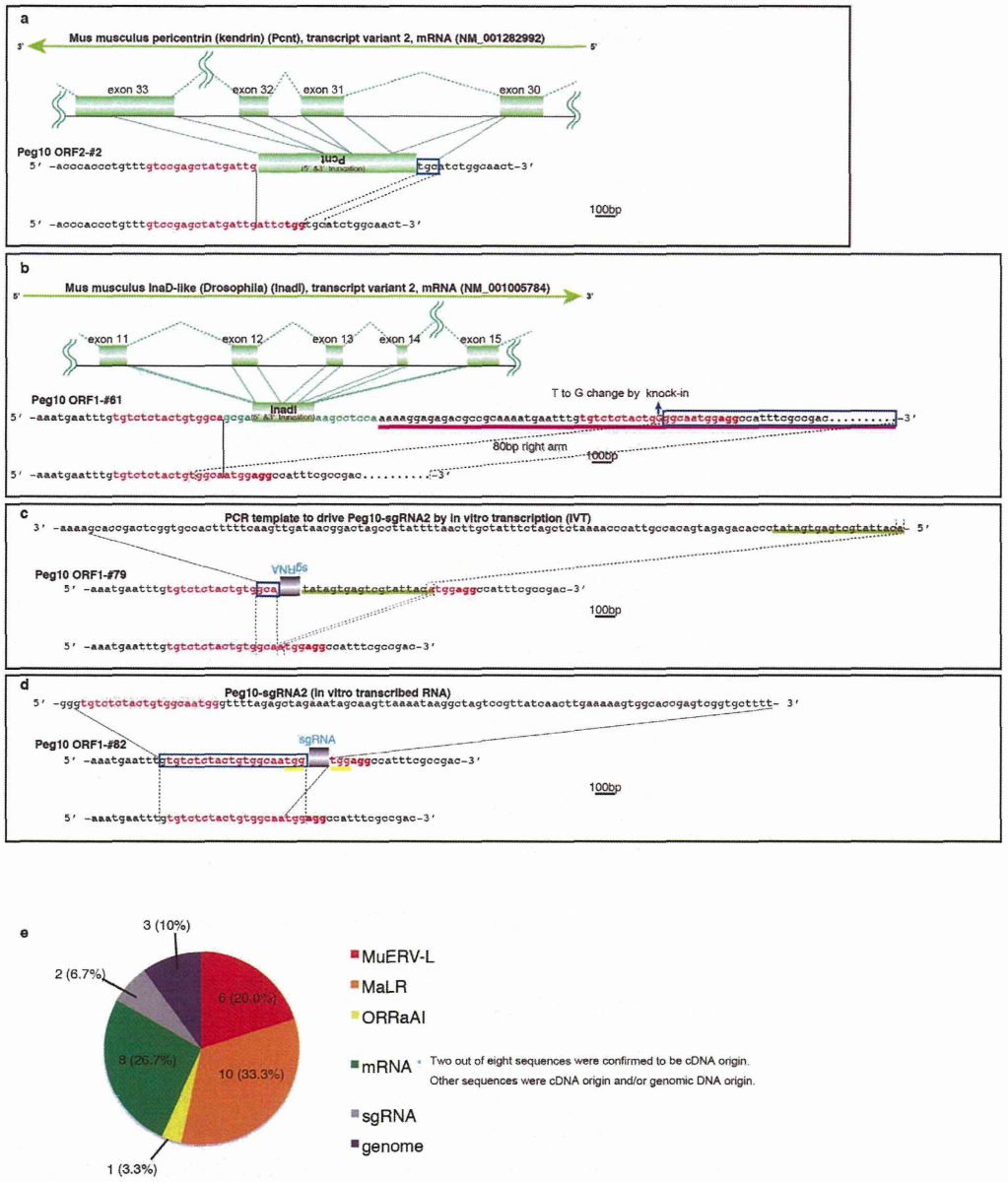

Figure 3. RT-product-mediated DSB repair (RMDR) and non-RT-product-mediated DSB repair (non-RMDR). (**a**) A partial sequence of the processed *Pcnt* gene (6803–7599 bp: NM_001282992) in reverse orientation was inserted into the DSB site mediated by RMDR. The integrated *Pcnt* fragment skipping introns 30–32 was inserted with a 3-bp microhomology. (**b**) A partial sequence of the processed *Inadl* gene (1527–1836 bp: NM_001005784) and a DNA oligo (pink bar) including a T to G mutation with 40 bp out of 53 bp in the 5′ homology arm were inserted into the DSB site. The 3′ side of the DSB site was repaired by HR with the long homology of the DNA oligo and 5′ side of the DSB site repaired by capture of the *Inadl* gene fragment, skipping introns 11–14 (RMDR). (**c**) A partial sgRNA sequence with T7 promoter (green bar) was inserted with 3-bp and 1-bp microhomologies mediated by non-RMDR. (**d**) A partial sgRNA in reverse orientation was inserted with a 21-bp microhomology, including a 20-bp guide RNA sequence. (**e**) Distribution of 30 insertion sequences at CRISPR/Cas DSB sites in 20 mice in which the DSBs were repaired by the capture of long DNA sequences. Approximately 57% of the inserted sequences were derived from LTR retrotransposons. The 8 insertions correspond to the exon regions of 8 genes. Two insertion sequences correspond to multiple exons, skipping introns, demonstrating that they are derived from cDNA.

