Sequence-specific retrotransposition of 28S rDNA-specific LINE R2Ol in human cells

YUTING SU,[1] NARISU NICHUGUTI,[1] AZUSA KUROKI-KAMI, and HARUHIKO FUJIWARA

Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8562, Japan

ABSTRACT

R2 is a long interspersed element (LINE) found in a specific sequence of the 28S rDNA among a wide variety of animals. Recently, we observed that R2Ol isolated from medaka fish, *Oryzias latipes*, retrotransposes sequence specifically into the target sequence of zebrafish. Because the 28S target and flanking regions are widely conserved among vertebrates, we examined whether R2Ol can also integrate in a sequence-specific manner in human cells. Using adenovirus-mediated expression of R2Ol constructs, we confirmed an accurate insertion of R2Ol into the 28S target of human 293T cells. However, the R2Ol mutant devoid of endonuclease (EN) activity showed no retrotransposition ability, suggesting that the sequence-specific integration of R2Ol into 28S rDNA occurs via the cleavage activity of EN. By introducing both R2Ol helper virus and donor plasmid in human cells, we succeeded in retrotransposing an exogenous EGFP gene into the 28S target site by the *trans*-complementation system, which enabled simplification of specific gene knock-in in a time-efficient manner. We believe that R2Ol may provide an alternative targeted gene knock-in method for practical applications such as gene therapy in future.

Keywords: long interspersed element (LINE); 28S rDNA specific element R2Ol; sequence-specific retrotransposition; targeted gene knock-in; human cell; gene therapy

INTRODUCTION

Precision targeted gene integration technologies are needed for improved safety, particularly in therapeutic applications involving human cells. Since 2013, CRISPR/Cas9-mediated genome editing has been trending because of its simplicity and efficiency (Lino et al. 2018; You et al. 2019). The CRISPR/Cas9 system repairs the DNA double-strand break (DSB) by the homology-directed repair (HDR) pathway resulting in gene knock-in at the target site in the presence of donor DNA. Although this HDR-mediated knock-in method has been successful in human cells, it still suffers from low HDR efficiency, off-target concern, and time and labor consuming of donor DNA designing (Liu et al. 2018; Lee and Kim 2019). Thus, another system that functions independently of HDR might provide a better alternative to targeted gene knock-in strategies.

Long interspersed elements (LINEs) are widely distributed in eukaryotes and move by a LINE-specific "copy-out and copy-in" mechanism. After transcription and translation of LINE mRNA, its protein and mRNA form the ribonu-cleoprotein (RNP) complex, move back to the nucleus, and insert the LINE cDNA into the target DNA by target primed reverse transcription (TPRT) (Luan et al. 1993). LINE encodes an endonuclease domain (EN) that nicks the target DNA and the reverse transcriptase (RT) domain, which reverse transcribes its own mRNA to cDNA onto the nicked site in 3′ to 5′ orientation. Most LINEs retrotranspose randomly into the host genome, such as human L1 elements, which demonstrate their ongoing activity in humans and cause various diseases (Kazazian and Moran 2017). Nevertheless, some elements integrate in a sequence-specific manner into repetitive genomic sequences, such as ribosomal DNA (rDNA), telomeric repeats, and microsatellites (Fujiwara 2015).

Phylogenetically, LINEs are classified into two large groups based on their structural and phylogenetic features: an ancient group encodes restriction enzyme-like endonucleases (RLEs) and a recently branched group encodes apurinic/apyrimidinic endonucleases (APEs). Many of the RLE-encoding elements (NeSL, R2, HERO, and CRE) retrotranspose in a sequence-specific fashion

(Eickbush and Eickbush 2015; Fujiwara 2015). Among the RLE elements, the 28S rDNA sequence-specific R2 elements are the most studied LINEs concerning retrotransposition mechanisms both in vitro (Luan et al. 1993; Bibillo and Eickbush 2004; Christensen et al. 2006; Eickbush and Eickbush 2010; Govindaraju et al. 2016) and in vivo (Eickbush et al. 2000). They are distributed in a diverse range of both invertebrates and vertebrates, such as fishes, birds, and reptiles, but not in mammals (Kojima et al. 2016). In addition, an approximately 300 base pair (bp) region around the R2 insertion site in the 28S rDNA is highly con-

served across a wide variety of eukaryotes, including humans (Fig. 1A; Kojima and Fujiwara 2005). Based on its high sequence-specific and efficient retrotransposition, R2 may be a candidate as an alternative tool for gene integration.

Recently, we isolated R2Ol from medaka fish, *Oryzias latipes,* comprising a 5′ untranslated region (5′ UTR), a single open reading frame (ORF) encoding EN and RT domains, and a 3′ UTR (Fig. 1B). We detected its sequence-specific retrotransposition into 28S rDNA both in insect, *Spodoptera frugiperda 9* (Sf9) cells (T Kawashima and H Fujiwara, unpubl.) and the zebrafish embryo (Kuroki-Kami et al. 2019). Recently, we also found that R2Ol encodes a self-cleavage hepatitis D virus (HDV)-like ribozyme in its 5′ UTR and that the cleavage site by the ribozyme is located within the 28S sequence, at 28 nt upstream from the 5′ junction between the 28S rDNA and the R2Ol element (F Shiba and H Fujiwara, unpubl.). We here tried to establish an adenovirus-based R2Ol retrotransposition system in human cells, because R2Ol seems to retain the retrotransposition activity among diverse organisms.

