17.7780-7795.2005 (2005).
11. Buzdin, A. et al. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of H. Genomics 80, 402–406 (2002).
12. Weber, M. J. Mammalian small nuclear RNAs are mobile genetic elements. Plas Genet 2, 205, doi: 10.1371/journal.pgen.0020205 (2006).
13. Wei, W. et al. Human L1 retrotransposition: cis preference versus trans complementation. Molecular and cellular biology 21, 1429–1439, doi: 10.1128/MCB.21.4.1429-1439.2001 (2001).
14. Esnault, C., Maestre, J. & Heidmann, T. LINE-mediated retrotransposons generate processed pseudogenes. Nature genetics 24, 363–367, doi: 10.1038/74184 (2000).
18. Ono, R. et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nature Genetics 38, 101–106, doi: 10.1038/ng1699 (2006).
19. Sekita, Y. et al. Role of retrotransposon-derived imprinted gene, Rtl1, in the fetomaternale interface of mouse placenta. Nature Genetics 40, 243–248, doi: 10.1038/ng.2007.51 (2008).
20. Dupressoir, A. et al. Syncytin-A knockout mice demonstrate the critical role in placentaion of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc Natl Acad Sci USA 106, 12127–12132, doi: 10.1073/pnas.0909225106 (2009).
21. Muotri, A. R. et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435, 903–910, doi: 10.1038/nature03663 (2005).
22. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. Nature genetics 23, 314–318, doi: 10.1038/ng.15490 (1999).
23. Dupressoir, A. et al. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. Proc Natl Acad Sci USA 108, E1614–16173, doi: 10.1073/pnas.1112304108 (2011).
24. Nakaya, Y., Koshi, K., Nakagawa, S., Hashizume, K. & Miyazawa, T. Fematrin-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 (2013).
25. Naruse, M. et al. Sirh7/1d0c1 knockout mice exhibit placental P4 overproduction and delayed parturition. Development 141, 4763–4771, doi: 10.1242/dev.114520 (2014).
26. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biol Reprod 68, 651–654 (2003).
27. Evsikov, A. V. et al. Systems biology of the 2-cell mouse embryo. Cytogenet Genome Res 105, 240–250, doi: 10.1159/000078195 (2004).
28. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7, 597–606, doi: 10.1016/j.devcel.2004.09.004 (2004).
Author Contributions
R.O. and M.I. conceived of the study and R.O., T.K.-I., J.K., M.Ik. and F.I. participated in the experimental design. R.O. performed most analyses. M.Is., Y.F. and M.Ik. analyzed Cxx1a, Cxx1b, Rgg1f, Ddx3y, Spaca5 and Rsp6a in DSB-induced mice. R.O., Y.F., T.U. and M.Ik. produced mutant and knock-in mice by CRISPR/Cas systems.

The authors thank J.V. Moran for critical reading of the manuscript. This work was supported by grants from Grants-in-Aid for Scientific Research on Innovative Areas (24113507), Grant-in-Aid for Scientific Research (C) (26430183), and Sumitomo Foundation to R.O.

Acknowledgments
We thank J.V. Moran for critical reading of the manuscript. This work was supported by grants from Grants-in-Aid for Scientific Research on Innovative Areas (24113507), Grant-in-Aid for Scientific Research (C) (26430183), and Sumitomo Foundation to R.O.

Author Contributions
R.O. and M.Is. conceived of the study and R.O., T.K.-I., J.K., M.Ik. and F.I. participated in the experimental design. R.O. performed most analyses. M.Is., Y.F. and M.Ik. analyzed Cxx1a, Cxx1b, Rgg1f, Ddx3y, Spaca5 and Rsp6a in DSB-induced mice. R.O., Y.F., T.U. and M.Ik. produced mutant and knock-in mice by CRISPR/Cas systems. R.O. wrote the manuscript. All authors read and approved the final manuscript.
Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ono, R. et al. 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 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/