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

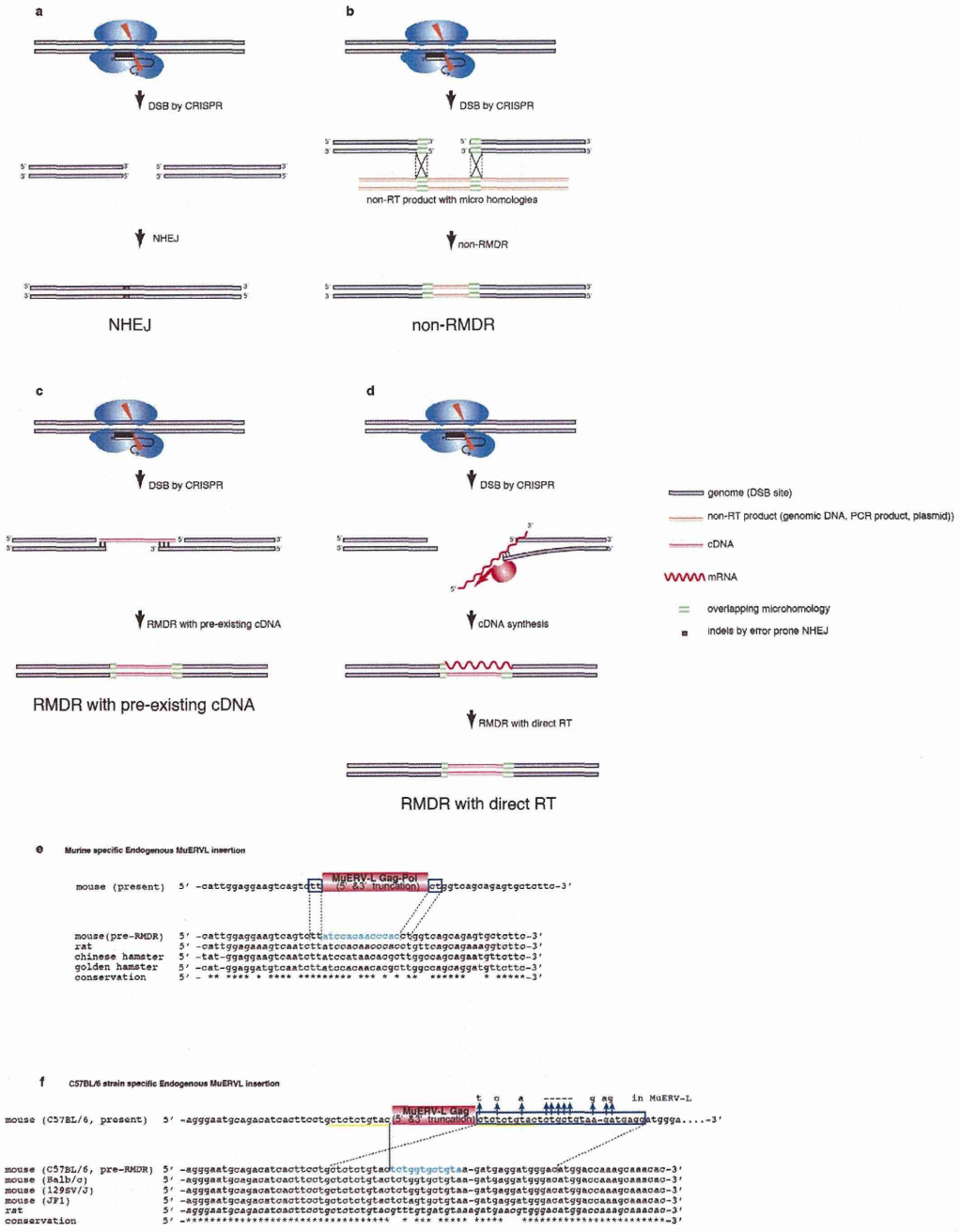

Figure 4. Possible mechanisms of RMDR, and RMDR under natural conditions. DSBs (the orange triangles) induced by CRISPR/CAS (blue sphere) are shown undergoing repair by NHEJ (**a**), non-RT-product-mediated DSB repair (non-RMDR) (**b**), RMDR with pre-existing cDNA (**c**) and RMDR with direct RT (**d**). 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[53,54]. 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[55]. 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* (*Tbcb*) 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 5 bp (rarely 6 bp) 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[17–24]. 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 genes 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)[5,32,33]. 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 ~500bp 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′- TGTCTCTACTGTGGCAATGG-3′)
Peg10-ORF2-sgRNA4 (5′- GTCCGAGCTATGATTGATTC-3′)