We confirmed that R2Ol retrotransposes in a sequence-specific manner into the human 28S target site. In addition, the exogenous gene of interest (EGFP) was retrotransposed into the 28S rDNA target site by *trans*-complementation, which might offer a convenient gene integration method. The R2Ol system uses TPRT, unlike the CRISPR/Cas9 gene knock-in system using HDR, which may provide an alternative tool with potential applications such as gene therapy.
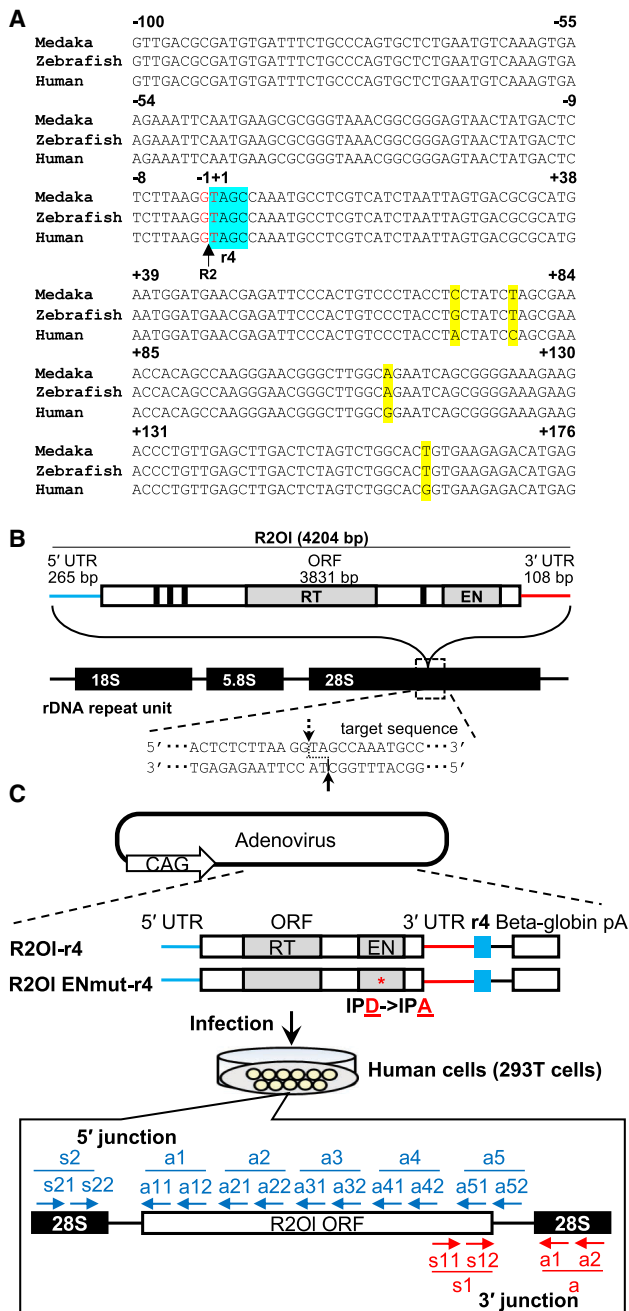
FIGURE 1. Schematic structure and insertion sites of R2Ol in the rDNA unit. (*A*) Comparison of the R2Ol insertion site sequence alignment of three vertebrates—namely, medaka fish (CP020779.1), zebrafish (BX537263.17), and human (M11167.1). An ~300-bp region around the R2 insertion site in 28S rDNA is strongly conserved except for differences at 4 nt (yellow). The insertion site (red) of R2 is found between −1 (G) and +1 (T). In the downstream 28S sequence of the insertion site, 1 to 4 (TAGC) nt are denoted as r4. (*B*) A schematic diagram of R2Ol (4204 bp), cloned from the medaka fish, *Oryzias latipes*, comprises a 5′ untranslated region (UTR; 265 bp, blue), a single open reading frame (ORF; 3831 bp) encoding EN and RT domains, and a 3′ UTR (108 bp, red). Vertical lines represent the zinc fingers. The rDNA unit is indicated by closed boxes. The double-stranded sequences of the precise insertion sites are indicated. The bottom and putative top strand cleavage sites are indicated by a solid and dashed arrow, respectively. (*C*) R2Ol retrotransposition assay in human cells. R2Ol-r4 contains the CAG promoter, a full-length R2Ol with the r4 downstream from its 3′-end, and a β-globin pA signal. R2Ol ENmut-r4 includes a mutation at the essential amino acid sites for EN (IPD → IPA). After 293T cells were infected by the R2Ol adenovirus for 72 h, retrotransposition events were detected by nested PCR by amplifying the 3′ and 5′ junction of the retrotransposed R2Ol in the 28S rDNA target. Horizontal arrows indicate the PCR primers used to detect retrotransposition events. Primer sets "s1" and "a" (red) were used for analyzing the 3′ junction, and primer sets "s2", "a1", "a2", "a3", "a4", and "a5" (blue) were used for analyzing the 5′ junction.
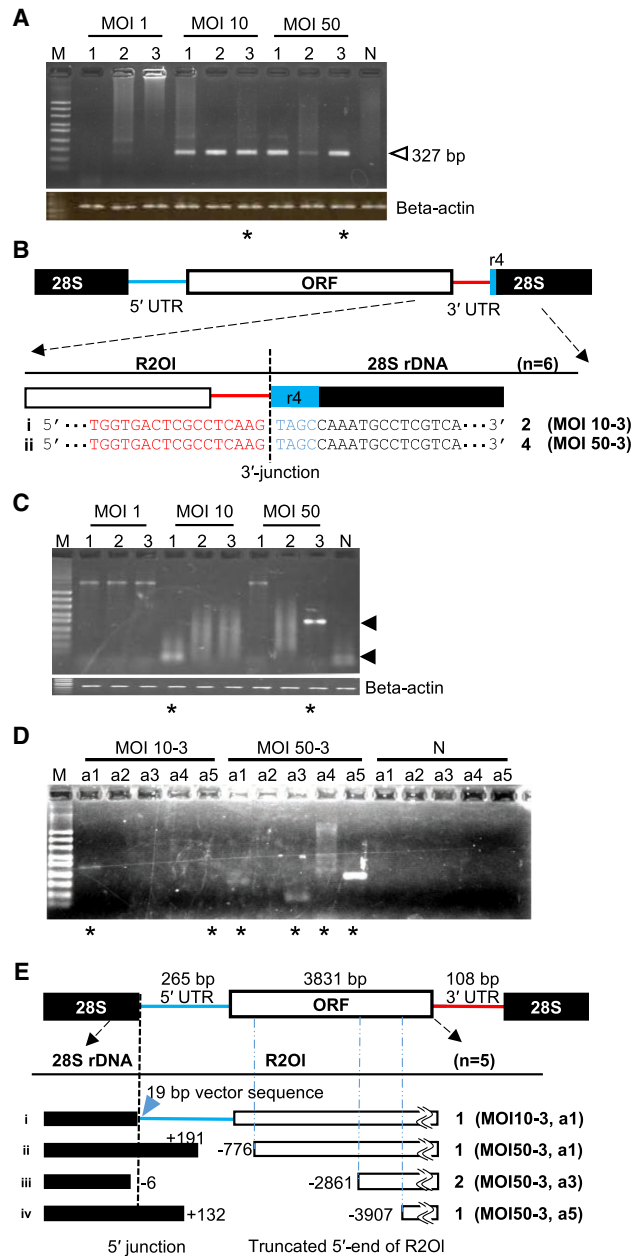
RESULTS