Cxx1a/b-sgRNA (5′-ACGGGATGGGGTTCCGCCGA-3′)
Rgag1-sgRNA (5′-GTGGTGTGGATGTCACTCCA-3′)
Ddx3y -sgRNA (5′-GAAAGATGCCTACAGCAGTTT-3′)
Rsph6a -sgRNA (5′-GGCTGGACCTCTGTGGCCAG-3′)
Spaca5 -sgRNA (5′-GCATGAAGGTCTGCAGCATTG-3′)

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

#1: AACCTACAGTTACTGCTCCCCAAAACATTCATTCACCCACAAGATTTAGAAACATAAA-ACGGCATAACTTCGTATAATGTATGCTATACGAAGTTATGCGGGGTGGGGGGGGAAGCTGAGG TCTCCGTGTAAACCTCACAAAGTCCGTAGCTGAAGGCTTC

#2: CTGTCCAAGGAAGAAAAGGAGAGACGCCGCAAAATGAATTTGTGTCTCTACTGGGGCA ATGGAGGCCATTTCGCCGACACGTGTCCAGCGAAAGCCTCCAAGAATTCGCCGCCGGGAAAC TCCCCGGCCCCGCT

Production of hCas9 mRNA and Peg10-ORF1-sgRNA. To produce the Cas9 mRNA, the T7 promoter was added to the Cas9 coding region of the pX330 plasmid by PCR amplification as previously reported[39]. The T7-Cas9 PCR product was gel purified and used as the template for *in vitro* transcription (IVT) using the mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). The T7 promoter was added to the Peg10-ORF1-sgRNA region of the pX330 plasmid by PCR purification using the primers listed below.