Adenovirus-mediated R2Ol retrotransposition system in human cells

To determine whether R2Ol retains its retrotransposon activity in human cells, we established a retrotransposition assay using an adenovirus-based expression system. We constructed an adenovirus R2Ol-r4 construct (wild type; WT) which contains the full-length unit of R2Ol (5′ UTR/ORF/3′ UTR) and a 4-bp 28S rDNA downstream flanking sequence at its 3′-end of 3′ UTR (r4; TAGC) under the control of CAG promoter and polyadenylation signal sequence (pA) (Fig. 1B,C). The r4 sequence at the 3′-end was expected to increase the retrotransposition efficiency and accuracy by annealing efficiently between the template RNA and target DNA (Kuroki-Kami et al. 2019). As a negative control, we constructed R2Ol ENmut-r4, which has a mutation (IP<u>D</u> > IP<u>A</u>) in the catalytic center that is essential for the endonuclease activity (Fig. 1C; Yang et al. 1999). After the recombinant R2Ol adenovirus infected the 293T cells for 72 h, the total genomic DNAs was extracted from the infected cells and the retrotransposition event was detected by nested PCR, amplifying the 3′ junction and 5′ junction sites between R2Ol and 28S rDNA with primer sets as shown in Figure 1C.

3′ and 5′ junction analyses of retrotransposed R2Ol in the 28S rDNA target

The R2Ol-r4 and R2Ol ENmut-r4 were introduced into 293T cells at varying (1, 10, and 50) multiplicity of infection (MOI). PCR-amplified bands of the expected size (327 bp), representing the precise integration of R2Ol at the 3′ junction, were observed at higher MOI—namely, 10 and 50—but not at MOI 1 (Fig. 2A, open triangle), indicating that higher expression of R2Ol from the adenovirus vector is necessary for efficient retrotransposition of the R2Ol element into the target site. To certify whether the PCR bands at MOI 10 and 50 represent the exact targeting of R2Ol, the PCR fragments from samples MOI 10-3 and MOI 50-3 were cloned and sequenced (Fig. 2A, asterisks). Among all clones analyzed, six clones indicated the precise integration of R2Ol at the 28S rDNA target region in the 3′ junction site (Fig. 2B). The remaining clones were nonspecific PCR artifacts originated from R2Ol or 28S rDNA sequences. In contrast, the PCR band representing a precise integration of R2Ol using R2Ol ENmut-r4 was not detected under any MOI conditions, whereas two extra bands were observed (filled triangles) (Fig. 2C). To analyze these extra bands in R2Ol ENmut-r4 samples, we cloned and sequenced PCR products MOI 10-1 and MOI 50-3 and observed that all the analyzed clones showed nonspecific PCR sequences, indicating no retrotransposition activity. Thus, the above results suggest that the