Peg10-ORF1-IVT-F (TGTAATACGACTCACTATAGGGTGTCTCTACTGTGGCAATGG), IVT-R (AAAAGCACCGACTCGGTGCC)

The T7-sgRNA PCR product was gel purified and used as the template for IVT using the MEGAshortscript T7 kit (Life Technologies). Both the Cas9 mRNA and Peg10-ORF1-sgRNA were DNAse treated to eliminate template DNA and purified using the MEGAclear kit (Life Technologies), and eluted into RNase-free water.

PCR and DNA sequencing. Genomic DNAs from the embryonic yolk sac or tail tip of pups were prepared using a Maxwll 16 system (Promega). Identification of the indels induced by DSB repair, were confirmed by PCR and subsequent DNA sequencing. The primers used for these DSB repair confirmations were Peg10-F (5′-GGAAGGTCTCAACCCAGACA-3′)/Peg10-R (5′-GTATCTCACGG TGGTCTCCC-3′), Cxx1a-F (5′-CAAATTACTTTGCTCCACTAACCCT-3′)/Cxx1a-R (5′-ATTCAG GAAGCCGTTGTAATCAT-3′), Cxx1b-F (5′-ATTGGGTAGCACTAAGGATTGTTGA-3′)/Cxx1b-R (5′-AGAGGCCTAGAAGTCCTCATCCT-3′), Rgag1-F (5′-TCTGTCCACACCTCTCATGG-3′)/Rgag1-R (5′-TGTTGCCGCTGTATCAGAAG-3′), Rsph6a-F (5′- AAGAATTCCAGGCAGGGTCCAGGATAGG-3′)/Rsph6a-R (5′- AAGGATCCCCTGGCTGAATATCTCATCC-3′), Ddx3y-F (5′- AAGAATTCCATGC CCTCATCTCAATATCCCATAAGGT-3′)/Ddx3y-R (5′- TTCTGCAGGGATAGCCATTGTTGGACTA GTTGGACA-3′), Spaca5-F (5′- aaGCTAGCCAGTGTCCTTATCCAATCTTTCTCCCTGC-3′)/Spaca5-R (5′- AAGGATCCCCTGGCTGAATATCTCATCC-3′). Each PCR product was purified with an S-400 spin column (GE Health Care) and sequenced with each F and R primer. The extra PCR bands (RMDR alleles) were extracted with a Gel-purification kit (Qiagen) and sub-cloned into a pGEM-T easy vector (Promega), and sequenced with each F and R primer.

HEK293T transfection and EGxxFP system. Five hundred ng of the pCAG-EGxxFP-target were mixed with 500ng of pX330 with/without the sgRNA sequences and then introduced into 4×10^5 HEK293T cells/well in a six well plate using Lipofectamine LTX (Life Technologies). The ratio of EGFP fluorescence positive cells/all cells (Hoechest 33342 positive nucleus) was monitored using a fluorescence microscopy EVOS cell counting system (Life Technologies) 48hrs 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 sperm, 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

GENETIC INFORMATION RESEARCH INSTITUTE (http://www.girinst.org/censor/index.php)[60], against the mouse genomes using each of the PCR products from the pX330-injected mice as a query.

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 2×10^5 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 10 μg/mL Blasticidin S (Life Technologies) and the other 10 μg/mL Blasticidin S and 50 μg/mL Azidothymidine (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/nrg3030 (2011).
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/ddr245 (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 l1. *Genomics* **80,** 402–406 (2002).
12. Weber, M. J. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* **2,** e205, 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. Human LINE retrotransposons generate processed pseudogenes. *Nature genetics* **24,** 363–367, doi: 10.1038/74184 (2000).
15. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics: TIG* **13,** 335–340 (1997).
16. Hancks, D. C. & Kazazian, H. H., Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22,** 191–203, doi: 10.1016/j.gde.2012.02.006 (2012).
17. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–789, doi: 10.1038/35001608 (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 feto-maternal 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 placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci USA* **106,** 12127–12132, doi: 10.1073/pnas.0902925106 (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: 10.1038/15490 (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,** E1164–1173, 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/Ldoc1 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).

29. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63, doi: 10.1038/nature11244 (2012).
30. Moore, J. K. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383,** 644–646, doi: 10.1038/383644a0 (1996).
31. Teng, S. C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **383,** 641–644, doi: 10.1038/383641a0 (1996).
32. Keeney, S. & Neale, M. J. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochemical Society transactions* **34,** 523–525, doi: 10.1042/BST0340523 (2006).
33. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual review of immunology* **24,** 541–570, doi: 10.1146/annurev.immunol.23.021704.115830 (2006).
34. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821, doi: 10.1126/science.1225829 (2012).
35. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109,** E2579–2586, doi: 10.1073/pnas.1208507109 (2012).
36. Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31,** 230–232, doi: 10.1038/nbt.2507 (2013).
37. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823, doi: 10.1126/science.1231143 (2013).
38. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826, doi: 10.1126/science.1232033 (2013).
39. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–918, doi: 10.1016/j.cell.2013.04.025 (2013).
40. Mashiko, D. *et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci Rep* **3,** 3355, doi: 10.1038/srep03355 (2013).
41. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. Ty elements transpose through an RNA intermediate. *Cell* **40,** 491–500 (1985).
42. Garfinkel, D. J., Boeke, J. D. & Fink, G. R. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42,** 507–517 (1985).
43. Eichinger, D. J. & Boeke, J. D. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* **54,** 955–966 (1988).
44. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics* **20,** 116–117, doi: 10.1038/2413 (1998).
45. Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G. & Heidmann, T. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nature genetics* **36,** 534–539, doi: 10.1038/ng1353 (2004).
46. Ribet, D., Dewannieux, M. & Heidmann, T. An active murine transposon family pair: retrotransposition of "master" MusD copies and ETn trans-mobilization. *Genome Res* **14,** 2261–2267, doi: 10.1101/gr.2924904 (2004).
47. Ko, M. S. *et al.* Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development* **127,** 1737–1749 (2000).
48. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343,** 193–196, doi: 10.1126/science.1245316 (2014).
49. Feng, Q., Moran, J. V., Kazazian, H. H., Jr. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87,** 905–916 (1996).
50. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci USA* **94,** 1872–1877 (1997).
51. Cost, G. J. & Boeke, J. D. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37,** 18081–18093 (1998).
52. Morrish, T. A. *et al.* Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446,** 208–212, doi: 10.1038/nature05560 (2007).
53. Piko, L. & Clegg, K. B. Quantitative changes in total RNA, total poly(A), and ribosomes in early mouse embryos. *Developmental biology* **89,** 362–378 (1982).
54. Schultz, R. M. Regulation of zygotic gene activation in the mouse. *BioEssays: news and reviews in molecular, cellular and developmental biology* **15,** 531–538, doi: 10.1002/bies.950150806 (1993).
55. Horii, T. *et al.* Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Sci Rep* **4,** 4513, doi: 10.1038/srep04513 (2014).
56. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature genetics* **31,** 159–165, doi: 10.1038/ng898 (2002).
57. Solyom, S. *et al.* Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res* **22,** 2328–2338, doi: 10.1101/gr.145235.112 (2012).
58. Vitullo, P., Sciamanna, I., Baiocchi, M., Sinibaldi-Vallebona, P. & Spadafora, C. LINE-1 retrotransposon copies are amplified during murine early embryo development. *Mol Reprod Dev* **79,** 118–127, doi: 10.1002/mrd.22003 (2012).
59. Dai, L., Huang, Q. & Boeke, J. D. Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem* **12,** 18, doi: 10.1186/1471-2091-12-18 (2011).
60. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6,** 11, doi: 10.1186/s13100-015-0041-9 (2015).

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., M.K. and M.Ik. analyzed *Cxx1a*, *Cxx1b*, *Rgag1*, *Ddx3y*, *Spaca5* and *Rsph6a* DSB-induced mice. R.O., Y.F., T.U. and M.Ik. produced mutant and knock-in mice by CRISPR/Cas. 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).