FIGURE 2. 3′ and 5′ junction analyses for retrotransposed R2Ol in human 293T cells. (*A*) PCR results of the 3′ junction of R2Ol-r4. The open triangle indicates accurate retrotransposition. β-actin was used as an internal control. The numbers marked above the lanes indicate independent experiments. (*B*) 3′ junction sequence of the R2Ol-r4 retrotransposition events. (*C*) PCR results of the 3′ junction of R2Ol ENmut-r4. Filled triangles indicate nonspecific bands. (*D*) PCR results of the 5′ junction of R2Ol-r4. (*E*) 5′ junction sequences of inserted R2Ol. The PCR products indicated by asterisks were TA cloned for further analyses. (N) Negative control PCR conducted with genomic DNA from noninfected cells, (closed box) 28S rDNA sequence, (open box) R2Ol ORF, (blue line) 5′ UTR, (red line) 3′ UTR, (blue triangle) vector sequences, (MOI) multiplicity of infection, (*n*) total number of sequenced clones. Truncated or deleted sites are indicated by nucleotide numbers (+1 as 5′end of R2Ol).

sequence-specific integration of R2Ol into 28S rDNA depends on the cleavage activity of self-encoded EN. Although we do not have experimental data using RT mutants in human cells, we speculate that the R2Ol integration observed here is caused by reverse transcription but not recombination events, because we observed the complete loss of the retrotransposition activity in both EN and RT mutants of R2Ol in zebrafish embryo (Kuroki-Kami et al. 2019).

We further analyzed the 5′ junctions between the retrotransposed R2Ol and the 28S rDNA target using five primer sets (Fig. 1C, blue). Using the same genomic DNA samples as those used for 3′ junction analyses (Fig. 2A, MOI 10-3 and MOI 50-3), we detected variable-sized PCR bands with the primer sets a1 and a5 in MOI 10 and a1, a3, a4, and a5 in MOI 50 (Fig. 2D, asterisk). After cloning and sequencing, we observed four types of integrated R2Ol (Fig. 2E). Notably, one type represented a full-length insertion (>4.2 kb) of R2Ol (Fig. 2E, i; Supplemental Table S2), suggesting that longer exogenous genes of interest could be knocked-in by using this system. The remaining three types showed 5′-truncated R2Ol, terminating its reverse transcription at various sites within the ORF (Fig. 2E, ii–iv; Supplemental Table S2). This type of 5′ truncation is often observed among LINEs. As a negative control, we used the genomic DNA from 293T cells without virus infection, which showed no PCR bands, as expected (Fig. 2D, N).

## R2Ol retrotransposes the exogenous EGFP gene into the 28S target by *trans*-complementation

Previously we showed that R2Ol recognizes its own 3′ UTR sequences to retrotranspose the exogenous gene (EGFP with the R2Ol 3′ UTR) in the zebrafish embryo by *trans*-complementation (Kuroki-Kami et al. 2019). Because it enables easier manipulation and designing of the transgene construct, we next tried to develop a sequence-specific gene integration system of R2Ol by *trans*-complementation in human cells. The adenovirus R2Ol-r4 or R2Ol ENmut-r4 (negative control) was used as a helper to retrotranspose two donor EGFP plasmids, EGFP-3U and EGFP-3U-r4 (Fig. 3A). The donor plasmid EGFP-3U includes a full length of EGFP gene that terminates with the 3′ UTR of R2Ol. We also constructed pEGFP-3U-r4 in which the r4 sequence was added to pEGFP-3U. The plasmids (0.5, 1, and 1.5 µg) were transfected into 293T cells, and subsequently the helper virus R2Ol-r4 infected cells at MOI 10 (Fig. 3A). It is hypothesized that the R2Ol ORF protein from the helper virus recognizes the R2Ol 3′ UTR in the donor mRNA, forms the RNP complex, and reverse transcribes the EGFP mRNA into cDNA on the 28S rDNA target. Retrotransposition of the EGFP gene was detected by nested PCR amplification of the 3′ junction between the 28S rDNA target and the EGFP gene using primer set s3 and a (Fig. 3A).

In case of EGFP-3U transfection, we observed several bands in lanes 1–3 around the predicted sized band (667 bp), representing the retrotransposition events (Fig. 3B). Cloning and sequencing of PCR bands (Fig. 3B, asterisks) revealed four types of clones representing a successful retrotransposition of donor EGFP-3U into the 28S rDNA target (Fig. 3C). All clones showed the 3′-truncated sequences in EGFP-3U, suggesting that the reverse transcription started from within the 3′ UTR (−71) or the vector sequence (−229, −243, −377), but not from the end of R2Ol 3′ UTR (Fig. 3C, left region; Supplemental Table S3). In addition, the R2Ol endonuclease digestion site of the bottom strand of the 28S rDNA at the 3′ junction was shifted to various positions from the exact target site (+203, +48, −142) (Fig. 3C, right region; Supplemental Table S3). However, when R2Ol ENmut-r4 was used as a helper, no clear band was observed (Fig. 3B).

In the EGFP-3U-r4 transfection, we observed a PCR band corresponding to approximately predicted size (667 bp) representing a retrotransposition event (Fig. 3D, R2Ol-r4, lane 2). After cloning and sequencing of the 3′ junction sites (Fig. 3D, asterisks), we observed five types of successful retrotransposition of the EGFP gene from EGFP-3U-r4 (Fig. 3E). It is noteworthy that two types of clones contained the complete 3′ UTR of R2Ol (Fig. 3E, i, ii; Supplemental Table S4), indicating that the reverse transcription started exactly from the 3′-end of R2Ol (Fig. 3E, left region; Supplemental Table S4). Because the EGFP-3U clones showed no accurate reverse transcription (Fig. 3C), addition of the r4 sequence to EGFP-3U-r4 enhanced the accuracy of reverse transcription in the *trans*-complementation system of R2Ol. This suggests that annealing of the r4 sequence to the 28S rDNA target is critical for determining the start point of reverse transcription. However, even in the case of EGFP-3U-r4, the digestion sites by R2Ol EN were shifted from the exact target site (Fig. 3E, right region; Supplemental Table S4), as shown in EGFP-3U. When R2Ol ENmut-r4 virus was used as a helper, the retrotransposition into the 28S rDNA target was not observed (Fig. 3D). Although we have tried to detect the 5′ junction sites, we failed to get PCR fragments probably because of the 5′-truncation at various sites in the upstream region of EGFP, or because of a very low efficiency of the *trans*-complementation retrotransposition system. The above results showed that EGFP in the donor plasmid could be retrotransposed into the 28S rDNA using the R2Ol *trans*-complementation system; however, the insertion site was not completely identical with the original target site of R2Ol.

DISCUSSION

A LINE-based tool for integrating exogenous genes in a sequence-specific manner in human cells has not been
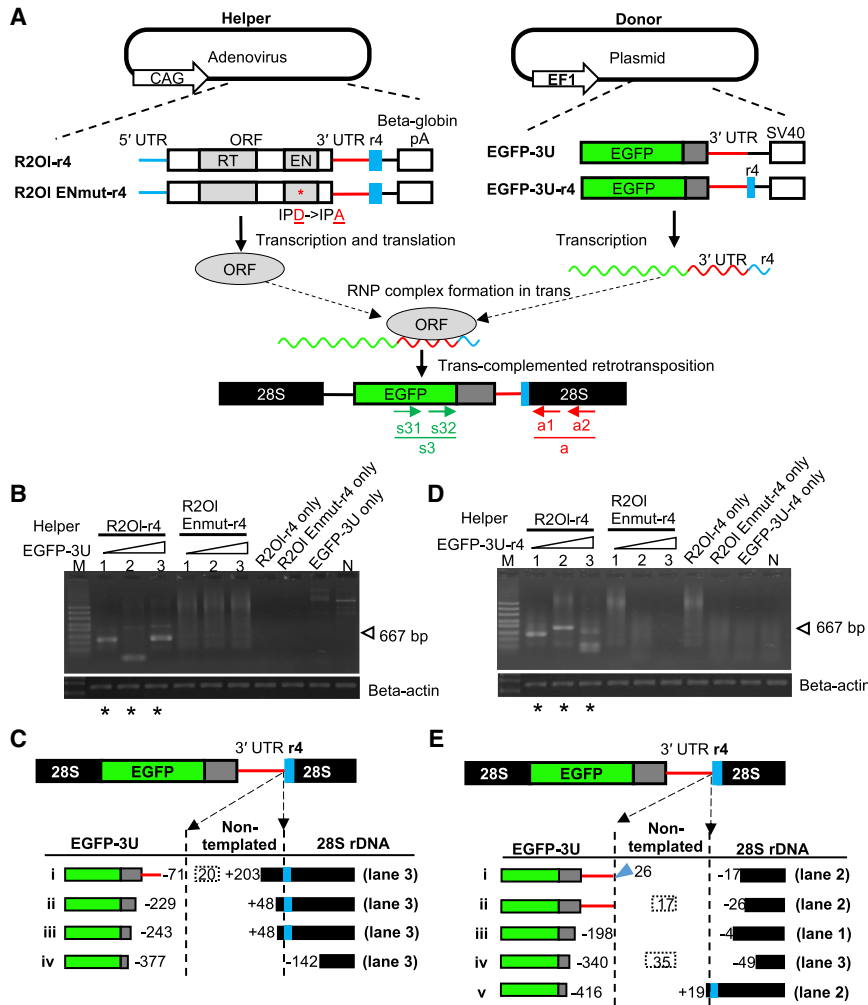
FIGURE 3. The EGFP gene fused with R2Ol 3′ UTR is retrotransposed into the 28S rDNA. (A) Schematic *trans*-complementation between the helper adenovirus and donor plasmids. The helpers are adenovirus R2Ol-r4 or R2Ol ENmut-r4; the donors are plasmids EGFP-3U or EGFP-3U-r4. The helper adenovirus provides the ORF protein to bind the donor mRNA to form a RNP complex. The 3′ junctions of retrotransposed copies were detected by PCR with the primer sets s3 and a. (B) PCR results of the 3′ junction of EGFP-3U. (C) 3′ junction sequence of inserted exogenous EGFP-3U. (D) PCR results of the 3′ junction of EGFP-3U-r4. (E) 3′ junction sequence of inserted exogenous EGFP-3U-r4. Lanes indicate independent experiments using different amounts of EGFP (0.5 µg in lane *1*, 1 µg in lane *2*, and 1.5 µg in lane *3*) with helper R2Ol-r4 or R2Ol ENMut-r4. The PCR products indicated by asterisks were TA cloned for further analyses. (R2Ol-r4 only or R2Ol ENmut-r4 only) Negative control PCR conducted with genomic DNA from the cells that only infected with adenovirus R2Ol-r4 or R2Ol ENmut-r4, (EGFP-3U only or EGFP-3U-r4 only) negative control PCR conducted with genomic DNA from cells that only transfected with plasmid EGFP-3U or EGFP-3U-r4, (N) negative control PCR conducted with genomic DNA from noninfected and nontransfected cells, (β-actin) internal control, (closed box) 28S rDNA sequence, (green box) exogenous EGFP sequence, (gray box) inverse SV40 sequence, (blue box) r4, (red line) 3′ UTR, (SV40) SV40 poly(A) signal sequence, (blue triangle) vector sequences, (dotted box) nonspecific sequences. Truncated or deleted sites in C and E are indicated by nucleotide numbers (*left* region, +1 as 5′-end of R2Ol; *right* region, +1 as 3′-end of R2Ol). The inverse SV40 sequence between EGFP and 3′ UTR of R2Ol (gray box) has no functional role for retrotransposition.