Scientific Report

SOURCE DATA | TRANSPARENT PROCESS

EMBO reports

Ubiquitin acetylation inhibits polyubiquitin chain elongation

Fumiaki Ohtake[1,*], Yasushi Saeki[2], Kensaku Sakamoto[3], Kazumasa Ohtake[3], Hiroyuki Nishikawa[4], Hikaru Tsuchiya[2], Tomohiko Ohta[5], Keiji Tanaka[2] & Jun Kanno[1]

Abstract

Ubiquitylation is a versatile post-translational modification (PTM). The diversity of ubiquitylation topologies, which encompasses different chain lengths and linkages, underlies its widespread cellular roles. Here, we show that endogenous ubiquitin is acetylated at lysine (K)-6 (AcK6) or K48. Acetylated ubiquitin does not affect substrate monoubiquitylation, but inhibits K11-, K48-, and K63-linked polyubiquitin chain elongation by several E2 enzymes *in vitro*. In cells, AcK6-mimetic ubiquitin stabilizes the monoubiquitylation of histone H2B—which we identify as an endogenous substrate of acetylated ubiquitin—and of artificial ubiquitin fusion degradation substrates. These results characterize a mechanism whereby ubiquitin, itself a PTM, is subject to another PTM to modulate mono- and polyubiquitylation, thus adding a new regulatory layer to ubiquitin biology.

Keywords acetylation; mechanism; post-translational modification; ubiquitin
Subject Categories Post-translational Modifications, Proteolysis & Proteomics
DOI 10.15252/embr.201439152 | Received 12 June 2014 | Revised 19 November 2014 | Accepted 20 November 2014 | Published online 19 December 2014
EMBO Reports (2015) 16: 192–201

Introduction

Ubiquitylation is an essential post-translational modification (PTM) involved in a variety of biological pathways including protein degradation, signal transduction, DNA repair, and gene regulation [1]. Ubiquitin is a 76 amino acid (A.A.) protein that is covalently conjugated to substrates through a cascade involving activating enzymes (E1), conjugating enzymes (E2), and ligases (E3). The widespread biological roles of the ubiquitin system are achieved by the diversity of ubiquitin topologies [2,3]. Ubiquitylation can comprise ubiquitin monomers or polymers. Monoubiquitylation regulates nondegradative events such as gene expression and endocytosis. In addition, ubiquitin can be polymerized through one of its seven lysine (K) residues or its N-terminus within a ubiquitin moiety, resulting in various polyubiquitin chains bearing different linkages. In addition to the best-studied role of K48-chains in the proteasome-mediated protein degradation, the roles of K63-chains in DNA repair and signal transduction, K11-chains in cell cycle regulation, and M1-linked (linear) chains in the NF-κB pathway are being unveiled [2,4,5]. Thus, diversity in the topologies of ubiquitylation constitutes a "ubiquitin code" that spells out the fate of substrates [3]. The ubiquitin code is highly dynamic and restricted to given cellular contexts, as exemplified in the signal-dependent and target-specific assembly of K63-chains or M1-chains.

The integrity of the ubiquitin code is maintained by the cooperation of factors that add, remove, or recognize the code. The addition of specific ubiquitin topologies is mediated by E2 enzymes or by HECT- and RBR-type E3s. CDC34, UBC13-UEV1a, or UBE2S specifically catalyze the elongation of K48-, K63-, or K11-linked chains, respectively [6]. UBCH5 mainly initiates the first (mono)ubiquitylation and also synthesizes multiple types of polyubiquitin linkages [7]. The trimming of ubiquitin topologies is regulated by certain classes of deubiquitylation enzymes (DUBs), which can remove specific linkages [8]. The resulting ubiquitin code is recognized by various decoding proteins that are selective for a specific ubiquitin topology. The hydrophobic patch within the ubiquitin moiety, consisting of Ile44, Leu8, and Val70, is frequently used as the interface with these ubiquitin-binding proteins [9].