the 3′- and 5′-ends of R2Ol were shown to be inserted into the exact target site of the human 28S rDNA. Although R2 elements had become extinct from mammalian genomes (Kojima et al. 2016), the above results indicate that R2Ol from a distantly related species retains the ability to retrotranspose into the same 28S rDNA target sequence even in case of human cells.

Although a full-length of R2Ol (>4.2 kb) could be inserted by the WT R2Ol virus construct (R2Ol-r4), a 5′ truncation due to the midway cessation of reverse transcription was often observed (Fig. 2E, ii–iv, right region; Supplemental Table S2), as reported in many other LINEs (Richardson et al. 2015). In contrast to the exact 3′ junction structure (Fig. 2B), various 5′ junction structures were often observed at the integration site of R2Ol (Fig. 2E,ii–iv; Supplemental Table S2). Because various 5′-end sequences of endogenous R2 are known to exist in the genome (Eickbush and Eickbush 2012), it is presumed that the structural variation at the 5′ junction is also caused by the intrinsic integration mechanism of the R2 element. It is reported that a ribozyme in the 5′ UTR of R2 elements cleaves the 5′-end of R2 mRNA cotranscribed with the 28S rRNA, and the cleaved site within 28S rRNA (9–36 nt of 28S flanking sequence) is correlated with the 5′ junction structure (Eickbush and Eickbush 2010; Eickbush et al. 2013). Notably, R2Ol encodes a ribozyme that cleaves the cotranscribed 28S rRNA at 28 nt upstream from the 28S-R2Ol 5′ junction site (F Shiba and H Fujiwara, unpubl.). At the time the experiments were conducted, the cleavage activity of the ribozyme was not characterized fully, and thus the 5′-end of the R2Ol transcript expressed in our constructs consisted of the 5′ UTR and no upstream 28S sequence. Furthermore, how the upstream 28S rDNA affects the retrotransposition events of R2Ol has not been clarified yet. The lack of the 28S flanking sequence at the 5′-end in the R2Ol-r4 construct, used in this study, might enhance the 5′ structural variation.

developed. In this study, we successfully demonstrated a 28S rDNA-specific retrotransposition of the R2 element (R2Ol) isolated from medaka fish in human cells. Both

In contrast to the retrotransposition of R2Ol (Fig. 2B), we could not obtain retrotransposed clones showing the exact 3′ junction structure in the *trans*-complementation system (Fig. 3C,E), except for observing intact 3′ UTR reverse transcription in EGFP-3U-r4 (Fig. 3E, i,ii, left region; Supplemental Table S4). Cleavage of the bottom strand of the 28S rDNA for insertion of the EGFP varied in structure even with the r4 sequence at the 3′-end of the construct (Fig. 3E, right region; Supplemental Table S4). Recently, we observed a decreased retrotransposition efficiency of R2Ol and an inaccurate 3′ junction structure of the integrated R2Ol in the zebrafish embryo, when a sequence unrelated to the target 28S rDNA, such as poly (A) tail, was added to the 3′-end of R2Ol mRNA (Kuroki-Kami et al. 2019). This suggests that the read-through 3′-region of EGFP mRNA transcribed from the donor plasmid vector, which is not related to the 28S sequence, may cause an inefficient and inaccurate integration of the EGFP gene. In addition, when the length of 28S rDNA sequence added to the 3′ UTR of R2Ol was elongated from 4 bp to 100 bp, the retrotransposition efficiency was drastically increased, even though the constructs ended with the unrelated sequence (Kuroki-Kami et al. 2019). This indicates that annealing of the downstream 28S rRNA sequence in the read-through transcript with the target 28S rDNA enhances the retrotransposition activity. Thus, the addition of a longer 3′-flanking 28S target sequence at the 3′-end of the exogenous gene in the donor plasmid is expected to improve the efficiency and accuracy of integration by *trans*-complementation.

Although we here succeeded in detecting the insertion of the donor sequence into the 28S rDNA target, more primer sets for PCR should be examined to exclude a possibility of the insertion in other target sites. Furthermore, the low efficiency of the *trans*-complemented retrotransposition of the donor sequence may be explained by competition with the WT helper construct, which retains the 3′ UTR of R2Ol and is retrotranspositionally competent itself. To elevate the efficiency and accuracy of in *trans*-retrotransposition system in human cells, at least the helper construct without 3′ UTR of R2Ol should be used in future. In addition, our recent study in zebrafish showed that the retrotransposition efficiency of R2Ol itself was much higher than that of the *trans*-complemented retrotransposition system, even using the helper construct without 3′ UTR (Kuroki-Kami et al. 2019). This indicates that further improvement may be necessary for establishing the efficient *trans*-complementation system, in addition to the block of retrotransposition of the helper sequence itself.

Establishment of the R2Ol-based *trans*-complementation system in human cells signals great progress for future advances in therapeutic engineering. In this system, the helper and donor constructs are separately incorporated and expressed in human cells, which enables a time-efficient and easier manipulation, because the same helper

R2Ol virus can be used in different cases. Here, we used plasmid constructs for donor by addition of the R2Ol 3′ UTR, which enabled an easier vector construction. This novel system presents the advantage that multiple genes of interest on different plasmids can be transfected simultaneously and integrated into multiple target sites in the rDNA tandem array, because most eukaryotes have retained more than one hundred copies of the rDNA unit (Salim and Gerton 2019). We consider that transgene integration into several rDNA units will have fewer effects on the host and thus the rDNA will be a good candidate target site for gene therapy.

A critical difference between the CRISPR/Cas9 and R2Ol systems lies in the process of integration; LINE uses reverse transcription of mRNA into cDNA using TPRT with relatively high efficiency and specificity. Although a further refinement and improvement of the R2Ol system are needed, we expect that R2Ol may provide an alternative targeted gene knock-in method for future practical applications such as gene therapy.

MATERIALS AND METHODS

Human cell culture

293T cells (RCB2202) provided from the RIKEN cell bank in Japan, were grown in DMEM medium (Sigma-Aldrich) supplemented with 10% FBS (Japan Bio Serum) and 1% penicillin–streptomycin (Thermo Fisher Scientific) at 37°C with 5% CO_2.

Plasmids

For the retrotransposition assay, a full-length R2Ol (pR2Ol-r4) and R2Ol ENmut (pR2Ol ENmut-r4), which had the 4 bp downstream sequence (TAGC) of the 28S target site r4 at the 3′-end, were digested with EcoRI and NotI and inserted between EcoRI and NotI of the pENTR plasmid (Invitrogen). This resulted in pENTR2Ol-r4 and pENTR2Ol ENmut-r4, which were used for recombinant adenovirus construction. For the *trans*-complementation assay, we designed donor plasmids, pEGFP-3U and pEGFP-3U-r4, as follows. For the construction of donor pEGFP-3U, a *Xenopus* EF1-α (EF1) promoter was inserted at KpnI and HindIII sites in the vector plasmid pBluescript II KS(+). The EGFP sequence was inserted into NocI and ClaI sites of the plasmid. Then the SV40 polyA signal sequence (in reverse orientation) was inserted into the ClaI and NotI site. The 3′ UTR of R201 unit was inserted into NotI site, ended with SV40 polyA signal sequence. Similarly, we constructed pEGFP-3U-r4, which had the r4 sequence at the 3′-end.

Recombinant adenovirus construction

Adenovirus expression vectors for R2Ol were constructed from pAxCAwtit (TaKaRa) containing the chicken β-actin promoter (CAG) promoter. Two plasmids, pENTR2Ol-r4 and pAxCAwtit (or pENTR2Ol ENmut-r4 and pAxCAwtit), were mixed with LR Clonase (Gateway Vector Conversion System, Invitrogen) to generate recombinant pAxCAwtit according to manufacturer's

instructions. The recombinant pAxCAwtit vector was digested with NspV, and the linearized adenovirus DNA was transfected into human 293T cells using FuGENE HD Transfection Reagent (Promega) according to the manufacturer's protocol to generate recombinant adenovirus.

Retrotransposition assay

A R2Ol retrotransposition assay in 293T cells was performed in 24-well plates. The recombinant R2Ol adenovirus infected at MOIs of 1, 10, and 50. In case of the *trans*-complementation retrotransposition assay, the donor plasmid pEGFP-3U or pEGFP-3U-r4 (0.5, 1, or 1.5 µg) was transfected in 293T cells with 60% confluence in 24-well plates using Lipofectamine 3000 (Invitrogen) following manufacturer's instructions. Twenty-four hours later, helper recombinant adenovirus of R2Ol was infected into the cells and incubated for 72 h.