Given the importance and complexity of ubiquitylation, we speculated that the ubiquitin code may be regulated by yet other diverse mechanisms. While polyubiquitin chain formation can be viewed as the ubiquitylation of ubiquitin itself, the existence and significance of other ubiquitin-directed PTMs are not well understood. In this study, we searched for possible PTMs on the ubiquitin moiety and found ubiquitin acetylation and phosphorylation. The data revealed a mechanism by which ubiquitin itself is modified by another PTM to regulate the mono- and polyubiquitylation of substrates. Thus, ubiquitin acetylation will add a new layer to the molecular basis of ubiquitin biology.

1 Division of Cellular and Molecular Toxicology, Biological Safety Research Center, National Institute of Health Sciences, Setagaya-ku, Tokyo, Japan
2 Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Sciences, Setagaya-ku, Tokyo, Japan
3 Division of Structural and Synthetic Biology, RIKEN Center for Life Science Technologies, Tsurumi, Yokohama, Japan
4 Institute of Advanced Medical Science, St. Marianna University Graduate School of Medicine, Kawasaki, Japan
5 Department of Translational Oncology, St. Marianna University Graduate School of Medicine, Kawasaki, Japan
*Corresponding author. Tel: +81 3 3700 9619; Fax: +81 3 3700 9647; E-mail: ohtake@nihs.go.jp

Results

The identification of acetylation and phosphorylation on endogenous ubiquitin

To identify possible PTMs directed to ubiquitin, endogenous ubiquitin conjugated to substrate proteins was immunoprecipitated from cultured human 293F cells (Fig 1A). The conjugated ubiquitin (> 70 kD in molecular weight) was subjected to trypsin digestion, and peptides were identified using liquid chromatography–mass spectrometry (LC-MS) [10]. A shotgun screening indicated a mass shift corresponding to acetylation at K6 (+42.0106 Da) or phosphorylation at T14 (+79.9663 Da). Next, we performed targeted acquisition of MS/MS spectra for possible acetylation and phosphorylation. As a result, we obtained MS/MS spectra for high-confidence acetylation sites, K6, K48, and K63, and phosphorylation sites, T14 and S65 (Supplementary Figs S1 and S2). These data revealed that ubiquitin was indeed modified by other PTMs (Fig 1B).

Among the identified ubiquitin PTMs, K6 and K48 residues are located close to the Leu8-Ile44-Val70 hydrophobic patch, while K63, T14, and S65 are at a distance from the patch (Fig 1C). We speculated that acetylation of K6 or K48 may modulate interactions involving the hydrophobic patch. Therefore, we explored the molecular function of ubiquitin acetylation at K6 and K48 (hereafter denoted AcK6 and AcK48). To unambiguously confirm K6 and K48 acetylation, standard AQUA (absolute quantification) peptides, harboring acetylation at the target residue and a stable isotope label at a leucine residue (Supplementary Fig S3A) [11,12], were mixed with endogenous ubiquitin-derived peptides before mass spectrometric analysis. The fragmentation patterns of the AQUA peptides were identical to those of the sample-derived peptides (Fig 1D and E). Moreover, extracted ion chromatograms revealed that the corresponding AQUA (heavy) and sample (light) peptides co-eluted at the same retention times (Fig 1F). These data successfully confirmed ubiquitin acetylation at K6 and K48.

Ubiquitin acetylation is regulated by multiple classes of histone deacetylases

Next, the levels of AcK6- and AcK48-ubiquitin in cells were quantified using the recently reported parallel reaction monitoring (PRM) approach, an MS/MS-based quantification method using a high-resolution MS instrument [10]. We prepared AQUA peptides as internal standards for AcK6, AcK48, and unmodified/GlyGly-bearing peptides, and selected MS/MS fragment ions were used for quantification (Supplementary Fig S3A). Standard curve analyses confirmed that AcK6 and AcK48 could be quantified from 0.1 to 100 fmol (Supplementary Fig S3B and C).