Genomic DNA was extracted from the cells infected with R2Ol viruses, and sequence-specific retrotransposition events were detected by nested polymerase chain reaction (PCR) to amplify the 3′ junction site and 5′ junction site, respectively. The 3′ junction of R2Ol integration was detected using a forward primer set within R2Ol (s1; s11 and s12) and a reverse primer set in the human 28S rDNA (a; a1 and a2) (Fig. 1C, red). The 5′ junction of R2Ol integration was detected using a forward primer set in the human 28S rDNA (s2: s21 and s22) and five reverse primer sets at different positions within R2Ol (a1: a11 and a12, a2: a21 and a22, a3: a31 and a32, a4: a41 and a42, a5: a51 and a52) (Fig. 1C, blue). To detect the 3′ junction integration of EGFP-3U and EGFP-3U-r4 by *trans*-complementation, a forward primer set s3 (s31 and s32) within EGFP and a reverse primer set a (a1 and a2) in Hm28S rDNA were used (Fig. 3A). All PCRs were performed with 35 cycles of amplification for 20 sec at 98°C, for 30 sec at 60°C, and for 30 sec at 72°C. The PCR products were TA cloned using pGEM-T Easy Vector Systems (Promega) and sequenced. The β-actin gene, used as an internal standard for equal genomic extraction, was amplified by primers human β-actin-F and human β-a ctin-R. All primer sequences are listed in Supplemental Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Bibillo A, Eickbush TH. 2004. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem* **279:** 14945–14953. doi:10.1074/jbc.M310450200

Christensen SM, Ye J, Eickbush TH. 2006. RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc Natl Acad Sci* **103:** 17602–17607. doi:10.1073/pnas.0605476103

Eickbush DG, Eickbush TH. 2010. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol* **30:** 3142–3150. doi:10.1128/MCB.00300-10

Eickbush DG, Eickbush TH. 2012. R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements. *Mob DNA* **3:** 10. doi:10.1186/1759-8753-3-10

Eickbush TH, Eickbush DG. 2015. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol Spectr* **3:** MDNA3-0011-2014. doi:10.1128/microbiolspec.MDNA3-0011-2014

Eickbush DG, Luan DD, Eickbush TH. 2000. Integration of *Bombyx mori* R2 sequences into the 28S ribosomal RNA genes of *Drosophila melanogaster*. *Mol Cell Biol* **20:** 213–223. doi:10.1128/MCB.20.1.213-223.2000

Eickbush DG, Burke WD, Eickbush TH. 2013. Evolution of the R2 retrotransposon ribozyme and its self-cleavage site. *PLoS ONE* **8:** e66441. doi:10.1371/journal.pone.0066441

Fujiwara H. 2015. Site-specific non-LTR retrotransposons. *Microbiol Spectr* **3:** MDNA3-0001-2014. doi:10.1128/microbiolspec.MDNA3-0001-2014

Govindaraju A, Cortez JD, Reveal B, Christensen SM. 2016. Endonuclease domain of non-LTR retrotransposons: loss-of-function mutants and modeling of the R2Bm endonuclease. *Nucleic Acids Res* **44:** 3276–3287. doi:10.1093/nar/gkw134

Kazazian HH, Moran JV. 2017. Mobile DNA in health and disease. *N Engl J Med* **377:** 361–370. doi:10.1056/NEJMra1510092

Kojima KK, Fujiwara H. 2005. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22:** 2157–2165. doi:10.1093/molbev/msi210

Kojima KK, Seto Y, Fujiwara H. 2016. The wide distribution and change of target specificity of R2 non-LTR retrotransposons in animals. *PLoS ONE* **11:** e0163496. doi:10.1371/journal.pone.0163496

Kuroki-Kami A, Nichuguti N, Yatabe H, Mizuno S, Kawamura S, Fujiwara H. 2019. Targeted gene knockin in zebrafish using the 28S rDNA-specific non-LTR-retrotransposon R2Ol. *Mob DNA* **10:** 23. doi:10.1186/s13100-019-0167-2

Lee M, Kim H. 2019. Therapeutic application of the CRISPR system: current issues and new prospects. *Hum Genet* **138:** 563–590. doi:10.1007/s00439-019-02028-2

Lino CA, Harper JC, Carney JP, Timlin JA. 2018. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv* **25:** 1234–1257. doi:10.1080/10717544.2018.1474964

Liu M, Rehman S, Tang X, Gu K, Fan Q, Chen D, Ma W. 2018. Methodologies for Improving HDR Efficiency. *Front Genet* **9:** 691. doi:10.3389/fgene.2018.00691

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72:** 595–605. doi:10.1016/0092-8674(93)90078-5

Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL, Moran J V. 2015. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol Spectr* **3:** MDNA3-0061-2014. doi:10.1128/microbiolspec.MDNA3-0061-2014

Salim D, Gerton JL. 2019. Ribosomal DNA instability and genome adaptability. *Chromosome Res* **27:** 73–87. doi:10.1007/s10577-018-9599-7

Yang J, Malik HS, Eickbush TH. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non–long terminal repeat retrotransposable elements. *Proc Natl Acad Sci* **96:** 7847–7852. doi:10.1073/pnas.96.14.7847

You L, Tong R, Li M, Liu Y, Xue J, Lu Y. 2019. Advancements and obstacles of CRISPR-Cas9 technology in translational research. *Mol Ther Methods Clin Dev* **13:** 359–370. doi:10.1016/j.omtm.2019.02.008