Quantification revealed that endogenous AcK6 and AcK48 represented approximately 0.03 and 0.01%, respectively, of total ubiquitin in 293F cells (Fig 2A). These proportions were comparable to those of the M1-linked chains (0.02%), which are known to be dynamically assembled in specific substrates and locations [4]. Importantly, ubiquitin acetylation could be quantified from the substrate-conjugated ubiquitin, indicating that "acetyl-ubiquitylation" occurs on substrates. Acetylation was also observed on ubiquitin monomers (Supplementary Fig S3D), suggesting the availability of free acetyl-ubiquitin for ubiquitylation.

Combined treatment of 293F cells with trichostatin A (TSA) and nicotinamide (NIA), which inhibit class I/II and class III histone deacetylases (HDACs), respectively, significantly increased levels of AcK6 and, to a lesser extent, AcK48 (Fig 2B). By contrast, individual treatment with class-specific HDAC inhibitors, MS-275 (class I), MC1568 (class II), or NIA (class III), only modestly increased ubiquitin acetylation. As expected, histone acetylation was increased markedly by the class I inhibitor, and to a lesser extent by the class II or III inhibitors (Fig 2C). These results indicated that ubiquitin acetylation is redundantly regulated by the three classes of HDACs in intact cells.

Functional dissection of ubiquitin acetylation using recombinant acetyl-ubiquitin

As characterized in histones, different PTMs on the same substrate cross talk with each other to constitute the protein PTM code [13]. Considering that K6 and K48 residues may modulate ubiquitin-E1/E2/E3 interactions, we asked whether ubiquitin acetylation could modulate its own ubiquitylation.

Recombinant acetyl-ubiquitin proteins were synthesized in *E. coli* with an expanded genetic code. Acetyllysine was incorporated at the positions specified with the UAG codon in the mutant genes [14]. MALDI-TOF MS analysis of ubiquitin proteins revealed a mass shift (~42 Da), confirming that the produced acetylated ubiquitin proteins contained one acetyllysine residue (Supplementary Fig S4A). Subsequent PRM analysis revealed that the recombinant acetyl-ubiquitins were specifically acetylated at the target residues (K6 or K48), while the wild-type ubiquitin was not (Fig 3A). The polyhistidine tag at the ubiquitin C-terminus was subsequently removed using USP2 [15] to produce intact ubiquitin proteins at near homogeneity (Fig 3B).

Using the recombinant acetyl-ubiquitins, we first assessed their ability to be charged to E1 or E2 enzymes. *In vitro* charging assays showed that the charging to E1 (UBA1) and E2 enzymes (UBCH5 or RAD6) was not significantly altered by acetylation (Fig 3C). Similarly, we tested the suitability of acetyl-ubiquitins for monoubiquitylation. Since histone H2B is well known as a monoubiquitylation substrate, an H2B ubiquitylation assay was performed using the cognate E2 RAD6 [16]. Ubiquitylation of H2B occurred almost exclusively as monoubiquitylation, and the levels of monoubiquitylation were not significantly affected by ubiquitin acetylation (Fig 3D). These data showed that acetyl-ubiquitins, at least in the contexts we tested, are essentially intact in charging to E1/E2 or acting in monoubiquitylation.

Acetylation of ubiquitin at K6 or K48 represses polyubiquitin chain elongation *in vitro*

We next asked whether ubiquitin acetylation affected polyubiquitin chain elongation. Chain elongating E2s (CDC34 (K48-link), UBE2K (K48-link), UBC13-Uev1a (K63-link), and UBE2S (K11-link)) and chain-unspecific E2s (UBCH5 and RAD6) [5,7,17–19] were tested in an *in vitro* ubiquitin chain elongation assay. AcK6 or AcK48 ubiquitins repressed polyubiquitin chain elongation by all of the E2s tested (Fig 4A). Assays using less ubiquitin or different antibodies (P4D1)